AI LAW MODEL
FOR ETHICAL LEGISLATION

STRATEGIC RECOMMENDATIONS FOR THE REGULATION OF ARTIFICIAL INTELLIGENCE

Oleksii KOSTENKO

2025

Oleksii Kostenko

AI LAW MODEL FOR ETHICAL LEGISLATION: STRATEGIC RECOMMENDATIONS FOR THE REGULATION OF ARTIFICIAL INTELLIGENCE

Monograph

SciFormat Publishing Inc.
Canada, 2025

AI Law Model for Ethical Legislation: Strategic Recommendations for The Regulation of Artificial Intelligence

Recommended for publication by the State Scientific Institution Institute of Information, Security and Law of the National Academy of Legal Sciences of Ukraine, Ukraine (Minutes No. 8 of the meeting of the State Scientific Institution Institute of Information, Security and Law of the National Academy of Legal Sciences of Ukraine dated 29.09.2025 p.)

Author:
Oleksii Kostenko, Ph.D. (Law) Senior Researcher, Associate Professor, State Scientific Institution, Institute of Information, Security and Law National Academy of Legal Sciences of Ukraine, Ukraine
ORCID ID: 0000-0002-2131-0281
Email: oleksii.kostenko@sciformat.com

Reviewers:
Dr. Siddhartha Paul Tiwari, PhD, FRAS, FRAI Head, Policy and Planning, Google Asia Pacific 70 Pasir Panjang Rd, Maple Business City, Singapore 389375.
Professor Adi Fahrudin, PhD Dean, Universitas Bhayangkara Jakarta Raya Jl. Raya Perjuangan, No. 81, Bekasi, West Java 17121, Indonesia.
Professor Nataliya Onishchenko, D.Sc. (Law), Professor, Academician of the National Academy of Legal Sciences of Ukraine, deputy director of the Institute of State and Law named after V.M. Koretsky of the National Academy of Sciences of Ukraine, Honoured Lawyer of Ukraine. Ukraine.

Published by: SciFormat Publishing Inc., 2734 17 Avenue Southwest Calgary, Alberta, Canada, T3E 0A7
Year of Publication: 2025
ISBN: 978-1-0690482-5-7 (eBook)
DOI: 10.69635/978-1-0690482-5-7

The monograph is devoted to a comprehensive analysis of the legal and ethical aspects of artificial intelligence. The paper considers the basic concepts, principles of functioning and classification of AI systems depending on the level of risk, as well as the issues of transparency, accountability, human rights protection and cybersecurity. The monograph is presented in the form of a Model of a legislative act in the field of AI regulation and is built on the principle of scientific and practical commentary, combining doctrinal analysis and practical guidelines. The publication is addressed to scientists, lawyers, legislators and everyone who studies the problems of modern digital regulation.

Keywords: Artificial Intelligence Law, Ethical Regulation, AI Governance, Interdisciplinary Approach, Legal Framework, Human-Centered AI, Transparency, Accountability, Non-Discrimination, Cyber Resilience, Risk Assessment, Public Trust, International Standards, Socio-Cultural Impact, Policy Recommendations, Preventive Regulation, Adaptive Regulation, AI Ethics, Technological Innovation, Legal Gaps.

TABLE OF CONTENTS

ABSTRACT

This study — the AI Law Model for Ethical Law: Strategic Recommendations for the Regulation of Artificial Intelligence — was created as a conscious and purposeful scientific paradigm for systemic flaws and hidden vulnerabilities identified in existing and draft regulations of the USA, EU, PRC and other leading jurisdictions in the field of artificial intelligence.

A comprehensive analysis of international practice shows that even the most ambitious and technologically advanced AI acts leave significant gaps — from vague and contradictory definitions of basic concepts, insufficiently elaborated and fragmented ethical requirements, to the complete absence or insufficiency of a deep assessment of the socio-psychological and cultural consequences of AI implementation. Such shortcomings and regulatory gaps can lead to serious legal conflicts, ethical crises, increased social tensions, as well as to the loss or undermining of public trust in digital institutions and public policy in general.

The purpose of the project "Model AI Law for Ethical Law: Strategic Recommendations for the Regulation of Artificial Intelligence" is not only to warn against mechanical and little critical copying of foreign regulatory models, but also to form a thorough approach adapted to the realities of national jurisdictions, based on deep interdisciplinary expertise and comprehensive risk analysis. This approach involves considering the legal, technical, ethical, social, cultural, and medical aspects of AI implementation, which makes it possible to consider LLM technology as a multidimensional phenomenon with long-term consequences. The key idea is that AI laws cannot be the product of the work of narrow-profile initiators who are far from a systematic understanding of the technology, or authors specializing exclusively in one field (law, technology, medicine, etc.) without proper involvement of related fields. Only a broad association of interdisciplinary lawyers, highly qualified technical specialists, sociologists, psychologists, doctors, ethicists, cybersecurity and risk management specialists as part of integrated research groups provides a real opportunity to cover the entire range of direct and indirect risks, as well as to predict the long-term social, economic and political consequences of the development and use of AI.

The structure of the study provides not only a formal statement of provisions, but also sections that define and disclose in detail the fundamental principles of human-centeredness, transparency, accountability, non-discrimination and cyber resilience, taking into account international standards and scientific approaches. At the same time, these sections are combined with analytical blocks, which systematically describe problem areas identified in other countries, provide examples of negative consequences of ignoring these principles, and formulate recommendations for avoiding such risks in Ukrainian legislation. In particular, the following critical aspects are emphasized:

- legal gaps and lack of clearly defined enforcement mechanisms that allow individual actors to avoid or minimize liability for the direct or indirect harmful consequences of the use of AI, including economic, ethical and social losses;
- the lack of a single, scientifically based and normatively fixed risk assessment scale, which complicates the objective identification of the level of danger of specific AI solutions and creates opportunities for conscious or unintentional manipulations in the classification of systems, in particular in order to avoid regulatory requirements or reduce the scope of verification measures;
- lack of effective, transparent and legally enshrined mechanisms of public control and independent audit, capable of ensuring systematic monitoring, objective assessment and public information about the compliance of AI systems with ethical and legal standards;
- ignoring a deep and comprehensive analysis of the psychological, emotional, and sociocultural impacts of AI on vulnerable populations, including children, the elderly, people with disabilities, and other socially sensitive categories, which can lead to deepening inequality, social exclusion, and negative transformations of cultural values;
- weak integration and insufficient regulatory consolidation of internationally recognized technical standards and protocols into legal regulation, which complicates the unification of requirements, reduces the compatibility of national solutions with global ecosystems, and creates risks of technical and legal fragmentation.

Thus, this draft law is not only a draft law in the traditional sense, but also a deeply thought-out analytical and regulatory warning designed to become a conceptual guide for future state policy in the field of artificial intelligence. Its task is to incorporate a multi-level preventive mechanism into AI legislation that would combine preventive, adaptive, and corrective regulatory tools that can flexibly respond to new technological challenges and ethical dilemmas. The implementation of this approach should guarantee not only a formal balance between legal regulation, technological innovation and public security, but also ensure the stability of this balance in the context of rapid transformations and global competition, avoiding strategic and tactical mistakes already made by world leaders in this area.

INTRODUCTION

In a period of global singularity, when intelligent systems are becoming the infrastructural core of socio-economic development, legal traditions and institutions face an unprecedented challenge: co-evolution with AI technologies not only changes the ways of knowledge production and organization of work, but also rebuilds the space of legally significant actions, creating new grounds for rights and obligations, new liability regimes and new forms of harm.

The regulatory discourse on AI has acquired signs of a global race: jurisdictions compete for leadership, simultaneously implementing large-scale legal acts. However, there is often a normative dissonance behind external dynamics: what looks like a complex paradigm often turns out to be a set of incompatible or insufficiently coordinated policies that do not cover critical risks or, on the contrary, give rise to new ones. Under these conditions, the state should not replicate other people's erroneous approaches; A scientifically grounded, interdisciplinary architecture of legislation is needed, synthesizing the best international standards with national legal tradition and institutional realities.

This study assumes that it is the singularity that generates the stage of conceptual design of norms and preliminarily determines the success of the further implementation of any AI act. Without a careful preliminary diagram of risks — legal, technical, social, psychological, cultural, security — any legal or technical regulation risks becoming either decorative or overly repressive.

In view of this, the object of the study is the social relations that are formed in the process of designing and implementing the legal regulation of AI technologies in the context of rapid technological evolution.

The subject of the study is methodological models and legal tools for ensuring an ethical, safe and innovatively compatible regulatory architecture built on the principles of human-centeredness, accountability, transparency, non-discrimination, cyber resilience and adaptive controllability. An important metaparameter is the concept of e-jurisdiction as a form of transnational legal interaction in digital environments, including the Metaverse, where legal dimensions intersect with protocols and data.

The purpose of the study is to formulate theoretical and methodological foundations and a set of practical recommendations for national legislation on AI, which will prevent the reproduction of identified international gaps and, thus, ensure the sustainability, legal certainty, legitimacy and accountability of technological solutions. To achieve the goal, the following tasks have been set:

- to carry out a systematic comparative legal analysis of the key acts of the USA, EU, PRC and other jurisdictions regarding definitions, classifications of risks, restrictions and prohibitions;
- identify critical vulnerabilities in impact assessment, confirmation of compliance and surveillance procedures;
- develop proposals for the structure of responsibility and accountability chains;
- integrate technical standards and engineering safety requirements into legal norms without deforming their content;
- to lay down institutional conditions for the formation of interdisciplinary expert councils and the implementation of permanent audit practices.

The methodological basis of the study is complex and combines the comparative legal method, system analysis, risk-oriented design of regulatory regimes, scenario modelling, and Algorithmic Impact Assessment tools [1, 2], Data Protection Impact Assessment [3, 4, 5], Fundamental Rights Impact Assessment [6, 7, 8]), as well as STS (Science, Technology and Society) approaches [9, 10], the doctrine of good governance and human rights research in the digital sphere. The empirical base includes international acts, national draft laws, case law, ISO/IEC and NIST standards, as well as expert interviews and Delphi surveys [11, 12]. Such complementarity of methods allows you to avoid reductionism and identify the systemic properties of AI as a sociotechnical phenomenon.

The scientific novelty lies in shifting the emphasis from a descriptive regulatory approach to a design-oriented, adaptive regulatory architecture that provides for built-in cycles of self-updating norms, regulatory

sandboxes, legal testing mechanisms in real conditions, and a meta-level of oversight with the design of accountability metrics and ethical quality indicators.

The ethical framework of the study proceeds from the priority of human dignity, autonomy, justice and non-discrimination, as well as environmental and social responsibility. Tests of ethical admissibility are proposed to be formalized in the form of mandatory procedures:

- notification of the fact of using AI;
- explainability and evidence of algorithmic conclusions;
- revision of data sources regarding their quality and legality;
- guarantee of human control over critical automated decisions;
- counteraction to manipulative and behavioural practices;
- protection of vulnerable groups.

Particular attention is paid to the psychological consequences of the mass use of AI in education, labour, healthcare, and public administration, as well as the risks of mass surveillance and digitalization of sanctions practices.

The risk taxonomy clarifies the classical categories (unacceptable, high, limited, minimal) by introducing the concepts of systemic, emergent, cross-border and interactional risk that arises in the process of integrating AI modules in different contexts. Mechanisms for dynamic risk reclassification based on operational metrics, conclusions of artificial intelligence audits (AI audits) and the "third line of defence" (independent agencies and public councils) are provided, which ensures accountable and prompt correction of the regulatory regime without waiting for full-scale changes in the law.

The institutional design project emphasizes the need to create a National Interdisciplinary Scientific and Expert Council on AI, with a mandate to form standards, accredit auditors, maintain public registers of high-risk systems, coordinate regulatory sandboxes, and ensure international interaction. The Council should act on the principles of transparency, conflict of interest management, open science and public participation. Emphasis is placed on compatibility with European law and international legal practice without losing national legal personality and digital sovereignty.

The cross-border dimension of the study recognizes the reality of global AI value chains: data, models, infrastructure, and users are crossing borders faster than norms are being updated. The answer for the State is a combination of the principle of mutual recognition of compliance, the launch of cross-border sandboxes and test legal regimes, as well as the introduction of the concept of electronic jurisdiction for dispute resolution and data management in digital environments. GDPR standards and related practices should be implemented in cooperation with national law on personal data protection and information security.

The working hypothesis is that an effective legislative architecture for AI should be structurally consistent with the OSI (Open Systems Interconnection) Model [13, 14, 15] — meta-level, adaptive and interdisciplinary. It builds on the synergy of legal regulations and technical standards integrated into cybersecurity and ethical oversight processes and ensures institutional accountability through transparent audit procedures. The expected scientific and practical contribution is to create a conceptual framework and a package of tools suitable for direct use in normative practice.

The legal support of AI is not reduced to a simple fixation of prohibitions and obligations. Its essence lies in the creation of manageable, predictable, ethical and at the same time compatible with innovations rules that form a space of trust, protect rights and promote development.

Each state has a chance to offer its own modern model of AI regulation, which does not mechanically copy, but creatively synthesizes the best of world practice with national institutional realities, setting both a tactical task — to secure the present, and a strategic mission — to determine the ethical vector of the digital future.

The formation of the theoretical, methodological and institutional framework of national legislation in the field of AI should be based on:

1) systematic comparative legal audit and analysis of acts of the USA, EU, PRC and other jurisdictions [16];

2) identifying vulnerabilities in definitions, risk classification, impact assessment and surveillance procedures [17];

3) designing a "responsibility matrix" and accountability chains [18, 19];

4) integration of technical standards (ISO/IEC, NIST) into legal regulations [20, 21];

5) institutionalization of interdisciplinary boards, audits and sandboxes [22, 23];

6) development of mechanisms for cross-border interaction and electronic jurisdiction [24, 25];

7) risk-based design of laws and their scenario modelling [26];

8) Algorithmic/Data/Fundamental Rights Impact Assessment, Approaches STS [27];

9) generalization of international standards ISO/IEC and NIST;

10) litigation, expert interviews and Delphi surveys;

11) contract and clause templates, algorithmic audit maps and DPIA/AIA/FRIA checklists;

12) models of responsibility sharing protocols of public participation, implementation roadmap and performance indicators.

A systematic comparative legal analysis of the regulation of artificial intelligence, in particular in the USA, EU, PRC and other jurisdictions, is presented in Table 1.

Basic approaches to regulation.

The analysis of current and draft acts of the USA, EU, PRC and other states shows a tendency to simplification, when the technocratic approach displaces an interdisciplinary vision, and narrow definitions do not cover the dynamics and interaction of systems.

However, AI acts reveal a number of systemic gaps:

- uncertainty in definitions (which allows you to bypass the scope of norms);
- inconsistent risk taxonomies and fuzzy triggers for GPAI/foundation models;
- weak and predominantly advisory mechanisms of oversight and accountability;
- insufficient institutional capacity for independent auditing;
- lack of unified AIA/DPIA/FRIA procedures and safety and quality metrics;
- opacity of origin and licensing of training data, uncertainty about TDM/IP, as well as problems of labelling and provenance of content;
- gaps in incident reporting procedures and lack of public registers;
- excessive or unjustified exceptions (in particular in the field of biometrics) and poor assessment of psychosocial consequences for vulnerable groups;
- lack of cross-border sandboxes and mechanisms for mutual recognition of compliance.

European Union (EU). The EU is implementing the most ambitious and systemic regulatory framework — the Artificial Intelligence Act (AI Act), based on a risk-based approach. The law classifies AI systems according to the level of risk (prohibited, high-risk, limited risk, and minimal risk) and establishes strict requirements for transparency, ethics, protection of fundamental rights, and security of citizens. The AI Act also forms a unified governance system that includes the European AI Office and national competent authorities, aimed at ensuring effective implementation and supervision.

USA. The U.S. is taking a more flexible, decentralized approach focused on supporting innovation. Regulation is carried out through industry standards, recommendations and current legislation, without a single nationwide act. The focus is on stimulating business development and research, as well as consumer protection through existing legal mechanisms.

China (PRC). The PRC applies a centralized, state approach aimed at the rapid introduction of innovations and strengthening control. Among the main regulations is the "Regulations on the Management of Algorithmic Recommendations in Internet Information Services" [28] and the Personal Information Protection Law [29]. China focuses on the development of technology, but this approach creates a risk of insufficient protection of citizens' rights due to the priority of state interests.

Table 1. Comparative table of key aspects

Jurisdiction	Basic approach	Key acts/initiatives	Priorities	Disadvantages	Source
EU	Risk-oriented, complex	Artificial Intelligence Act, GDPR	Protection of rights, ethics, transparency	Inhibition of innovations is possible, there is no control of law enforcement agencies	[30, 31, 32, 33, 34, 35, 36, 37]
United States	Flexible, industry-specific	American AI Initiative, industry standards	Innovation, business, consumers	Lack of a unified system	[38, 39]
PRC	Centralized, public	Algorithm Recommendation Rules, PIPL	Innovations, state control	Insufficient protection of civil rights	[40, 41]

Mutual influence and global trends: The EU seeks to establish its approaches as a global standard in the field of AI, by analogy with the impact of the GDPR, and is already shaping policies in the United States, China and other countries (the so-called "Brussels effect"). At the same time, excessive detail of definitions and strictness of requirements can limit the global spread of EU standards, especially in states that defend technological sovereignty and develop their own approaches.

Modern AI regulation in the world is developing under the influence of three main models: European (ethics and protection of rights), American (priority of innovation and business) and Chinese (state control). The differences between these models create significant challenges for international harmonization, but at the same time stimulate the search for an optimal balance between innovative development and human rights guarantees.

SECTION I. AI MODEL LAW FOR ETHICAL LAW. BASIS

CHAPTER 1. SUBJECT OF REGULATION

This Law establishes the legal basis for the creation, development, training, testing, application, integration, operation, audit, control, termination, export and neutralization of artificial intelligence (AI) systems on the territory of the state, as well as regulates legal relations arising in connection with their impact on fundamental human rights and freedoms, national security, public administration, digital economy, ethics, environment and international integration.

The regulatory mechanism performs the function of establishing the boundaries and rules of legal relations regulated by the Law. Its use causes:

mandatory compliance with the Law in case of any use of artificial intelligence systems:

− in the activities of public authorities, in the provision of administrative services, in law enforcement, justice, public policy formation and in the field of digital governance;

− in the private sector, including business, education, healthcare, social platforms, advertising analytics, banking and insurance services;

− in the field of national security and defence, in particular in the development and use of autonomous systems, combat platforms, means of information and psychological influence and digital weapons;

− in the cross-border digital space, when processing personal and biometric data of citizens of the State, as well as when using national AI models or the results of their functioning outside the national jurisdiction.

Creation of a universal regulatory core covering the entire life cycle of an AI system — from idea and architectural design to deployment, modernization, neutralization, or shutdown [42, 43]. This regulatory core defines the duties and responsibilities of the following categories of subjects:

− developers and engineers who form the software, technical and semantic architecture of AI systems [44, 45] and are responsible for its safety and legality;

− providers and aggregators of training and operational data, including personal, biometric, and behavioural data [46], which are obliged to ensure their quality, legality and transparency of use;

− controllers and operators of AI systems, who implement them in a real environment and are responsible for compliance with ethical, security and legal requirements [47];

− independent auditors who carry out systematic verification of algorithms, results and compliance with established regulations [48];

− end users who have the right to be informed about interaction with the AI system and use mechanisms for control, appeal, and rejection of automated decisions [49].

Have direct effect on public authorities, which are obliged to:

− implement only those AI solutions that comply with the principles of ethics, transparency, non-discrimination and legal certainty [50];

− guarantee open and full public access to information on algorithmic tools used in administrative processes, administration of justice or political management;

− to ensure the human-centricity of digital services through the mandatory availability of an alternative without the use of AI to receive administrative services [51];

− implement internal mechanisms of ethical control, systematic monitoring of the results of the functioning of AI solutions and carry out public reporting on the practice of their application [52].

Delimitation of the jurisdictional effect of the Law in the case of the use of AI tools of a cross-border nature, in particular in the following cases:

− if foreign companies offer AI services to citizens of the State, regardless of the location of servers or the governing body;

− if the results of the activities of AI systems created or trained in the State are used outside its borders, but affect the rights of citizens of the State or national security;

− if cross-border processing of personal, biometric or other sensitive data of citizens of the State is carried out by AI models operating or controlled from abroad;

− if AI systems are used to make or support political decisions that have an international effect for the State.

In these cases, the subjects are subject to the effect of this Law on the extraterritorial principle, which is based on jurisdiction related to the impact on national interests, information sovereignty and human rights.

Ethical standards — for the first time, the obligation to comply with the following principles is introduced into national legislation:

"ethics-by-design" is the concept of developing AI systems in such a way that ethical principles such as fairness, dignity, non-discrimination, the right to privacy and transparency are integrated at every stage of the AI lifecycle [53, 54]: from planning the algorithm architecture to its testing, implementation and control of results [55]. This principle obliges developers to anticipate potential ethical risks [56], create built-in abuse prevention mechanisms [57], ensure proper human oversight, and guarantee the user the right to opt out of an automated decision [58].

"human-centric AI" is the concept of human-centricity, according to which artificial intelligence systems must be designed, implemented and applied in such a way as to guarantee the protection of autonomy, dignity, well-being and fundamental human rights. This principle implies:

− giving priority to human decision in case of any automated influence on a person;

− ensuring the right of a person to refuse to use AI without restricting access to basic services;

− ensuring the transparency of algorithms and the obligation to explain the logic of decision-making in an understandable form;

− the availability of effective mechanisms for restoring violated rights if the decision of the AI has caused negative consequences for the person;

− mandatory human participation at all stages of the functioning of high-risk systems ("human-in-the-loop" — a person in the decision-making process or "human-on-the-loop" — a person in the control function);

− implementation of regular ethical and legal audits on the impact of AI systems on human rights.

− "do-no-harm principle" — the principle according to which any application of artificial intelligence should exclude the possibility of harm to the physical, psychological or digital integrity of a person, violation of public order, undermining democratic institutions, intentional or unintentional damage to the environment. This principle imposes an obligation on developers, operators and regulators:

− assess risks for different categories of users at the stage of design and testing of AI systems;

− take measures to prevent data distortions, toxic results, discriminatory decisions and erroneous autonomous actions;

− establish technical and organizational control barriers, as well as protocols for shutting down systems in case of damage detection;

− carry out mandatory continuous monitoring and regular reporting of incidents that have had or may have negative consequences for humans, society or the environment.

− "explainability and transparency" is the principle according to which algorithmic decisions made by AI systems must be understandable, logically reproducible and available for interpretation by the user, regulator or auditor.

This principle implies:

− mandatory technical documentation of the logic, structure and parameters of the model;

− providing clear reports on the causes and factors that led to a specific result or decision;

− providing access to information about data sources that influenced the conclusions of the AI system;

− development of interfaces and formats for presenting results that take into account the level of digital awareness of the user;

− conducting an independent audit to identify discrimination, bias or unintended effects;

ensuring the application of international standards and recommendations (in particular, NIST AI RMF, ISO/IEC 24028 and EU approaches to high-risk systems).

The requirements apply to the following key areas:

- public administration and justice;
- healthcare, education, social protection;
- national defence and security;
- economics, finance, advertising, labour market;
- information space and media;
- applied research, as well as the development and application of military technologies.

Compliance with international models:

The Law of the State structurally corresponds to Sections 1 of the AI Act (EU), which determines that the act applies to all operators, developers, suppliers and users of AI systems, including those located outside the European Union, but provide services or affect persons within the EU. This approach forms the principle of extraterritoriality, which is also taken into account in the Law of the State: its provisions apply to any AI system that has an impact on the national interests, rights of citizens or information security of the state, regardless of the country of origin or jurisdiction of the operator.

Takes into account the provisions of Executive Order 13960 "Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government" (USA), which enshrines the principle of end-to-end management of the life cycle of AI systems — from the research and design stage to implementation, continuous monitoring, neutralization or shutdown. The decree establishes the obligation of public authorities and private companies to ensure security, responsibility, ethical consistency and transparency in all phases of AI development and use. In addition, it provides for interagency coordination, systemic risk management, public participation and consideration of human rights impacts. The relevant provisions have been adapted and integrated into this Law of the State.

It partially adapts the approach enshrined in the Personal Information Protection Law **(**PIPL), which establishes systematic control and regulation of the processing of personal and biometric data. PIPL enshrines the principles of legality, transparency, data minimization, intended use and prevention of harm to an individual.

The law also imposes on the entities using AI systems the obligation to obtain the user's information consent, guarantees the right of a person to refuse automated decision-making and the right to appeal against them, and determines increased ethical and social responsibility for the consequences of the functioning of algorithms — both at the individual and societal levels.

In the national law, these approaches are integrated by strengthening the personal data protection regime, creating ethical supervisory boards and introducing the principle of presumption of possible risk of harm in automated information processing.

SECTION II. SCOPE OF THE LAW

The Law applies to all natural and legal persons, regardless of the form of ownership and organizational and legal form, who develop, implement or use AI systems on the territory of the State. The Law also applies to entities that apply AI systems to citizens or residents of the State. In addition, the Law applies to foreign entities in cases where the functioning of AI systems causes or may cause an impact on the information space, economy, social sphere, defence or other national interests of the State.

2.1 This Law applies to all subjects of public and private law that are directly or indirectly involved in the life cycle of artificial intelligence systems, including the stages of research, development, design, training, testing, implementation, commercialization, modification, deactivation or use in any sphere of public life, including economic, defence, cultural and administrative.

This Law applies to:

developers, operators, suppliers, auditors and end users of AI systems located or operating in the territory of the State;

institutes, organizations, startups or other entities that train, adapt or localize AI algorithms using national data or infrastructure;

companies that integrate AI into their business models, digital products, logistics systems, decision-making systems, or human resources;

government agencies that use AI in digital governance, data analysis, law enforcement, risk management, automated assessment, access control, or border surveillance.

In addition, this Law applies to any AI systems that:

cause or may cause a direct or indirect effect on natural or natural or legal person **s**under the jurisdiction of the State;

are used to process, generate, analyse, classify, predict or model information relevant to the public interest, the protection of human rights, public administration, national security or the digital economy of the State.

2.2 Effect of the Law on Foreign Entities. This Law applies to any foreign individuals and legal entities, including companies and authorities, that carry out activities related to the use of AI systems outside the territory of the State, if such activities cause or may cause a direct or indirect impact on the citizens of the State, its national security, economic or informational interests.

In particular, this Law covers foreign entities if:

a) their AI systems process or store personal data, including biometric, behavioural, financial or other sensitive data identifying citizens of the State;

b) automated decisions made by AI systems affect the rights, obligations or legal status of individuals and legal entities of the State;

c) they provide digital services or deploy algorithmic models on platforms accessible within the State's territory, regardless of the location of the cloud infrastructure, servers, computing power or legal address of the company.

In case of confirmation of the influence of a foreign entity on the objects of legal regulation of this Law, the jurisdiction of the State may be applied to it in accordance with the principle of protection of national sovereignty in the digital space. Determination of the presence of such an impact is carried out through the digital monitoring and oversight, expert opinions and consultations with the international partners of the State.

2.3 Areas of application of the Law. This Law applies to the main sectors of public, state, security and economic life, in which the use of AI systems has a significant impact on the rights, interests and security of citizens and the State, in particular:

− in the field of public administration: automated platforms for the provision of administrative services, e-governance, digital regulation of public processes, big data analysis for public decision-making;

− in the field of justice: algorithmic assistance in the formation of court decisions, digital processing of evidence, identification of patterns in criminal and administrative proceedings, electronic justice;

− in the field of security, defence and intelligence: autonomous combat systems, control systems for unmanned aerial vehicles (drones), threat detection systems and technologies, open-source analytics (OSINT), tactical forecasting, cyber operations, digital protection systems for critical infrastructure;

− in the field of healthcare: predictive diagnostics, clinical decision support systems, personalized treatment, drug management, medical image recognition, virtual physician assistants;

− in the field of education: individualized training programs, automated assessment systems, language models to support distance learning, analytics of educational trajectories;

− in the field of business: forecasting market fluctuations, optimization of logistics, automation of financial reporting, automated systems for recruiting and evaluating personnel, detection of fraudulent schemes;

− in the field of advertising, media and information influence: content generation, creation of overly personalized advertising, automated content moderation and control systems, information manipulation;

− in the field of science and innovation: modelling complex physical, biological or social processes, generating hypotheses, supporting experimental design, discovering new materials, analysing scientific publications.

The use of AI systems in these areas requires differentiated legal regulation, considering the specifics of risks, the level of impact on humans, ethical standards, and the need to ensure human control.

2.4 Cases of non-application of this Law. This Law does not apply to:

− systems that are created or used exclusively for the personal needs of individuals, for household affairs, personal planning, hobbies or non-commercial communication, if there is no systemic impact on the public space, public security or information environment of the State;

− systems that are used exclusively for research or academic purposes, are under development, do not interact with real users, do not participate in the decision-making process and do not have access to public or commercial platforms, until their official registration or launch;

− cases where international treaties that have higher legal force than national legislation in accordance with Article 9 of the Constitution of the State are applied, in cases where such treaties establish different rules regarding artificial intelligence, data processing standards or the jurisdictional affiliation of technical solutions.

In each individual case, the application of exceptions must be justified based on technical expertise, a potential impact assessment and a decision of the relevant digital surveillance body.

2.5 Priority of the provisions of this Law. The provisions of this Law are special norms in terms of regulating relations related to the development, use, implementation, control, audit and termination of artificial intelligence systems. In cases where other legislative acts of the State contain general provisions on information activities, protection of personal or biometric data, cybersecurity, digital transformation, administrative proceedings, labour or ethical relations, the provisions of this Law shall apply as a priority.

In case of conflicts between the provisions of this Law and other normative legal acts, unless otherwise provided by international treaties ratified by the State. Such a ratio is aimed at ensuring the integrity of the regulation of the field of artificial intelligence within the framework of a single logic of digital law and preventing fragmentation of jurisdictional competence between different regulatory systems.

CHAPTER 3. BASIC TERMS AND THEIR DEFINITIONS

3.1. In this Law, the terms are used in the following meaning:

1) **Electronic** is the one associated with the use of electronic devices or channels for creating, transmitting, storing, processing or reproducing information in electrical or digital form. Electronic systems are constituent parts of digital infrastructure, but have a wider application range, including analog or hardware-oriented systems.

2) **Digital** is related to the representation, processing and storage of information in the form of numerical (discrete) values. It defines everything that operates with information in the form of bits, bytes, numbers, code or signatures; digital is the principle of modelling, representation, transformation of reality using mathematical models, algorithms and computational processes.

3) **Artificial intelligence (AI)** is a system capable of carrying out processes of analysis, forecasting, decision-making, generating information or actions based on given or independently detected patterns, which are similar in nature to human cognitive functions synthetic speech model, autonomous control.

4) **An artificial intelligence system** is a technological AI system developed using machine learning, logic programming, statistics, heuristics, modelling, which: a) performs certain functions without direct human intervention; b) able to adapt to new inputs; c) functions in accordance with algorithmically determined goals.

5) **The AI life cycle** is a set of stages that include conceptualization, design, architecture, training, validation, testing, use, monitoring, refinement, updating, decommissioning, and destruction of an AI system.

6) **Generative AI** is a system capable of creating textual, visual, audiovisual, code, or multimodal content using artificial intelligence models, including large language models (LLMs), transformers, or diffusion networks.

7) **A high-risk AI system** is a system the use of which: a) may affect the rights, freedoms or security of a person; b) relates to public administration, education, health, justice, security, labour relations or finance; c) is subject to mandatory registration and audit in accordance with the legislation.

8) **AI audit** is the process of verifying the compliance of the AI system with established standards of safety, ethics, non-discrimination, effectiveness, data protection and human rights, carried out by independent experts or authorized bodies.

9) **An algorithmic decision** is a decision formed by the AI system in whole or in part, with or without the possibility of human intervention in the result.

10) **Human-in-the-loop** is the principle that critical decisions affecting human rights or security should be made solely with the participation of a person who has the authority to approve, correct, or reject an automated outcome.

11) **Digital surveillance** is a system of state or independent surveillance that includes institutions for monitoring, incident investigation, abuse response, and public reporting on the functioning of AI systems.

12) **AI transparency** is the ability of an AI system to be technically, legally and ethically understandable, predictable, explainable and open to verification of the results of its activities.

13) **The Register of AI Systems** is a centralized database maintained by an authorized body and contains information about registered high-risk, generative and military-civilian artificial intelligence systems, their operators, functional purpose, risk level, and audit results.

14) **Ethical assessment of AI** is a procedure for a systematic analysis of the goals, means, impact, and consequences of the AI system from the perspective of human rights, sustainable development, and institutional justice, which includes the participation of an independent expert council.

15) **Biometric data** is digital data obtained because of processing a person's physiological or behavioural characteristics that provide his/her unique identification (e.g., face, fingerprints, voice, gait, eye movements).

16) **Digital sovereignty** is the ability of the state to ensure control over critical digital resources, infrastructure, data, and artificial intelligence systems that affect national security, economic independence, and citizens' rights in cyberspace.

17) **A high autonomy model** is an AI system that can make decisions independently without human intervention in a changing context, including learning from new data in real time (online learning) and changing the algorithm of actions without prior human programmatic intervention.

18) **The National AI Ethics Council** is an advisory body established under the executive body of the State, whose functions include addressing ethical challenges, recommending AI policies, coordinating critical applications of artificial intelligence systems, and participating in international digital ethics control initiatives.

19) **Critical infrastructure** is facilities, systems, networks and services that are critical to national security, defines, health, economy and society, the functioning of which largely depends on the reliability, resilience and security of digital and algorithmic technologies.

20) **AI-based autonomous weapons** are a system capable of independently identifying, tracking, evaluating and hitting a target without direct intervention of a human operator, using computer vision algorithms, data analysis, behavioural models and combat simulation.

21) **A Chief AI Officer** is an authorized person in a government agency or legal entity who is responsible for strategic implementation, ethical support, risk assessment, auditing, transparency, and interaction with the regulator regarding implemented AI systems.

22) **Institutional control over AI** is a set of organizational, legal, administrative, and procedural measures aimed at ensuring legality, ethics, transparency, responsibility, and accountability during the creation, implementation, management, operation, and termination of AI systems at the level of public authorities, business, and civil society.

23) **The principle of preventive liability** is the principle according to which the developer or operator of an AI system is obliged to take measures to prevent harm that may potentially arise because of the operation of this system, even if the damage has not yet occurred or is hypothetical.

24) **A digital avatar** is a visually simulated image that represents a person in a digital or virtual environment, reproducing their appearance, facial expressions, gestures, body movements, or a stylized projection of their appearance. Digital avatars can be either an exact biometric copy or a stylized or abstract reproduction that serves as a user's identification or self-presentation in online spaces, virtual environments, metaverses, or immersive platforms.

25) **A voice clone\synthetic voice** a digital model that imitates or synthesizes the voice, accent, timbre, intonation, rhythm of speech, vocabulary, speech turns or communication style of a particular person educational or memorial purposes.

26) **A virtual agent (representative, digital assistant, chatbot**) is a software system that autonomously or semi-autonomously communicates on behalf of a person using his/her style, rhetoric, intentions or predetermined parameters. Such an agent can be integrated into social networks, digital government services, corporate platforms, media platforms or online court procedures for the purpose of delegated participation, information exchange, moderation and representation.

27) **A digital shadow** is an indirect digital representation of a person that is formed algorithmically based on collected data on their behaviour, preferences, interactions, click history, purchases, navigation, ratings, time rhythm of activity, etc. The digital shadow does not have an autonomous visual or voice form but is used by artificial intelligence systems to predict behaviour, personalize interfaces, recommendations, or verify the user without their direct involvement.

28) **Digital reconstruction (hologram, 3D models, posthumous digital personality ('post-personality'))** is a highly accurate or stylized digital replica of a person, which can be created during life or after death on the basis of biometric data, audio and video materials, written texts, archives, social profiles. Such reconstructions are used in cultural, historical, educational, memorial, artistic or religious contexts, as well as in inheritance digital law systems.

29) **Artificial intelligence system in the field of justice (hereinafter referred to as the AI system)** is a technological complex that carries out automated data processing, the formation of analytical conclusions, indicative recommendations or information messages that can be used by a judge, investigating judge, prosecutor or parties in the course of procedural activities, while such a system is not empowered to make final procedural decisions.

30) **Decisive human participation** is the mandatory and real ability of an authorized official (judge/investigating judge/prosecutor) to independently evaluate evidence, change or reject the result of the AI system, and be responsible for the final decision.

31) **Human controllability (human-in/on-the-loop)** is an organizational and technical mode of application of the AI system, in which a person retains control over critical stages of analysis and decision-making, has the means to stop, review, and adjust the results of the system.

32) **Algorithmic recommendation** is an informational message, assessment, rating, hint, forecast or other non-final conclusion formed by the AI system to support decision-making without binding legal force for the subject of authority.

33) **Ethical certification** is a procedure of preliminary and periodic assessment of the AI system, carried out by an authorized body to verify compliance with human rights and freedoms, non-discrimination, transparency, human controllability, safety and proportionality.

34) **Independent algorithmic audit** is a third-party verification of the AI system for quality, stability, absence of prohibited bias, security and compliance with the declared purpose with the obligatory fixation of methods and results obtained.

35) **Local explainability** is the provision of a clear justification for a specific result of the AI system in a particular case (including key factors of influence), sufficient for a procedural check.

36) **Prohibited characteristics** are any sensitive characteristics of an individual, the use of which in AI systems, directly or indirectly (through derived or correlated indicators), may lead to discrimination. Such characteristics include race, skin colour, ethnic or social origin, language, religion, political or other opinions, health status, disability, sexual orientation, age, trade union affiliation, and other characteristics determined by law.

37) **A proxy feature** is a variable that is not directly a prohibited feature but has a strong correlation with it and reproduces a discriminatory effect in the results of the AI system.

38) **AI System Operator** — a public authority, other state body or institution/organization designated by it in accordance with the law that operates the AI system in proceedings, ensures its integration, security, logging and compliance with the requirements of this Article.

39) **Supplier (developer)** of an AI system is a business entity or other person that creates, implements, supplies or provides support for an AI system and bears legal responsibility for its quality, functional properties, defects and updates.

40) **A critical conclusion of the AI system** is any result of data processing generated by the AI system that can significantly affect the procedural rights and freedoms or the scope of legal responsibility of a person (in particular, risk assessment, proposal of a measure of restraint, analysis of evidence).

41) **Logging means an automated and secure record** of the use of the AI system in a case (identifier and model version, call time, nature of the task, a brief description of the input/request, information about considering/rejecting a human recommendation).

42) **Methodological documentation of the AI system** is a set of documents that disclose the purpose, architecture, training and validation data, limitations, known risks, test and audit results, including "model passport" and "data set passports".

43) **The Register of High-Risk AI Systems** is a state information system for accounting for AI systems approved for use in the field of justice, with information about their version, certification, audit results, and incidents.

44) **An AI security incident** is any event that indicates a compromise of the integrity, confidentiality, or correctness of the AI system (including "hint" attacks, "jailbreaks", uncontrolled updates, data leaks), which requires mandatory notification and response.

45) **Reproducibility of results** is the provision of the ability to re-obtain the relevant result of the AI system in the same case under the conditions of fixing the version, parameters and data sources.

46) **"Data room"** is a special regime of controlled access of the court, parties and experts to confidential methodological information about the AI system with mandatory logging of all actions and with a ban on disclosure or use of such information beyond the purposes specified by law.

47) **Human rights impact assessment** is a documented procedure for identifying, measuring and minimizing risks to human rights and freedoms from the use of AI in justice.

48) **Educational subjectivity** is the ability of a pupil, student, student, teacher and other participants in the educational process to independently make decisions on learning, development and assessment, preservation of academic dignity and autonomy in the context of the use of digital technologies.

49) **The AI system in education (hereinafter referred to as the AI system)** is a technological tool that carries out automated processing of educational data and generates recommendations, hints or other results to support teaching, learning, assessment or administration, without replacing the decisive role of the teacher.

50) **Pedagogical sovereignty** is the exclusive competence of the teacher enshrined in law to determine the methodology, scope and methods of assessment, interpret learning outcomes and be responsible for final decisions.

51) **A person in the loop (science: human-in/on-the-loop)** is a mode in which the teacher/examiner retains control over the critical stages of learning and assessment, has the means to review, adjust and reject algorithmic results.

52) **Adaptive learning** is the individualization of the content and pace of learning based on algorithmic analysis of educational data, provided that pedagogical sovereignty and procedural fairness are preserved.

53) **Social scoring** is an automated assignment of an integral assessment of "academic performance", "reliability" or "riskiness" to a person to influence access to educational services, scholarships, dormitories or other rights.

54) **Emotional recognition** is the use of biometric or behavioural signals to infer the emotional states, motivation or personality traits of the student, including micro expressions, gaze, heart rate, etc.

55) **Academic analytics is** the systematic collection and analysis of educational data to support learning and manage the quality of education in compliance with the requirements of minimization, proportionality and safety.

56) **Educational profile** is a set of personal and educational data on educational achievements, learning style, interests and other characteristics of an applicant, which is formed, processed and stored exclusively for educational purposes and cannot be used for purposes other than provided for by this Law, without a separate legal basis determined by law or the consent of the data subject.

57) **Inclusive design** is the design of AI systems considering the accessibility and needs of people with disabilities and other vulnerable groups.

58) **AI operator** is an individual or legal entity, public authority or other entity defined by law, which uses, manages, integrates or otherwise operates the AI system in accordance with the specified purposes of its application and is responsible for compliance with the requirements of this Law.

59) **AI Provider** — a natural or legal person, public authority, or other entity that creates, distributes, provides, implements, or maintains an AI system and is responsible for its compliance with the requirements of this Law, including its characteristics, security, quality, and updating.

60) **Child Data** means any information about the child, including device and network identifiers, educational, medical, biometric, location, social data, and derived inferences and profiles created by AI systems.

61) **Processing of children's data** is any operation or a set of operations with children's data (collection, storage, use, transfer, analysis, profiling, depersonalization, deletion, etc., or other actions with such data), which is subject to enhanced protection regardless of the means and place of implementation.

62) **AI child profiling** is any form of AI processing of a child's data, which consists in using such data to assess personal aspects or predict the behaviour, inclinations, academic performance or social characteristics of a child.

63) **Emotional tracking** is the AI identification, analysis, or prediction of a child's emotional states by analysing the face, voice, gaze, postures, physiological signals, or other signs, including derived features.

64) **A predictive profile ("future profile")** is a set of inferences or scoring scores generated by the AI system to predict future decisions, educational or career trajectories, risks, preferences, or "life paths" of a child.

65) **Targeted ads for children** are ads that are personalized based on the child's data (including psychotype, neurobehavioral signs, predicted emotional states), as well as data collected from third parties or through cross-platform tracking.

66) **Child information invisibility** is a mode of using the service without creating personal profiles, without cross-platform tracking, and with minimal short-term logs that do not allow you to restore the child's identity.

67) **Child Impact Assessment** is a documented ex ante procedure for identifying and minimizing risks to the rights and freedoms of a child from the functions of the AI system, indicating control measures, review periods, and responsible persons.

68) **Child Safety Data Sheet** is a public document of the operator containing a description of the functions of the AI system, data categories, risks, prohibited practices, technical and organizational controls, storage periods, incident procedures, and contact of the DPO/responsible person.

69) **Age assurance** is a method of establishing the age of a user using the principles of data minimization and prohibition of storing excessive information.

70) **An emergency kill switch** is a technical mechanism for immediately stopping certain functions of the AI system in the event of significant harm to the child.

71) **A simulacrum (face simulation)** is a representation of an individual's identity (image, voice, manner of speech, style of thinking/behaviour, "digital twin", avatar, memorial bot) created or modified with the help of an AI system, which can be perceived as related to a specific person or his/her "will".

72) **Speech/behavioural/communicative patterns** are patterns of speech, intonation, gestures, communication style, reactions, preferences identified or synthesized with the help of the AI system, which allow you to reproduce characteristic features of a person.

73) **Simulacrum monetization** is any economic benefit (direct payments, advertising, paid licenses, sale of access/API, donations, revenue share) from creating, distributing, or using a simulacrum or personality patterns.

3.2. All other terms not defined in this article shall be construed in accordance with the provisions of:

a) the legislation of the State in the field of information, personal data protection, cybersecurity, digital transformation, national security and international law;

b) international regulations, standards and recommendations, in particular NIST AI Risk Management Framework (USA) [59], *OECD* Council Recommendation on Artificial Intelligence (2019), Regulation (EU) 2024/1689 — AI Act, [60], UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) [61], AI Bill of Rights (USA, 2022), ISO/IEC 22989:2022 Artificial Intelligence — Concepts and terminology [62] and other officially recognized documents;

c) official interpretations, clarifications or doctrinal approaches or practices provided by authorized bodies or ethics councils in the field of AI at the national or international levels.

In cases where there is no common understanding of the term, preference shall be given to the interpretation that best ensures the highest level of protection of human dignity, security, digital sovereignty and sustainable development, in accordance with the principles laid down in this Law.

3.3. The interpretation of the terms used in this Law is carried out in accordance with the principle of maximum protection of human rights and freedoms, human dignity, non-discrimination, digital sovereignty and national security of the State, as well as on the principles of transparency, ethical responsibility, technological neutrality, interoperability and social justice.

In case of discrepancies in the interpretation or ambiguity of the term, its meaning is determined taking into account the priority of protecting human rights, human dignity and security of a democratic society, the prevention of abuse or covert influence on decision-making, as well as taking into account the best international practices of digital governance, recommendations of UNESCO, OECD, the Council of Europe and the United Nations and other internationally recognized documents in the field of artificial intelligence.

CHAPTER 4. PRINCIPLES OF ARTIFICIAL INTELLIGENCE REGULATION

The regulation of artificial intelligence systems in the State is based on the principles of the rule of law, which ensure a balance between technological development, human rights protection, security, ethics and digital sovereignty. In the event of a conflict between innovative development and human rights, human rights shall prevail.

The principle of technological neutrality means that legal regulation focuses not on technologies, but on the functional impact of AI systems. The law should not create advantages for individual platforms or architectures, promoting fair competition and equal market access.

The human-centered principle assumes that AI systems must function with human dignity and autonomy in mind. Individuals should be able to know about the impact of AI, appeal its decision, get an explanation, and maintain control. Full automation of critical decisions without human intervention is prohibited.

The principle of transparency and explainability obliges to provide access to technical documentation, logs, justifications, and application limits. All AI systems should inform the user about their algorithmic nature, and the results should be available for auditing.

The principle of responsibility establishes personal legal, administrative, ethical, and civil responsibility of all participants in the AI life cycle. This includes preventing damage, identifying responsible links, liability agreements, and indemnifying damages.

The principle of ethics implies compliance with moral standards, the prohibition of manipulative practices, compliance with international codes, the presence of ethical expertise and influence on vulnerable groups. AI systems should not only avoid harm but also promote well-being and trust.

The principle of preventive risk management requires threat assessment prior to the implementation of AI, monitoring during operation, auditing, scenario modelling, and prevention of negative consequences.

The principle of institutional accountability provides for the creation of independent supervisory and ethical bodies, the introduction of internal control systems, mandatory reporting and international cooperation in the field of digital governance.

The principle of digital sovereignty guarantees state control over critical digital resources, infrastructure, data, and algorithms. Data localization, independence from foreign platforms, and restrictions on the transfer of AI components abroad are required.

The principle of adaptability means the flexibility of the regulatory system — updating legislation, creating "sandboxes", delegating functions to councils with public participation, using artificial intelligence tools in rulemaking.

The principle of interoperability and harmonization ensures consistency with international standards (AI Act, NIST, ISO, UNESCO, IEEE), includes compatibility of audit, licensing, certification procedures and the creation of conformity recognition mechanisms.

The principle of digital diplomacy enshrines the regulation of AI as a foreign policy tool. The state participates in the formation of global standards, protects the interests of citizens, concludes international agreements, and prepares diplomats and digital experts for negotiations on AI issues.

4.1 The principle of technological neutrality: the legal regulation of artificial intelligence systems should not be based on specific tools, software solutions, architectures or types of models (e.g. neural networks, transformers, statistical models, etc.), but solely on their functional impact, context of application, level of autonomy, potential risks to human rights, security, the environment and democratic governance [63, 64].

This principle ensures that regulatory requirements must be adaptive to technological developments, ensuring regulatory fairness for emerging solutions that may arise outside of established technical categories [65, 66].

Technological neutrality also means that the law should not favour individual vendors, platforms, or architectures, promoting innovative competition and adherence to the principle of equal market access regardless of the technological origin of the AI solution [67].

4.2 The principle of "human-centred principle": the development, implementation, use and termination of artificial intelligence systems should be carried out with the unconditional priority of human dignity, fundamental human rights and freedoms, protection of autonomy of will and personal integrity [68].

This principle assumes that a person remains a key subject of legal relations, who has a guaranteed right to know about the functioning of the AI system, the impact of its decisions on himself or his environment, as well as to have really legal and technical means to control, review, explain, appeal or terminate algorithmic results [69].

The implementation of this principle requires:

- availability of human-in-the-loop mechanisms for high-risk systems [70];
- the introduction of independent ethics panels, digital ombudsmen, and specialized hotlines to address complaints related to the operation of AI systems [71];

ensuring the availability and effectiveness of administrative and judicial procedures for reviewing, explaining and appealing automated decisions by persons affected by them;

- mandatory consideration of social vulnerability, age, gender, linguistic, cultural and physical characteristics of persons interacting with AI systems [72];

imperative legislative prohibition on the full delegation of critical management, justice, security, healthcare, or oversight functions to AI systems without the participation of an authorized human person.

The principle of transparency and explainability: developers, operators, supervisors and end users should have a legally guaranteed opportunity to understand the logic of the functioning of the AI system, its goals, input and output parameters, data sources, decision-making models, training structure, risks and limits of applicability.

This principle provides for ensuring the availability of the following elements:

- technical documentation (model architecture, sources of training data);
- logs of decisions made and interpretations of algorithms (explainability);
- models of justification of results (for example, contextual hints for generative models);
- defined limits of the model's functioning, including scenarios in which its use is prohibited or restricted.

Transparency also means that any automated system that interacts with citizens must have a mandatory label or interface that communicates its algorithmic nature, and the results of decision-making must be available for verification, review and audit by third parties.

Principle of responsibility: each operator, developer, implementer, or user of an AI system bears personal legal, administrative, ethical, technological, and civil responsibility for the functioning of the AI system, its results, impact on the environment, individuals, and institutions, within limits proportional to the level of autonomy, risk class, scope, and access to the internal logic or configuration of the system.

This principle includes:

responsibility for preventing the adoption of knowingly harmful or discriminatory decisions, even if they are generated algorithmically;

implementation of responsible AI governance mechanisms;

a clear legal definition of roles and responsibilities in the chain of responsibility: developer — supplier — operator — end user;

ensuring the availability of internal liability policies, agreements between the parties and a system of regular internal audits;

application of the principle of preventive responsibility in the public sector and in high-risk areas of activity.

In case of damage to an individual or legal entity, the environment, society or violation of fundamental rights, the responsible parties are obliged to compensate for the losses in full, conduct a public audit and ensure the update, modification or complete decommissioning of the AI system that caused the negative impact.

The principle of ethics: all stages of development, implementation, operation and decommissioning of artificial intelligence systems must comply with fundamental moral and ethical norms based on respect for human dignity, solidarity, social justice, non-discrimination, inclusion and sustainable development.

This principle implies:

a) prohibition of the use of AI systems for systematic mass surveillance, manipulation of citizens' behaviour, digital segregation, censorship in the digital environment, psychological pressure or human rights violations;

b) ensuring compliance of AI systems with international codes of ethics and recommendations, in particular: Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021), IEEE *Ethically Aligned Design*, CAHAI (Council of Europe), AI4People;

c) mandatory implementation of independent ethical expertise, which is mandatory for high-risk and generative models;

d) creation of ethical audit bodies under state authorities, business structures and scientific institutions;

e) conducting a preliminary *Human Rights Impact Assessment* for systems used in the public sphere and against vulnerable groups.

The ethics of AI means not only the absence of abuse or harm, but also the presence of an active ethical focus: promoting public welfare, increasing transparency, strengthening trust in digital systems, and preventing social risks of the future.

4.3 The principle of preventive risk management. All AI systems, regardless of risk class or scope, are subject to a comprehensive assessment of potential threats and negative consequences before they are implemented in the application environment, as well as continuous or periodic monitoring after commissioning [73,74,75].

This principle includes:

− development and implementation of internal risk management systems at all stages of the AI system life cycle;

− conducting mandatory ethical due diligence prior to implementation, taking into account the impact on human rights, ecosystem, social relations and public trust;

− independent technical and ethical audit of high-risk systems, generative models and systems of critical importance after commissioning;

− publication of a public report on the results of risk management and identified threat vectors;

− providing continuous monitoring of the behaviour of the AI system in real time in order to early detect unintended effects, algorithmic distortions or technical failures.

The purpose of the preventive approach is not only to respond to negative events, but also to actively anticipate them, simulate possible threat scenarios and implement preventive mechanisms that minimize damage or make it impossible for it to occur.

The principle of institutional accountability. A multi-level system of supervision, control and reporting on the creation, implementation, use and decommissioning of artificial intelligence systems is provided, with special attention to high-risk, generative and critical infrastructure applications.

This principle requires:

establish independent oversight bodies, ethics committees, and institutions empowered to review AI systems in the public and private sectors, as well as apply the necessary response measures;

implementation of internal accountability mechanisms in organizations that develop or use AI systems, with the mandatory appointment of Chief AI Officers;

providing the public with open access to information on the use of AI systems in the areas of public administration, justice, education, health care and security;

mandatory reporting of all negative incidents related to malfunctions, manifestations of discrimination, making unexplained decisions or failures of AI systems;

development of international cooperation with the oversight mechanisms of the Council of Europe, OECD, UNESCO and other intergovernmental organizations that form the principles of global ethical digital governance.

Accountability is not a procedural formality, but a key democratic principle that guarantees the controlled nature of the impact of AI systems, provides an opportunity for its verification and public review.

4.4 The principle of digital sovereignty. The State is obliged to ensure full control and strategic management of critical digital assets, infrastructure, data, algorithms, computing power, cloud environments and software and hardware architecture cause or may cause a direct or indirect effect on natural or indirect used or integrated into the functioning of artificial intelligence systems on the territory of the State or in the interests of its citizens [76,77,78].

This principle is ensured by:

prohibiting the dependence of critical solutions based on artificial intelligence on foreign uncontrolled platforms, data centres, APIs or systems located outside the jurisdiction of the State or its strategic partners;

localization of data necessary for training, validation and operation of AI systems serving state, security, medical, educational and other critical sectors on the territory of the State or in safe environments under its jurisdiction;

creation and maintenance of national computing power and open models capable of ensuring digital independence in the fields of generative AI, natural language processing, computer vision and other key areas;

legislative restriction of the transfer of algorithms, models and training data to third countries without special permission from the competent authorities, taking into account national security, ethical control and citizens' rights;

strategic participation of the State in international initiatives to form digital sovereignty as a basis for democratic governance by artificial intelligence at the global level.

Digital sovereignty is the basis for preserving democratic governance by artificial intelligence and protecting the interests of citizens in the global digital environment.

4.5 The principle of adaptability. The national system of legal regulation in the field of artificial intelligence should be flexible, dynamic and capable of evolutionary updating in order to quickly respond to the emergence of new technologies, challenges, interaction models and social expectations.

This principle includes:

a) regular updating of legislation through digital consultation mechanisms, forecasting technological developments (foresight), modelling of future scenarios and risk analytics;

b) creation of regulatory sandboxes for testing innovative AI models in controlled legal environments with temporary status, feedback and support of expert commissions;

c) delegating, within the limits determined by law, the right to adapt the updating of by-laws to ethics and regulatory councils functioning with the participation of the public, academics, business and civil society institutions;

d) application of technology transparency tools, artificial intelligence in rule-making, and expert digital dashboards for accelerated policy updates.

Adaptability is a guarantee of the resilience and viability of the regulatory system in the face of digital transformations, ensuring a balance between innovative development and the protection of human rights in a changing algorithmic reality.

4.6 The principle of interoperability and harmonization means that AI law should be built taking into account international standards, principles of mutual recognition, compatibility of platforms, models and regulatory structures. The system of legal regulation of artificial intelligence should be aligned with key international and cross-border legal acts, technical standards, security protocols, as well as the principles of mutual recognition, openness and interoperability of digital infrastructures in order to ensure the integration of the State into global digital ecosystems [79,80].

This principle includes:

− taking into account and implementing EU norms (AI Act), USA (NIST AI RMF, Executive Orders), UNESCO, OECD, ISO/IEC, IEEE and other influential international frameworks;

− ensuring compatibility between national registers, audits and licensing procedures, and those of foreign AI certification systems;

− implementation of technical interoperability standards that define data formats, training protocols, algorithmic compliance markers;

− creation of internationally compatible mechanisms for inspections, exchange of ethical solutions and mutual recognition of compliance with high-risk systems;

− providing guarantees for the participation of national developers, scientists and government agencies in the processes of developing global standards and policies in the field of AI.

Interoperability and harmonization in the field of AI encompass not only the technical compatibility of systems, protocols and standards, but also the legal, institutional and ethical integration of the State into the international digital community of democratic countries. This approach ensures not only technological interaction, but also the commonality of values, procedures and guarantees for the protection of human rights in global digital ecosystems[81].

4.7 The principle of digital diplomacy: The state recognizes the regulation of artificial intelligence as a key element of its foreign policy, a tool for promoting digital sovereignty, supporting global ethical governance, and protecting human rights in the digital age[82,83,84].

This principle requires:

− active participation in the formation of international norms, standards and framework conventions on AI issues within the framework of the UN, the G7, the Council of Europe, UNESCO, OECD, ISO, ITU and other global institutions;

− concluding bilateral and multilateral treaties, memorandums and strategic alliances in the field of ethics, security, interoperability and research in the field of AI;

− protection of the interests of citizens of the State in case of violation or threat of violation of their rights, freedoms or data by foreign artificial intelligence systems through digital consulting, the use of diplomatic mechanisms and representation in global digital structures;

− training of diplomatic personnel and digital experts for the representation of the State in international negotiations and political coordination in the field of AI regulation.

Digital diplomacy is a continuation of technological sovereignty by means of international law and a new paradigm of global cooperation in the era of algorithmic states[85, 86, 87].

This principle includes:

− taking into account the norms of the EU (AI Act), USA (NIST AI RMF, Executive Orders), UNESCO, OECD, ISO/IEC, IEEE and other influential international frameworks;

− ensuring compatibility between national registers, audits, licensing procedures and foreign AI certification systems;

− implementation of technical interoperability standards for data formats, training protocols and algorithmic compliance tokens;

− creation of internationally compatible mechanisms for inspections, exchange of ethical solutions and mutual recognition of compliance with high-risk systems;

guaranteeing the participation of national developers and government agencies in the development of global standards and policies in the field of artificial intelligence.

CHAPTER 5. CLASSIFICATION OF AI SYSTEMS BY RISK LEVEL

For regulatory and supervisory purposes, AI systems are classified into four risk categories. These categories define the scope of requirements for security, transparency, auditing, ethics and oversight.

Prohibited AI systems are systems that qualify as posing an unacceptable level of risk and are not allowed to be used. These include: systems that violate human rights or national security, create digital forms of discrimination, carry out uncontrolled biometric surveillance, or operate in autonomous decision-making mode with physical influence without human intervention. Their development, implementation or use is prohibited and subject to legal liability.

High-risk systems require a special legal regime. These are systems that directly affect the implementation of basic human rights and access to socially significant services, the functioning of critical infrastructure, public security, justice and defence. Such systems are subject to mandatory registration, ethical and technical audits, annual reporting, implementation of a human-in-the-loop regime and the appointment of a responsible officer.

Limited risk systems have an indirect impact on the rights of individuals or the information environment. This category includes marketing, communication, human resources (HR) and other service systems. For their use, mandatory labelling, informing users, providing an opportunity for opt-out, conducting a basic audit and adherence to the principles of transparency and non-discrimination are provided. Prior authorization for the use of such systems is not required.

Minimal risk systems are artificial intelligence systems that perform support functions, do not make decisions about users, and do not significantly affect security or human rights. These are spam filters, autocomplete, translation systems, basic voice interfaces. Such systems are regulated by general ethical standards, do not require special permits, but in case of their mass application, they are subject to a declaration of conformity.

The classification of the risk level of AI systems is carried out comprehensively, considering: sources and quality of training data; level of autonomy; areas of application; possibilities of human intervention; potential unpredictable effects. At the same time, the principle of foresight and proportionality between risk and public benefit is applied.

Risk categories and relevant requirements may be updated by the authorized public authority in the field of AI, taking into account international standards (AI Act, NIST AI RMF, ISO/IEC 23894:2023). At the same time, legal, technical and ethical harmonization with global standards and certification procedures is ensured, including the implementation of interoperability protocols and the creation of mechanisms for mutual recognition of conformity.

5.1 For regulatory and supervisory purposes, AI systems are classified into four risk categories. Each of them defines the scope of requirements for security, transparency, auditing, ethics and oversight.

Risk categories of AI systems:

1. Prohibited AI systems (unacceptable risk AI) are a category of artificial intelligence systems whose use is prohibited in any form. Such systems are considered unacceptable because they contradict the fundamental values of a democratic society, violate human rights, or pose an excessive threat to the social order, ethical norms, and national security. This category includes [88,89]:

systems that carry out algorithmic tracking, monitoring or evaluation of people's behaviour with subsequent decision-making that affect the rights, status or opportunities of a person (for example, social ratings);

models that use subconscious, emotional, or neurobehavioral manipulative techniques to influence user choices without their consent or awareness of such influence;

technologies used for mass biometric surveillance in real time in the public space without an explicit legitimate purpose, judicial control or clear proportionality criteria;

any AI systems designed or capable of autonomous physical impact (for example, autonomous weapons), which does not guarantee human participation in critical decisions;

systems that create conditions of digital discrimination on the grounds of race, age, gender, language, religion, disability, national origin, political beliefs or social status, without transparent compensatory mechanisms or the right to review decisions.

The use of such systems on the territory of the State is prohibited in any form by both the State and the private sector. Their implementation, distribution or testing is subject to criminal, administrative or civil liability in accordance with the law.

High-risk systems are a category of artificial intelligence systems that requires enhanced regulatory, ethical, and technical controls. Such systems have a direct and significant impact on the rights, security, well-being or life opportunities of citizens, as well as on the functioning of the state, national security and critical infrastructure [90, 91, 92]. This category includes:

systems used to make or prepare decisions that directly affect an individual's access to education, healthcare, employment, social security, justice, housing rights or freedom of movement (including systems for predicting recidivism, assessing creditworthiness, assigning social benefits);

technologies used in the management, monitoring or operation of critical infrastructure (energy, transport, communications, digital platforms, banking system, healthcare), where a malfunction or manipulation can cause mass failures, man-made disasters or serious violations of public order;

systems integrated into public administration, public security, military structures, law enforcement, digital security, surveillance or allocation of resources, which are capable of shaping or changing the legal status of a person without proper judicial or institutional control;

systems that autonomously process personal data, carry out biometric recognition or mass profiling on a scale that creates a risk of discrimination, manipulation or abuse.

Such systems are subject to:

- mandatory registration in the National Register of High-Risk AI;
- independent technical and ethical audit before commissioning;
- annual reporting of incidents, failures or complaints;
- mandatory human-in-the-loop mode for critical decisions;
- appointing a responsible digital officer (Chief AI Officer) with supervisory powers.

State bodies that implement artificial intelligence systems are obliged to ensure the maximum possible transparency of procedures, the involvement of independent experts in assessment and control processes, as well as public access to information on algorithmic logic and justification of decisions.

Limited risk systems are used in the field of business, communications and services and do not have a direct impact on the fundamental rights or security of an individual. At the same time, such systems can influence user behaviour, consumer choice, digital reputation, or the information environment. They are widely used in marketing, e-commerce, customer service, media, entertainment, communications, and human resources management. The use of limited risk systems is allowed provided that: informing users about their algorithmic nature; providing users with the opportunity to opt out of interaction; conducting a simplified compliance audit; providing ethical support.

Such systems include:

- chatbot systems, chatbots, virtual assistants, digital interfaces that automate communication;
- consumer behaviour assessment systems, personalized recommendation algorithms and AI-based marketing platforms;
- digital assistants (managers) who facilitate household or organizational decision-making (scheduling planning, communication, navigation, booking);
- HR tools used at the stage of pre-selection, emotion analysis, resume selection without making final decisions on employment.

Regulation of these systems includes:

mandatory labelling of the system indicating its algorithmic (AI) nature;

informing the user about the possible influence of the system on his choices, decisions or behaviour;

providing users with an accessible opt-out function or an alternative, non-algorithmic interface;

conducting periodic simplified audits with an assessment of the risks of information distortion, covert manipulation or discrimination;

mandatory compliance with the principles of transparency, explainability, non-discrimination and responsible use.

Limited risk systems are not subject to prior administrative authorization (licensing), but must meet established standards of ethical responsibility and public trust requirements.

2. Minimal risk systems: AI tools with little or no impact on rights, health, security, such as spam filters, recommendation systems, or translation. Subject to a general standard of ethics without mandatory audits[93].

Minimal risk systems: These are artificial intelligent systems that have little or no impact on the rights, health, safety or well-being of individuals, and do not pose significant risks to the environment, democratic processes or social stability. This category includes systems that predominantly perform technical, supportive, service, or household functions without autonomous user decision-making[94].

Examples of such systems are:

- spam filters in emails;
- automatic translation tools without deep semantic analysis;
- spelling and style correction programs;
- algorithms for pre-sorting emails or requests;
- basic voice interfaces for household devices;
- autofill systems in browsers or instant messengers;
- game or educational programs without a personalized assessment function.

Regulation of minimal risk systems includes:

- adherence to general ethical standards, prevention of misinformation, covert data collection or discriminatory behaviour;
- does not require mandatory audits, but encourages voluntary compliance checks with AI ethics best practices;
- in cases of large-scale applications (more than one million users);
- submission of a voluntary declaration of conformity to the AI Ethics Council.

These systems are not subject to registration or authorization, but manufacturers and operators are obliged to ensure a prompt response in the event of unexpected impacts, failures or complaints that may lead to a revision of the risk classification.

5.2 The classification of the risk level of an artificial intelligence system is carried out comprehensively on the basis of a set of qualitative and quantitative criteria that reflect the potential impact of such a system on the fundamental rights and freedoms of man and citizen, public security, national interests, critical infrastructure and the environment [95, 96].

The main factors of assessment include:

Source and quality of training data: volume and representativeness, presence or absence of biases; transparency of origin; legality of use (availability of licenses); the potential to reproduce discriminatory or manipulative practices.

The level of autonomy of the system: the degree of independence from a person in decision-making; availability of mechanisms for controlling, confirming or cancelling AI actions; the possibility of checking the logic of the system's functioning by authorized entities; degree of explainability.

Sphere of influence and context of application: level of sensitivity of the sphere (medicine, education, justice, national defence, security); scale of application (local, national, cross-border); type of interaction with people (passive, active, autonomous).

Mechanisms of human intervention: availability of procedures of preliminary (ex-ante), current (in-process) and subsequent (ex-post) control; the reality and effectiveness of human influence on the course of decision-making; technical possibility of suspending or modifying AI actions in critical situations.

Likelihood of unintended effects or discriminatory behaviour: the ability of the system to create new risks or threats due to complexity, opacity, or interaction with other digital systems; the ability to produce new types of vulnerability, in particular vulnerabilities for certain social groups.

These criteria are applied considering the principle of foresight ("precautionary principle") and the principle of proportionality, which determines the ratio of the scale of risk and public benefit.

5.3 The list of categories can be updated by the public authority responsible for the field of AI upon the submission of the National Ethics Council on AI based on international framework acts.

The principle of interoperability and harmonization means not only technical compatibility, but also the legal and ethical integration of the State into the ecosystem of digital democracies.

This principle includes:

considering the norms of the EU (AI Act), USA (NIST AI RMF, Executive Orders), UNESCO, OECD, ISO/IEC, IEEE and other influential international regulations;

ensuring compatibility between national registers, audits, licensing procedures and foreign AI certification systems;

implementation of technical interoperability standards for data formats, training protocols, algorithmic compliance tokens;

creation of internationally compatible mechanisms for inspections, exchange of ethical decisions, mutual recognition of compliance with high-risk systems;

guaranteeing the participation of national developers and government agencies in the development of global standards and policies for artificial intelligence.

Interoperability and harmonization are not only a matter of technical compatibility, but also a process of legal and ethical integration into the ecosystem of digital democracies.

CHAPTER 6. GENERATIVE AI SYSTEMS: SPECIAL REGULATION MODE

Generative AI systems constitute a separate category of high-performance models capable of independently generating digital content (texts, images, sound, video, code, 3D objects, etc.) based on input data without direct human intervention. Such systems are characterized by autonomy, variability and scalability. They can function both in open and closed modes and are accompanied by ethical, legal and social risks due to a significant impact on the information space, copyright and public opinion, as well as through the transformation of the ontological boundaries of creativity.

Generative AI systems are subject to a special regime of legal regulation, which provides: mandatory state registration in the National Register of Generative AI; mandatory labelling of created content as AI-generated; the use of technical means of verification (watermarks, crypto tags, tracing methods, etc.); prohibiting the creation of manipulative, discriminatory or false content without the consent of the person or in violation of the public interest; supervision of state, public and professional bodies over the use of generative AI in sensitive areas; periodic certification, ethical reporting and implementation of responsible use policies.

It is prohibited to use generative artificial intelligence systems to create fake news, deepfake content and simulated images; political campaigning without the appropriate permit (license); distribution of pseudoscientific materials; generation of discriminatory content; automated disinformation; Creating content that incites violence or hostility.

Developers, providers and operators of generative systems are obliged to provide transparent mechanisms for controlling the results (audit trail, automatic labelling, alarm); implement internal editing and blocking policies; organize independent ethical verification; report to the authorized bodies; to carry out copyright screening and ensure the protection of personal non-property rights to images and author's style.

Generation systems operating in the State or processing data of citizens of the State are subject to mandatory audits by independent accredited structures. The audit includes: assessment of manipulative impact; identification of biases; analysis of the context of use; Verification of compliance with privacy, copyright and transparency requirements. If risks are identified, the auditor has the right to request blocking or modification of the system.

The use of generative AI in education, science, culture and creativity is allowed only if: compliance with academic integrity, copyright and social ethics; mandatory disclosure of the degree of AI participation; prohibition of use for dissertations without indication of participation AI; blocking content that reproduces the unique style of the author without his consent; consolidation of the legal responsibility of the human author or curatorial institution; implementation of codes of ethics and training programs for the use of generative AI.

The state body responsible for the field of artificial intelligence approves a separate regulation on generative systems, which includes national and international standards (AI Act, NIST, UNESCO, ISO/IEC, IEEE, etc.); registration and audit procedures; mechanisms for the exchange of ethical decisions; rules for participation in the development of global policies in the field of artificial intelligence.

6.1 Generative AI systems are a category of high-performance AI models capable of creating new, original digital content based on incoming textual, audiovisual or structured data. Such content includes, but is not limited to, text documents, graphics, soundtracks, video files, 3D models, code, virtual objects, and other digital products [97, 98, 99].

Generation is carried out without direct human intervention at the stage of creating the result, and the process itself is characterized by the following features:

autonomy (the ability of the system to independently decide on the structure and content of the content created);

variability (the ability to create unique results for each run based on the same inputs);

scalability (the ability to process large amounts of data and generate significant amounts of content in a short period of time);

adaptability (ability to modify results depending on the context or conditions of application).

Such systems can operate both in open mode (interaction with public users) and in closed mode (corporate or government use) and are accompanied by potential ethical, legal and societal risks due to their high level of influence on the information space, public opinion, aesthetic parameters of the digital environment, copyright and ontological boundaries of creativity.

Generative AI systems are subject to special legal supervision due to their ability to reproduce or change the cognitive and behavioural patterns of society, structure media flows, create the risk of substitution of authorship, form the algorithmic reputation of individuals, or carry out large-scale digital simulations of human interaction.

A special legal regime is applied to generative systems, which establishes a set of legal, ethical and technical requirements aimed at ensuring transparency, accountability and safe functioning of such systems [100,101]. This mode includes:

mandatory state registration of all generative systems operating on the territory of the State or processing citizens' data in the National Register of Generative AI Systems. Registration is a prerequisite for their legitimate use in the commercial, educational, governmental or media sphere;

mandatory labelling of each generation result using an explicit visual and/or technical identifier that signals that the content was created with the participation of AI. This approach avoids mixing generative and human content, especially in sensitive areas — journalism, science, education;

the use of mandatory technical means of verification, in particular digital watermarks, cryptographic tags, data trace history and metadata, which record the generation process and exclude forgery. It provides independent auditing, copyright protection and prevention of disinformation;

Imposing restrictions on the creation of content that contains potentially harmful, discriminatory, manipulative, or intentionally false information. It is prohibited to automatically generate images of persons without their consent, as well as to create simulations of political figures or experts in contexts that may mislead or form false public opinion;

implementation of state, public and professional supervision over the use of generative systems in highly sensitive sectors — education, politics, law, healthcare, public administration, where such tools can influence consciousness, choices, behaviour or access to services. Such oversight includes periodic certification of systems, the participation of ethics boards, reporting obligations of providers, and the implementation of responsible use policies.

6.2 The use of generative AI systems for purposes that violate public order, human rights, ethical norms and the principle of legal certainty is prohibited. In particular, it is not allowed [102,103]:

- creating and distributing fake news, deepfake content, simulated visual or audio images that can mislead the public, especially if such content is not clearly labelled and is not accompanied by explicit disclosure of how it was created;
- the use of generative systems in political campaigning, election campaigns, information and psychological operations without obtaining a special state license, without transparency, neutrality and independent monitoring procedures. Any use of AI to influence the will of citizens is subject to careful state control and mandatory public disclosure;
- generation of pseudoscientific information, forged or automatically generated scientific articles, expert opinions, reviews, diplomas, certificates or other documents of legal or academic significance, as such actions undermine academic integrity, legal stability and trust in institutions;
- creating images of real or fictional persons in a pornographic or discriminatory context, as well as in forms that humiliate or insult human dignity, without their explicit consent;
- using generative AI systems to mass-produce disinformation, including fake experts, fake news platforms, or imitation of authoritative sources;
- automated creation of content capable of intensifying hatred, interethnic or interfaith conflicts, calls for violence or hatred contrary to the Constitution of the State and international human rights standards.

6.3 Developers, providers, and operators of generative AI systems are required to adhere to the principles of ethical responsibility, transparency, security, and legal protection at all stages of the life cycle of content creation, processing, and distribution.

In particular, they must:

provide transparent mechanisms for monitoring the results of generation, including algorithmic tracing (audit trail), automatic labelling of content and mechanisms for signalling vulnerable or harmful content, as well as technical means for immediate termination of generation in real time;

develop internal rules (policies) for editing, moderation, temporary blocking and final termination of the work of generative mechanisms in case of detection of violations of ethical or legal standards, subject to mandatory verification and confirmation of such actions by an independent ethics council;

ensure regular reporting to the authorized state body on the functioning of generative models, in particular on the number of generated content units, their topics, audience, distribution tools and methods of complaint processing. Reports should contain separate sections on cases of distortion of facts, influence on public opinion or probable conflicts with public policy norms;

implement comprehensive copyright verification and content filtering procedures that prevent copyright, image rights, commercial and personal property rights violations. Particular attention should be paid to the prohibition of automatic copying or reproduction of unique stylistic features of artists, journalists, experts without a license or a clear indication of the original source.

6.4 All large-scale content generation systems operating on the territory of the State or processing personal, behavioural or culturally significant data of citizens of the State are subject to mandatory preliminary and periodic audits for compliance with ethical, legal and social standards.

The audit is carried out by independent authorized bodies or accredited structures with the involvement of experts in the fields of ethics, human rights, information security, psychology of influence and cognitive sciences.

The mandatory elements of the audit are:

− assessment of the level of manipulative potential of content and its ability to shape or change public opinion;

− identification of systemic or structural biases in the initial data or generation algorithms;

− analysis of contexts for the use of content (political, cultural, economic) to prevent targeted information influences;

− verification of mechanisms for observing confidentiality, copyright and the right to privacy;

− establishing the level of transparency, explainability and controllability of systems in cases of complaints, incidents or public criticism.

In case of a high level of risk or potential harmful impact, the auditor has the right to demand the termination of the system, changes to its architecture or temporary blocking until the violations are eliminated.

6.5 The use of generative AI systems in the field of education, science, culture and creativity is carried out exclusively in compliance with the principles of academic ethics, copyright and social responsibility [104, 105].

Such use is allowed provided that:

mandatory indication of the participation or use of AI in the creation of each product, result of creativity or publication. Information on the degree of AI participation should be displayed in an open, understandable and unambiguous form;

compliance with standards of academic integrity, including a ban on the use of AI for the preparation of diploma, scientific, dissertation works without explicitly indicating the AI identifier, logs of the system, and confirmation of human-centric participation;

ensuring the preservation of copyrights of persons whose work is used as educational material for AI or whose stylistic features are reproduced. If a reproduction of the style of a particular author is detected without a license or consent, the relevant content is subject to blocking;

definition of collective intellectual responsibility: in the case of co-authorship between AI and a person, the individual author or the institution that initiated the generation remains legally responsible. In such cases, special licensing or legal consolidation of the status of responsible content curator is provided;

adherence to cultural and social ethics, which means the prohibited use of generative AI to create images that may offend national dignity, religious feelings, historical memory or certain social groups;

introduction of ethical codes for the use of generative AI in educational, scientific and artistic institutions, with mandatory training of staff and students on the ethical and legal application of such systems.

SECTION III. BASIC PRINCIPLES OF REGULATION AND ETHICS

CHAPTER 7. PRINCIPLES OF RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE

The use of artificial intelligence systems must be carried out in compliance with the principle of responsibility. The development and application of such technologies is possible only if ethical standards, security requirements, human rights and the interests of the state are observed. Entities that develop, implement or use AI systems are obliged to ensure compliance with legislative, ethical and technological requirements, transparency and explainability of algorithms, as well as to prevent the use of AI in a way that is contrary to the Constitution of the State, international law or public interest.

The state and authorized bodies ensure the functioning of mechanisms for preventing harm, mandatory involvement of a person in critical decision-making, respect for the dignity of the person, non-discrimination, protection of privacy, as well as compliance with the principles of technological neutrality, legal and technical interoperability and environmental responsibility.

The principle of responsible use is mandatory for all sectors in which AI affects or may affect fundamental human rights, national security, the digital economy, the information environment or democratic processes. Its violation entails the application of measures in the form of blocking the relevant system, revocation of certificates, cancellation of registration, initiation of legal proceedings and public reporting of incidents.

The state ensures the formation of a culture of responsible use of AI as a component of digital transformation by integrating ethical modules into the education system, conducting information campaigns, developing ethical standards and codes, introducing voluntary ethical certification, and creating support centres.

The AI Ethics Authority functions as an independent institution operating based on the law and accountable to the parliament. He is vested with the right to carry out inspections, conduct examinations, issue mandatory orders, initiate proceedings and coordinate the work of ethics councils in key areas of public life.

7.1. Responsible use of artificial intelligence systems is a fundamental condition for technological progress, which is carried out in compliance with human rights, fundamental freedoms, national interests and ethical norms.

This principle includes:

− full responsibility of developers, operators, users and owners of AI systems for compliance with the requirements of legislation, ethics, security and transparency at all stages of the life cycle of AI systems - from development and deployment to use;

− the obligation to ensure the explainability of algorithms, i.e. the possibility of explaining the logic, stages and criteria of decision-making by the AI system in a form accessible to a person, regardless of the level of his technical training; explainability should include a description of training data, internal decision-making models, weighting factors, as well as potential errors and limits of the algorithm;

− prohibiting the use of AI systems for purposes that violate the Constitution of the State, international obligations, human rights or the public interest;

− introducing harm prevention mechanisms, including monitoring, feedback, complaint procedures, as well as the possibility of suspending or modifying the system in cases of threats or violations;

− adherence to the principle of "human-centric", which provides for human participation in critical decisions, the right of an individual to appeal or review the results adopted with the participation of AI;

− the inadmissibility of discrimination or biased behaviour of AI systems against any social group on the basis of race, sex, age, language, disability, religion, political belief or social origin;

− protection of privacy, confidentiality, copyright, the right to one's own image and digital dignity of each person;

− adherence to the principles of technological neutrality and interoperability as the basis of competitiveness, security and international cooperation;

− considering the long-term environmental impacts of deploying large-scale AI systems, including energy consumption, carbon footprint, and disposal of digital infrastructures.

7.2. The principle of responsible use of AI systems is mandatory in all sectors where AI has or may have an impact on fundamental human rights, public and national security, ecosystem, information environment, digital economy, political stability, cultural heritage or democratic processes.

This principle is of particular importance in the context of large-scale digital influence, automated decision-making, personalized content, as well as in the use of AI in the areas of law enforcement, defence, healthcare, justice, social policy, electoral systems, and security initiatives.

Violation of the principle of responsible use of AI, through uncontrolled implementation, lack of human control mechanisms, ignoring ethical norms or creating systems designed to manipulate society or individuals, is grounds for:

− temporary or permanent blocking of the relevant system;

− cancellation of registration and revocation of certificates of conformity;

− initiation of administrative or civil proceedings against entities that manage AI systems;

− reporting incidents to the public and bringing officials to justice in case of systematic violation of ethical norms of regulation or undermining public trust.

7.3. The state ensures the formation of a sustainable culture of responsible use of artificial intelligence systems as one of the key areas of digital transformation of society. Such a culture is based on the principles of ethics, security, respect for human rights and freedoms, as well as the promotion of innovative development.

This culture is implemented by:

− integration of modules on AI ethics and law into educational programs at all levels — from general education institutions to training programs for civil servants, lawyers, developers, media specialists and technical specialists;

− implementation of national information and education campaigns aimed at increasing the digital literacy of the population and awareness of the opportunities and risks of using AI;

− development and implementation of national ethical standards, sectoral professional codes and intersectoral charters of ethics in the field of AI;

− introduction of a system of voluntary ethical certification of developers, digital platforms and technology companies;

− creation of specialized support centres on the responsible use of AI for citizens, businesses, educational and cultural institutions.

The AI Ethics Authority functions as an independent institution with advisory, control, and sanctioning powers. The Authorized Body for AI Ethics has the right to:

− carry out scheduled and unscheduled inspections of the use of AI systems;

− to conduct examinations on the activities of entities in the field of AI in accordance with the established ethical standards;

− issue binding orders;

− initiate administrative, civil or disciplinary proceedings in case of violation of the principles of ethics of artificial intelligence;

− to coordinate the creation and maintenance of ethics councils in the fields of education, science, defence, justice, health care and culture.

CHAPTER 8. PRINCIPLES OF TRANSPARENCY, EXPLAINABILITY AND OPENNESS OF ARTIFICIAL INTELLIGENCE SYSTEMS

Artificial intelligence systems that potentially affect human rights, freedoms, security, or well-being are subject to mandatory identification, traceability, explainability, and accountability. The use of such systems is allowed only if the principles of transparency, explainability and openness are observed.

Subjects of AI implementation or use are obliged to inform users about the use of AI, ensure its clear identification in the public environment, disclose technical documentation and sources of training data, create interfaces for verifying decisions, explain the logic of the system functioning in an accessible form, and indicate the degree of AI participation in decision-making.

Explainability is ensured through the implementation of mechanisms for reproducing algorithmic logic, decision logging, interpretation of results, detection of errors or bias, as well as by providing independent access to the results for ethical and legal auditors.

Openness in the field of artificial intelligence is ensured by regularly informing the public about policies and risks, publishing reports on the impact of AI systems in public spheres, providing open access to training materials, ethical documents, contract templates, codes of conduct, as well as databases with recorded incidents, errors or complaints within the limits that do not contradict the protection of security and intellectual property rights.

Violation of the principles of transparency, explainability or openness is grounds for the application of sanctions, including suspension of activities, imposition of fines, revocation of certification or license.

Compliance with these principles is monitored by the Authorized Body for AI Ethics and Security, which maintains an open register of systems that meet the established requirements and audits verification procedures in key sectors.

The state authority in the field of AI sets standards of transparency and explainability depending on the level of risk, including mandatory requirements for high-risk systems, generative models, autonomous systems used in the field of security, healthcare, justice and digital identification.

8.1 Principle transparency requires that all AI systems that have a potential impact on human rights, freedoms, safety, or well-being be designed, implemented, and used in a way that ensures their identification, traceability, explainability, and accountability.

8.2 Entities that develop or use AI systems are obliged to:

clearly and unambiguously notify users of the use of AI before the interaction begins. [106, 107];

ensure the identification of AI systems in public space [108] by using clear markings (markings, interface messages, audio or visual indicators) if the results of their activities can be perceived as human-created [109];

ensure the availability of technical documentation, description of the system architecture, sources of training data and interfaces for testing AI solutions within the limits that do not violate the protection of trade secrets and security [110];

explain the logic of the system functioning in a form understandable to non-professional users, using the standards of the accessible language [111];

indicate the degree and nature of AI's participation in the decision-making process (automatic, semi-automatic, or consultative mode) [112].

8.3 Explainability requires developers to:

create mechanisms that provide the ability to reproduce the logic of the algorithm, including decision logs, data processing processes and model parameters [113, 114];

ensure that the results are interpreted, allowing users and independent auditors to identify errors, biases or inconsistencies [115, 116];

create independent access interfaces for ethical and legal auditors to the results of the system in order to verify compliance, with protection against unauthorized interference and maintaining data confidentiality.

8.4 The principle of openness implies:

regularly informing the public about policies, technologies and risks associated with the use of AI, using open online resources and public reports [117];

mandatory publication of open reports on the impact of AI systems in the public sector (public services, digital governance, education, healthcare, security) [118];

ensuring open access to sources of learning, ethical guidelines, model contracts and codes of conduct that do not contain restricted information;

creation and maintenance of open databases for assessing AI errors, incidents or user feedback, in compliance with the requirements of personal data protection and state security [119].

8.5 Violation of the principles of transparency, explainability and openness is the basis for the application of regulatory sanctions, in particular: suspension of activities, imposition of fines, suspension of certification, and, in cases of systematic or gross violations, deprivation of license. Sanctions are applied in the manner prescribed by law, ensuring the right to administrative and judicial appeal.

8.6 The Authorized Body for AI Ethics and Security monitors compliance with these principles, maintains an open state register of transparent AI systems, and conducts an independent audit of verification procedures in key sectors of public and private governance.

8.7 The Authorized Body for Regulation in the Field of AI of the State develops and implements standards of transparency and explainability for various risk categories, including high-risk systems, generative AI, autonomous systems in the field of security, healthcare, justice and digital identification, harmonized with international standards.

CHAPTER 9. PRINCIPLES OF SAFETY AND RELIABILITY OF ARTIFICIAL INTELLIGENCE SYSTEMS

Security and reliability are prerequisites for the admission of AI systems to the market, implementation, use and scaling, especially in high-risk areas, in particular: healthcare, defence, justice and law enforcement, transport, energy, cybersecurity and public administration. Such systems are subject to state and independent control, certification, ethical examination and independent verification for compliance with safety, reliability and legality requirements.

The security of AI systems encompasses technical, organizational and procedural measures aimed at stable, predictable and controlled operation. Mandatory are: resistance to failures, hacker attacks, data processing errors; the availability of mechanisms for crashing, backup, and operator intervention; multi-level data protection; minimizing the risks of unsupervised learning; preliminary laboratory and field testing in critical scenarios with independent verification.

The reliability of AI systems means the ability to operate for a long time without significant failures, while maintaining functionality and safety even with changes in the external environment. Mandatory characteristics include: resistance to configuration changes, adaptability without loss of controllability, replication of results, testing on edge and anomalous cases, compliance with declared characteristics.

All AI systems must be provided with a risk management policy approved by the provider and agreed with the authorized body, which includes the identification of technical, legal and social risks, incident prevention and response procedures, identification of responsible persons and resilience measures: redundancy, forced shutdown, functional decomposition, control of the level of autonomy.

For high-risk systems, the provider is obliged to provide a full package of supporting documentation before commissioning: a certificate of conformity, an external audit report, a description of recorded incidents, an architectural justification, risk and constraint indicators, as well as other documents specified by regulations.

If dangerous or unstable behaviour is detected, the system is subject to immediate blocking or termination. The decision on this is made by the authorized body, cybersecurity authorities or the court within their competence. Such decisions must be motivated, with the conditions for re-admission to use, and must be entered into the public register.

The supreme collegial executive body of the State ensures the establishment of minimum requirements for the safety and reliability of AI systems, determines evaluation criteria, approves audit accreditation procedures and ensures the maintenance of the state register of certificates of conformity.

9.1 The security and reliability of artificial intelligence systems are critical and mandatory conditions for their legal implementation, use and scaling in the territory of the State. Special requirements are established for high-risk areas, in particular: healthcare, national defence and security, justice and law enforcement, transport and critical infrastructure, energy, cybersecurity and public administration [120].

In order to prevent man-made, legal or social disasters, AI systems used in high-risk areas are subject to state control, mandatory certification, independent ethical expertise and must provide the possibility of independent verification of the results of their functioning for compliance with safety, reliability and legality requirements [121].

9.2 The security of AI systems includes a set of technical, organizational and procedural measures aimed at ensuring stable, predictable, controlled and non-destructive (safe for humans, society and the environment) functioning of systems in a real environment, considering the dynamics of risks and emerging threats [122].

Mandatory safety characteristics of AI systems include:

−technical resistance to internal and external errors, software or hardware failures, malicious intrusions, cyberattacks, manipulation of parameters or code, as well as to anomalous behaviour due to incorrect or unexpected inputs and other risks identified by the regulator;

−the presence of built-in mechanisms for crashing system actions (fail-safe), functions for backing up critical data and automatic recovery (self-healing), as well as opportunities for human operator intervention in case of an emergency;

−Multi-layered protection of data confidentiality, integrity and availability, including encryption, digital signatures, perimeter security, access control, isolation of environments and data integrity verification at input and output;

−preventive minimization of the risks of runaway training, false self-learning due to distorted data, catastrophic forgetting, cross-system interference or incorrect interaction between the same type or related AI modules;

−Conducting mandatory pre-deployment and field testing before putting the system into operation, with simulation of worst-case scenarios, peak loads, atypical situations and unethical behaviour.

Such testing must be transparent, documented and verified by an independent accredited examination.

9.3 The reliability of AI systems implies the ability to function for a long time without significant failures, loss of efficiency or incorrect adaptation, regardless of changes in the external environment, scale of use or nature of input data. Reliability means maintaining predictable and controlled system behaviour that meets declared technical standards, legal requirements and user expectations [123].

Key reliability parameters include [124]:

− stability of functioning in conditions of changes in the configuration of the environment, type of input data, operation parameters and load, ensuring accuracy and efficiency;

− adaptability to new situations or previously unknown data without loss of general controllability, including self-diagnosis, automatic self-assessment and detection of deviations from reference behaviour;

− replication of results — repeatability or reproducibility of the system's responses in conditions of identical input data, which is the basis for trust, verification and forensic evidence;

− testing on marginal, marginal or anomalous cases, including using a simulated environment, to identify the risks of loss of accuracy, the system going beyond the established technical and legal parameters of behaviour or the occurrence of an incorrect or dangerous reaction of the system, the results of which are subject to documentation and accredited independent verification;

− compliance with the declared technical characteristics, functionality and declared limitations, including open documentation, reference metrics and independent testing results, subject to analysis, verification and comparison in state-defined registers or databases.

9.4 All AI systems should have comprehensively designed, officially approved, and documented risk management policies that are the foundation for decision-making to prevent incidents, mitigate the effects of failures, prevent crisis escalation, and ensure long-term functional and organizational resilience. Such a policy must be approved at the level of the governing body of the responsible organization or enterprise that deploys or uses the AI system[125].

The risk management policy necessarily covers [126]:

− systematic identification of potential technical risks (system failure, computer failures, incorrect training), legal risks (human rights violations, discrimination, unauthorized use of personal data), social risks (manipulation, reduced trust in public institutions, information distortions), as well as other risks identified by the authorized body;

− clearly defined procedures for preventing, detecting, responding to and eliminating incidents, in particular, algorithms of actions when detecting dangerous behaviour of the system, crisis response protocols, internal notification schemes and the involvement of external experts, including government agencies and accredited independent organizations;

− identification of responsible persons, departments or external partners authorized to carry out continuous security monitoring and responsible for each stage of the life cycle of the AI system — from development to decommissioning;

− Integrated measures to increase the system's resilience to failures, including multi-level redundancy, functional decomposition of components, limiting the level of autonomy of solutions, the use

of timeouts, tools for forced shutdowns, as well as periodic rotation of key modules in order to identify vulnerabilities.

9.5 In case of high-risk systems or their use in critical infrastructure, the provider (system developer or operator) is obliged to provide a full package of documentation confirming the compliance of the system with established national and international standards of security, reliability, controllability and ethical requirements. Such a requirement is mandatory before putting the system into operation and is subject to periodic updating based on the results of monitoring and auditing [127].

This package must include [128]:

− a certificate of conformity issued by an institution accredited in accordance with the established procedure, confirming the compliance of the system with the requirements of national standards or international safety protocols (ISO/IEC 23894, NIST AI RMF, OECD AI Principles or other standards recognized by the authorized body), with the obligatory indication of the level of risk and scope of application;

− an independent external audit report covering technical aspects (vulnerabilities, stability, disaster termination), organizational procedures (monitoring, access, conflict management), as well as legal and ethical risks (discrimination, privacy violations, transparency of decisions);

− a comprehensive register and report of all technical incidents, failures, deviations, machine learning errors, incidents of unexpected behaviour, or any issues recorded during development, testing, beta, or pre-implementation of the system;

− An architectural justification of the adopted technical solutions, containing a description of the overall architecture of the system, autonomy levels, human-in-the-loop procedures, human intervention modules, control channels, protocols for recording system actions, risk assessment criteria, fault containment measures, limitations of interaction with other digital systems, and potential impact on critical infrastructure components.

9.6 In case of detection of systems artificial intelligence, which exhibit unreliable, unstable or potentially dangerous behaviour, such a system is subject to immediate suspension, blocking, isolation or withdrawal from public access. The grounds for the application of such measures are: violation of security or reliability requirements, presence of critical technical vulnerabilities, destructive behaviour due to failure, detected human rights violations, significant decrease in the level of user trust (confirmed monitoring results), the risk of causing significant harm to life, health or fundamental human rights, as well as other cases determined by the authorized body [129].

The decision to stop operation or block the system can be made:

an ethics, security or digital regulatory authority;

by the National Cybersecurity Centre or the State Security Service — in cases of threats to critical infrastructure, defence or national security;

by a judicial body — at the request of a state body or upon a complaint by a user, an injured person or a public organization.

The decision to block should be transparent, reasoned and include the terms of isolation, the conditions for re-admission of the system to operation (re-entry protocol), a list of detected violations and a step-by-step plan for eliminating deficiencies.

The authorized body ensures the maintenance of a public register of such decisions, indicating the grounds, date of adoption, status of the system and the results of the re-audit, while restricting access to technical details that can be used by attackers.

9.7 Government of the State, upon the submission of the authorized bodies, approves the list of minimum requirements for the safety and reliability of systems AI, criteria for assessing their compliance, as well as procedures for accreditation of bodies carrying out audits and maintaining the state register of certificates of conformity.

CHAPTER 10. PRINCIPLES OF FAIRNESS AND NON-DISCRIMINATION IN ARTIFICIAL INTELLIGENCE SYSTEMS

Equity in AI systems involves the responsibilities of developers, suppliers, operators, and users to prevent structural biases in training data, algorithms, and decision logic, ensure equitable and non-discriminatory access to technology, and eliminate any form of unfair impact on certain groups of individuals. AI systems should not create or reinforce discriminatory practices, marginalization, or exclusion of vulnerable populations.

It is prohibited to use any AI systems that directly or indirectly lead to discrimination, unequal treatment or violation of the dignity of a person. This applies to automated and semi-automated solutions in the fields of education, healthcare, employment, finance, justice, social security and other areas of public importance. It is prohibited to use models that form segregated or biased digital profiles, scoring systems or "black lists" that negatively affect the social, legal or economic rights of a person.

Developers, vendors, and users of AI systems are required to implement equity policies that include auditing training data, applying rebalancing techniques, using models that can mitigate bias risks, and continuously monitoring neutrality across social, demographic, cultural, and regional lines. Technical mechanisms for identifying and correcting discriminatory decisions should be created, as well as the user's right to review, explain and appeal the results of AI's work should be ensured.

The AI Ethics Authority carries out permanent, independent interdisciplinary monitoring of algorithmic discrimination, maintains a public register of complaints, analyses the impact of systems on demographic groups, initiates technical and legal inspections, has the authority to suspend the functioning of suspicious systems, and provides mechanisms for restoring the rights of persons affected by discriminatory decisions.

The Authorized Body in the field of AI approves and implements fairness audit methodologies focused on identifying signs of discrimination, algorithmic asymmetry or excessive differentiation, considering international standards of the EU, OECD and UNESCO.

Violation of the principles of fairness and non-discrimination entails the suspension of the functioning of the system, administrative or civil liability, the obligation to eliminate discriminatory consequences, as well as other measures provided for by law. Systematic violation may be the basis for a judicial ban on the use of the AI system in certain areas.

10.1 Justice In AI systems, it involves not only formal adherence to the principle of equality, but also active opposition to structural biases that may be implicitly embedded in algorithms, training data, or decision-making logic. This concept covers the obligation of developers, operators and users of such systems to identify and correct any form of disproportionate, unfair or unethical influence on certain groups of persons.

The principle of fairness in the application of AI systems [130, 131] includes:

- ensuring equal and non-discriminatory access to AI systems without restrictions on social or demographic criteria [132];
- prevention of automated discrimination, marginalization or exclusion (stigmatization) of representatives of minorities or vulnerable groups [133];
- mandatory testing of algorithms for hidden bias, feedback loops, and excessive reliance on historical data that can be distorted [134];
- assessment of the fairness of AI decisions, considering the social, cultural, legal and economic context of their application;

creation of mechanisms of legal protection and social recovery in case of injustice due to the functioning of the system.

10.2 The use of artificial intelligence systems that directly or indirectly lead to discrimination, unequal treatment or violation of the dignity of a person is prohibited. The ban applies to both automated decisions and decision support systems in the fields of education, labour, service delivery, healthcare, finance, housing, justice, law enforcement, and social security.

In this Act, discrimination is considered to be any restriction or advantage in the treatment, classification, access to services or rights on the grounds of race, colour, ethnic or social origin, language, religion, sex, gender identity, sexual orientation, age, disability, marital status, citizenship, political opinion, refugee or internally displaced person status, place of residence, economic status, education or other characteristics determined by international and national law. Right.

It is prohibited to use machine learning models that lead to segregation, exclusion, digital blacklists, biased individual risk ratings or opaque scoring that harm a person or group of persons in social, legal or economic dimensions.

10.3 Developers, vendors, and users of AI systems have a shared responsibility for preventing, detecting, and eliminating discriminatory bias at all stages of the system's lifecycle—from data collection and model construction to implementation, use, update, and decommissioning. They are required to implement comprehensive policies and equity procedures that include:

− carrying out mandatory preliminary audit of training data to identify structural or contextual bias that may arise from historical, geographical, socio-cultural or gender characteristics of sources [135];

− application of methods for equalizing imbalances in data (rebalancing), including oversampling, under sampling, synthetically balanced datasets, as well as the use of fairness-aware machine learning models;

− ensuring the absence of direct or indirect discriminatory influences on protected grounds (gender, ethnicity, age, religion, language, socio-economic status, regional origin, etc.) with constant verification of the results [136];

creation of functional mechanisms for automated detection and signalling of discriminatory decisions, ensuring the right of the user or human operator to initiate their review;

- providing users with understandable information about potential risks of bias, training criteria for the system, possible consequences of discriminatory decisions and available procedures for appealing them [137];

- documenting all changes to models and algorithms aimed at eliminating discriminatory manifestations or correcting biased results, with a mandatory audit of the impact of such changes on the quality and fairness of the system.

10.4 The Authorized Body for Ethics of Artificial Intelligence is obliged to ensure comprehensive and independent monitoring of algorithmic discrimination that may arise as a result of the functioning of AI systems in various spheres of public life. Monitoring should be carried out on a regular and cross-sectoral basis using evidence-based methods of data analysis and with the mandatory involvement of experts in the field of law, sociology, computer science and human rights.

The functions of the Authorized Body include:

- creation and maintenance of a centralized public register of complaints about discrimination, with the possibility of filing appeals by citizens, non-governmental organizations (NGOs), media or international institutions [138];

- conducting regular analysis of the impact of algorithmic systems on different demographic groups of the population, using indicators of social vulnerability, geographical disproportion and modelling of discriminatory scenarios based on real cases [139];

- initiating inspections, technical audits and legal investigations, as well as temporarily suspending the use of AI systems suspected of discriminatory behaviour, with the right to publicly disclose the results of such inspections [140];

- ensuring effective mechanisms for restoring the violated rights of persons affected by unfair or discriminatory decisions of AI systems, including the provision of legal aid, compensation for damages, restoration of rights, and cancellation or revision of automated digital solutions [141];

- introduction of mandatory public reporting on all cases of recorded algorithmic discrimination, with recommendations for the elimination of repeated violations and revision of fairness policies at the national level [142].

10.5 The authorized public authority ensures the development and implementation of special methodologies for auditing the fairness of AI systems. Such audits include regular tests for non-discrimination, detection of algorithmic asymmetry and disproportionate impact on certain groups of individuals. Audit methodologies are formed considering international practices and standards of the European Union, OECD and UNESCO, and the results of audits are subject to public disclosure in accordance with the established procedure.

10.6 Detection of a violation of the principles of fairness and non-discrimination in the functioning of the artificial intelligence system is the basis for the immediate suspension of its work. Responsible persons are subject to administrative or civil liability. In addition, they are obliged to publicly apologize to the affected persons and ensure that the discriminatory consequences are addressed through the restoration of violated rights, compensation for damages or other appropriate measures.

CHAPTER 11. THE PRINCIPLE OF CONFIDENTIALITY AND PROTECTION OF PERSONAL DATA IN ARTIFICIAL INTELLIGENCE SYSTEMS

All artificial intelligence systems that process personal data or other information that directly or indirectly identifies an individual are required to strictly comply with the requirements of confidentiality, integrity, availability and security of data. The processing of such information is carried out in accordance with the legislation of the State, the EU General Data Protection Regulation (GDPR), the Council of Europe Recommendations and international standards in the field of privacy.

Any unauthorized access to personal data, including intra-organizational access without proper legal grounds, is prohibited. The use of the data must be purposeful, limited to a specific and legitimate purpose and meet the reasonable expectations of the data subject as defined by the law and the terms of consent. AI system operators are obliged to ensure the protection of data at all stages of their processing by implementing appropriate technical and organizational measures, including monitoring and auditing.

Providers and developers of AI systems are required to implement the principles of privacy by design and privacy by default, apply appropriate technical and organizational measures, including encryption, anonymization, pseudonymization, access control mechanisms and auditing of data actions, as well as guarantee the implementation of the rights of the data subject, including the right to erasure (right to be forgotten) and the right to appeal automated decisions. In the event of security incidents, the system operator shall immediately, but not later than within the period specified by law, notify the authorized body and interested parties of the nature and consequences of such a breach.

Processing of personal data without proper legal basis, non-transparent or mass identification or tracking of persons, creation of behavioural profiles, cross-border data transfer without proper legal support, as well as re-identification (de-anonymization) of anonymized data without legal grounds are prohibited. It is also prohibited to use personal data for commercial purposes (including for sale, advertising, targeted marketing or profiling) or for the purpose of influence that restricts or distorts the user's freedom of choice, without the user's explicit and informed consent.

Control over compliance with the principles of confidentiality in AI systems is carried out by the authorized body by conducting inspections, audits, issuing mandatory orders, applying administrative and financial sanctions, as well as temporarily suspending or restricting data processing. The authorized body is obliged to inform the public about significant violations by publishing official notices and annual reports.

In case of systematic or particularly serious violations, the authorized body has the right to apply to the court with a demand for a complete ban on the operation of the relevant system or the establishment of restrictions on its functioning.

Violation of the requirements for the protection of personal data, in particular failure to comply with consent procedures, failure to comply with security requirements, refusal to exercise the rights of the data subject or unlawful use of data in high-risk areas, entails administrative, civil or criminal liability in the manner provided for by the legislation of the State, as well as the obligation to compensate for the damage caused to the data subject.

11.1 All artificial intelligence systems that process personal data or other information that directly or indirectly identifies an individual are required to ensure compliance with the principles of confidentiality, integrity, availability and security of data. Such systems must implement appropriate legal, technical and organizational measures in accordance with the legislation of the State, the EU General Data Protection Regulation (GDPR), the Council of Europe Recommendations, as well as international standards, in particular ISO/IEC 27701 and the NIST Privacy Framework, to the extent that it does not contradict national law.

The principle of confidentiality is implemented through compliance with the following requirements:

prevention of any unauthorized access to personal data, including intra-organizational, beyond the limits of official necessity [143, 144];

a clear, specific, lawful and transparent definition of the purpose of the processing of personal data, prohibiting their use outside this purpose [145];

limiting the amount of personal data (data minimization) in accordance with the specific and legitimate purpose of processing [146];

processing personal data in a way that meets the legitimate and reasonable expectations of the data subject, while respecting his/her privacy and dignity;

The System Operator is obliged to ensure that all algorithmic processes in which personal data are used are carried out in compliance with appropriate technical and organizational mechanisms for the protection, monitoring and audit of confidentiality [147].

11.2 Personal data cannot be used for training, improvement, validation or testing of AI models without a clearly defined and properly documented legal basis that complies with the principles of legality, good faith and transparency established by the legislation of the State and international standards. Such a legal basis is, in particular:

obtaining informed, voluntary, unambiguous, explicit and properly documented consent of the personal data subject with a clear explanation of the purposes, terms, scope and possible consequences of the processing;

other forms of legitimate interest defined by law, subject to a Data Protection Impact Assessment, which confirms that such interests do not violate the rights and freedoms of the data subject;

use of data in an impersonal (anonymized or non-personalized) form, provided that it is impossible to restore the identity of the person;

the presence of a special regulatory act that directly allows the training of AI systems based on certain categories of data (for example, in the field of healthcare, education, public administration) and defines the procedures for control and supervision of such processing.

Any use of personal data without proper legal grounds, as well as in a way that does not comply with the principles of legality, good faith, transparency and proportionality, is considered illegal processing of personal data and entails administrative, civil or criminal liability in accordance with the procedure established by law.

11.3 Developers and suppliers of AI systems are obliged to ensure comprehensive technical, organizational and legal protection of personal data processed in their systems, in compliance with the principles of confidentiality, responsibility, transparency and control by data subjects. In particular, they are obliged to:

implement a system architecture that implements the principles of privacy protection from the design stage (Privacy by Design) and by default (Privacy by Default), by defining the proportional and minimum required amount of data to be processed, limiting access at all levels and preventing excessive aggregation of information [148];

ensure the use of modern methods of end-to-end encryption, reliable anonymization or pseudonymization of data, taking into account the risks of re-identification and other forms of deanonymization;

create an effective system of logging access to data with detailed records of the subject of access, purpose of use, time, scope and legal basis of access, ensuring the protection of logs from unauthorized modification and the possibility of independent audit;

guarantee the implementation of the full range of rights of the data subject: the right to information, access, rectification, deletion (right to be forgotten), restriction of processing, objection to processing, data *portability*, as well as the right to review and appeal against automated decisions;

immediately, but no later than within the time limits specified by law, inform the authorized authorities and relevant persons about any data breaches, providing information about the scope and nature of the incident, possible consequences, risks and measures taken to eliminate them and prevent their recurrence.

11.4 Prohibited:

− processing of personal data in AI systems without a clearly defined and transparently declared purpose, proper legal basis, a defined storage period and clearly established rules for their use, transfer, updating and destruction [149];

− the use of AI systems for mass or covert surveillance, systematic identification of persons, facial recognition, behavioural profiling, risk forecasting or the formation of social scoring is prohibited, except in cases expressly and unambiguously provided for by a special law, confirmed by a court decision or a justified and proven state necessity, carried out in accordance with the principles of proportionality and legality and under the control of an authorized body [150];

− any attempts to de-anonymize or re-identify a person from previously anonymized or pseudonymized data without a direct legal basis and permission of the authorized body [151];

− automated transfer of personal or sensitive data outside the State without duly concluded contracts, cross-border processing agreements or adequate safeguards that are not lower than the requirements of the legislation of the State and the European Union;

− the use of personal data for non-transparent commercialization, manipulative influence, targeted advertising or the creation of personalized information flows without the explicit and clearly recorded consent of the user.

11.5 The Authorized Body for Personal Data Protection carries out permanent, independent and interdisciplinary and evidence-based control over compliance with the principles of confidentiality in artificial intelligence systems operating in the territory of the State or processing data of citizens of the State [152, 153] regardless of their location. Such control is carried out within the limits of the powers determined by law and includes:

− conducting scheduled and unscheduled inspections of technical, organizational and legal aspects of the functioning of AI systems, in particular the processes of collection, storage, transformation, use, transfer and deletion (destruction) of personal data;

− mandatory initiation of an audit of high-risk systems by independent accredited structures that have state-confirmed technical, legal and ethical competence;

− issuance of orders to eliminate the identified violations with the determination of specific measures and terms of their implementation, mandatory for the addressee;

− application of administrative sanctions, including fines, official warnings, temporary restrictions on access to databases or infrastructure, suspension or complete suspension of the processing of personal data until the violations are completely eliminated;

− mandatory informing the public about the results of inspections and audits, in cases where the detected violations were of a significant public or security nature, in compliance with the principle of proportionality and balance between the public interest and the rights of data subjects.

In case of systematic or particularly serious violations, the authorized body has the right to apply to the court with a demand to prohibit the operation of the relevant AI system or impose restrictions on its use in certain areas (in particular, public administration, finance, education, healthcare).

11.6 Violation of the principles of confidentiality and protection of personal data [154] in the context of the functioning of artificial intelligence systems, including, but not limited to, the unlawful collection, storage, use, transfer or disclosure of personal data, entails bringing responsible persons to administrative, civil or criminal liability in accordance with the procedure established by the legislation of the State.

The types of violations include:

− intentional violation or gross negligence in the implementation of security measures, which led to leakage or unauthorized access to personal data;

− failure to comply with the procedure for obtaining the consent of the data subject or falsification of the legal basis for data processing;

− processing of personal data without conducting a mandatory *Data Protection Impact Assessment* or contrary to its results;

− the use of AI systems that violate confidentiality requirements in areas with a high level of risk to human rights and freedoms (in particular, in the field of healthcare, education, public administration, finance);

− refusal to provide the data subject with access to his/her information or failure to comply with a request for rectification, deletion, restriction of processing or transfer of data.

In case of detection of such violations, the authorized bodies are obliged to initiate an appropriate investigation, record the fact of the offense and apply sanctions, in accordance with the national codified acts of the State (in particular in the areas of civil, criminal, administrative law and personal data protection), and in common law jurisdictions — in the relevant statutes and judicial precedents.

CHAPTER 12. THE PRINCIPLE OF TRANSPARENCY

In the field of development, implementation, operation, control and monitoring of artificial intelligence systems, entities are obliged to ensure the transparency of the relevant processes. Transparency implies accessibility, explainability, openness and clarity of information about the system architecture, algorithmic logic, data sources, decision-making methods, level of autonomy, risks and consequences of application.

The principle of transparency applies at all stages of the life cycle of an artificial intelligence system. The use of the system without notifying the user, without explaining the mechanisms of operation, without the possibility of verifying, appealing or reviewing the results is a violation of the requirements of this Law.

In the case of using high-risk AI systems, entities are obliged to: publish a system passport indicating its technical characteristics, goals, limitations and risks; ensure the maintenance of an open register of such systems; record changes, incidents, and decisions in the appropriate log; appoint an official responsible for transparency and ethics; initiate public discussion in case of implementation in high-risk areas.

The information disclosed in accordance with the principle of transparency should be adapted to the target audience, including considering digital literacy, social context and the legal status of the addressee. It is presented in an accessible, understandable and inclusive format, with the possibility of multimodal presentation.

The principle of transparency is mandatory for AI systems, regardless of ownership, source of funding, technical architecture, or jurisdiction of creation, if the results of their activities have legal or factual consequences for individuals or legal entities, in the areas of public administration, medicine, education, justice, security, migration, finance or social security.

Violation of the principle of transparency entails administrative, disciplinary, financial or civil liability. Until the violation is eliminated, the authorized body has the right to suspend the functioning of the system or restrict its use. In case of a systemic or intentional violation, the system is withdrawn from circulation, recognized as dangerous and subject to a ban on use in critical areas.

Control over compliance with the principle of transparency is carried out by the authorized body through audits, inspections and, if necessary, applying to the court with a demand to restrict or terminate the activities of the relevant artificial intelligence system.

12.1 The principle of transparency in the field of artificial intelligence systems provides for the obligation of all entities involved in the development, implementation and operation of such systems to ensure openness, explainability, clarity and availability of information on the technical structure, logic of algorithms, development goals, decision-making methods, levels of autonomy, probable errors, potential risks, expected social and legal consequences of the use of AI systems [155].

This principle applies to the entire life cycle of AI systems — from training to decision-making and their subsequent monitoring. The information should be accessible not only to technical specialists, but also to end users, persons affected by AI decisions, public authorities and oversight, researchers, human rights institutions and the media [156].

Transparency includes:

description of the processes by which recommendations or decisions are formed;

explanation of cases and parameters in which the system may make errors or false results;

availability of mechanisms for control, verification, moderation and appeal of decisions.

Systems that do not provide transparency are considered potentially dangerous and are subject to enhanced regulation or temporary suspension or restriction of operation in accordance with the provisions of this Law.

Transparency covers the following aspects:

notifying the user about the use of the AI system before interacting with it, including appropriate labelling (for example, "this content is generated by AI" or "decisions are made automatically");

justification of decisions made by AI, in the form of an understandable explanation of the logic of the algorithms' functioning, indicating key variables, factors, sources of influence, data sets used, weighting factors (if possible), as well as information about the level of accuracy and the probability of errors;

documentation of all parameters of model training, its technical architecture, adaptation and change processes, in particular changes in the logic of decision-making after additional training (fine-tuning);

open access to policies and protocols for risk assessment, ethical testing, human rights impacts, and incident response mechanisms (including reporting of malfunctions, unforeseen behaviour, or ethical violations);

availability of information about user rights, including:

- the right to appeal;
- the right to request human intervention;
- the right to an alternative assessment;
- the right to restriction of data processing;
- the right to delete the results;

the right to protection from discriminatory decisions.

12.2 For high-risk systems, transparency implies additional, more stringent requirements in view of their potential impact on human rights, security, economic stability, the political system, or the environment[157, 158]. Such requirements are:

mandatory publication of the AI System Card, which contains information about the purposes of application, scope, sources and formats of training data, description of algorithms, level of autonomy, interaction interfaces, test methods used, limitations, potential risks, as well as data on the developer, customer and certification of conformity;

maintaining a state or departmental register of high-risk systems with open access, including data on the right holder or operator, history of changes, audit results, complaints, cases of suspension or termination of use;

mandatory updating of information about any changes in the system, including modification of models, expansion of functionality, change of data sources, emergence of new risks or areas of application, indicating their impact on the results;

maintaining and storing system decision logs (decision logs) in a secure form throughout the entire life cycle with the ability to check, audit, retrospectively analyse and record atypical or erroneous behaviour;

appointment of an authorized official (transparency and ethics officer) responsible for regular monitoring of the system's compliance with the principles of transparency and explainability;

conducting public discussions and consultations on the feasibility of using a specific AI system in high-risk areas (e.g., education, healthcare, justice, law enforcement).

12.3 Information provided within the framework of the principle of transparency should be adapted to a specific target audience — end users, authorities, human rights defenders, journalists, technical specialists and other interested actors — considering their level of digital literacy, expectations, legal status and social context [159].

The information must be submitted in a form that complies with the following principles:

accessibility — simple and intuitive access to information without the need for specialized technical knowledge;

linguistic clarity — the use of the state language, avoiding excessive terminology, explaining key concepts through examples;

visual explainability — the use of infographics, diagrams, icons, step-by-step instructions;

inclusiveness — considering the needs of people with disabilities and other users with perceptual disabilities (visual impairment, hearing, cognitive disabilities);

multimodality — simultaneous presentation of data in text, audio and visual formats.

AI system operators are obliged to provide this data in an adapted form not only upon request, but also in an open and standardized mode, which allows the user to familiarize themselves with the information in advance before interacting with the system.

The principle of transparency applies to both public and private AI systems, regardless of ownership, source of funding, or technical architecture [160] and is mandatory in the following cases:

making decisions that have legal or significant factual consequences for a person, in particular regarding access to education, employment, medical services, social benefits, the granting of licenses, court decisions or police intervention;

use in judicial, administrative, medical, educational, social, migration, financial, criminal or security spheres, where the consequences of the application of automated decisions may affect the dignity, well-being or rights of a person;

conducting investigations, auditing, controlling, supervising or evaluating the behavior of individuals or legal entities, including through forecasting, classification, scoring or monitoring.

In these cases, operators are obliged to:

to ensure documentary recording of the algorithmic logic of decision-making, as well as access to the justification of the result, data on its accuracy and associated risks;

inform the subject in respect of which a decision has been made about the right to appeal to a person, file an appeal, express objections or request a review;

provide a clear description of the functioning of the system, its data sources, evaluation criteria and determination of legal responsibility for errors.

Violation of the principle of transparency constitutes a significant violation of this Law and entails the following consequences:

imposition of administrative, financial or disciplinary sanctions on the system operator and/or responsible officials;

suspension or restriction of the AI system until all detected violations of the transparency principle are eliminated;

the obligation to publish a public explanation of the nature of the detected violations, the causes of their occurrence, measures to eliminate them, as well as information about the introduced changes in the functioning of the system;

in case of a systemic or deliberate violation — temporary or permanent withdrawal of the system from circulation, its inclusion in the list of dangerous technologies, prohibition of operation in areas classified as critical.

Control over compliance with the principle of transparency is carried out by an authorized body that has the right to conduct audits, inspections and apply to the court with a demand to limit or terminate the activities of the relevant AI system.

CHAPTER 13. THE PRINCIPLE OF ACCOUNTABILITY

The principle of accountability provides that all entities involved in the design, development, implementation, supply, use, maintenance or control of artificial intelligence systems are required to bear legal responsibility, technical, organizational, financial and ethical responsibility for the compliance of these systems with national legislation, international obligations, industry standards and principles of human rights protection.

This includes legal responsibility for the consequences of the system, clarity and transparency of the chain of responsibility between all participants in the life cycle, access to appeal, compensation and review mechanisms, as well as a policy of escalation of responsibility in case of delegation of authority to autonomous systems. Accountability is an essential condition for the legitimate application of AI and cannot be transferred to a machine or algorithm.

For each AI system, an operator must be determined — an individual or legal entity who is personally responsible for the system's compliance with the law, its safe operation, the appointment of responsible persons, providing access to technical information to state authorities, and communication with users in case of complaints or violations. The operator must be identified in all documents, contracts and product labelling.

Accountability includes a set of measures to prevent and respond to violations: documenting decisions and algorithms; appointing responsible persons, including a Data Protection Officer and an AI Governance Officer; conducting internal and external audits; the creation of response channels, including a hotline and a crisis group; implementation of internal liability policies and backup scenarios in case of failures or loss of control.

In case of violation of the principle of accountability, the controlling entity is responsible for the consequences, regardless of the source of the error. It can be administrative, civil, disciplinary or criminal liability. If the damage is caused to vulnerable groups or in sensitive areas, the liability is aggravating and provides for public redress measures.

The state is obliged to ensure the creation and functioning of an institutional accountability infrastructure, including government agencies, the national ethics council, an open register of high-risk systems, public complaint platforms, expert certification programs, regional centres at higher education institutions. Their financing is carried out at the expense of the state budget and international assistance, and the activities must be transparent, controlled by the parliament and open to public oversight.

The principle of accountability assumes that all actors [161], involved in the design, development, implementation, supply, use, maintenance, or control of AI systems [162], are obliged to bear legal, technical, organizational, financial and ethical responsibility for the compliance of these systems with national legislation, international obligations, industry standards, principles of integrity, innovative security and protection of human rights and freedoms [163]. This includes:

liability for results that have legal, social or physical consequences for users or third parties [164];

clarity and transparency of the chain of responsibility between the developer, operator, user and supervisory authority and other participants determined by law [165];

providing access to mechanisms for appealing, indemnifying and reviewing automated decisions;

the obligation to introduce a mechanism for escalating responsibility in cases where decision-making is delegated to autonomous systems.

Accountability is an essential condition for the legitimate application of AI in the public and private sectors and cannot be transferred or delegated solely to a machine or algorithm.

13.1 For each artificial intelligence system, an operator must be determined — a legal entity, an individual entrepreneur or an individual acting in a special status provided for by law (for example, within the framework of an experimental legal regime), which is an officially registered business entity or has the technical and legal capacity to independently fulfil the obligations provided for by this Law.

Such an operator is personally responsible for:

compliance of the functioning of the AI system with the requirements of the legislation of the State and international standards;

ensuring the supervision of the system at all stages of its life cycle — from the stage of training to the stage of use and decommissioning;

appointment of responsible officials authorized to carry out control, monitoring, audit, response and reporting;

keeping records and documenting changes, incidents and updates of the AI system;

providing access to technical and organizational information about the AI system for authorized state bodies, inspectors or auditors;

organization of communication with users, data subjects, and other parties affected by AI activities in case of complaints, risks, or violations.

The operator must be identified in all official registers, contracts, technical documentation and product labelling. In the case of distributed responsibility, the operator ensures a clear consolidation of authority between all subjects of the supply chain or operation of the AI system.

13.2 Accountability includes a set of measures aimed at preventing violations, prompt response and protecting users' rights in case of risks or failures in the operation of AI systems [166, 167]:

recording and documenting all decisions, algorithmic scenarios, training datasets, action logs and changes related to the functioning of AI in order to ensure evidence, verification and audit [168];

appointing responsible persons (Data Protection Officer, AI Governance Officer, Ethics Officer) who have sufficient autonomy, authority, access to information, and financial resources to influence the life cycle of the AI system, with regular reporting to senior management;

conducting internal audits at least once a year, and external audits with the participation of independent certified experts at least once every two years, with mandatory publication of the results in the public domain in cases where the system is recognized as having a socially significant impact;

creation of rapid response mechanisms, including a hotline, a crisis response team, protocols for automatic suspension of the AI system in case of a violation, tools for cancelling or reviewing decisions that have legal consequences;

implementing internal liability policies that determine the degree of personal and corporate responsibility of each unit for certain AI functions, and creating fallback procedures in case of failure, unexpected behaviour, or loss of control over the system.

13.3 In the event of a violation of the principle of accountability, including failure to comply with the requirements of transparency, confidentiality, due diligence or ethical testing, the entity exercising control over the AI system is legally responsible regardless of the source of the error, including human error, technical failure, imperfect model training, external interference or loss of control over the system.

Legal liability may arise in the form of administrative, civil, disciplinary or criminal liability — depending on the degree of violation, the consequences caused, and the legal status of persons affected by the decisions or actions of such a system.

In cases of significant damage (physical, moral, reputational, economic) caused to vulnerable groups of the population or in areas of special regulation (medicine, justice, national security, personal data protection), liability is qualified as aggravating and entails the obligation of the operator or owner of the AI system to take public measures to compensate for the damage and guarantees to prevent recurrence.

13.4 The state ensures the creation of an independent, multi-level institutional infrastructure designed to monitor, verify, certify, audit, and ensure the accountability of artificial intelligence systems. Such infrastructure includes:

central executive bodies or specialized state agencies with powers of a supervisory, permitting, control and analytical nature;

the National Council on Ethics and Accountability in the Field of Artificial Intelligence, which includes representatives of civil society, academia, business, media and human rights organizations;

a register of high-risk AI systems with open access to basic information, mandatory entry of data on operators and the results of independent audits;

public platforms for filing complaints, applications and requests about violations of the principle of accountability, followed by a mandatory response of the authorities;

certification and training programs for independent experts and AI accountability officers;

regional centres for ethical control operating at higher education institutions and scientific institutions.

Financing of this infrastructure is carried out at the expense of the State Budget of the State and international technical assistance. The activities of institutions should be transparent, impartial, controlled by the parliament through the relevant committee on digital transformation, and open to public oversight.

CHAPTER 14. PRECAUTIONARY PRINCIPLE

In development, implementation and operation of artificial intelligence systems, the precautionary principle is applied, which provides for mandatory preventive assessment, prevention and minimization of risks that may pose a threat to life, health, human rights, environment, public security, economic stability, functioning of the state, critical infrastructure or national defence.

This principle applies even in the event of scientific or technical uncertainty about the scale of possible consequences. If there are doubts about the safety of scaling or the implementation of AI, it should be suspended, postponed, or carried out only under the control of a state or independent authorized body.

All participants in the life cycle of AI systems are required to implement a risk-based approach, conduct human rights impact assessments, refrain from using non-certified systems in sensitive areas, provide mechanisms for responding to failures and emergency protocols, and openly declare the limits of the use of such systems. The use of systems with a high degree of autonomy without proper supervision is unacceptable.

In case of uncertainty, lack of complete data or conflict of interest, the decision on further use of the system is made solely for the purpose of preventing harm. Priority is given to protecting life, the environment, democratic institutions, preventing technological escalation and undermining legal certainty.

The Government of the State ensures the implementation of the national methodology for prudent risk management, which is based on international standards (in particular, NIST AI RMF, ISO/IEC 23894, OECD, IEEE, as well as recommendations of UNESCO and the Council of Europe), takes into account national priorities, market structure and contains risk assessment criteria, response mechanisms, indicators, tools for consultation, monitoring and integration with state and international platforms.

The application of the precautionary principle is mandatory in high-risk areas such as biomedicine, autonomous transportation, defence, digital identity, behaviour analytics, e-justice, cybersecurity, biometric processing, critical infrastructure management, and information policy.

Public authorities have the right to temporarily or permanently restrict the functioning of any AI system in case of violation of the precautionary principle, if a threat to life, human rights, national security is established, or technical malfunctions, inaccurate information or unpredictable autonomous behaviour of the system are detected. The ban is applied in accordance with the administrative procedure determined by law, and in emergency cases — immediately, with further revision.

14.1 The precautionary principle in the development, testing, implementation and operation of artificial intelligence systems [169] provides for mandatory preventive detection, detailed analysis, assessment, as well as prevention and minimization of potential risks, that may cause direct or indirect harm to life, health, human rights and freedoms, public security, the environment, economic stability, the functioning of state institutions and systems of national defence, critical infrastructure and cybersecurity [170, 171].

The precautionary principle is activated even in conditions of incomplete scientific, empirical or technical knowledge about the nature, probability or scale of possible negative consequences. In case of uncertainty or doubts about the security of the AI system, by default, the decision to use, scale or integrate it should be made in favor of restraint, postponement or modification under the control of a state or independent authority [172, 173].

All entities involved in the life cycle of AI systems (developers, integrators, suppliers, operators, users, government agencies) are obliged to act in accordance with the precautionary principle by:

implementation of a risk-based approach at all stages — from initial design, data acquisition and testing — to launch, scaling, operation, upgrade, and decommissioning [174];

systematic assessment of the potential consequences of the implementation of the system for fundamental human rights, public and institutional trust in technology, social justice and the environment [175];

prevention of the use of experimental, untested or uncertified AI systems in high-risk areas, especially in cases involving vulnerable groups of the population (children, the elderly, people with disabilities, prisoners, patients, migrants) [176];

development and implementation of mechanisms for prompt detection, fixation, analysis and elimination of critical failures, as well as emergency protocols (automatic blocking, transfer to safe mode, notification of responsible persons) with mandatory registration of such incidents in the appropriate log;

an open declaration of the limits of the use of AI systems, in conditions of limited human control, lack of verification of input data, high uncertainty or a high degree of autonomy of the system;

Human Rights Impact Assessment before each large-scale implementation, as well as after each significant update or change in algorithms, documenting and publishing the main findings.

14.2 In all cases of uncertainty, incompleteness of data or potential conflict of interest regarding the safety, legality, legality or ethics of the use of artificial intelligence systems, the final decision on their further use, launch, scaling or preservation should be made solely considering the principle of priority prevention of harm.

Priority is given to:

- protection of human life and health;
- protection of human rights and freedoms;
- preservation of ecological balance and the environment;
- ensuring stability and public trust in the systems of justice, democracy and public administration;
- preventing the escalation of technological risks beyond the limits of control;
- avoiding the creation of precedents that jeopardize the legal certainty and ethical responsibility of developers and operators.

In cases of serious uncertainty, the introduction or use of AI should be temporarily suspended until scientifically and technically justified confirmation of its safety and compliance with international standards is received.

The Government of the State is obliged to ensure the approval and implementation of a comprehensive methodology for the careful management of risks associated with the use of artificial intelligence systems. Such a methodology has:

be based on internationally recognized standards for AI risk management, in particular NIST AI Risk Management Framework (RMF), ISO/IEC 23894, OECD AI Principles, IEEE 7000, ISO/IEC 42001 and other generally recognized international standards;

consider national priorities, legal specifics, the level of digital development and the sectoral structure of the national market;

determine risk categories, methods of their quantitative and qualitative assessment, typical scenarios of failures and errors, mechanisms for containment, recovery and documentation of violations;

establish requirements for the annual update of risk profiles for high-risk systems and adaptation of policies and procedures to changes in the technological environment;

contain tools for public consultations, public participation, external peer review, and multi-level approval of recommendations;

provide the possibility of integration with state registers, data exchange platforms and international information systems for monitoring and analysis of technological risks.

14.5. The application of the precautionary principle is mandatory in high-risk areas, i.e. those where the impact of decisions or actions of AI systems can cause significant negative consequences for human life and health, human rights and freedoms, national security, public order or sustainable development. High-risk areas include:

biomedical systems, including diagnostics, treatment, medical device management, and healthcare decision support systems [177];

autonomous transportation (cars, drones, rail and sea systems), especially in conditions of mass transportation of people or goods of strategic importance;

development, application, management or control of weapons and means of military influence, including autonomous or semi-autonomous AI systems [178];

digital identities, authentication, e-voting, and e-certification systems that may affect civil rights and individual security [179];

behaviour prediction and social scoring systems (including predictive police analytics, social media analysis, behavioural modelling) that may pose risks of discrimination or invasion of privacy;

digital justice, in particular automated decision-making assistance systems in judicial and administrative jurisdictions [180, 181];

protection of cybersecurity, including detection, response and counteraction of threats in government systems, banking, telecommunications, defence;

processing of biometric and personal data, including video surveillance, face, voice, emotion or fingerprint recognition [182];

critical infrastructure management: energy, water supply, telecommunications, environmental protection, logistics;

information policy, including algorithmic content moderation, automated news generation and systems for influencing the formation of public opinion [183].

14.6. Public authorities that supervise compliance with legislation in the field of artificial intelligence have the right, within the limits of their powers, to temporarily suspend, restrict or completely prohibit the functioning of any AI system in case of violation of the requirements of the precautionary principle.

The grounds for such intervention are:

- real or potential threat to human life or health, fundamental human rights and freedoms or public welfare;
- negative impact on critical infrastructure, information resources of the state or national security;
- creation of uncontrolled or unpredictable risks caused by autonomous behaviour or lack of effective control mechanisms;
- detection of significant technical malfunctions, use of non-certified components, submission of false or false information about the purpose, functionality or characteristics of the system.

The decision on a temporary or final ban is made in accordance with the established administrative procedure, notifying the system operator, indicating the identified risks and setting a deadline for eliminating violations. In the event of a threat to life or national security, the ban is applied immediately, followed by judicial or administrative review.

CHAPTER 15. THE PRINCIPLE OF CONTINUOUS SUPERVISION

All AI systems, regardless of their scope, level of autonomy, or origin, are subject to continuous, systemic, and multi-level monitoring at all stages of the life cycle, from development to decommissioning. Such supervision covers the technical, legal, ethical, social and security planes and is carried out in real time or according to a certain schedule using automated means and with the involvement of authorized persons.

Entities responsible for AI systems are required to implement procedures for continuous monitoring of efficiency, accuracy, compliance of algorithms with stated goals and non-discrimination of results, as well as tracking updates, integrations and changes in the logic of systems. For socially significant systems, regular audits and publication of their results are mandatory. Notifications of failures, incidents or identified risks must be available to internal users of the system and authorized external bodies or observers.

Supervision of high-risk systems is carried out by specialized state bodies in cooperation with accredited experts, auditors, representatives of the academic and public sectors. Proactive control tools are in place, including audits, stress tests and requests for clarification, as well as response measures, including suspension of operation, cancellation or revocation of the certificate of conformity, imposition of fines or public investigations. In case of transnational influence, it is planned to inform the partner structures of the EU, NATO and relevant international organizations.

High-risk AI systems are subject to mandatory certification at least every two years, unscheduled inspections in case of incidents, reporting on changes in functionality, and mandatory risk reassessment in the event of a change in context or scope.

In case of detection of new risks or anomalies, the operator is obliged to immediately stop the operation of the system, initiate a technical and ethical review, notify the authorized authorities, users and interested parties, record the incident in the state register and refrain from further application without updated certification. Violation of this procedure entails administrative or criminal liability.

The state creates and maintains a national AI monitoring infrastructure based on openness, reliability, international compatibility and transparency. It includes: an incident platform, an audit register, a bank of typical failures, a library of risk indicators, and modules for integration with cyber defence systems. The infrastructure operates under the control of an independent body overseeing the ethics and safety of AI, provided with adequate funding, parliamentary oversight and openness to international exchange and independent expertise.

15.1 The principle of continuous oversight implies that all AI systems, regardless of their scope, level of autonomy or origin, are subject to continuous, systemic and multi-level monitoring throughout their entire life cycle — from conceptual design, training, testing, deployment, modernization to complete decommissioning.

Monitoring covers:

- technical control (accuracy, reliability, resistance to errors, deviations and incorrect behaviour)[184];
- legal control (compliance with the law, ensuring legal access to data, compliance with the requirements established by licenses and permits) [185,186];
- ethical supervision (detection of discriminatory behaviour, algorithmic bias and assessment of the ethical consequences of decisions made) [187];
- social supervision (assessment of the impact on public opinion, social groups, labor relations, as well as protection of the rights of vulnerable categories of persons) [188,189];
- security oversight (risks to state security, defence capability and digital sovereignty) [190].

Such supervision is carried out in real time or at a certain frequency, using automated diagnostic systems, expert reviews, algorithms for analysing behavioural patterns and early warning systems about risks.

15.2 Entities responsible for the AI system are obliged to:

introduce procedures for continuous monitoring of the efficiency, accuracy, reliability and compliance of algorithms with stated goals and regulatory requirements;

provide automated control and documentation of updates, changes in model logic, integration with other systems, and maintaining parameter consistency;

conduct regular (at least once every six months) internal audits and external audits, the results of which are subject to mandatory entry into the open register in cases where the system has a socially significant impact;

Establish mechanisms for reporting failures, errors, or threats to the security, functioning, or rights of users, which can be initiated by both internal users and external observers.

15.3 Supervision of high-risk artificial intelligence systems is carried out by specialized state bodies, whose competence includes control over digital technologies, national security, human rights protection, compliance with digital ethics and the functioning of critical infrastructure. Such supervision is carried out in cooperation with authorized independent experts, technical auditors, representatives of the academic community, ethical oversight institutions and public organizations that have passed state accreditation.

Supervisory authorities are obliged to apply both proactive control tools (audits, risk assessment, stress tests, requests for clarification) and response mechanisms (decision to temporarily suspend operation, revoke a certificate, impose a fine, public investigation). In case of detection of violations that may have a transnational effect or violate the international obligations of the State, the relevant authorities shall immediately notify the partner structures of the EU, NATO or international standard-making organizations.

High-risk systems are subject to:

− mandatory institutional certification at least once every two years, which includes technical, legal, ethical and security audits;

− unscheduled inspections in case of user complaints, recorded incidents, threat reports or deviations from expected behaviour;

− mandatory submission of a report on changes in functionality, architecture or decision-making logic within 15 calendar days after such changes;

− periodic reassessment of risks, especially in the event of a change in the context of use, introduction of new data or expansion of the scope of application.

The state is obliged to ensure the transparency of supervisory procedures, the publicity of certification conclusions and the involvement of representatives of the public and the scientific community in the assessment of particularly sensitive or influential systems.

15.4 In case of detection of new unforeseen risks, anomalies in the behaviour of the system or significant deviations from the declared characteristics, the operator is obliged to:

− immediately suspend the use of the relevant module, functionality or operation of the entire system;

− initiate a review of the technical, ethical and safety compliance of the system by conducting an internal analysis and independent assessment;

− notify the authorized state authorities in writing (including electronic) and, in case of a high threat, inform users and stakeholders about the identified risk;

− ensure that the incident is recorded in the state register of incidents in the field of security of AI systems;

− refrain from re-putting the system into circulation without updated certification or a positive opinion of the competent authority.

Failure to comply with this procedure is considered as a gross violation that entails administrative, and in case of significant damage, criminal liability in accordance with the current legislation.

15.5 The state ensures the creation and continuous improvement of the national infrastructure for analytical monitoring of artificial intelligence, which operates on the principles of openness, reliability, transparency, transnational compatibility and compliance with international standards.

Such infrastructure includes:

− a single national platform for recording and analysing incidents with automated triage, impact assessment, report generation and follow-up response functions;

− a state register of audits and certification opinions on high- and medium-risk AI systems, available to regulators, scientists, stakeholders and the public;

− the National Bank of Typical Failures and Anomalies, which contains examples of abnormal behavior, the impact of updates, and algorithm conflicts, designed for training, testing, and revisions by developers;

− a library of early warning indicators on risks and threats, based on international experience, NATO, EU, ISO/IEC practices;

− modules for integration with cyber defence platforms, digital forensics, threat registers and cross-border exchange databases.

The infrastructure is subordinated to an independent state body or an authorized agency for oversight of the ethics and safety of AI, with adequate funding, annual reporting to the parliament and open channels for public monitoring. It should be open to international exchange, integrated with standardization systems, and able to provide independent peer review and public audit.

CHAPTER 16. THE PRINCIPLE OF ETHICAL RESPONSIBILITY

Ethical responsibility is a fundamental principle of the legal regulation of artificial intelligence in the State. It consists in the mandatory integration of moral and philosophical principles, standards of integrity, social justice, transparency, inclusiveness, environmental and cultural sensitivity at all stages of the life cycle of artificial intelligence systems — from their design and training to post-marketing support and completion of use. All persons, institutions, enterprises, organizations and public authorities involved in the creation, adaptation, use or regulation of AI systems are obliged to act within the framework of ethical requirements, ensuring respect for human dignity, observance of human rights and freedoms, avoidance of discrimination, preservation of individual autonomy, predictability of consequences and validity of decisions, as well as effective human control at critical stages of the system's functioning. Ethical responsibility also encompasses the constant updating of knowledge, critical ethical reflection, openness to external ethical auditing, and a willingness to rethink approaches to AI regulation and design in the context of their evolution.

All actors creating, exploiting, or using artificial intelligence systems are required to adhere to universal moral and ethical standards, respect the dignity and autonomy of the individual, ensure that harm to life, health, honour, freedoms and human rights is not harmed, avoid discriminatory practices, manipulation or disproportionate impact on personal freedom of choice, and contribute to increasing public trust in technology through transparency and accountability.

Ethical responsibility covers the following key areas:

algorithmic design ethics: inclusion of ethical due diligence procedures at the design stage of AI systems, with a ban on the implementation of functionality that has discriminatory, manipulative, privacy-threatening, or socially distorting effects;

ethics of use: prohibition of the use of AI without the informed consent of a person, use of his/her data without proper legal grounds, as well as creation of dependence, exclusion or psychological pressure;

ethics of consequences: conducting a mandatory analysis of the social, cultural, legal, gender, economic and psychological consequences of the use of AI, involving interdisciplinary expertise and modelling of potential risks;

Ethics of interaction: ensuring friendliness, intuitiveness and accessibility, non-discrimination and respect for the digital dignity of the user when designing AI interfaces, including the right to refuse interaction with algorithmic agents;

ethics of responsibility: clear legal and organizational consolidation of personal or institutional responsibility for decisions made or formed using AI systems, taking into account the role of the developer, operator, owner, auditor, user and state customer.

The state ensures the implementation of ethical standards through the formation of a national ethical infrastructure, which includes: the development and approval of the National Code of Ethics for Artificial Intelligence; creation of independent ethics councils at regulatory bodies, ministries, defence structures, scientific institutions and universities; introduction of mandatory interdisciplinary training courses on digital ethics for all categories of participants in the AI life cycle; mandatory certification for ethical compliance as a condition for the use of AI in the public, defence and critical infrastructure sectors.

In case of a gross or systematic violation of ethical standards, in particular the use of AI systems for the purpose of spreading propaganda, disinformation, manipulation of consciousness without the informed consent of individuals, masking or legitimizing human rights violations, distortion of legal processes or encroachment on the principles of human dignity, justice, equality or environmental ethics, such a system is subject to temporary withdrawal from circulation and the opening of a public ethical investigation procedure. The decision to remove and apply sanctions is made on the basis of the conclusion of the AI Ethics Council in coordination with the central executive body for digital policy.

Based on the results of the investigation, guilty persons are subject to disciplinary, administrative or criminal liability depending on the severity of the violation, in particular in cases of interference in democratic processes, spreading enmity or creating threats to national security.

All AI systems used in the fields of education, justice, healthcare, defence, national security and social policy are subject to mandatory ethical assessment at the implementation stage and before application. Ethical assessment includes: verification of compliance with the basic ethical principles defined in this Code; conducting interdisciplinary expertise with the involvement of representatives of the public and specialists in ethics, law, technology and social sciences; risk assessment and identification of measures to minimize them with proper justification.

The results of such an assessment are open, subject to mandatory publication on the official state portal on digital policy and are included in the National Register of Ethical Analysis of Artificial Intelligence, which provides free access for researchers, educators and the public.

16.1 Ethical responsibility is a fundamental principle of the legal regulation of artificial intelligence in the State and consists of in the integration of moral and philosophical foundations, standards of universal integrity, social justice, transparency, inclusiveness, environmental and cultural sensitivity at all stages of the life cycle of artificial intelligence systems — from design and training to post-marketing support and decommissioning.

This principle provides that all persons, institutions and public authorities involved in the creation, adaptation, use or regulation of AI systems are morally and legally responsible for the compliance of their actions with the principles of respect for human dignity, ensuring non-discrimination, autonomy of choice, the principle of algorithmic justice, predictability of consequences, reasonableness of decisions, prevention of harm to life, health, honour, dignity and human rights, as well as ensuring an adequate level of human control.

In addition, ethical responsibility involves constant updating of knowledge, ethical reflection, openness to external ethical audit, and readiness to critically review and rethink approaches to the development and application of AI as an evolutionary technology that directly affects the structure of values of society.

16.2 All entities that create, apply or operate artificial intelligence systems are required to:

− to adhere to universal moral and ethical standards;

− to respect the dignity and autonomy of the person;

− to ensure the prevention of harm to life, health, honour, dignity, freedoms and human rights;

− avoid discrimination, manipulation and any manipulative or disproportionate impact on freedom of personal choice;

− contribute to strengthening public trust in technology by ensuring transparency and accountability.

Ethical responsibility encompasses a set of principles and areas that determine the moral quality of the functioning of artificial intelligence as a socially significant technology. It includes:

algorithmic design ethics — the introduction of ethical review procedures at the stage of model design; inclusion in the terms of reference provisions on the prevention of discriminatory goals, manipulative behaviour, violation of privacy or distortion of social reality;

ethics of use — requires that no AI system be used in a way that forces a person to interact without his free, conscious and unambiguously expressed consent; prohibits the use of personal data without the legal basis or consent of the subject; and also makes it impossible to create conditions of dependence, exclusion or psychological pressure.

ethics of consequences — provides for a mandatory analysis of the social, cultural, legal, gender, economic and psychological consequences of the use of AI systems with risk modelling and the involvement of multidisciplinary expertise;

interaction ethics — requires the design of AI interfaces in such a way that they are friendly, understandable, non-discriminatory, respect the digital dignity of the user, do not contain manipulative messages, and provide the right to refuse interaction with the algorithmic subject;

ethics of responsibility — provides for a clear normative definition and consolidation of personal or institutional responsibility for decisions made with the participation or influence of AI systems, including the developer, operator, owner, auditor, user or government customer.

16.3 The state ensures the institutional implementation of ethical standards through the development of a system of public ethical infrastructure, which includes:

development and adoption of the National Code of Ethics for AI, a regulatory document that establishes minimum ethical standards, principles of algorithmic integrity, procedures for responding to violations and mechanisms of public control;

creation of independent ethics councils and commissions under regulatory bodies, relevant ministries, defence and security agencies, research institutions and higher education institutions. Such councils are empowered to carry out an independent assessment, provide advice, investigate violations and carry out public examinations;

introducing mandatory interdisciplinary courses and advanced training programs on digital ethics, AI ethics, and responsible algorithmic design for all categories of people involved in the AI lifecycle — including programmers, project managers, civil servants, military, educators, doctors, and judges;

development of procedures for mandatory certification for ethical compliance as a separate criterion for allowing the use of the AI system in the state, defence and critical infrastructure sectors. Such certification provides for ethical examination, public scrutiny, and a transparent procedure for appealing the results of the analysis.

16.4 In case of gross or systematic violation of ethical standards, including the use of the AI system: for the purpose of disseminating propaganda or disinformation materials; as an instrument of manipulative influence on the consciousness or behaviour of persons without their knowledge and consent; to conceal human rights violations or distort legal processes; in a way that violates the foundations of human dignity, justice, equality, or the principles of environmental ethics and sustainable development — such a system is subject to mandatory temporary removal from use and access in public space with simultaneous opening procedures of public ethical investigation with the participation of representatives of academia, civil society, regulatory authorities and international observers.

In case of a proven violation, responsible individuals or legal entities are subject to prosecution:

- disciplinary liability in accordance with the internal acts of the institution;
- administrative liability, in particular in the form of fines and bans on activities;
- criminal liability, in case of corpus delicti, in particular for spreading hatred, interference in democratic processes or undermining national security.

The decision to apply temporary withdrawal and impose sanctions is made on the basis of the conclusion of the AI Ethics Council in coordination with the authorized body for national digital policy.

16.5 All artificial intelligence systems used in the fields of education, justice, healthcare, defence, national security or social policy are subject to mandatory prior ethical assessment, which includes the following elements:

- analysis of compliance with the key ethical principles defined in Article 18 of this Code;
- interdisciplinary expertise involving members of the public, specialists in law, ethics, technology and social sciences;
- risk assessment and development of mechanisms for their minimization.

The results of the ethical assessment are open, subject to mandatory disclosure, posted on the official digital policy portal of the State and stored in the National Register of Ethical Analysis of Artificial Intelligence (AI) with open access for researchers, educators and journalists and the public.

CHAPTER 17. THE PRINCIPLE OF ETHICAL RESPONSIBILITY OF AI SYSTEMS

All artificial intelligence systems, regardless of their architecture, level of autonomy, functional purpose or field of implementation, are subject to the principle of ethical responsibility. This principle provides for the creation, testing, use and maintenance of such systems in full compliance with the fundamental values of an open democratic society, including respect for human dignity, autonomy, mental and physical integrity, freedom of will, legal equality, non-discrimination, social justice, privacy, environmental sustainability and intergenerational ethics.

Ethical responsibility is an operational responsibility of all actors involved in the AI life cycle — developers, implementers, administrators, operators, data owners (controllers), state regulators — and requires clear risk management procedures, internal ethical audit mechanisms, human rights impact monitoring systems, as well as preventive harm prevention and compensatory measures to restore rights and compensate for harm in case of negative consequences.

Within the framework of the implementation of this principle, entities that create, implement or operate artificial intelligence systems are obliged to ensure:

implementation of ethical design focused on maintaining the autonomy of the user, his psycho-emotional comfort, the right to an informed decision and the prevention of manipulative influence;

conducting an independent preliminary ethical review of each new functionality, architectural change or algorithm update that may affect human rights, public trust or environmental balance;

institutionalization of independent ethics commissions with the right to suspend a project, appeal to competent supervisory authorities, initiate a public audit or regulatory intervention;

ensuring algorithmic validity, proportionality, non-discrimination, intercultural sensitivity and social context in all decision-making processes.

It is prohibited to create, disseminate and use AI systems that knowingly or due to side effects manipulate user behaviour, cause digital addiction, emotional exhaustion, loss of trust, suppress autonomy or promote discrimination on any grounds. It is prohibited to use AI to restrict freedom of expression, control the information space without a proper legal basis, distort public opinion, illegally monitor political activity or public protests, and support dehumanization processes.

Ethical responsibility is a mandatory criterion for certification of high-risk AI systems, especially in the areas of justice, healthcare, education, social management, and public service delivery. In case of violation of the principle of ethical responsibility, the Central Executive Body in the field of AI has the authority to initiate a temporary suspension of the operation of the system, demand the mandatory removal or modification of unethical functionality, apply penalties, restrict access to state resources, tenders or data, revoke permits or completely stop the operation of the system. Such actions are accompanied by a public justification, an official report on the detected violations and public consultations on the ways of ethical reconstruction.

Developers and operators are required to annually publish the Annual AI Ethics Impact Report, which assesses the impact of the system on human rights, psycho-emotional health, social equality, trust in society, environmental sustainability, and intercultural harmony. Reports should include information on the ethical procedures implemented, identified risks, results of complaints consideration, examples of controversial decisions, ethical audit findings, user feedback, and the projected ethical impact of the next stages of system development. Such reports are posted in the public domain, subject to public comment and independent expert analysis.

The principle of ethical responsibility is a key guarantee that the use of artificial intelligence does not go beyond humanitarian legitimacy, preserves human dignity as an inviolable value, and serves as a source of public trust in the age of algorithms.

17.1 The principle of ethical responsibility means that every artificial intelligence system — regardless of its type, level of autonomy, functional purpose, or scope of application — must be created, tested, implemented, and operated in strict accordance with the moral and ethical norms adopted in an open democratic society [191].

These norms encompass the values of human dignity, personal autonomy, physical and mental integrity, freedom of will, justice, legal equality, non-discrimination, solidarity, respect for privacy, environmental responsibility, as well as intergenerational ethics [192].

Ethical responsibility is not a declarative, but an operational category that is implemented through internal risk management processes, human rights control systems, harm prevention mechanisms, and procedures for restoring rights and reparations in the event of negative consequences [193].

AI systems that do not meet ethical standards cannot be considered legitimate, even if technical or legal criteria are formally met. Thus, ethical responsibility is the highest level of requirements for artificial intelligence, integrating the legal, social, cultural and humanitarian perspectives of its application.

17.2 All entities involved in the life cycle of an AI system (developers, suppliers, operators, administrators, data owners/controllers, supervisory authorities) are obliged to:

implement the principles of ethical design (ethics by design), which includes not only the avoidance of manipulation, bias or violation of the autonomy of the individual, but also the proactive development of functional solutions that support the integrity of the individual, autonomous decision-making, psycho-emotional comfort of the user and social well-being;

carry out a preliminary independent ethical examination of each new functionality, technological update or change in the architecture of the AI system, if they can change the nature of interaction with a person, affect his rights, dignity, trust, psychological safety, social structure or ecological balance [194];

institutionalize permanent internal ethics commissions (ethics review boards), which consist of independent experts — ethicists, human rights defenders, sociologists, psychologists, engineers and user representatives. Such commissions are empowered to initiate an ethical review, suspend the implementation of the project, or apply to the competent supervisory authorities in case of ethical threats [195];

ensure that each algorithmic decision complies with the following principles: rational reasonableness (clear explanation taking into account all relevant factors), proportionality (fit for purpose and avoidance of excessive interference);

non-discrimination and equal treatment (neutrality in relation to groups based on gender, race, age, health, social status, etc.);

intercultural sensitivity (recognition and respect for cultural differences);

social context (analysis and consideration of the impact of the decision on specific communities or groups of the population).

17.3 The creation, distribution and implementation of AI systems is prohibited[196,197], that:

purposefully, covertly or through side effects, manipulate the user's behaviour to change his beliefs, preferences or decisions, cause digital dependence, emotional exhaustion, loss of concentration, trust or autonomy, or other forms of psycho-emotional harm - without ensuring full informed consent and the possibility of refusal;

create or maintain discriminatory algorithmic patterns that directly or indirectly lead to prejudice against individuals or groups based on race, ethnic origin, gender, age, religion, political or philosophical beliefs, social or economic status, disability, health status or place of residence, or any other characteristic that may be grounds for discrimination;

are used by public or private actors to suppress freedom of expression, control over the information space, illegal or disproportionate monitoring of citizens, distort public opinion, influence electoral processes, protest activity or social mobilization;

contribute to the processes of dehumanization — that is, the depersonalization of a person, reducing their status exclusively to a set of data or targets; reproduce structural prejudices, historical injustices, or maintain inequality through the uncritical use of learning models detached from the social context and ethical adjustments.

17.4 Ethical responsibility is recognized as a critical and mandatory criterion for the certification of high-risk AI systems operating in the fields of justice, healthcare, education, and the provision of public services, algorithmic control of behavioural practices or social processes, as well as in all cases in which algorithmic decisions can cause legal, social or psychological consequences for the person [198].

In case of a violation of the principles of ethical responsibility — both at the design stage and during implementation or operation — the Central Executive Body in the field of AI has the authority to apply a set of administrative and regulatory measures of influence. In particular, it can:

− issue an order on temporary suspension of operation or temporary blocking of the functioning of the system until the violations are eliminated;

− demand a complete revision or mandatory removal of functionality that does not meet ethical requirements, even if it is technically efficient or commercially feasible;

− apply penalties, temporary or permanent restriction of access to government data and resources, restriction of participation in public tenders, revocation of permits, as well as exclusion from pilot programs or regulatory sandboxes;

− in cases of special danger, prohibit further implementation or operation of the system.

The application of such measures should be accompanied by a public justification, an official report on the detected violations and public consultations on the restoration of the ethical compliance of the system.

17.5 Developers and operators of AI systems are required to publish AI Ethics Impact Report at least once a year, which are public documents aimed at ensuring transparency, accountability, and increasing the level of ethical compliance of AI systems. Submission of such reports is a prerequisite for certification and further operation of high-risk systems.

An ethics report should contain:

− a detailed assessment of the impact of the system on human rights, vulnerable social groups, mental health of users, social justice, cultural diversity, the environment, the information space and trust systems in society;

− description of the implemented internal ethical procedures: functioning of the ethics committee, conducting audits, reviewing incidents, staff training, ethical design principles;

− a documented and confirmed list of risks identified during the audit or use of the system, examples of conflicting decisions and complaints from users, as well as response measures implemented to prevent the recurrence of violations;

− qualitative and quantitative user reviews, reflecting the level of trust in the system, its transparency, explainability, reliability, impartiality and user orientation;

− the results of an independent ethical audit (if any), indicating the source of the audit, verification criteria, conclusions and recommendations for improvement;

− forecast of ethical impacts on the next cycle of system development, taking into account planned updates, scaling, new functionalities and changes in the external environment (legal, social, technological or environmental).

Ethical reports should be submitted in an accessible and understandable form, posted on the official resources of the developer or operator, be open to public comment, independent expert analysis and the possibility of submitting alternative positions and proposals.

CHAPTER 18. THE PRINCIPLE OF LEGAL RESPONSIBILITY AND ACCOUNTABILITY OF ARTIFICIAL INTELLIGENCE SYSTEMS

The principle of legal responsibility and accountability in the AI ecosystem provides for the mandatory establishment of a clear and public subject of responsibility at all stages of the life cycle of an AI system, including development, implementation, administration, technical support, ethical assessment, auditing and decision-making that affect human rights, public interest or public safety.

Responsibility is distributed in proportion to the role, degree of control, access to data, and ability to influence the results of the system's functioning. Including:

AI suppliers and developers are responsible for architectural solutions, model quality, and compliance with safety standards.

Users are responsible for the proper use of the systems within their intended purpose.

Operators are responsible for developing and implementing internal control policies, informing individuals, monitoring the functioning and responding to failures.

Auditors are responsible for the validity of conclusions and the prevention of a formal, biased or selective approach to audits.

Legal liability in the field of artificial intelligence covers civil, administrative, and criminal consequences.

Civil liability arises in cases of material or moral damage to persons who suffered because of algorithmic errors or negligent use of AI.

Administrative liability arises in case of non-compliance with regulatory requirements, security standards, lack of transparency of decisions, improper audit or failure to comply with the obligation to preserve logs.

Criminal liability arises for serious violations related to the manipulation of public opinion, interference in electoral processes, discriminatory practices, violation of children's rights or the creation of systems that pose a threat to the life and safety of people.

For all high-risk AI systems, regular external audits, preservation of event and decision logs, creation of public appeal and immediate response mechanisms are mandatory. Suppliers, operators and users are obliged to implement and maintain incident tracking systems and immediately notify the National Agency for AI Ethics and Oversight of critical failures, deviations or detected human rights violations.

Public authorities have increased public accountability for the implementation of AI systems. They are obliged to inform the public about the purpose of using the system, the legal basis, the amount of data processed, the impact on human rights, as well as to publish annual reports containing analytical data on effectiveness, inclusiveness and the results of external and internal audits. In the event of a violation or risk, the public authority is obliged to immediately suspend the functioning of the system, provide access to alternative mechanisms, initiate an independent examination and inform the regulatory body.

Each public AI system is subject to an ethical assessment at least once every two years with the participation of members of the public, experts in the field of law, digital technologies and human rights. Public authorities are obliged to conduct regular training of personnel on the responsible use of AI systems and ensure information accessibility on the principles of algorithms, their impact on citizens' rights and available legal remedies.

18.1 The principle of legal responsibility and accountability forms the basis of democratic control over AI systems. It provides that any use of AI must be accompanied by the mandatory definition and fixation of the subject of responsibility for the development, implementation, operation and results of the system.

18.2 The principle of delimitation of legal responsibility.

Legal responsibility for the activities of AI systems is distributed among the entities involved in different stages of the life cycle of such systems, considering their role, level of control, amount of access to data, management tools and ability to influence the results of functioning:

suppliers and developers are responsible for errors in architectural solutions, deficiencies in model training, violations of quality, safety and transparency requirements, as well as for failure to comply with the obligation to have and maintain technical and ethical documentation;

users are responsible for inappropriate or negligent use of the system in areas where AI does not provide sufficient reliability (including without human control), for the use of AI beyond the intended purpose, as well as for failure to comply with the obligation to inform third parties about the use of AI;

system owners and/or operators are responsible for the lack of internal policies, procedures and access controls, failure to comply with the obligation to regularly check the system, lack of proper training of personnel to interact with AI, as well as for refusing to shut down the system in case of failures or risks;

auditing, certification and supervisory bodies acting as independent third parties are responsible for improper conformity assessment, the issuance of certificates without due diligence, as well as for the formal approach to assessing transparency, security, respect for human rights and the impact of the AI system on society.

Each subject is responsible within its competence and sphere of influence. The principle of proportionate accountability is introduced, which considers both formal obligations and the actual influence of the subject on the results of the AI system functioning.

18.3 Areas of responsibility.

Legal responsibility in the field of development, implementation and operation of artificial intelligence systems provides for several levels of legal consequences, which differ in nature, amount of damage, mechanisms of implementation and impact on subjects:

Civil liability arises in cases of material or moral damage to individuals or legal entities as a result of incorrect operation, errors or failures of the AI system. It can occur both in the form of an individual claim from the victim, and in the form of collective protection of rights in cases of damage to a large number of persons. Civil liability may also cover reputational damage, breach of contractual obligations or provision of unreliable results of automated decisions.

Administrative liability arises in case of violation of regulatory requirements, safety standards, ethical standards, or failure to fulfil the obligations imposed on the user, supplier or operator in accordance with the legislation on high-risk AI systems. It may include the imposition of fines, the suspension of the license, the obligation to remedy violations, as well as the obligation to inform supervisory authorities.

Criminal liability arises for the acts (action or inaction) of subjects that fall under the signs of a crime directly or indirectly related to the use of AI. Such crimes include:

manipulation or other unlawful interference in electoral processes and referendums, as well as in the formation and expression of public opinion using artificial intelligence (AI) systems;

creating or spreading disinformation using AI, which is harmful to national security;

committing discriminatory actions using AI systems on the grounds of race, gender, religion, belief or other characteristics protected by law;

developing, distributing or exploiting AI systems that harm human life or health, violate the right to privacy, the rights of the child or international obligations of the State.

Judicial practice in the field of responsibility for the use of AI is developing considering precedents, assessment of the degree of autonomy of the system and the actual possibility of exercising control over **its** functioning.

18.4 Audit, investigative and monitoring mechanisms.

All high-risk AI systems are required to undergo multi-level verification, evaluation, and monitoring procedures to ensure transparency, efficiency, and abuse prevention. Such mechanisms should be organizationally and functionally independent of the developer or operator of the system and guarantee that the interests of both users and affected persons are taken into account.

A mandatory periodic audit includes an external assessment of the compliance of the AI system with established standards of ethical safety, technical sustainability, human rights protection, gender equality, non-discrimination and explainability of results. The audit covers the analysis of algorithms, data sources, level of autonomy, control mechanisms, and rollbacks.

It is mandatory to keep and keep logs of all actions of the system and its interactions with users, which is a prerequisite for proper investigation of incidents, prevention of manipulation and ensuring transparency of functioning. Storage is carried out considering the requirements of personal data protection using hashing and digital signature technologies.

Public appeal of decisions provides for the creation of procedures according to which a person against whom a decision was made with the involvement of AI has the right to file a complaint, receive an explanation of the logic of the system's functioning, request a review of the decision by an authorized person, as well as appeal the results in court or administrative.

Incident reporting is the responsibility of providers, operators, and users of AI systems. In case of failures, incorrect behaviour, violation of rights or integrity of data and infrastructure, the relevant information is subject to immediate notification to the National Agency for the Oversight of Artificial Intelligence Systems. High-risk systems are required to have built-in self-alert mechanisms and digital risk tracking tools.

The state ensures regular analytics on the functioning of AI systems and the level of compliance with liability standards. The results of audits and monitoring are subject to disclosure in compliance with restrictions on state, commercial or personal secrets.

18.5 Accountability of public authorities in the field of AI application.

Public authorities that implement or use artificial intelligence systems in the field of public administration, security, education, healthcare, justice or social protection have a special responsibility to comply with the principles of transparency, legality and compliance with basic human rights.

Public openness means that each body is obliged to inform the public in advance about:

a) the purpose and functionality of the AI system;
b) the legal basis for its application;
c) volumes and types of data processed;
d) the impact of automated decisions on the rights and obligations of individuals;
e) a contact person or unit responsible for compliance with ethical and legal standards.

Annual reporting includes:

a) quantitative and qualitative indicators of system efficiency;
b) indicators of fairness, equality and non-discrimination;
c) a description of all cases of appeals or complaints about the actions of AI systems;
d) the results of internal and external audits;
e) recommendations for improving or suspending the use of systems.

Such a report shall be published on the official web portal of the body.

Immediate response mechanism: the responsible authority is obliged to:

a) suspend the use of the AI system in case of detection of human rights violations or a significant risk of harm;

b) to provide a person with the opportunity to exercise his/her rights through an alternative mechanism that is carried out without the use of automated systems;

c) initiate an independent audit within a period not exceeding five working days;

immediately notify the National Agency for Ethics and Supervision of AI of the measures taken.

Ethical Audit and Social Oversight.

Each AI system in the public sector is subject to an ethical assessment at least once every two years with the participation of members of the public and independent experts in the field of law, digital technologies and human rights. The results of such an assessment are subject to mandatory publication and are accompanied by an official comment from the management of the relevant institution.

Educational and explanatory duties.

Public authorities are obliged to provide regular training of employees on the responsible and lawful use of artificial intelligence systems, as well as to provide citizens with access to official explanations about the principles of operation of such systems, their impact on human rights and legal remedies in case of their misuse.

CHAPTER 19. THE PRINCIPLE OF TRANSPARENCY, OPENNESS AND EXPLAINABILITY OF ARTIFICIAL INTELLIGENCE SYSTEMS

All artificial intelligence systems operating on the territory of the State or used in the field of public decisions are subject to mandatory compliance with the principle of transparency and explainability. The principle of transparency and explainability guarantees, that the architecture of the system, algorithmic logic, functionality, declared constraints, data sources, learning processes, purpose, and potential implications of AI operation are understood, accessible, verifiable, and explanatory at every stage of the system's lifecycle.

The principle of transparency, openness and explainability in artificial intelligence systems provides for the mandatory provision of clear, accessible and timely information about the use of AI, its role in the decision-making process, the sources and nature of input data, as well as the logic of functioning and the potential consequences of decisions made by the system. No person can be exposed to AI without first informing about the algorithmic nature of interaction, the limits of autonomy of such a system, the level of human control, and available mechanisms for appealing its results. Covert, non-transparent or manipulative use of AI systems in processes that have legal, social, economic, political or other socially significant consequences is prohibited.

Transparency of AI systems should be ensured for all stakeholders, including:

developers — by maintaining internal documentation, knowledge management models, change logs and verification reports;

users — through accessible interfaces that explain the logic of the system, taking into account the level of digital literacy and socio-cultural context;

regulators — by providing a complete technical, methodological, legal and ethical dossier, including architectural schemes, model training protocols, risk analysis and accompanying analytical materials, with access in a machine-readable format;

third parties affected by AI — by creating mechanisms for public information, access to explanations, appeal, appeal, and independent monitoring.

Ensuring transparency includes mandatory disclosure of information about data sources, the legal status of the developer, the intended purpose of the system, its functional boundaries and declared technical and operational reliability limitations, as well as clearly and proactively informing the user about the degree of AI involvement in decision-making. It is mandatory to have an audit log with the recording of key events, configurations, changes and interventions of the human operator, as well as the implementation of regular internal and external transparency audits with proper access to the technical dossier.

The explainability of AI systems is the ability to provide a meaningful, structured, and contextually relevant interpretation of decision-making logic, covering the technical model, the significance (weight) of variable parameters, the method of calculation, the degree of human influence, and the level of uncertainty. The explanation should be adapted to the individual characteristics of the user or interested person, provided in an accessible language without excessive technocratic jargon, considering ethical, social and psychological factors. The system must provide multi-level explainability — technical, functional and ethical. Explainability of decisions is a key element of trust in AI systems and means the ability of a system or responsible actor to provide a reasonable, logical and understandable explanation of the causes, factors, mechanisms and identified limitations that led to a specific outcome. Such an explanation should include: a description of the key variables, data sources, the role of the person in the process, the identified limitations of the model, and the available appeal mechanisms. The explanations are adapted to the level of digital literacy of the user and, in the public sector, have legal force defined by law.

In the public sector, as well as in cases where AI decisions affect an individual's rights, status, access to resources, or social reputation, the lack of mechanisms for explaining or presenting a formal, obscure, or technocratic explanation is considered a violation of the principles of good governance. In such cases, the Central Executive Authority authorized in the field of artificial intelligence has the authority to stop the

operation of the system, initiate an investigation, revoke the permission to use and provide a public justification for the measures taken with guarantees of appeal.

Any AI system in the public or commercial sector should be accompanied by a complete information support adapted to the target audience, which discloses: the fact of its application, the name of the operator or administrator, its functional purpose and scope of use, the characteristics of the input data, the description of the principles and logic of decision-making, the mechanisms of human control, external audit, model updates and ensuring fairness. Such support of the AI system is provided in a user-friendly format, in an open machine-readable form, and is subject to mandatory publication in the National Register of Algorithm Transparency.

For high-risk AI systems, the availability of comprehensive documentation is a prerequisite for certification and ensuring explainability. The documentation should contain decision-making architecture, description of the main algorithms and their functional logic, human control protocols, risk mapping, logging systems (audit trails), model quality indicators (accuracy, reliability, absence of bias) and incident response scenarios. All comprehensive technical documentation is an integral part of the technical passport of the AI system and must be available in a differentiated manner: to the competent authorities — in full for the purposes of supervision, audit and certification; users — to the extent necessary for safe, conscious and responsible use; to victims — to the extent necessary for the effective protection of their rights and legitimate interests.

Each person has an inalienable right to identify the AI system and the responsible subject, access to information about the owner, operator or administrator of such a system, information about the scope, purpose and terms of processing of their personal data and ways to protect them, as well as the possibility of correcting or deleting such data (the right to be forgotten). This right applies to all forms of interaction with AI systems, regardless of the interface, level of anthropomorphism, or technological complexity. Suppliers and operators of AI systems are obliged to provide technical, organizational and legal conditions for the implementation of this right on an ongoing basis.

The authorized central executive body in the field of AI ensures the functioning of the National Register of Algorithm Transparency, a public digital platform that collects, stores and provides access to technical data sheets of systems, explainability models, audit results, API documentation, instructions for explaining decisions, indicators of trust and user ratings. The Register is a tool for democratic oversight, scientific analysis and legal control over the activities of algorithmic systems. Failure to comply with the obligation to submit information to the National Register of Algorithm Transparency entails legal liability.

To unify approaches to transparency and explainability of AI systems, the Central Executive Body in the field of AI develops and periodically updates the National Methodological Recommendations from explainability. Such recommendations define categories of models according to the degree of explainability, requirements for their documentation, examples of application in sensitive areas, templates of multilevel explanations, communication protocols with different audiences, and mechanisms for responding to complaints or requests. Compliance with these recommendations is mandatory during the certification of high-risk systems and is used as a guide by courts, auditors and human rights organizations when checking the legitimacy of algorithmic decisions.

19.1 The principle of transparency and explainability is one of the key ethical and legal foundations for the functioning of artificial intelligence systems. It requires that the architecture and logic of the system, its functionalities and limitations, data sources, learning processes, the purpose of application and the potential consequences of the work be clear, available for analysis and verification at every stage of its life cycle.

Transparency and explainability must be ensured by:

for developers — in the form of clear internal documentation, knowledge management models, change logging, ensuring replication of results, and independent verification;

for users — through interfaces and instructions that explain how the system works, especially in cases where AI decisions or recommendations affect their rights, interests or actions;

for regulatory authorities — by providing full access to technical, methodological, legal and ethical information about the system, including data on the models used, justification of algorithmic decisions and risk analysis;

for third parties that are directly or indirectly affected by AI decisions (e.g. consumers, patients, public associations), through notification, access to explanation, complaint and public monitoring mechanisms.

Transparency should not be limited to technical documentation only but should include mechanisms for social explanation presented in user-friendly language, considering cultural, professional and emotional contexts.

19.2 Transparency involves a set of actions that ensure the openness, accessibility and verifiability of all key aspects of the functioning of the artificial intelligence system for various groups of stakeholders (stakeholders).

In particular, the principle of transparency includes:

mandatory disclosure of information about the system developer or developers, data sources used to train models (indicating the origin, representativeness, legitimacy of collection and use), applied algorithmic architectures, the intended purpose of the system (use case), its functional limits, verified application scenarios and declared limitations (including statistical error probabilities and high uncertainty scenarios);

clearly informing users, regulators and third parties that a decision or recommendation is made or supported by an AI system, indicating the role of AI in this process: whether it is a fully automated decision, solution support or only auxiliary analytics;

creation and maintenance of a log of system decisions (audit log), which contains a detailed recording of key parameters, events, configuration changes, data used, as well as human interventions, with the possibility of retrospective analysis of system actions in a certain time period;

implementation of mandatory audit mechanisms — both internal (at the level of the developer, operator or authorized regulatory body) and external independent, with access to the full technical dossier, training and testing documentation, risk scenarios, as well as ethical conclusions in cases where they are mandatory for high-risk applications.

19.3 Transparency as a basic requirement in the field of artificial intelligence means the availability of clear, timely and accessible information to users, consumers or third parties regarding the fact of using the AI system, its role in the decision-making process, and the possible consequences of such participation. Transparency implies that every person who interacts with or is under the influence of an AI system must be clearly, directly, proactively and in an accessible form properly informed about the:

the use of AI systems in the process of providing services, decision-making or interaction;

the limits of the autonomy of the system and the nature and scope of human participation in making a final decision that has legal or factual consequences;

potential risks associated with algorithmic computation, including the likelihood of errors, the risk of algorithmic bias, and existing model limitations;

the mechanism for obtaining additional information, submitting an appeal to the operator or the competent authority and the appeal procedure.

Any covert, improperly identified, or manipulative use of AI in public or private processes, especially those with legal, social, economic, or political implications, is deemed unacceptable and entails administrative or criminal liability under this Code, depending on the severity of the violation.

19.4 Every AI system, regardless of risk level, should be accompanied by comprehensive information support that ensures transparency of its functioning, the ability to understand algorithmic logic and verify the validity of decisions. Such information should be accessible, understandable, adapted to the target audience (in particular, for persons without technical training), published in formats convenient for machine reading, and updated on an ongoing basis.

Mandatory components of this support are:

a clear indication of the fact of using the AI system in any interaction with a person, institution or community;

name, contact details, country of registration and legal responsibility of the operator or administrator of the system;

the functional purpose of the system, its scope, the type of tasks to be solved, and the contexts in which its use is allowed or restricted;

characteristics of the input data: categories, sources, legal bases for processing, methods of anonymization or pseudonymization, frequency of updates and restrictions on reuse;

a description of the logic of the decision-making process, including indicative variables and general principles of the structure of weighting factors, as well as the heuristics or other methods of information analysis used;

technical and organizational mechanisms for monitoring the quality and fairness of decisions, including procedures for human intervention, external auditing and determining the conditions for the use of self-learning mechanisms or updating models.

The information shall be stored in the National Register of Algorithm Transparency and shall be available in the public domain, subject to the restrictions defined by this Code, on an ongoing basis and with regular updates.

19.5 For high-risk AI systems, a mandatory is the availability of comprehensive, up-to-date, and properly published technical documentation, which is the basis for examination, certification, supervision and ensuring the explainability of decisions. Such documentation should be updated on an ongoing basis and include:

a detailed diagram of the decision-making architecture, including levels of data processing, interaction between subsystems, human intervention points and control protocols;

structured description of algorithmic logic: machine learning models used, basic hyperparameters (within general limits), validation methods, nature of loss functions, approaches to optimization, methods of training and adaptation to environmental changes;

a full-fledged system of logging and tracing actions (audit trail), which allows you to establish cause-and-effect relationships between input data, algorithm interpretation, intermediate calculations and the final result, with the provision of protection against forgery and the possibility of independent auditing;

formalized human intervention procedures that determine the moments of mandatory human control, the degree of autonomy of the system, the roles of responsible persons, as well as methods for suspending or changing the result;

description and mapping of risks: technical (for example, classification errors), social (for example, discriminatory consequences), legal (for example, violation of rights), using methods of quantitative and qualitative assessment, indicating the limits of permissible use and conditions for deactivation of the system;

detailed technical data on training processes: data sources (with confirmation of their legality and compliance with intellectual property rights), volumes and quality of sets, methods of data cleaning, test results, accuracy indicators (accuracy), recall, precision, F1-score, as well as other relevant metrics of robustness and reliability defined by international standards, with an analysis of the system's ability to generalize results, risks of overtraining and manifestations algorithmic bias.

All the specified documentation is an integral part of the technical passport of the system and must be

up-to-date, reliable and provided upon request:

to state bodies — in full;

users — to the extent necessary for safe and informed use;

victims — to the extent necessary to protect their rights, in accordance with the transparency and openness procedures established by this Code.

19.6 Explainability is a property of the AI system or the responsible entity to provide a clear, accessible and meaningful interpretation of the logic, structure and basis of decisions taken both for users and for supervisory authorities, auditors or persons affected by such decisions. Explainability of results is

a key element of trust in AI systems and implies the ability of the system or its operator to provide a clear, logical and understandable justification for any result, decision or action that affects the rights, freedoms or interests of an individual, with an explanation of the main factors and variables that led to this result, in an understandable form.

Explainability implies:

− the ability of the responsible actor or the AI system itself to formulate a structured, logical, relevant and contextually sound explanation regarding the logic of decision-making, the significance (weight) of variable parameters, the type of model, the method of calculating the result, the role of human intervention and the nature of uncertainties;

− adaptation of the explanation to the individual characteristics of the user: level of digital literacy, specialization, language, age, cognitive abilities and psycho-emotional state. The explanation must not only be technically accurate, but also understandable, objective and unbiased, stated without excessive technocratic jargon;

− provision of multi-level explainability, covering: technical level (type of model, values of coefficients, decision-making mechanism); functional level (goals, rules, logic of business processes, restrictions, exceptions); ethical level (assessment of potential risk, social impact, fairness and non-discrimination, justification of choice and rejected alternatives).

All decisions made by automated AI systems that have legal, financial, reputational, or other significant consequences must be accompanied by proper documentation in the form of a digital report that contains:

− explanation of the main factors that became key in determining the result, with a description of how each of them influenced the final decision;

− indication of the sources, categories and technical characteristics of the input data, methods of obtaining them, including information on relevance, accuracy, possible biases or limitations;

− a description of the person's role in the process — whether they were an observer, confirmer, reviewer or final decision-maker, with a record of the moments where there was or could have been an intervention;

− mechanisms for reviewing, editing, appealing or reversing the decision, with a clear definition of the deadlines for submitting the complaint, the procedures for its consideration and the contact information of the responsible authority.

Explanations should be adapted to the perception of not only technical specialists, but also ordinary users, taking into account the principle of "algorithmic literacy" of the population. In the public sector, the justification of AI decisions is an integral part of an administrative act, must have legal force and be subject to appeal in court.

19.7 In all cases, the person interacting with the AI system has an inalienable right for identification algorithmic subject, legal transparency regarding its owner and full awareness of the processing of their personal data. This right is a fundamental element of digital dignity and legal certainty in an algorithmic society.

Individual rights to transparency and data control when interacting with AI:

− the person must be clearly and proactively informed that they are interacting with an automated or semi-automated system and not with a human — regardless of the level of anthropomorphism, design, or voice interface imitation;

− information about the person, organization or authority that is the owners, customers or administrators of the relevant AI system must be provided, indicating their legal authority, area of responsibility, contact details and communication mechanisms, in a form accessible to the user, including in an open machine-readable format;

− a person has the right to find out what his/her personal, biometric, behavioural or other data has been collected by the AI system, on what legal grounds this has occurred, how the data has been processed, with whom it has been or may be shared, what protection mechanisms and restrictions on their use are in place, as well as what possibilities exist for viewing, editing or deleting such data within the time limits, determined by the legislation of the State.

These provisions are mandatory for all providers and operators of AI systems and are subject to continuous monitoring by the authorized supervisory body for the ethics and transparency of artificial intelligence, with the application of responsibility for their violation.

19.8 In the public sector and also in all cases, when decisions of an AI system have direct or indirect legally significant consequences for an individual or group of persons, ensuring explainability is an unconditional prerequisite for the admissibility of using such a system. In particular, we are talking about decisions that affect the rights, obligations, social status, access to services, level of security or reputation of a person.

The lack of mechanisms for explaining AI decisions or presenting a formal, overly technocratic or insufficiently understandable explanation that does not allow a person to understand the logic of the decision, its grounds, as well as the possibility of appeal or revision, is recognized as a violation of the principle of good governance.

In such cases, the Authorized Central Executive Body in the field of AI, in accordance with the supervision procedures and in accordance with the procedure established by law, has the right to apply a temporary suspension, administrative suspension of operation or a complete ban on the operation of the relevant system, with guarantees of appeal and public review (revision) of decisions of the competent authority. Methodological recommendations should include:

− categorization of types of AI models according to the level of explainability (for example, "white-box" — open models, "interpretable black-box" — interpreted closed models, "opaque systems" — opaque systems) and the corresponding requirements for explanations;

− sectoral examples of technical, functional and ethical explainability with standard documentation templates, examples of complaints and clarifications;

− requirements for the communication form of explanations for different audiences (citizens, civil servants, judicial authorities, journalists, auditors);

− typical samples of multilevel explanations based on specific examples (in particular, in the areas of social benefits, criminal justice, credit scoring, recommendation advertising);

− algorithms for interacting with feedback mechanisms, procedures for appealing and correcting decisions based on explanations.

These guidelines are mandatory for consideration during the certification of high-risk AI systems and are used as a guide by courts, supervisory authorities, auditors, and human rights institutions when analysing the legitimacy of decisions made by artificial intelligence systems.

19.9 The authorized central executive body in the field of AI is obliged to create and ensure the functioning of the National Register of Algorithm Transparency, a public digital platform that collects, stores, systematizes and provides open access to key parameters of the functioning of algorithmic systems.

The following must be published in this register:

− technical data sheets of AI systems, which contain data on the architecture, types of models, purposes of use, limitations, risks and limits of application;

− transparency and explainability models, including typical decision scenarios, logic principles, and impact assessments;

− the results of ethical audits, social testing, legal assessment and verification of compliance with fundamental rights;

− open APIs and documentation for them — taking into account security and intellectual property protection regimes;

− digital instructions for explaining decisions — in the form of graphic diagrams, text explanations, videos or interactive visualizations;

− trust and verification indicators, including risk indicators, expert assessments, feedback from users, as well as the methodology for calculating them.

The National Register of Algorithm Transparency operates on an ongoing basis, is subject to regular updates and is a tool for democratic oversight, legal control, scientific analysis and accountability of AI system providers to society and the state.

CHAPTER 20. THE PRINCIPLE OF FAIRNESS, NON-DISCRIMINATION AND INCLUSIVENESS OF AI SYSTEMS

The principle of fairness, non-discrimination and inclusion establishes a fundamental legal and ethical requirement for the design and development of, implementing and using AI systems in a way that makes it impossible to consolidate, automate, or reproduce social inequality, discrimination, social exclusion, or algorithmic bias.

No AI system may create or exacerbate unfair practices, either directly or indirectly, based on race, ethnicity, national origin, sex, language, religion, political or philosophical beliefs, sexual orientation, gender identity, age, health, disability, social or economic status, or any other protected ground provided for by international and national law.

All AI systems should be designed considering the concept of "equity by default", the principle of diversity of social experiences, the participation of vulnerable groups in the development process, as well as structural inclusion — linguistic, visual, cultural and functional.

It is strictly forbidden to implement AI systems that demonstrate or create discriminatory effects, regardless *of the developer's intention or the nature of algorithm errors. Such systems include models built on distorted or non-representative data; algorithms that have a disproportionate negative impact on vulnerable groups; architectures that ignore cultural specifics; systems that have not been audited for algorithmic bias or do not contain mechanisms for its correction.*

All AI developers, suppliers, and operators are required to implement effective and publicly documented measures to monitor, prevent, and eliminate any form of discrimination or exclusion. Such measures include: applying equity metrics considering socio-demographic diversity; regular auditing of data and models; mechanisms for attracting representatives of vulnerable groups; integration of ethical constraints into the logic of the model at the design stage.

In the public sector, the use of any AI systems that lead to a disproportionate negative impact on protected populations is prohibited, including by exclusion, distortion of representation or restriction of access to services. Each system intended for implementation in the public sector is subject to a preliminary comprehensive assessment of the impact on human rights, social inclusion, gender equality and risks of discrimination, which is carried out with the participation of independent experts, representatives of civil society and human rights organizations. The results of the assessment are subject to public disclosure and independent review. The lack of assessment or negative results is a sufficient legal basis for a complete ban on the implementation and operation of such a system until the risks that threaten the principle of fairness are eliminated.

The National Ethics and Non-Discrimination Authority is obliged to prepare and publish the AI Fairness Report every year, which is a strategic tool for overseeing compliance with the principle of non-discrimination. The report should include analysis of detected incidents; classification by technology, sectors, severity of consequences and legal status of victims; assessment of the effectiveness of response measures; Recommendations for improving the regulatory framework.

The report is submitted to national legislative and executive authorities, authorized bodies for artificial intelligence and non-discrimination, as well as to civil society, professional communities and relevant international organizations. The conclusions of the Report are considered when updating national strategies in the field of digital human rights, ethical and legal regulation of artificial intelligence.

20.1 The principle of fairness in the field of artificial intelligence defines the normative and ethical requirement to prevent the automation of social injustice, the reproduction of historical discrimination, the legitimation of biased social or power or systemic exclusion. This implies that AI systems must be designed, learned, and implemented in such a way that:

− not to create or consolidate new forms of discrimination on any of the grounds prohibited by international and national law;

− ensure equal access to opportunities, services and public participation regardless of the social status, physical or cognitive characteristics, origin or beliefs of the person;

− be inclusive in their architecture, structure, semantic design, language support, visual adaptability, and use cases;

− to consider the needs of groups that are traditionally in a state of social vulnerability or structural exclusion through their participation in the process of development, testing, public consultations and ethical examination.

The principle of fairness is implemented through the concepts of fairness-by-default, plurality of inputs, and participatory AI design, and is a prerequisite for certification of high-risk AI systems.

It is prohibited to use AI systems that directly or indirectly cause discriminatory effects, lead to marginalization, exclusion or bias — regardless of whether these consequences are the result of intentional programming, poor training, or uncontrolled application.

Prohibited systems include systems that:

− create, deepen or legitimize discriminatory practices on any of the grounds protected by international treaties and national legislation, in particular on the grounds of race, ethnic or national origin, colour, sex, language, religion or creed, political or other opinion, membership of national minorities, property or social status, age, disability, sexual orientation, gender identity, state of health, citizenship or other legal basis status determined by law;

− allow systemic or individual bias against persons in vulnerable situations, including children, the elderly, people with disabilities, refugees, internally displaced persons, members of ethnic, religious and linguistic minorities, as well as persons belonging to LGBTIQ+ communities;

− do not consider or distort the cultural, linguistic, religious, social, gender or regional characteristics of the respective communities, which leads to the exclusion, lack of representation or misinterpretation of the interests of such groups;

− built on training data that contains detected or potential systematic biases (algorithmic bias), without proper protocols for detecting them, without mechanisms for correcting models, without the use of multidimensional fairness tests or without proper auditing for bias.

20.2 All AI vendors and operators are required to implement systemic, verifiable, and properly documented and publicized measures, are aimed at preventing discrimination, correcting algorithmic bias and strengthening inclusivity in the processes of creating and using artificial intelligence systems. Mandatory measures include:

− application of fairness metrics that consider the demographic, social, regional and cultural characteristics of target audiences, through multi-group accuracy assessment, FPR/FNR parity, group justice metrics, equivalent impact indices and other quantitative indicators of fair distribution of results;

− conducting regular audits of data sets, model architecture, training, validation and decision-making processes to identify direct or hidden biases (bias), with mandatory recording of conclusions and publication of reports on corrective actions taken;

− ensuring the participation of representatives of various social, gender, ethnic, age and vulnerable groups in the processes of testing, approbation, public discussion and ethical examination, which allows reflecting the diversity of social experiences and avoiding exclusionary or discriminatory decisions;

− adherence to the principle of "fairness-by-design", which provides for the inclusion of ethical, legal and social safeguards in the algorithmic core of the system at the design stage, including the use of fair algorithms, restrictions on the use of highly sensitive variables, procedures for ensuring inclusivity in sample data, as well as mechanisms for liability for unfair results.

20.3 In the public sector, the use of artificial intelligence systems that directly or indirectly have a disproportionate negative impact on certain social, ethnic, demographic or other legally protected groups, regardless of the developer's intentions or the nature of the application, is prohibited. Such impacts include not only discriminatory consequences, but also systemic inequalities of access, restriction of rights, distortion of representation or undue restriction of user access to services.

Any AI system intended for use in the public sector is subject to a mandatory prior impact assessment on human rights, gender equality, cultural sensitivity, social inclusion and discrimination risks (AI impact assessment) before its implementation. Such an assessment includes:

− analysis of direct and indirect impact on different social groups;
− modelling of typical application scenarios from the point of view of legal consequences;
− assessment of the validity of the choice of training data and algorithmic decisions;
− participation of representatives of civil society, human rights organizations, sociologists and experts on non-discrimination in the evaluation commission;
− public disclosure of the results of the assessment with the possibility of independent peer review.

The lack of proper assessment or obtaining unsatisfactory results is a sufficient legal basis for a complete ban on the use of such a system in the public sector until the identified risks are eliminated.

20.4 The National Ethics and Non-Discrimination Authority is obliged to: annually compile and publicly publish the Artificial Intelligence Equity Report (AI Fairness Report), which is a strategic document for monitoring, analysis and prevention of discriminatory or exclusionary consequences caused by the use of artificial intelligence systems in the territory of the State.

The report should contain:

− statistics and analytical description of detected cases of discrimination, bias, inequality, disproportionate impact or social exclusion due to the operation of AI systems;
− classification of incidents by type of technology, sectors of application (public administration, education, medicine, social services, commercial sector, etc.), severity and scale of consequences, as well as the legal status of the affected persons;
− review of the effectiveness of the response measures taken by operators, suppliers or government authorities after the detection of such incidents;
− recommendations for updating policies, regulations, procedures for the development and implementation of AI systems, considering the principles of ethics, fairness and inclusiveness;
− consolidated national indicators on algorithmic fairness, group equality, adaptability of appeal procedures, effectiveness of public oversight and public engagement.

The report is submitted to the national legislative and executive authorities of the State, authorized bodies for human rights and artificial intelligence, as well as to international organizations, civil society and the expert community, and is used as an official basis for the annual update of national strategies in the field of digital human rights and ethical regulation of artificial intelligence.

CHAPTER 21. THE PRINCIPLE OF SAFETY, RELIABILITY AND STABILITY OF AI SYSTEMS

The principle of safety, reliability and resilience of artificial intelligence systems establishes mandatory requirements for design, deployment, and operation, updating and maintaining AI systems in order to ensure physical, psychological, economic, digital, environmental and social security of an individual, society and the state.

AI systems should be developed and implemented in such a way as to minimize the risks of harm to life and health, psycho-emotional state of a person, guarantee the integrity of personal and critical data, exclude harmful effects on the infrastructure and exclude undesirable or unpredictable scenarios of algorithm behaviour.

Functional security, cybersecurity, social security, and context-sensitive adaptability are components of integrated security implemented through technical, organizational, procedural, and ethical mechanisms.

Each AI system must have a defined area of application within which it has been tested, certified and monitored. Self-learning without proper control and verification, automatic architecture changes, non-transparent decision-making, or transferring logic to new contexts without adaptation are prohibited.

A secure system should provide multi-factor detection of failures, emergency shutdown in case of a threat, restoration of functionality after an error, audit of the trace of decisions made, as well as the possibility of a complete transfer of control to a person.

The reliability of the AI system is determined by its ability to consistently perform the assigned functions with a high level of accuracy, repeatability, and predictability — both in normal and stressful conditions. This requires the introduction of mechanisms of technical redundancy, self-diagnosis, continuous monitoring of parameters, backups and compliance with reliability standards confirmed by certificates.

The stability of the AI system lies in its ability to function without losing critical characteristics in the face of external attacks, internal errors, crises, sabotage or force majeure. Ensuring resiliency requires built-in mechanisms for multi-layered protection, authentication, cryptographic protection, anomaly detection, access control, regular security updates, and isolation and previous state restoration protocols.

Entities responsible for developing, integrating, implementing, operating, or decommissioning an AI system are required to implement risk management systems that include threat identification, risk matrices, preventive and compensatory measures, early incident warning, and crisis response.

All incidents that threaten human life, health, human rights or critical infrastructure are subject to immediate reporting, but no later than 24 hours from the moment of detection to the National AI Safety Monitoring Center, with the mandatory submission of a technical assessment, analysis of the causes, description of the measures taken and a plan to prevent recurrence.

The central executive body in the field of artificial intelligence approves and updates the National Catalog of Safety, Reliability and Resilience Standards, which is mandatory for all high-risk AI systems operating in the territory of the State. The catalog contains technical regulations, testing procedures, accreditation criteria, documentation templates, and control mechanisms. Failure to comply with its provisions is the basis for the application of administrative, financial or other liability provided for by the legislation of the State in the field of digital security and data protection.

21.1 Safety is a fundamental principle that provides for the systematic provision of physical, psychological, economic, digital, environmental and social protection of an individual, society and the state during the development, implementation and operation of artificial intelligence systems.

The security of an artificial intelligence system includes a combination of the following requirements:

minimizing the risk of harm to human life, health or psycho-emotional state, in particular due to improper AI decisions, exposure to visual, audiovisual or other multimedia content, as well as behavioural or cognitive effects of a manipulative nature [199];

prevention of negative economic impact due to algorithmic errors, fraudulent scenarios or failures in automated financial transactions [200];

prevention of threats to national security, defence capability or continuity of critical infrastructure, if AI is used in defence, energy, transport, communications, etc. [201];

guaranteeing digital security, i.e. protecting personal, biometric, medical and financial data from loss, theft, unauthorized processing or transfer [202];

prevention of algorithmic creation of informational influence or emotional and behavioural pressure, especially on children, persons with disabilities, military personnel, patients or prisoners [203, 204].

The security of the AI system should be ensured by technical (architectural), organizational (control), legal (regulatory) and behavioural (ethical) mechanisms, which should be adapted to each application context and regularly updated in accordance with technological and social changes [205].

21.2 All AI systems, regardless of their application, are required to comply with a set of integrated safety standards that cover the following components:

functional *safety* — the ability of the system to guarantee the correctness and predictability of operation in accordance with certain parameters;

cybersecurity — protection against external and internal threats aimed at violating the integrity, availability or confidentiality of data;

social safety — prevention of negative impacts on a person, society and the state, including information manipulation or discriminatory practices;

context-aware *robustness* — the ability of the system to act correctly in changing conditions, taking into account new application contexts.

In particular, the AI system should:

act exclusively within a predetermined area of application (safe scope enforcement), avoiding uncontrolled transfer of logic to new environments or situations without confirmation of adaptability [206];

have multi-level mechanisms for detecting anomalies, predicting failures, actively responding and, if necessary, automatically limiting or suspending activities in case of errors, hacker attacks or deviations from expected behaviour [207];

prevent unsupervised self-modification outside of established parameters or without proper human supervision (unsupervised self-modification), which may cause unpredictable or dangerous behaviour of the system [208];

provide full traceability of the life cycle of models, changes in architecture, algorithms, parameters and solutions (traceability & auditability), which allows for retrospective analysis of each result;

be designed with built-in backup failure response scenarios, including Graceful Degradation, Emergency Shutdown, Human Takeover, and Recovery Protocols [209].

Each of these mechanisms must comply with current international standards and be confirmed by certification or independent expert opinions.

21.3 Reliability of the AI system means its ability to guarantee the performance of certain functions with a high level of accuracy, reliability and stability throughout the entire life cycle of operation - both within the predicted scenarios and in case of possible deviations from them. This parameter is critically important for high-risk systems used in the field of healthcare, justice, energy, transport, public administration and other critical or socially significant industries.

All high-risk AI systems are required to:

undergo mandatory reliability certification in accordance with technical standards and national regulations, with confirmation of the stability of results based on simulations, field tests and stress tests [210, 211];

have built-in automated self-diagnosis and functional monitoring modules that continuously monitor productivity, accuracy, resource consumption and compliance of results with established limits, with mandatory informing of the operator in case of detection of critical deviations [212, 213];

Ensure redundancy by using hot-standby systems, multi-tier request processing architecture, data and module duplication, and stabilization protocols to support operations in the event of congestion, cyberattacks, or critical infrastructure failures [214, 215].

Reliability is not only a technical, but also an organizational category, encompassing:

− availability of procedures for regular validation and reassessment of risks;

− implementation of disaster recovery *plans*;

− systematic logging of incidents and failures;

− training and advanced training of personnel responsible for the operation and supervision of the system.

21.4 Resilience of AI systems — This is their ability to maintain integrity, stability and functionality in the face of external cyberattacks, internal errors, emergencies, sabotage, destructive interference or radical changes in the operating environment. This is a key characteristic of the viability of high-risk digital systems, which guarantees the continuity of their operation even in crisis situations.

To achieve sustainability, the AI system must:

include multi-layered preventive and reactive (defence-in-depth) measures that combine request verification and authentication, cryptographic encryption mechanisms, isolation of data transmission channels, digital identification of components, and integrity verification of algorithmic modules before each launch [216, 217];

have behavioural anomaly detection algorithms capable of real-time identification of non-standard, harmful or potentially dangerous activity of a user, administrator or external digital agent, as well as detect indicators of internal sabotage (including deliberate training of the system on toxic or false data) [218];

provide access control to critical functions, using multi-factor authentication, role and authority matrices, mandatory logging of all actions, automatic notifications of attempts to violate access policies, and the ability to immediately block the account [219];

provide for regular security updates implemented through automatic patch distribution systems, compatibility testing of new modules, saving backups of stable versions, mechanisms for rolling back changes in case of vulnerability detection, and automatic scanning for zero-day exploits [220].

Resilience is a mandatory criterion for the accreditation of critical digital systems, and their level of resilience is subject to annual confirmation through independent audits and the publication of public reports on all confirmed incidents that occurred during the reporting period.

21.5 Manufacturers, suppliers, integrators and operators of AI systems have direct legal, organizational and reputational responsibility for the implementation and continuous improvement of risk management systems covering the entire life cycle of the system — from the conceptual design phase through the development, testing, launch, operation, upgrade phase to the final retirement phase.

The risk management system should include:

− identification and classification of potential threats at each stage of the life cycle;

− development of mechanisms for reducing and levelling risks (mitigation strategies);

− regular updating of risk matrices in accordance with new technological conditions, threats or operational experience;

− appointing responsible persons for risk management in each key unit or chain of operation;

− creation of internal mechanisms for early warning of incidents (early warning systems), as well as crisis response procedures.

All actual incidents and detected circumstances that could potentially pose a threat to the life, health, rights, freedoms or safety of users, society or the state are subject to mandatory immediate notification to the National Centre for Monitoring the Safety of Artificial Intelligence.

The message must contain:

− initial technical assessment of the incident;

− analysis of the causes and vulnerabilities that caused the incident;

− description of the urgent measures taken to neutralize and prevent the consequences;

− an action plan to prevent the recurrence of similar incidents in the future.

The fact of failure to submit or intentional concealment of information about the incident is the basis for the application of sanctions, revocation of certificates, temporary suspension of activities and inclusion

in the register of violators in the field of critical digital technologies in accordance with this Code and other national laws.

21.6 The central executive body in the field of AI is obliged to to develop, regularly update and implement the National Catalog of Safety Standards in accordance with the established procedure, reliability and stability of AI systems is an integrated regulatory and technical act containing mandatory norms, methods, procedures, technical regulations and control criteria.

The catalog includes:

− mandatory standards for the architecture, testing, auditing, and maintenance of high-risk AI systems;

− methods of stress testing, threat modelling, functional viability, operational and cyber resilience testing;

− requirements for incident management, failure investigation procedures, reporting and response protocols;

− recommendations for interoperability, secure updates, encryption and configuration of redundant mechanisms;

− typical templates of technical and legal documentation, certificates of conformity, certification scenarios, ISO/IEC compliance matrices.

The National Catalog is mandatory for all suppliers, developers, integrators and operators of high-risk AI systems, as well as for state bodies that exercise control in this area, operating on the territory of the State. Failure to comply with it entails administrative, financial or institutional liability in accordance with digital security legislation.

CHAPTER 22. THE PRINCIPLE OF ACCOUNTABILITY AND RESPONSIBILITY IN THE FIELD OF AI

The principle of accountability ensures the mandatory legal, ethical and social responsibility of the actors involved in the creation, use and supervision of artificial intelligence systems. It provides for the establishment of clear, transparent, formalized and personalized responsibilities regarding the results and impact of the functioning of algorithmic systems, including legal, technological, ethical and social risks.

At each stage of the AI system's life cycle — from exploratory design, data acquisition and processing, model training, testing and validation, to implementation, practical use, modifications, maintenance and decommissioning — the responsible entity or group of entities must be clearly defined with the fixation of the boundaries of responsibility, role in the system architecture and the level of access to the parameters of algorithmic decisions.

All persons, organizations, institutions or enterprises involved in the creation, integration, supply or operation of AI systems are required to implement and ensure the functioning of structured accountability mechanisms. This includes: the creation of responsible departments (responsibility office) in the internal organizational structure; maintaining complete documentation of each phase of the life cycle; systematic ethical and technical reporting, on human rights, safety and non-discrimination; ensuring transparent interaction with state regulators, consumers, affected persons, supervisory authorities and civil society representatives.

The accountability of an AI system goes beyond a technical or administrative tool and includes legal, ethical, social, and reputational responsibility for every decision made by or based on an algorithmic system. In case of damage to an individual, legal entity or public or state interests, liability arises within the relevant stage of the AI life cycle and depends on the role, control, responsibilities and degree of prudence of each of the participants.

Legal responsibility is distributed between the developer (architecture, algorithms, data); supplier (configuration, installation, compliance with instructions); operator (operation, parameter changes, supervision); by the subject that manages the data (data owner) — for the legality, quality and objectivity of the data sets used. In cases of joint liability, a joint or proportional distribution model with the right of recourse claims between subjects is allowed.

A centralized Register of responsible subjects of artificial intelligence systems is created and operates in the State. It must record legal entities and individuals — developers, suppliers, operators; authorized persons for liability management; results of audits, certifications, ethics control; structural descriptions of areas of responsibility within the organization; information on documented violations, complaints, regulatory measures and sanctions, indicating dates, responsible persons and decisions made. The Register is open for public access in the part that does not contain personal or protected data, with a secure digital interface and category search functionality.

Failure to comply with the principle of accountability entails legal liability in the form of civil (compensation for damages), administrative (fines, orders, suspension of activities), disciplinary (dismissal, deprivation of special clearance, revocation of permits or certificates) and criminal (in case of fraud, discrimination, violation of human rights or harm to life, health or safety), depending on the severity of the offense. In addition, by the decision of the authorized body, additional measures of influence may be applied, in particular: prohibition of the operation of the relevant system, inclusion of the subject in the Register of Digital Rights Violators, blocking participation in public tenders or international partnerships, revision or cancellation of previously issued permits, grants or contracts.

The principle of accountability ensures the functioning of the artificial intelligence system within the framework of democratic control, social trust, transparency and the rule of law.

The principle of accountability implies the presence of clear, transparent and structured responsibility of all actors involved in the creation, deployment, use and maintenance of artificial intelligence systems. This means that at every stage of the AI system's life cycle — from exploratory design, data collection, processing, and legitimacy, model training, testing, and validation, to implementation, update and adaptation, daily operation, and modification — there must be a clearly identified entity responsible for the consequences of the algorithmic system's action or inaction.

Accountability encompasses:

responsibility for the content, quality and legality of the data used to train models [221];

legal and regulatory responsibility for design decisions in the system architecture that have a direct impact on its behaviour and consequences [222];

ethical and social responsibility for the way AI systems are implemented and applied in environments where risks to human rights or public welfare may arise;

reputational and institutional responsibility for transparency, openness to oversight, providing explanations and remedying harm if it occurs [223].

The principle of accountability requires the implementation of a systemic multi-level management model, where each participant in the AI ecosystem acts within its authority, but is responsible to society, regulator, user and other stakeholders for their own role in the functioning of the algorithmic environment.

All entities involved in the creation of, implementation, maintenance or commercialization of artificial intelligence systems, are obliged to ensure the effective functioning of accountability mechanisms by taking the following measures:

detailed documentation of each phase of the AI system lifecycle, including the stage of data collection and cleaning, architectural design, rationale for the choice of algorithms, training, testing, implementation, modification, update, and decommissioning processes. Documentation must be stored in secure audit repositories and be available for supervision by authorized bodies;

creation of a separate accountability office (office) in the structure of each organization working with AI systems, who has defined powers to monitor algorithmic decisions, assess risks, handle complaints, maintain an ethical journal of decisions (AI Ethics Logbook) and interaction with the regulator and the public;

providing effective and accessible feedback procedures for users, affected or stakeholders, including:

(a) a system for receiving complaints and requests;

(b) procedures for reviewing or cancelling results;

(c) clearly defined mechanisms for compensating for damage caused by the incorrect operation of artificial intelligence systems;

(d) the ability to apply to an independent ombudsman for artificial intelligence;

mandatory external independent accountability audits (accountability audits) at least once a year, with the involvement of certified conformity assessment bodies that carry out the audit:

(a) effectiveness of the risk management system;

(b) compliance with legislation and ethical standards;

(c) the level of public transparency, accessibility of reporting and incident response mechanisms.

The results of audits are published in the public domain, considering confidentiality and security requirements.

22.1 Legal liability for damage caused to individuals, legal entities or public law institutions because of the action or inaction of the artificial intelligence system is determined according to the principle of differentiated liability attribution and depends on the stage of the life cycle and as a result of whose actions or omissions this damage occurred. The distribution of responsibility is carried out considering the role, control over the system, due diligence and the level of impact on the parameters of the AI system [224, 225].

Including:

per developer (developer of the software core, architecture, algorithms, language models or machine learning tools) — in cases where the damage is caused by design errors, flaws in the algorithm's logic, inexplicability or opacity of the model structure, lack of bias tests, or the use of defective, unrepresentative or distorted training data;

for a supplier (technological integrator, vendor, service distributor) — in case of poor-quality installation, incorrect configuration, non-compliance with technical instructions, non-compliance with certification or standardization requirements, violation of licensing conditions or placement of the system in an inappropriate environment without proper testing;

to the operator (institution, enterprise, public or private entity that manages the AI system in the process of its practical use) — if the damage occurred as a result of violation of the instructions for use, change of initial parameters without additional certification, inadequate response to risk warnings, or due to the absence or inadequate human oversight, where it was necessary [226];

to the data owner or provider — in cases where illegally collected, technically distorted, knowingly falsified, biased, or insufficiently representative data was used, which led to erroneous or discriminatory decisions by the AI system.

In the case of joint fault or interrelated violations, liability is defined as joint and several or proportional depending on the level of influence of each of the participants. The division of responsibility does not exclude the right of the victim to apply to any of the perpetrators with a subsequent recourse claim for compensation (right of recourse) between the responsible subjects of the chain of creation and operation of the AI system [227].

22.2 A nationwide Register of Responsible Subjects of Artificial Intelligence Systems is being created, maintained and implemented — a centralized digital database containing up-to-date, verified and structured information on all entities responsible for creating, implementing, operating or auditing AI systems in the State.

This Register must record:

legal entities, individual entrepreneurs, institutions and organizations that are developers, suppliers, operators or administrators of AI systems;

authorized persons for liability management, their official means of communication and defined powers;

information on the completion of conformity assessment procedures, audits, certifications, licensing, as well as the results of accountability and ethics audits, indicating the dates and bodies that carried them out;

documented structures of responsibility within the organization (subordination, functional distribution, control mechanisms);

Officially recorded cases of incidents, violations, recourse claims, or administrative sanctions applied to the subject during the last three years, indicating the legal bases.

The Register is open to the public through a secure digital portal integrated into the system of the National Register of e-Governance, in the part that does not contain personal or protected data, in accordance with the legislation on personal data protection and state secrets. Public oversight institutions, state regulatory authorities, consumers and human rights organizations have the right to view and filter data into certain categories.

The inclusion of a record of the responsible entity in the Register is a prerequisite for obtaining permission to deploy a high-risk AI system in the public sector or any area that has a significant public impact (medicine, education, justice, law enforcement, etc.).

22.3 Failure to comply with the principle of accountability in the field of artificial intelligence entails individualized legal responsibility in the forms provided for by the current legislation of the State and international legal acts. Types of liability include:

civil liability (compensation for material or moral damage, compensation for losses, compensation for lost profits, restitution);

administrative liability (fines, orders to eliminate violations, temporary suspension or restriction of activities);

disciplinary liability (dismissal of responsible persons, deprivation of access to critical infrastructures, cancellation of certificates of conformity or permits);

criminal liability (for intentional actions using AI systems for the purpose of fraud, discrimination, violation of human rights, illegal data collection, or harm to life, health or safety).

In addition, depending on the degree of threat and the nature of the violation:

a temporary or indefinite ban on the operation of the relevant AI system may be applied;

by the decision of the authorized body, licenses, certificates of conformity and permits for access to high-risk applications may be revoked or revoked;

the subject must be entered into the Register of Violators of Digital Law, with the publication of information about the type of violation, decisions of regulatory authorities and the consequences that have occurred;

The presence of an entry in the Register of Violators is a legal basis for temporarily blocking the entity's participation in public tenders, international partnerships, as well as for initiating procedures for reviewing already issued permits or grants.

CHAPTER 23. THE PRINCIPLE OF HUMAN-CENTRICITY IN AI SYSTEMS

The principle of human-centeredness is a fundamental ethical and legal basis for the functioning of artificial intelligence systems, which guarantees the priority of dignity, Autonomy, human rights, freedoms, privacy, security, and well-being in the processes of designing, deploying, using, and monitoring AI systems. No artificial intelligence system can replace or eliminate human moral and legal responsibility in contexts with profound social, legal, or humanitarian implications.

AI systems should be designed and function as auxiliary tools that reinforce human decision in all areas of societal importance, rather than replace it. This approach requires adaptability to the individual needs of users, sensitivity to the context of interaction, and respect for cultural, linguistic, age, gender, psycho-emotional and social diversity. In all cases, AI systems must support a meaningful human decision, especially in the areas of health, justice, education, social protection, and security.

Any decision formed or supported by the AI system must remain under effective, timely and legally enshrined human control with the right to intervene, suspend, cancel or appeal. The use of artificial intelligence systems in full autonomy modes in high-risk human rights contexts is prohibited. Human-in-the-loop, human-on-the-loop or human-in-command interaction models or other equivalent models of human supervision are mandatory. The use of such systems in public administration, justice, social policy, health or security is allowed only if there are documented control procedures, intervention mechanisms and effective appeal channels.

Each person is guaranteed the following digital rights according to the principle of human-centeredness:

the right to be informed in advance about the participation of the AI system in decision-making that may affect its rights, freedoms or obligations (AI disclosure);

the right to an understandable explanation of algorithmic logic and influencing factors in an accessible form, considering the level of awareness of the user (right to explanation);

the right to prohibit the use of their personal or biometric data in the absence of proper legal basis or consent;

the right to protection from hidden neurobehavioral influences, manipulative targeting, or digital exploitation of user vulnerabilities;

the right to an alternative human decision (human review), as well as to fair compensation and restoration of violated rights.

In the field of development and implementation of AI systems, functioning in socially sensitive industries, the implementation of the principle of human-centeredness requires the involvement of future users, NGOs, vulnerable groups and professional communities to the processes of design, testing, model fairness analysis and social impact assessment. Systematic ethical examination is necessarily carried out at all stages of the life cycle of the system. Only a depersonalized approach to the user as a statistical object without considering his individual characteristics is prohibited. Empathic design is recognized as an indispensable element of such systems.

Confirmation of human-centricity is recognized as a prerequisite for certification, licensing and operation of AI systems in the State. For this purpose, the National Methodology for Assessing Human-Centeredness is created and applied, which contains criteria for compliance with human rights, ethical standards of international organizations, as well as requirements for the functionality of AI systems. These requirements include transparent informing users; functions of safe self-shutdown and effective human intervention; interfaces adapted and inclusive to the characteristics of different user groups; mechanisms for the participation of human rights organizations and public institutions in certification procedures.

Human-centricity is recognized as a basic condition for the legitimacy of artificial intelligence technologies in a democratic society and a fundamental principle of ethical law and order in the digital environment.

23.1 The principle of human-centricity is recognized as a fundamental ethical and legal basis for the functioning of artificial intelligence systems, which establishes the obligation that all actions related to

the development, implementation, use and supervision of AI are primarily focused on respecting and strengthening human dignity, autonomy, rights, freedoms, privacy, security and well-being of humans [228].

It is prohibited to use AI systems in a way that depersonalizes, devalues or replaces the role of a person in decision-making processes, especially in areas that have significant moral, ethical, legal or social consequences (justice, healthcare, education, social protection, law enforcement, etc.) [229].

A human-cantered approach requires ensuring that AI systems are sensitive to human needs, individual differences, cultural contexts, as well as the ability to transparently, understandably interact and adapt to the user. In all cases, AI systems should remain supporting rather than replacing human decision-making, especially when it comes to issues of life, health, freedom, justice, or dignity [230].

All decisions made by artificial intelligence systems must remain under effective and legally enshrined human control, be subject to verification, evaluation, and possible appeal. It is prohibited to use artificial intelligence systems in a mode when decisions with high legal, social, ethical or psycho-emotional consequences are made completely automatically without human involvement at any stage.

The following approaches are mandatory:

- Human-in-the-loop — a person is directly involved in the decision-making process [231, 232];
- Human-on-the-loop — a person monitors and can intervene or stop a decision[233];
- human-in-command — the person retains priority and has the final right to make or reject decisions [234, 235];
- or other equivalent models of human oversight recognized by international standards.

AI systems operating in the areas of public administration, justice, healthcare, security, child protection, social services, or human behaviour assessment are not allowed to be used without a documented human control procedure, suspension mechanisms, and effective appeals channels.

Human supervision must be real and effective, provide the ability to influence the outcome, ask questions, analyse the logic of the decision made, have the authority to cancel or change it, and be responsible for the intervention carried out.

23.2 The principle of human-centeredness guarantees every person who interacts with an artificial intelligence system inalienable digital rights aimed at protecting their autonomy, freedom of choice and security. In particular, all users, consumers or individuals affected by AI solutions have the right to:

clear, timely and understandable information about the use of the AI system in the decision-making process, with the obligatory indication that the result is formed by an automated tool, the definition of the responsible person and the specific amount of algorithmic participation (right to AI disclosure) [236];

Full access to the explanation of the logic, principles, variables, evaluation methods and justifications behind the automated decision in a form understandable to a non-professional user, with the provision of a multi-level explanation - technical, ethical and practical (right to explain) [237];

the ability to reasonably or unconditionally opt out of the use of their personal, sensitive or biometric data for the purpose of training, adapting or improving AI models, unless such processing is based on a direct legal basis, public necessity or voluntary informed consent;

protection against covert influences, neurobehavioral manipulation, subliminal stimuli, or forms of behavioural targeting by AI systems, especially in educational, medical, political, or law enforcement environments;

a guaranteed alternative — a re-examination of the automated decision by a competent person (human review), the right to file a complaint and receive a reasoned response, as well as the right to a fair, proportionate and humane restoration of the violated right or condition in case of negative consequences of the AI system [238, 239].

23.3 When developing, implementing and testing artificial intelligence systems designed to function in socially sensitive areas — such as healthcare, education, justice, social protection, protection of children's rights — the principle of human-centeredness is implemented not declaratively, but through specific procedural and institutional mechanisms that guarantee that human needs are taken into account at each stage of the life cycle of the system.

In particular, this involves:

ensuring the active participation of future users, representatives of civil society, vulnerable groups, patients, students, teachers, human rights defenders in the process of design thinking, testing, validation of models and analysis of the impact of decisions, which is the implementation of the principle of participatory design as an ethical and democratic standard;

systematic ethical review at all stages of the AI life cycle — from functionality planning to algorithm updates and adaptations, with the involvement of independent multidisciplinary commissions consisting of experts in the field of ethics, psychology, law, technology, and representatives of the public (ethics by design);

prohibition of depersonalized decisions that treat a person exclusively as a carrier of parameters or a statistical unit, without considering his individuality, context and dignity. Models should consider individual effects on a specific individual, not just an aggregated statistical group;

introduction of standards of empathic interaction, including sensitivity to language, cultural background, social status, age, gender, psycho-emotional state of the user;

providing interfaces that are accessible, user-friendly, respectful and inclusive of the cultural, linguistic, social, age and gender diversity of users.

23.4 The human-centricity of AI systems is subject to mandatory confirmation at each stage of certification, registration and verification of compliance by applying clearly defined indicators and procedures. To this end, the Central Executive Body in the field of AI approves, implements and ensures the application of the National Methodology for Assessing Human-Centricity, which is a mandatory element of the overall process of licensing, certification and risk assessment.

The methodology includes the following key elements:

verification of the compliance of the AI system with the principles and standards of human rights, the norms of the Constitution of the State, the provisions of the Universal Declaration of Human Rights, the European Convention on Human Rights, as well as recommendations and guiding documents of the Council of Europe, UNESCO and the UN on the ethics of artificial intelligence;

- expert analysis of the clarity, intuitiveness and accessibility of user interfaces, considering the age, language, educational, mental and physical characteristics of different social groups, with a special focus on ensuring accessibility for people with disabilities, the elderly, minors and users with low digital literacy;
- availability of functionality in the AI system that provides:

(a) emergency self-shutdown/suspension of AI at the request of a user or administrator (emergency off) [240];

(b) guaranteed possibility of human intervention at every critical stage of decision-making (human override) [241];

(c) Ensuring transparent communication on data sources, processing methods, decision logic, and possible risks (transparent feedback) [242];

- participation of representatives of civil society, human rights organizations, ethics councils, academic institutions, as well as users with experience with AI systems as a member of the certification commission, with the right of an advisory vote and the possibility of submitting a separate opinion, which is subject to mandatory consideration in the materials of the certification process.

CHAPTER 24. THE PRINCIPLE OF LEGITIMACY AND CONSTITUTIONALITY

The use of artificial intelligence systems on the territory of any State is allowed only if they comply with the Constitution or the Basic Law of this State, laws and other regulations adopted in accordance with it, international treaties, consent to be bound by which has been granted by the relevant national authorities, as well as fundamental principles of law, in particular the rule of law, legal certainty, prohibition of abuse of law and good faith.

It is prohibited to introduce, use or support the functioning of any artificial intelligence system if its activities violate human rights and freedoms guaranteed by the Constitution or the Basic Law of the state, lead to the undermining of the foundations of the constitutional order, attempts to usurp power, interfere with the independence of the judiciary or manipulate democratic procedures, and are carried out without an adequate legal basis, preliminary impact assessment, independent expert opinion and effective mechanisms public control.

No public authority at the national or local level, as well as no legal or natural person, regardless of the form of ownership and organizational and legal status, has the right to use artificial intelligence systems in a way that contradicts the principles of the separation of powers, the independence of the judiciary, the autonomy of local or regional governments, as well as undermines or jeopardizes the independence and proper functioning of representative bodies of state power, including national parliaments or other higher representative bodies.

24.1 The use of artificial intelligence systems on the territory of the state is allowed only under the conditions of compliance with all current norms of national legislation, including the Constitution as the Basic Law that has the highest legal force, as well as international legal obligations undertaken by the state in accordance with international treaties, the consent to be binding of which was granted in accordance with the established procedure. The use of artificial intelligence systems must fully comply with the principles of the rule of law, which implies the subordination of all subjects — both public authorities and private individuals — to legal norms, ensuring stability and predictability of legal regulation, as well as guaranteeing judicial protection in case of violation of rights.

In addition, the functioning of such systems must comply with the principle of legal certainty, which requires clear and transparent rules that regulate the implementation, operation and responsibility for the consequences of the use of AI [243, 244]. At the same time, any use of artificial intelligence aimed at abuse of law, manipulative interpretation of norms or creating situations of legal uncertainty is prohibited. All participants in legal relations in the field of artificial intelligence are obliged to act in good faith, respecting the rights of other persons and strictly adhering to the principle of the presumption of good faith, according to which each action of the subject is considered lawful until proven otherwise in accordance with the procedure established by law [245, 246].

24.2 It is prohibited to introduce, use or support the functioning of any artificial intelligence system on the territory of the state if such a system directly or indirectly violates human rights and freedoms guaranteed by the Constitution, as well as undermines or threatens the foundations of the constitutional order, including the principles of democracy, the rule of law, the independence of the judiciary, the separation of state power into legislative, executive and judicial, as well as other fundamental principles of the democratic system [247, 248, 249].

In particular, it is unacceptable to use AI systems in a way that may lead to usurpation of power or unconstitutional redistribution of powers between branches of government, manipulation of electoral and electoral processes using digital technologies, undermining the independence of the court, restricting or distorting access to justice, or establishing monopolies on information in the digital environment [250, 251]. Especially dangerous are cases when AI systems are used without a proper legal basis, i.e. without clear regulatory regulation of their functioning, without conducting a preliminary comprehensive assessment of legal, ethical, security and social risks, without an independent expert opinion on their compliance with constitutional standards, as well as without a proper procedure for public control, open discussion and involvement of stakeholders [252, 253].

The use of artificial intelligence systems in the field of public administration without ensuring openness, accountability, and citizen participation undermines the legitimacy of state decisions, the transparency of the functioning of democratic institutions, and trust in public authorities. The ban on the introduction of AI systems that violate constitutional values is unconditional, does not allow exceptions and is subject to mandatory control both by authorized state bodies and through judicial protection and civil supervision [254].

24.3 No public authority or public administration body at the national or local level, nor any legal or natural person — regardless of the form of ownership, subordination, or sources of funding, has no right to use artificial intelligence systems in a manner that is contrary to the principle of separation of powers, undermines the independence of the judiciary, limits the autonomy of local or regional governments, or encroaches on the independence and proper functioning of representative bodies of state power, including national parliaments or other higher representative institutions.

It is unacceptable to delegate critical functions of governance, law-making or judicial proceedings to autonomous or semi-autonomous digital systems if this leads to the formal or de facto replacement of the subject of authority with a digital agent that is not under the control of society, law and constitutional oversight [255]. This applies to algorithmic decisions on the distribution of budget funds, administration of social services, adoption of administrative or judicial decisions, processing of electoral information, assessment of the effectiveness of public authorities, etc[256].

The use of AI systems for modelling, forecasting or influencing political processes, parliamentary activities, the course of court proceedings or the activities of local or regional government bodies is allowed only in the form of an auxiliary tool that does not go beyond the competence of the relevant bodies and does not interfere with the decision-making process that belongs to the exclusive competence of a person as a bearer of political, judicial or administrative responsibility [257, 258, 259]. Thus, AI cannot replace a member of parliament or other supreme representative body, nor a judge, nor the head of a local or regional government body, including in the form of a consultation or information algorithm, if this contradicts the constitutional nature of their functions [260, 261, 262].

24.4 The state guarantees the existence of effective legal mechanisms that provide any individual or legal entity, including public associations, journalists, human rights defenders and scientists, with the opportunity to initiate an independent legal review of the legality of the implementation, use or operation of any artificial intelligence system in the public sector. Such an audit may have an administrative, judicial or expert-analytical form and is provided on accessible terms in case of reasonable suspicion of violation of constitutional human rights and freedoms, principles of good governance, transparency or accountability.

Initiation of the check can be carried out by:

submission of an appeal to the authorized national body for the protection of human rights or digital rights regarding the violation of a person's rights or improper human oversight of digital processes;

appealing administratively or judicially against decisions of public authorities regarding the implementation of AI systems;

making requests for public information on the use of AI in the activities of public authorities, including access to descriptions of algorithmic processes, impact assessments and technical documentation;

use of parliamentary or representative control mechanisms, in particular by initiating hearings, appeals of members of parliament or the creation of temporary control commissions;

initiating investigations or analytical checks through public or expert institutions.

Everyone has the right to protection against opaque or potentially harmful algorithmic practices, including in areas such as the assignment and administration of social benefits, medical or educational decision-making, automatic sorting of documents in administrative processes, and the use of AI in the field of security and control. Such controls should ensure the effective participation of civil society in digital governance, maintaining a balance between innovation and accountability.

CHAPTER 25. THE PRINCIPLE OF HUMAN DIGNITY

Human dignity is recognized as an inalienable, supreme, and unconditional value in the digital age and is subject to unquestioning protection at all stages and in all forms of the life cycle of artificial intelligence systems — from conceptual design, technical development, model training, operation, maintenance, modernization to their decommissioning.

Any artificial intelligence system operating or proposed for use in the state must be created, tested and applied with full respect for the human personality, his bodily, psychological, moral and digital integrity, privacy, freedom of self-determination, autonomy of will and cultural uniqueness.

The use of artificial intelligence systems is prohibited:

in order to humiliate human dignity, dehumanize a person, devalue him, discredit him or reduce (reduce) the individual exclusively to digital parameters (profiles, ratings, classifications);

for the creation, modelling, generation or public use of so-called "digital twins" (virtual reconstructions or simulations of a real person) without the explicit, voluntary, informed and personalized consent of such a person; in the case of posthumous modelling, without an expression of will recorded during lifetime or the permission of a legal representative, if such use may harm the reputation or dignity of a person;

to collect, analyse, process, store or transmit intimate, deeply personalized, biometric or other sensitive data without the explicit, informed and documented consent of the individual, based on a full understanding of the purposes and consequences of the processing.

To establish the highest value of human dignity in the digital environment, the state undertakes:

ensure the creation and implementation of codes of ethics, standards of digital mutual respect, ethical responsibility procedures and mechanisms for legal response to violations of dignity in the field of artificial intelligence;

to promote the development of a digital culture of dignity in educational institutions, public administration bodies, the media, on electronic platforms and in the field of open data;

to support the formation of a responsible technological environment, where each stage of the creation and use of AI is evaluated through the prism of human dignity as the highest constitutional value.

25.1 Human dignity is recognized as a supreme, inalienable and inviolable value that has absolute priority in any process related to the design, training, implementation, use or improvement of artificial intelligence systems. It cannot be compromised for the sake of innovative efficiency, convenience, resource optimization or technological dominance. In the digital environment, this means that neither public nor private actors can create or exploit algorithms that directly or indirectly violate, nullify or ignore the dignity of a person as a unique, morally autonomous entity [263, 264].

Within the framework of this principle, dignity is not reduced to the status of only a legal category but appears as a fundamental reference point of digital humanism, which determines the permissible limits of the development of artificial intelligence in society. Its protection means not only a formal ban on humiliation of a person, but also a proactive obligation of the state, developers and users of AI to ensure conditions under which no person will become the object of technical reduction, psychological depreciation or symbolic exclusion [265].

25.2 Any artificial intelligence systems that have the potential to interact with the physical, psychological or informational characteristics of a person must be created and operated with full respect for the bodily, psychological, emotional and digital integrity of everyone. Such respect includes guaranteed refraining from any actions that may cause humiliation, disorientation, emotional exhaustion, or create a sense of vulnerability or helplessness of the person before the algorithmic system [266].

Interface design is carried out in such a way as to avoid implicit influence on user behaviour through framing (hidden choice construction), manipulative UX design, emotional reading without consent, or exploitation of the user's cognitive and neuropsychological vulnerabilities [267]. Special protection is provided to vulnerable categories of persons — children, persons with mental illnesses or disorders, victims of violence and other groups for whom any form of digital contact can become a source of retraumatization or a new form of humiliation [268].

In addition, each artificial intelligence system or algorithmic module must contain a mechanism for assessing the impact on dignity — regardless of whether it is a chatbot, a generative model, or video surveillance systems. Such an assessment is a prerequisite for the legitimacy of the functioning of the AI system, and its absence is the basis for termination of the operation of the relevant system [269].

25.3 To prevent the dehumanizing use of AI, the following prohibitions are directly established:

1. It is prohibited to use artificial intelligence systems for the purpose of humiliating, discrediting, satirically ridiculing or objectifying a person, his body, voice, language, sex, ethnic origin, nationality, accent, age, sexual orientation, profession, religion or political beliefs. This applies to both text content and visual models, deepfake videos, generative avatars, or other forms of digital reconstruction. The admissibility of such use is assessed taking into account the public interest and the need to ensure a balance of rights.

2. It is prohibited to create and use so-called "digital twins" of a person (including his face, gait, voice, intonations, manners, favourite expressions, gestures or writing style) without the direct, voluntary, specially confirmed consent of this person - during his lifetime or through his legal heir or successor after death. The enhanced legal protection regime applies in cases where a digital replica of a person is created in the context of memorialization, historical reconstruction, public impact or commercial exploitation. Such practices are recognized as incompatible with human dignity if they lead to memory distortion, distortion of reputation, or manipulation of the audience's emotions.

3. It is prohibited to process intimate, sexually sensitive, biometric, medical, religious or ideological, as well as psycho-emotional data without the direct, specially confirmed, informed and lawfully documented consent of the person. Such consent cannot be implicit, general or conditional. It must consider the specific purpose, duration, scope and level of algorithmic processing. The use of such data in generative models, behavioural forecasting systems, advertising or risk assessment is legal only if a high level of ethical transparency is ensured, the right to withdraw consent is guaranteed, and the subject is fully controlled [270, 271].

25.4 The state undertakes to develop and implement a comprehensive policy for the establishment, guarantee and protection of human dignity in the digital environment, covering the educational, administrative, institutional, communication and technical and legal levels. For this purpose, the state provides:

1. Development and implementation of codes of digital integrity that form ethical guidelines for developers, providers, and users of AI systems, as well as establish standards of respect, digital restraint, non-violent speech, and humane interaction in the online space. Such codes should have regulatory support and public accessibility, be subject to annual independent peer review, and be mandatory for implementation in the public sector.

2. Institutionalization of digital ethics as an interdisciplinary element of education, in particular the integration of the principles of protection of dignity, autonomy, privacy, and psycho-emotional well-being into educational programs at all levels, from school to higher. The pedagogical community should be prepared to teach ethical dilemmas related to AI based on science-based and culturally sensitive scenarios.

3. Creating a public culture of dignity in the digital space, which includes information campaigns, popularization of ethical models of digital behaviour, promotion of the principle of respect for the individual in public discourse, media and visual environment. Such a culture includes the responsibility of civil servants, public figures, bloggers and influencers for the algorithmic impact they have on the audience.

4. Formation of mechanisms for monitoring and reporting on the implementation of the principle of dignity in digital policy. This includes the creation of an independent Council on Digital Dignity at the National Human Rights Institute or other equivalent body, the introduction of indicators of compliance of digital services with the principle of human dignity, as well as annual public reporting to the government or other competent executive body on the state of protection of dignity in the field of artificial intelligence.

5. Ensuring that no state or municipal AI system can be implemented without a prior assessment of its impact on human dignity with the participation of civil society representatives, ethics councils and experts in the field of digital rights. Such an assessment should be public, open to discussion and independent analysis.

CHAPTER 26. THE PRINCIPLE OF CONTINUOUS ETHICAL EXAMINATION

All artificial intelligence systems that are used in areas that can affect human rights, public safety, public policy, democratic processes, health, education, the environment or human dignity are subject to mandatory ethical examination.

Ethical due diligence of AI systems should be a continuous process that accompanies and controls all stages of their life cycle — from design and training to deployment, scaling, upgrade, and decommissioning.

It is the responsibility of the AI developer, supplier, and operator to ensure the involvement of ethics boards, interdisciplinary commissions, or independent experts at every stage of using systems that pose a high ethical risk. Ethical expertise covers:

risks of manipulation, control and distortion of behaviour;

risks of interference in personal life;

disproportionate or disproportionate impact on certain social groups;

disturbing the balance between innovation and public trust.

The results of the ethical examination should be transparent, publicly available, documented in the form of digital audits, ethical reports or certificates of conformity, as well as contain clear indicators of ethical admissibility, identification of risks and the level of interference with human autonomy.

26.1 Ethical examination is not an optional or consultative procedure, but acts as a mandatory tool for preventing harm, human rights violations, discrimination, social destabilization or dehumanization [272]. It applies to systems that affect the life, health, human dignity, privacy, security or expression of the will of individuals. Ethical expertise is also applied to AI systems that operate in the field of justice, defense, information management, education, ecology and other socially significant environments [273, 274].

The traditional model of one-time approval at the launch stage is recognized as insufficient. Instead, a mandatory model of ethical monitoring is established, which includes:

- preliminary (pre-operational) examination,
- periodic (current) revaluation,
- retrospective (post-operational) verification of the impact of the AI system on the real living conditions and activities of individuals.

This provision is intended to ensure that even minor changes to the architecture, training data, or application modes of the system do not cause ethically unacceptable consequences.

26.2 There is a regulatory obligation for developers, suppliers, distributors and operators to involve interdisciplinary commissions in the process of ethical support, which should include specialists in law, ethics, sociology, philosophy, psychology, cybersecurity, ecology and protection of digital rights and data. For high-risk systems, the involvement of external independent experts is mandatory.

The results of the ethical examination cannot be hidden or restricted in access to the public. They should be drawn up in a standardized form:

- ethical passport of the system;
- certificate of compliance with ethical standards;
- digital register of ethical status;
- risk indicator system (traffic light model) [275].

Such documents are used as a means of public control, the subject of legal analysis, as well as a criterion for public procurement, accreditation and certification [276].

CHAPTER 27. THE PRINCIPLE OF EQUALITY AND IMPARTIALITY IN ACCESS TO ALGORITHMIC OPPORTUNITIES

All persons staying on the territory of the State, regardless of citizenship, social status, place of residence, language affiliation or other characteristics, have the right to equal access to technological opportunities, digital services, information resources, educational platforms and algorithmic solutions of public importance, created using artificial intelligence systems.

The right to equal access is guaranteed regardless of citizenship, sex, age, state of health, language, region of residence, economic situation or any other characteristic. Any actions, decisions or policies that directly or indirectly restrict access to AI systems are prohibited.

It is prohibited to deliberately narrow functionality for certain groups, create interfaces or architectures that make it impossible or significantly difficult to use for certain categories of people, as well as introduce discriminatory tariffs, quotas, geo-blocking or closed access interfaces (APIs) that lead to unequal access or de facto segregation of users.

Providers and operators of AI systems are required to ensure universal design of digital services and compliance with internationally recognized accessibility standards, including alternative modes of interaction for persons with disabilities and the elderly. Interfaces and support materials should be multilingual, including in languages of national minorities; The elimination of language barriers cannot depend on the payment of the service. Basic access to functions of public importance (medical, educational, environmental, legal) is provided on transparent and non-discriminatory terms.

It is prohibited to take any action or implement policies that directly or indirectly:

- *restrict individuals' access to AI technologies based on economic status, regional affiliation, language barriers, or social background;*
- *consist in the creation of systems, architectures or interfaces that intentionally exclude or restrict access of certain categories of persons to the use of AI;*
- *provide for unequal access to algorithmic opportunities by establishing corporate, political or organizational preferences that violate the principle of an open and equal innovation environment.*

In order to ensure the principle of equality in access to AI systems, the state ensures:

1) implementing sustainable digital inclusion programs aimed at supporting people with limited access to technological resources or digital skills;

2) development of open and transparent technological platforms available to public authorities, educational and scientific institutions, civil society institutions and individuals;

3) adoption and implementation of digital accessibility standards for persons with disabilities, the elderly and users with sensory, motor or cognitive disabilities;

4) introduction of guaranteed quotas of free access to systems of increased social importance (medical, educational, environmental, legal) for socially vulnerable groups of the population.

27.1 General rule and prohibitions. Every person who is on the territory of the State has an inalienable right to equal access to technological opportunities, digital services, information resources, educational platforms and algorithmic solutions of public importance, created with the use of artificial intelligence systems [277]. Any actions or policies that directly or indirectly restrict access to AI systems on the basis of economic status, regional affiliation, language barriers, social origin, disability, age, gender, migration status or other characteristics are prohibited; provide for the deliberate exclusion of certain categories of persons by means of architectural solutions or interfaces; establish unequal access through corporate, political or institutional preferences, including through closed APIs, discriminatory tariffs, geo-blocking or other forms of indirect discrimination [278, 279].

Typical barriers and risks of discrimination in access to AI systems. For the purposes of law enforcement, the barriers to access to be identified and eliminated include:

1) digital poverty/digital *divide* — lack of devices, connectivity, limitation of mobile traffic;

2) language and cultural barriers — lack of localization of interfaces, educational materials, support for languages of national minorities and regional languages;

3) accessibility barriers for persons with disabilities and the elderly — non-compliance with international accessibility standards, lack of alternative modes of access and interaction;

4) technical and organizational barriers — closed standards, vendor lock-in, disproportionate identification/verification requirements, discriminatory limits on computing resources or API quotas;

5) algorithmic biases — lower quality of service for linguistic, regional, or social groups; discriminatory scoring, moderation, personalization models; manipulative interface practices (*dark patterns*);

6) economic practices that exclude users — excessive paywalls, mandatory paid packages for basic functions, artificially reducing the quality of free modes.

27.2 Obligations of the State, Suppliers and Platforms to Ensure Equal Access:

1) The state implements sustainable digital inclusion programs for persons with limited access to digital resources or skills; develops open and transparent technological platforms for public authorities, education, scientific institutions, civil society institutions and individuals; adopts and implements digital accessibility standards for persons with disabilities and the elderly; introduces guaranteed quotas of free access to socially significant systems (medical, educational, environmental, legal) for socially vulnerable groups of the population.

2) Suppliers and platforms provide: universal design and compliance with international accessibility standards; multilingual interfaces and support materials; open and non-discriminatory conditions for access to APIs and basic functions; transparent billing rules without hidden restrictions; providing an alternative that does not involve algorithmic influence or has reduced requirements, where possible; regular Equality and Accessibility Impact Assessment with the publication of indicators and plans to remove identified barriers.

3) The National Centre for Digital Inclusion is being established as an independent coordination structure to monitor equality of access, conduct audits, support localization and inclusive design, address complaints and promptly remove barriers.

4) Violation of this principle entails the application of legal and regulatory measures, including orders to eliminate violations, administrative and economic sanctions, restriction of participation in public procurement, and, in case of systematicity, temporary restriction or suspension of access to the state technological infrastructure and registers.

CHAPTER 28. THE PRINCIPLE OF INFORMATION ENVIRONMENT PROTECTION

The use of artificial intelligence systems to create, spread or legitimize disinformation, carry out information attacks, manipulate mass consciousness, undermine democratic processes or elements of the information sovereignty of the State *is prohibited.*

The use of artificial intelligence in the public sphere is recognized as an activity of increased risk and is always subject to assessment in view of the possible undermining of public trust and distortion of facts.

The design, training, commercialization, distribution, maintenance, and use of AI systems designed for automated generation or mass targeted dissemination of false or manipulative information are prohibited. It is prohibited to create synthetic or artificially generated audio, photo and video materials, including deepfake content, voice clones or images, without reliable marking of origin and context. It is also prohibited to use systems to simulate or stage unreliable socially significant events to distort political processes, undermine trust in democratic institutions, incite hatred or undermine the legitimacy of state or international organizations.

Providers of digital communication services operating within the jurisdictions of the States are obliged to: implement technically secure, sustainable and integral labelling of content created using AI systems, indicating the source, type of model and time of creation; ensure the availability of tools for verifying the authenticity of digital information, including built-in fact-checking mechanisms, assessment of the reliability of sources, and alarms regarding potentially manipulative materials; maintain systems for traceability of the origin of content (provenance), maintain reliable and independently verifiable audit logs; store digital evidence in a way that guarantees its integrity and suitability for use in official and judicial investigations.

Providers also carry out continuous cooperation with national authorities responsible for information resilience and cybersecurity (in particular, specialized centres or agencies), and, if necessary, also with international institutions. They implement risk management procedures, ensure timely notification of incidents to competent authorities, and take prompt and proportionate measures to neutralize threats to digital infrastructure, key democratic processes, national security and public trust. Providers are obliged to publish regular and publicly available and sufficiently detailed reports on moderation activities, as well as to provide timely, clear and motivated explanations to users and competent authorities regarding the restriction or removal of content.

Everyone has the right to be informed about interactions with content created or modified by AI systems, to receive information about its origin, to appeal against the platform's decisions, to request the correction or removal of false information, and to use an effective complaint and appeal procedure. Moderation measures cannot lead to arbitrary restrictions on freedom of expression and must be based on clear, public and non-discriminatory rules.

The implementation of this Article is ensured by state supervision and control powers of authorized bodies, which have the right to issue orders to eliminate violations, apply administrative and economic sanctions, demand urgent restriction of access to dangerous content, temporarily suspend the operation of certain services in case of systematic violations, and restrict the admission of violators to public procurement. Such measures should be applied based on the principles of legality, necessity and proportionality.

In case of doubt, the provisions of this article shall be interpreted considering the need to preserve the integrity of the digital public sphere, protect democratic processes and respect fundamental human rights.

It is prohibited to design, train, commercialize, distribute, maintain, or use AI systems designed to:

1) automated generation and dissemination of false or manipulative information;

2) creation of synthetic audio, photo or video materials, including using technologies of image synthesis, voice cloning or deepfake modelling, without clear and technically secure marking of the source and circumstances of creation;

3) modelling or staging of unreliable socially significant events, carried out with the aim of manipulative distortion of political processes, undermining the legitimacy of democratic institutions, inciting enmity or delegitimating of state authorities and international institutions.

All providers of digital communication services — including social networks, instant messengers, streaming platforms, video hosting, forums, blogs, podcast platforms and news aggregators — operating in the territory of the State are obliged to:

1) implement technically secure and sustainable mechanisms for automatic labelling of content created using AI systems, indicating the source, type of AI model and time of creation;

2) ensure the availability of tools for verifying the authenticity of digital information, including through built-in functions of fact-checking, transparent assessment of the reliability of sources and alarms for potentially manipulative materials;

3) to carry out continuous cooperation with national authorities responsible for information resilience and cybersecurity (in particular, the Centre for Information Resilience of the State), as well as, if necessary, with international institutions, to identify and neutralize artificial intelligence threats aimed at digital infrastructure, democratic processes, national security and public trust.

28.1 Artificial intelligence in the field of public communications is recognized as a high-risk technology: it cannot be considered neutral if the consequence of its application is to undermine the truthfulness, trust and integrity of the digital public sphere. This norm protects the information sovereignty of the State, the security of democratic processes (including electoral processes), the stability of institutions and the right of every person to access reliable information. For the purposes of this article, the "information environment" covers the infrastructure for creating, storing, circulating and consuming content, as well as ranking, personalization and moderation algorithms that affect the formation of public opinion [280, 281].

The ban applies to the entire life cycle of systems and tools if their dominant purpose or intended impact is disinformation or manipulation: fundamental models, specialized generators of text, images, audio and video; "orchestrators" of botnets and mass mailing plugins; voice and face cloning tools; modules for scenario simulation of events; services for legitimizing untruths (masking the origin of content, bypassing labelling, creating falsified or manipulative metadata). Prohibited activities include designing, training, testing, deploying, commercializing, maintaining, and providing access to third parties if, under normal operating conditions, this leads to the erosion of trust and the actual destabilization of public communication [282].

At the same time, the existence of "dual use" in the field of public communications is recognized. Research or creative tools themselves are not prohibited if the provider or developer has proven effective safeguards and safe use modes:

- built-in and non-removable marking (watermark/provenance) at the model and interface level;
- logging and saving interaction logs in a way that provides the possibility of independent auditing; controlled API quotas and speed limits;
- limiting mass operations and automated distribution;
- customer verification for high-risk scenarios (KYC/KYB);
- regular red-teaming, updated lists of prohibited tips, and bypass prevention mechanisms.

Artistic, educational and scientific experiments are allowed only if there is an explicit warning label, an unambiguous disclosure of the artificial nature of the material and no real risk of mixing with the facts.

The application of this norm is based on a set of material criteria:

"public harm test": whether there is a significant risk of mass misleading or destabilization of processes of public importance[283];

"intent and/or impact test": whether there is intent (deliberate intention) or actually confirmed and duly documented effect of undermining public trust, stability of the information environment or legitimacy of democratic institutions, regardless of the declared intention;

"scale and automation test": whether coordinated automated or semi-automated execution of actions aimed at destabilizing the information environment is provided, including through the use of botnets, mass automated creation of requests, automatic generation and auto-publication of content, as well as other forms of mass targeted distribution;

"Test of deception of authenticity": whether synthetic or artificially generated content is passed off as reliable evidence or facts without marking, including imitation of specific persons, institutions or attributes of official authenticity (logos, signatures, verification marks);

"Predictability and manageability test": whether the damage could reasonably have been (reasonably) foreseen and whether appropriate and proportionate control measures were taken, including risk assessment and implementation of procedures to minimize possible negative consequences;

"Test of vulnerable periods": whether artificial intelligence systems are used in election campaigns, during martial law, emergencies, riots or other crisis circumstances, when the threshold conditions for the application of regulatory measures are lowered, and the obligations of providers regarding preventive mechanisms are increased.

The balance with freedom of expression is ensured by the doctrine of necessity, proportionality and minimal interference: this norm does not aim to censor assessments, opinions or journalistic investigations, and does not cover bona fide satire and works of fiction if their artificial nature is clearly exposed. The ban is aimed at systematic, technically organized deceptive reproduction of "reality signals" (images, voices, documents, data, attributes of authenticity) and their mass replication in order to influence the expression of will, security and trust. In case of doubt, priority is given to preserving the integrity of the public sphere and minimizing the risk of large-scale disorientation of the audience, provided that the least burdensome means for freedom of speech are used [284].

For high-risk scenarios, preventive obligations are established: mandatory registration of mass content generation and distribution functions in a special register of the national regulator; ex ante risk assessment with submission of results to national authorities responsible for information resilience (in particular the Information Resilience Centre), including analysis of potential harm scenarios and mitigation measures; availability of a technical mechanism for emergency shutdown of the "kill switch" system and emergency curtailment procedures; prompt notification of incidents to the competent authorities within the established period and cooperation with authorized entities; publication of regular reports on the measures taken within the time limits set by the national authorities.

Failure to comply with these requirements qualifies as circumvention of the ban and entails the application of increased administrative, financial and technological sanctions.

This paragraph details the typical threats posed using generative models[285]. Including:

- Disinformation encompasses messages that are systematically generated or disseminated by automated means, including through news feeds, campaigning chatbots, and other forms of personalized automated content;
- Deepfake falsifications (synthetic audio, video and photo materials) are recognized as one of the most dangerous distortions of reality, which can significantly undermine citizens' trust in public authorities, the media and expert institutions;
- Scenario simulations of unreliable or artificially simulated events ("fake") — for example, simulated attacks, disasters, voting, etc. — pose a particular threat to digital and public security and, in cases where they are aimed at destabilizing society or undermining the security of the state, can be qualified as a form of digital terrorism;
- Manipulation is defined as the algorithmically organized mass imposition or amplification of distorted narratives through targeting, personalization, ranking, automated information injections, as well as masking the artificial nature of messages.

Priority categories of threats include:

a) content falsifications — synthetic texts, images, audio and video (deepfakes, cloning of voice or face, imitation of real sources, fabrication of official documents or pretended "expert" opinions), including with fake metadata;

b) automated influence operations — coordinated manipulative activity (botnets, account farms, automated SEO poisoning, mass auto-publications, generative "whitening" of content and its translation while preserving manipulative narratives);

c) simulated scenarios of events — staging or imitation of socially significant facts (attacks, catastrophes, "official" statements or voting results) presented as reliable facts;

d) legitimization of falsehoods — services designed to remove or circumvent labeling, falsify origin (provenance), create false fact-checking mechanisms, or pseudo-sources designed to give falsehood the appearance of truth.

Signs of attribution of content or activity to the scope of this paragraph are:

- the systematic nature and scale of automated operations;
- hiding the artificial nature of content and/or imitating specific individuals, institutions or media;
- significant risk of mass misrepresentation;
- targeted targeting of vulnerable groups of the population or interference in critical socio-political processes (elections, referendums, martial law, emergencies);
- organizational connection with botnets, simulation accounts ("doubles") and mirrored domains ("mirrors");
- manipulation of recommendation and ranking mechanisms;
- deliberate distortion or fabrication of metadata, which makes reliable attribution impossible.

If there is a combination of these features, the content is considered to have been created or distributed using AI systems for manipulative purposes.

Risk levels are established for law enforcement:

- *increased* — risk of local misrepresentation or limited exposure;
- *high* — the possibility of large-scale distribution, cross-platform coordination, and the occurrence of secondary waves of propagation;
- *critical* — the risk of immediate and significant harm to public security, electoral processes or international stability.

The intensity of preventive and reaction measures, response times, as well as the obligation to keep event logs and provide digital evidence to authorized bodies depend on the level of risk.

In view of the rapprochement with the criminal [286, 287], Cyber [288, 289]- and Information Law [290], Content and activities that have signs of digital terrorism (for example, the deliberate creation or dissemination of falsified reports of fictitious attacks or disasters that provoke panic, harm the life and health of people, or interfere with the operation of critical infrastructure) qualify as socially dangerous acts with appropriate legal qualification [291]. Platforms, intermediaries and providers that algorithmically amplify such materials or avoid the implementation of obvious safeguards (labelling, restriction of mass transactions, detection of botnets) are subject to public liability within the competences defined by law and international treaties.

Exceptions apply to bona fide journalism, scientific research, satire and works of fiction, provided that there is a clear and conspicuous warning label of the artificial or staged nature of the material, its unambiguous separation from factual messages, and there is no risk of mixing with real data in contexts of public importance. Any dual-use project is allowed only if there are effective technical and organizational safeguards — built-in and immutable (non-removable) labelling, controlled API quotas and speed limits, logging, KYC/KYB procedures for access to mass operations, regular red-teaming, and bypass prevention mechanisms. The effectiveness of such fuses is subject to independent audit by the authorized bodies.

28.2 This clause establishes the obligation of platforms and providers of digital communication services not only to provide technical infrastructure for the dissemination of information, but also to bear public responsibility for the transparency, reliability and verifiability of content generated or modified using artificial intelligence systems. The position of "simple placement" of content cannot be recognized as unacceptable in the case when the technological infrastructure is used as a channel for mass or systemic manipulation or disinformation campaigns.

Platforms are required to implement standardized systems for automatic and mandatory labelling of AI-generated content, indicating the source, type of model, time and conditions of creation. Such labelling must be embedded in metadata, non-removable by the user, and available for verification by third-party tools. For high-risk content (including content of socio-political significance, content that may affect

election processes or security), advanced labelling is applied, with additional authentication and contextual attribution verification.

In addition to labelling, platforms are required to provide information verification tools, including integrated or connected fact-checking functions, automated assessment of the reliability of sources, an automated warning system for suspicious material, risk indicators, and access to independent databases for content verification. Every user has the right to transparent reporting of interactions with AI content, access to information about its origin, as well as tools for appealing or requesting correction of false materials.

To prevent mass manipulations, platforms are required to implement risk management systems: limiting the rate of mass distribution of suspicious materials, identifying and blocking botnets, regular independent audits of ranking algorithms and recommendations, as well as publishing transparent reports on moderation actions. Such reports should include information about the amount of content removed or flagged, the reasons for the intervention and the number of appeals, and the average response time.

Cooperation with the Centre for Information Resilience of the State is recognized as mandatory and provides for regular exchange of data on new forms of manipulation, urgent notification of incidents within a specified timeframe, participation in joint trainings and crisis communication activities. The Centre performs not only a reactionary, but also an analytical function: it formulates recommendations, conducts scenario modelling of information attacks, and coordinates preventive actions of government agencies, platforms, and public organizations.

The responsibilities of platforms are both technical (provision of tools and mechanisms) and legal (responsibility for non-compliance) nature. Insufficiency of verification measures or lack of proper labelling is considered a violation of this Article and entails administrative and economic sanctions, including fines, restriction of participation in public procurement, an order to eliminate deficiencies, and in case of systematic violations — temporary restriction or suspension of the service in the territory of the State.

The principle of transparency of the origin of information is recognized as a new standard for ethical management of platforms, which has the force of a regulatory obligation. Its content is to provide users with the right to be informed, to prevent the possibility of hidden manipulative influence and to guarantee the integrity of the digital public sphere. All restrictive measures should be carried out in a manner proportionate to the intended purpose, respecting freedom of expression and ensuring effective appeal procedures.

CHAPTER 29. THE PRINCIPLE OF ACCESSIBILITY TO ALGORITHMIC EDUCATION

Everyone has an inalienable right to access knowledge about the functioning of artificial intelligence systems, their capabilities, limitations, decision-making mechanisms, ethical aspects and risks. The state undertakes to ensure the development of algorithmic literacy (AI literacy) *as a priority of educational policy. For this purpose, international approaches are used, which require education systems to go beyond basic digital literacy and are aimed at forming the ability to confidently navigate in an algorithmic environment. Algorithmic literacy is recognized as a mandatory element of education for everyone — not only for IT specialists, but also for all participants in the educational process.*

Within the framework of the implementation of this principle, the State creates the following conditions:

– courses on digital literacy and the basics of algorithmic thinking are integrated into educational programs of all levels, covering an understanding of the use of algorithms and the ability of a person to create and manage them;

– interdisciplinary programs are being developed and implemented for pupils, students, civil servants and workers in critical areas, which include AI ethics, principles of algorithmic justice, legal frameworks and technical aspects, as AI literacy must be integrated into all disciplines;

– public educational platforms are being created with open access to materials about AI, adapted to age, profession and level of training, with the possibility of using different formats and languages to meet the needs of the community;

– linguistic and cultural accessibility of educational materials is ensured both at the national and regional levels, including through translation into the languages of national minorities and adaptation for different age, professional and social groups;

– measures are being taken to include persons with disabilities, in particular through the adaptation of content using assistive technologies, which have proven to be effective in creating personalized and equitable learning environments;

– all educational initiatives must be free from political, religious or commercial bias and based on ethical standards, principles of academic integrity and evidence-based scientific information.

The principle of accessibility to algorithmic education is based on the understanding that modern societies are increasingly dependent on algorithmic solutions. The right to gain knowledge about AI systems includes the study of their impact on the economy, culture, public policy and social processes. International studies emphasize that algorithmic literacy (AI literacy) should develop not only technical skills, but also critical thinking, creativity, and ethical responsibility. The introduction of open educational platforms provides access to learning resources for all segments of society, including those who do not have the opportunity to attend traditional educational institutions. Public and private educational initiatives must work together in partnership to ensure that learning materials are user-friendly, adaptable, and inclusive.

29.1 Everyone has an inalienable right to access knowledge about the functioning of artificial intelligence systems, their capabilities, limitations, decision-making mechanisms, ethical aspects and risks. The state undertakes to ensure the development of algorithmic literacy (AI literacy) as a priority of educational policy, since modern education must go beyond basic digital literacy and include algorithmic competence.

29.2 Compulsory courses on digital literacy and the basics of algorithmic thinking are integrated into educational programs of all levels. Such courses are aimed at building knowledge and skills on how algorithms work, how to create and analyze them, as well as how to critically evaluate the results of their work, which is in line with the recommendations of international initiatives in the field of education and artificial intelligence.

29.3 Interdisciplinary programs are developed and implemented for pupils, students, civil servants and workers in critical areas. Such programs combine technical, humanitarian, and legal disciplines and reflect the structure of the AI Literacy Framework (engagement, creation, management, design).).

29.4 Public educational platforms are being created with open access to training materials about AI, covering different levels of training and adapted to the age, profession and needs of users. Such initiatives are based on the successful international experience of open resources that provide free materials for teachers and students.

29.5 Educational materials on algorithmic literacy (AI literacy) should be available in minority languages and adapted to cultural characteristics. UNESCO emphasizes that algorithmic literacy programs should be inclusive and create resources in multiple languages and formats (textual, audiovisual, interactive).

29.6 Educational initiatives ensure the inclusion of people with disabilities. Materials should be adapted to different types of impairments using *assistive technologies*, as the use of AI and other technologies can "break down barriers and make education more accessible and equitable».

29.7 Educational materials about AI should be free from political, religious, or commercial bias and based on ethical standards and evidence-based scientific knowledge. They should consider the age, professional and social characteristics of different groups, forming critical skills in evaluating algorithms and their results, as well as developing creativity and ethical consciousness.

CHAPTER 30. THE PRINCIPLE OF ETHICAL INTEGRATION OF AI INTO THE JUSTICE SYSTEM

The presumption of human motivation means that the final factual circumstances and legal conclusions in the case are formed by the judge, and not by an automated tool. Algorithmic results or hints are exclusively advisory in nature and cannot be recognized as accepted by default without a separate assessment by the judge. Doubts about the reliability, completeness or impartiality of the results of AI systems are interpreted in favour of the person.

The accessibility and equality of the parties ensure that neither side is worse off due to a lack of technical resources or specialized knowledge. The court is obliged to provide an explanation of the role of AI, provide procedural time to familiarize itself with the materials of the algorithm's transparency, and in case of proven financial insolvency of the party, to ensure a free independent examination of the algorithm.

The principle of data minimization and secrecy. In the model, it is allowed to use only those data that are necessary and proportional to the goals of the proceedings. It is prohibited to import commercial scorings; data collected from social networks or biometric parameters without a clear legal basis and a court order. The chain of custody for all algorithmic artifacts is subject to mandatory documentation.

Stability and reproducibility of algorithmic results. They are provided by fixing versions, parameters and the runtime. "Silent" updates during the case are prohibited. If the result cannot be reproduced under the conditions of the call and processing log, its evidentiary value is subject to reduction, or such a result is recognized as inadmissible evidence.

Extraterritorial Providers and Data Transfers. The use of AI systems developed or maintained by foreign entities is allowed only if all disputes are subject to the jurisdiction of national courts, the storage of logs and critical data on the territory of the State or in jurisdictions with an adequate level of protection, as well as the prohibition of cross-border data transfers without legal basis and due process guarantees.

The principle of human justice. Every person has the right to have his/her case considered and resolved by an independent judge who personally and independently evaluates evidence, forms legal conclusions and is responsible for the decision made. Algorithmic processing of information can be used exclusively as an auxiliary tool and cannot replace judicial judgment, procedural autonomy and personal motivation of a court decision.

The principle of procedural autonomy means the possibility of participating in proceedings without mandatory consent to the use of artificial intelligence systems. The refusal to use such systems or the requirement for human review may not worsen the procedural situation of the participant in the proceedings, restrict access to justice or reduce the scope of his rights and guarantees.

No technological system has the right to interpret disagreement with the use of AI, challenging algorithmic conclusions, or the absence of digital traces as a manifestation of bad faith, "riskiness" or a basis for a negative assessment of a person. Any social scoring, the use of prohibited or proxy features and other forms of discriminatory classification that affect the rights and obligations of the participants in the process are prohibited.

Any person has the right to be informed about the use of AI in his/her case, to receive an understandable and accessible explanation of algorithmic results, to have access to methodological information in the modes provided for by law, as well as the right to initiate a review by a judge or an authorized person and verification of the relevance, admissibility and weight of such results in evidence. The implementation of these rights guarantees the equality of the parties and ensures the effective implementation of protection.

Every AI system integrated into justice should function exclusively according to the principle of "human decision-making": with mandatory ethical certification, transparent methodology, mandatory logging of use, reproducibility of results, managed updates, and means of immediate termination (kill switch). Automatic confirmation or generation of the motivating part of court decisions without personal and individual assessment by the judge is prohibited.

30.1 Prohibited practices [292, 293].

1. Fully automated sentencing or other final procedural decision in criminal, administrative, civil, economic or any other public law or quasi-judicial proceeding without decisive human participation is prohibited.

a) It is prohibited to use the modes of "auto-confirmation", automatic generation and signing of the operative part of the procedural act, as well as program triggers (automatically executed instructions that trigger legal consequences) that entail the imposition of a fine, seizure, blocking of assets or any other restriction of the rights, freedoms or legitimate interests of a person without a separate human decision [294].

b) "Decisive participation of a person" means a real and full opportunity of a judge or other authorized person to: independently re-evaluate the evidence; change or reject the result of the AI system functioning; make a final decision that cannot be automatically overridden or replaced by the system.

c) The case file must indicate which recommendations of the artificial intelligence system were taken into account or rejected and for what reasons.

Evidence of compliance with the requirement of decisive human participation must be confirmed and stored in the case, in particular: model call logs, system ID and version, task description, human review mark, and the judge's or other authorized person's own motives. Such data must be kept in a secure form for at least five years after the conclusion of the proceedings and be available for inspection and audit.

2. Biased (discriminatory) assessment of the risk of recidivism, the "danger" of a person, his/her social "value" or "trust", as well as any forms of social scoring are prohibited [295, 296].

a) Direct or indirect (through proxy) processing of sensitive personal data: race, ethnic origin, religion, political opinion, health status, disability, sexual orientation, trade union affiliation and other categories, the processing of which is expressly prohibited by law, is prohibited.

b) It is prohibited to use data from commercial scorings, social media arrays, broad behavioural trackers, geolocation and biometric parameters without a direct legal basis and a court order. The use of such data without proper grounds is recognized as a violation of human rights.

c) Risk assessment models based on non-transparent algorithms ("black boxes") cannot be applied if it is not possible to provide a local explanation of a specific conclusion in the case, verify the absence of prohibited bias and ensure an independent audit.

d) Insufficient representativeness of training data, lack of calibration by groups or detected statistical discrimination are grounds for prohibiting the use of the results of such a model. The conclusions obtained from such models are recognized as improper evidence and cannot influence procedural decisions.

3. Substitution of procedural analysis with statistical classifiers or template "motivational constructors" is prohibited [297].

a) Automatically generated texts cannot constitute the motivating part of a court decision without an individual legal assessment by the judge. Only their auxiliary use for editorial and technical purposes (formatting, numbering, structuring) is allowed, subject to mandatory verification and approval of the content by a judge.

b) It is prohibited to rely on the "probable" or "average" conclusions of the model to establish facts, assess the reliability of testimony, resolve conflicts of evidence, determine the measure of punishment or preventive measure. Any statistical assumptions cannot replace the direct assessment of evidence by the court.

c) References to integral indices, ratings, scorings or other indicators that do not have a verifiable methodology, independent verification and local explainability are excluded from the motives of a court decision and are recognized as improper motivation.

4. Anti-abuse fuse. The prohibitions established by this section cannot be circumvented by: formally renaming the goals or functions of the AI system (for example, "assistant", "hint", "template"); making algorithmically formed conclusions outside the text of the verdict or ruling, if such conclusions actually affected the content of the decision; the use of algorithmic recommendations in the form of reference materials, draft texts or consultations, if they were accepted by the judge without critical verification and independent legal assessment.

All these cases are recognized as a violation of the principle of decisive participation of a person and entail the invalidity of the relevant part of the decision.

30.2 Mandatory requirements for AI systems in the field of justice [298].

1. Ethical certification and periodic supervision [299].

a) Any AI system before its use in proceedings is subject to preliminary ethical certification by an authorized body with the participation of the ethics council and human rights experts. The certificate is valid for no more than 12 months; After this period, a second assessment is carried out. The decision on the issuance, suspension or cancellation of the certificate shall be entered into the public register.

b) The composition of the assessment includes:

(i) assessment of the impact on human rights and freedoms;

(ii) Algorithmic Impact and Risk Assessment (AIA), including screening of training data, models, and results for discriminatory effects;

(iii) audit of human controllability, which implies the presence of a "stop button", mechanisms for reviewing and cancelling automated decisions;

(iv) checking the reliability, stability and information security of the system.

The certificate can be suspended or revoked in case of critical defects, security incidents, failure to comply with transparency conditions or implementation of "silent" updates without the approval of a court or supervisory authority. Revocation of the certificate entails the immediate termination of the use of the system in all proceedings.

2. Transparency and methodological openness.

The operator and supplier of the AI system ensure the availability and availability of constant updating of the "model card" and "datasheets for datasets", which should contain information about the purpose, version, dates and volume of data used for training and validation, known limitations and risks.

Both global explainability (the general logic of the system) and local explainability in a particular case (key factors that influenced the result) must be ensured. Results that do not have a verifiable explanation are recognized as improper evidence and cannot form the basis for procedural decisions.

A brief profile of the system (purpose, scope, main limitations) is made public to the general public. Extended materials are provided to the court, parties and appointed experts in a "safe room" mode with access logging and non-disclosure obligations, considering the protection of personal data, the secrecy of the investigation and national security.

3. Human controllability (human-in/on-the-loop).

The system interface should provide for clear labelling of algorithmic results, means of rejecting or editing recommendations, as well as a forced stop function.

Any "automatic confirmation" is prohibited; The application of the recommendation is allowed only after explicit confirmation by the judge indicating the reasons in the case materials.

All users of the system (judges, prosecutors, investigators, court secretaries) undergo mandatory training in the basics of artificial intelligence and digital literacy; At least once every 24 months, they are re-certified.

Data management and minimization principle.

Only those data that are necessary and proportionate to the declared purpose are used; The processing of sensitive data is carried out exclusively on the basis of the law and in the presence of a court ruling with a separate assessment of proportionality.

It is prohibited to import commercial scorings, social media data sets, broad behavioural trackers, geolocation and biometrics data, unless such use is expressly permitted by law and court within a specific case.

Proper data quality (relevance, completeness, reliability) is ensured; the chain of custody must be documented; Clear retention periods are defined, as well as procedures for depersonalization and/or deletion of data.

4. Registry, logging, and reproducibility.

Each installation of the AI system is subject to entry into the State Register of High-Risk Systems, indicating the version, certification data, audit results, and recorded incidents.

All model calls within the proceedings are subject to mandatory logging, which includes: time, version, instance ID, task description, generalized description of input/request, result obtained, and human decision to take the result into account.

Reproducibility of results (deterministic replay) based on fixed versions and parameters is ensured; Journals are kept for at least 5 (five) years from the date of completion of the proceedings.

Security and integrity.

The supplier and operator implement technical and organizational measures to protect against prompt injection, jailbreaking, model inversion, data poisoning, leaks, and unauthorized changes.

Regular security testing (penetration testing, red teaming), vulnerability management, and integrity checks of runtime environments are carried out.

Updates of system versions are carried out exclusively according to the approved protocol with notification of the court and the parties in open proceedings; "Silent" updates are not allowed.

Critical incidents shall be reported to the authorized body and the court no later than 72 hours from the moment of their detection; An analysis of the causes is carried out and a plan of corrective actions is drawn up.

Testing, validation and quality control.

Prior to system deployment, mandatory functional, stress, and acceptance tests are performed on control datasets; In a productive environment, continuous quality monitoring and data/model drift detection are carried out.

Minimum integrity metrics include: disparate impact, equalized odds / TPRFPR parity, calibration by groups, stability over time, assessment of errors of the second kind in relation to vulnerable groups.

Detection of a significant decrease in quality or signs of discriminatory influence is the basis for immediate suspension of the use of the system in proceedings until the identified violations are eliminated.

5. Procurement and contract conditions.

All contracts with suppliers of artificial intelligence systems must contain: the customer's rights to audit and access to complete methodological and technical documentation; requirements for localization of logs and data in accordance with the law; the supplier's obligation to timely report incident reporting; prohibition of the involvement of sub processors without prior written consent with the customer.

Technological escrow of critical components (models, scales, configurations) is provided in order to ensure reproducibility, forensic examination and continued use in the event of a dispute or termination of support by the supplier.

The jurisdiction of all disputes is determined by the courts of the State; applicable law — national legislation.

6. Interface Labelling and User Warranties.

The interface of the artificial intelligence system is obliged to directly inform the user about the fact of using AI, display the level of confidence (if any), and warn about known limitations and risks.

Checklists are introduced for users to check the relevance and completeness of the entered data, as well as the absence of prohibited features or discriminatory characteristics in the input data and results.

7. Interaction with experts of the parties.

At the justified request of the party, the court provides its experts with access to transparency materials (documentation, journals, explanations) in the "safe room" mode, as well as provides procedural time for an independent review.

a) Failure by an operator to comply with transparency requirements or unjustified denial of access entails procedural consequences, including a reduction in the weight of proof or recognition of the results of the artificial intelligence system as inadmissible as evidence.

8. Suspension and prohibition of use.

If there are signs of violation of this Article or a threat to human rights and freedoms, a court or an authorized body may immediately suspend the use of the system in all proceedings or in individual cases.

a) Resumption of use is possible only after the elimination of violations and based on the results of an extraordinary audit and re-certification.

30.3 Procedural rights of participants and remedies.

1. Right to report (right to be notified).

The parties and the person in respect of whom the issue is being decided must be notified in writing before the first use of the AI system in the case, specifying: the purpose of the system, its name, version, supplier, role (advisory, search, etc.), the list of data, access modes and methods of appeal.

a) Failure to report or report after a fact ("post factum") entails the inadmissibility of the relevant results of the AI system as evidence.

The right to access information about the AI system.

At the request of the party, the operator is obliged to provide materials that ensure transparency, including local explanations of specific results in the case.

Access is provided in the form of copies or in a "safe room" with mandatory logging; Refusal is allowed only on the grounds of secrecy protected by law, provided that sufficient generalization is provided, which is necessary to ensure the right to defence.

The deadline for submission is no later than 10 days from the date of receipt of the petition, unless otherwise established by the court (in urgent proceedings — up to 48 hours).

2. Right to challenge the use of AI systems.

A party has the right to file a motion for:

(i) recognition of the results of the AI system as inadmissible evidence;

(ii) reducing their evidentiary weight;

(iii) conducting an independent algorithmic examination;

(iv) Suspension of the use of the system in the case until violations are eliminated.

The burden of proving compliance with transparency, certification, security, and human-in-the-loop requirements rests with the operator or supplier of the AI system.

The submission of such a motion cannot be considered as an abuse of procedural rights and does not entail negative procedural consequences for the applicant.

3. Human review [300].

At the request of the party, the court ensures that the person re-evaluates the algorithmic conclusion with a written motivation for accepting or rejecting the recommendation of the AI system.

If the algorithmic result cannot be properly explained or reproduced, it cannot be used as the basis for the reasoning part of the court decision.

4. Right to examination and assistance in defence.

The court, upon request, appoints an independent examination of AI algorithmic systems; The costs are borne by the party that lost the dispute based on the result of the examination, unless otherwise determined by the court.

In case of proven financial insolvency of a person, the court may order an examination at the expense of the state.

The parties' experts have the right to ask questions to the supplier or system operator regarding the methodology, data and metrics; Refusal without proper grounds entails procedural sanctions.

5. Right to data protection and privacy.

A person has the right to request the exclusion from the model of excessive or such data processed without proper legal grounds, as well as their correction or anonymization; The dispute is resolved by the court in simplified proceedings.

a) The use of sensitive data without a proper legal basis entails the inadmissibility of the relevant results and the possibility of filing a separate complaint with the authorized data protection body.

6. Standards of admissibility and evidentiary weight of algorithmic results.

The algorithmic result can be considered by the court only if it is proven:

(i) model validity for the stated task;

(ii) of proper quality and sufficient representativeness of the data;

(iii) absence of prohibited bias;

(iv) reproducibility of a specific conclusion.

Probabilistic scores cannot by themselves serve as a basis for establishing facts or determining the measure of punishment or preventive measure.

7. Remedies in case of violations.

Procedural: inadmissibility of evidence, reduction of its evidentiary weight, exclusion of references to the AI system from the reasoning part of the decision, postponement or suspension of the consideration of the case until the elimination of violations.

Material: cancellation of the decision and appointment of a new trial; revision according to newly discovered circumstances in case of detection of critical defects in the model.

Liability of the operator or supplier: fine, cancellation or suspension of the certificate, exclusion from the state register, civil liability for the damage caused.

Disciplinary: responsibility of officials for the use of uncertified systems or failure to comply with transparency obligations.

8. Stopping effect (suspensive effect) and timing.

Appealing against decisions on the use of the AI system in the case has a suspending effect (suspensive effect) on the further use of the relevant system until the complaint is considered on the merits, unless otherwise determined by the court for reasons of urgency.

Appeals against rulings on the use of the AI system are considered immediately, within a shortened timeframe; Missed deadlines can be renewed if the omission was due to the untimely provision of materials that ensure transparency.

9. Fixation in the court decision.

The court notes in the judgment whether the AI system was used, its name and version, the role in the process and the way in which the recommendations were considered or rejected.

If the algorithmic result is considered, the court is obliged to provide the key factors (local explainability) that influenced its conclusion.

10. Additional guarantees for vulnerable persons.

For minors, persons with disabilities and other vulnerable groups, increased standards of transparency and explainability are applied, the use of aggregated social indicators and behavioural scoring is prohibited.

The Court shall provide an explanation of the role of the AI system in an accessible and adapted form and shall provide additional procedural time for the exercise of the rights provided for in this Article.

11. Inadmissibility of negative consequences for the exercise of rights.

The exercise of the rights provided for by this Article (request for information, appeal, request for review by a person) cannot be the basis for conclusions on bad faith or abuse of procedural rights.

SECTION IV. SOVEREIGNTY, STATEHOOD AND PUBLIC INTEREST

CHAPTER 31. THE PRINCIPLE OF TECHNOLOGICAL SOVEREIGNTY

Technological sovereignty implies the ability of the State to exercise strategic control, management, audit, localization and full responsibility for the use and development of artificial intelligence systems without creating critical dependence on foreign states, transnational corporations or unfriendly jurisdictions.

The content of this principle includes: political autonomy in the field of AI; full legal control over the national artificial intelligence ecosystem; ethical and normative self-determination in matters of technologies affecting human rights; digital security of critical components; scientific and technological independence in the formation of standards, research priorities and directions of innovation.

The state guarantees the availability of national infrastructure for the deployment of AI systems, including independent data centres, cloud environments, computing power, and testing facilities. The development of technical regulations and standards in the field of AI is based on international practices, taking into account the national context. The state develops its own innovation ecosystem by supporting research centers, startups, educational programs and public regulatory sandboxes, as well as protecting digital sovereignty at the international level.

It is prohibited to use artificial intelligence systems on the territory of the State if it is impossible to conduct a full-fledged independent audit, there is no control over critical components, there are risks of bookmarks or backdoors, there is a critical technological dependence, or there are reasonable suspicions of facilitating the activities of foreign intelligence or military structures. Such systems may be prohibited by the authorized body on the basis of an expert opinion and entered into a special register.

The state has the right to apply special measures to protect strategic control in the field of AI, in particular: limit the export of sensitive technologies; to establish priority in public procurement according to the criteria of national security and support for domestic developers; approve foreign investments in critical AI infrastructure; carry out digital due diligence of cross-border transactions related to the transfer of key data, models or algorithms.

The central executive body that ensures the formation and implementation of state policy in the field of artificial intelligence implements this principle through the National Strategy for the Development of Artificial Intelligence, which determines research priorities, import substitution policy, coordination between sectors, support for innovations and international partnerships.

31.1 The principle of technological sovereignty is to guarantee the ability of the State to exercise strategic control, management, audit, localization and full responsibility and control over the use, implementation, development and decommissioning of artificial intelligence systems on the territory of the State without critical technological dependence on foreign states, corporations or unfriendly jurisdictions[301,302].

This principle includes:

- political autonomy in the field of AI, which ensures the right of the state to adopt its own development strategy regardless of geopolitical influences;
- legal control over the AI ecosystem — from rule-making to judicial supervision;
- ethical self-determination in the issues of admissibility of technologies that affect human rights, digital dignity, privacy and freedoms of citizens;
- digital security and control over critical components — in particular, large language models, computing infrastructure, update channels, databases, neural network architecture;
- scientific independence and sovereignty in the formation of research priorities, development of standards, interoperability mechanisms and digital innovations.

31.2 The state is obliged to [303, 304]:

- ensure the availability of critical digital infrastructure for the secure deployment and operation of AI systems without dependence on foreign servers, platforms or control nodes;
- maintain independent computing power, data centres, cloud environments and testing facilities of domestic and open origin;

- to create and maintain state standards, protocols and technical regulations in the field of AI, harmonized with international ones, but taking into account the specifics of the legal regime of the State;
- promote the development of a local innovation ecosystem, in particular through funding for research centers, start-ups, educational programs, public regulatory sandboxes, and institutional infrastructure for the development of AI systems;
- to protect national digital sovereignty in international organizations, forums and trade agreements, upholding the principles of transparency, technological neutrality, interoperability and respect for national jurisdiction.

31.3 In order to realize technological sovereignty, the central executive body that ensures the formation and implementation of state policy in the field of artificial intelligence develops and regularly updates the National Strategy for the Development of Artificial Intelligence, which defines[305]:

−priorities of research work;

−directions of import substitution and localization;

−mechanisms for stimulating public and private investments in the field of AI;

−the principles of interagency coordination, including the defence, civilian, educational, economic and security sectors;

−approaches to regional development and international partnership.

31.4 It is prohibited to introduce, disseminate or use on the territory of the State artificial intelligence systems, in respect of which [306, 307]:

the possibility of conducting an independent comprehensive technical, legal and ethical audit, revision of the model architecture, verification of training data and certification in accordance with national standards is not provided;

there is no or limited direct control by national actors over critical components of the system, including source codes, algorithms, update mechanisms, inputs (including training data), control interfaces or security modules;

identified risks of unilateral technological dependence, critical vulnerabilities, undocumented functionality or built-in mechanisms for influencing user behaviour, including the so-called "hidden features" or algorithmic backdoors;

there are reasonable suspicions that the system is directly or indirectly used for the benefit of foreign intelligence agencies, military or corporate structures whose activities are incompatible with the national interest, the security policy of the State or international law.

The decision to prohibit the implementation of such systems is made by the authorized state body on the basis of the conclusions of an independent examination with subsequent inclusion in a special register of prohibited or unwanted AI systems.

31.5 The state has the right to take special regulatory, institutional and security measures to maintain strategic control over critical technologies in the field of artificial intelligence, in particular[308,309]:

restrict or completely prohibit the export of critical solutions based on artificial intelligence, models, training samples, optimization algorithms or dual-use technologies that can be used in the defence, intelligence or infrastructure fields; their export is allowed only after a preliminary analysis of threats and obtaining a special permit from the relevant national security body;

establish priority in public procurement according to the criteria of origin, security, openness of code, certification and compliance with national interests, with a special emphasis on supporting domestic scientific institutions, start-ups, the military-industrial complex, higher education institutions and centres of expertise;

thoroughly vet and approve foreign investments, mergers, acquisitions, and joint ventures in the field of strategic AI infrastructure, including computing clusters, data centres, data analytics platforms, cyber defence systems, development of basic models and language architectures;

conduct mandatory digital due diligence of any cross-border transactions related to the transfer or processing of large data sets, components of AI architecture, logic modules, learning models or control subsystems, taking into account the requirements of national security, digital ethics, preservation of intellectual potential and scientific independence of the State.

CHAPTER 32. THE PRINCIPLE OF DIGITAL NEUTRALITY

On the territory of the State, it is prohibited to develop, implement, integrate and use artificial intelligence systems that directly or indirectly violate the principle of digital neutrality give an unreasonable advantage to some or harm other states, companies, ideologies, groups or individuals by distorting information processes.

This principle applies to all aspects of the functioning of AI systems: query processing, filtering, ranking, content personalization, resource management, access to digital services, and interaction with other AI systems.

The artificial intelligence infrastructure must ensure platform, content, operational neutrality, as well as information openness. Algorithms must not contain bias, elements of hidden censorship, or discriminatory practices that influence the formation of users' opinions or restrict access to information based on language, origin, social status, or beliefs.

AI system operators are required to document the logic of decision-making, provide users with clear explanations, and ensure that an independent audit is conducted to ensure compliance with the principle of digital neutrality.

Established cases of digital discrimination, algorithmic manipulation or distortion of algorithms are recognized as a threat to national security and digital sovereignty. In such cases, mandatory notification of the regulator, interdisciplinary audit, possible blocking of the system, suspension of certification and application of sanctions are provided.

Information about violations is subject to entry into the public Register of Digital Neutrality Incidents.

The Government of the State ensures the development and regular updating of the White Paper on Digital Neutrality of the State as a strategic document defining the principles of neutrality, key risks, typical examples of violations, monitoring methods and response mechanisms. The document is updated at least once every two years and undergoes public discussion.

32.1 The principle of digital neutrality is that no artificial intelligence system developed, implemented, integrated or used on the territory of the State should directly or indirectly favour, restrict, discriminate, distort or manipulate digital processes, information flows, algorithmic decisions or data processing results for the benefit of some subjects and/or to the detriment of others, in particular states, transnational companies, political ideologies, religious teachings, financial groupings, national, ethnic or social groups, professional categories or individuals [310, 311].

This principle applies to all aspects of AI systems, including:

- the logic of processing requests;
- ranking, filtering and personalization of content;
- management of computing resources;
- regulation of access to government, educational, medical or administrative digital services;
- interaction with other artificial intelligence systems, including transnational or hybrid solutions.

Any deviation from the principle of digital neutrality is considered a violation of democratic principles, the principle of equal and non-discriminatory access to information and technology, and also poses a threat to the digital sovereignty of the State.

32.2 All digital infrastructures related to the use of AI are required to function in accordance with the following principles:

- platform neutrality — no dependence on a single vendor (vendor lock-in);
- content neutrality — the absence of politically or commercially motivated moderation, except for cases expressly defined by law;
- operational neutrality — ensuring equality in the performance of computing tasks, access to data and updates on non-discriminatory terms;
- information openness — prevention of artificial restriction or privileged dissemination of information.

32.3 AI system operators are required to ensure that internal decision-making logic, recommendation algorithms, classification, filtering, ranking, personalization, and other related processes do not contain built-in bias, hidden discrimination, elements of automated censorship, or external interference capable of, among other things,:

- influence the formation of political beliefs, religious or philosophical views of the user;
- restrict access to information based on language, ethnic origin, citizenship, social status, political affiliations, gender identity or other sensitive characteristics;
- create "digital bubbles" or "algorithmic echo chambers" that limit critical thinking and pluralism of thought;
- distort search results, recommendations of goods, services, news or public services in favour of specific subjects of economic or political power, which constitutes a manifestation of discrimination or unfair competition.

In order to ensure transparency, operators are obliged to:

- document the logic of decision-making and version-locking model updates;
- provide users with clear explanations of the reasons for the decisions made;
- provide the possibility of independent auditing and periodic verification of algorithms for compliance with digital neutrality and non-discrimination.

32.4 Any detected cases of violation of the principle of digital neutrality, including factually recorded cases of digital discrimination, information manipulation, algorithmic distortion or deliberate privilege or restriction of certain subjects or ideas, pose a direct threat to national information security, democratic order and digital sovereignty of the State [312].

In case of detection of such violations:

- the operator is obliged to notify the national AI and digital regulator within 24 hours with a detailed description of the incident;
- the regulator initiates an independent interdisciplinary technical, legal and ethical audit with the involvement of civil society experts and international partners;
- for the period of investigation, the regulator has the right to introduce a temporary restriction or complete blocking of access to the relevant system, as well as to suspend or revoke the certificate of conformity;
- in case of confirmation of intentional violation or systemic negligence, administrative, financial, and, if there is damage, criminal sanctions are applied to legal entities and individuals in accordance with national legislation and international obligations of the State.

Information about such cases is subject to inclusion in the National Register of Digital Neutrality Incidents with open access for scientists, media and the public, except for data constituting state secrets or information protected by law.

32.5 The Authorized Body of Public Administration in the field of AI, in cooperation with specialized scientific institutions, independent expert centres, digital rights institutions and representatives of international technical assistance, organizes the development, regular updating and implementation of the "White Paper on Digital Neutrality of the State" — a strategic conceptual document of a recommendatory and regulatory nature, which establishes [313]:

- basic principles of digital neutrality as a component of the national digital policy;
- classification and description of potential risks of digital discrimination, manipulation and distortion of information;
- models of application of the principle of digital neutrality in various areas (education, healthcare, justice, defence, cybersecurity, social networks);
- indicators, criteria and monitoring techniques, including automated tools, independent examination mechanisms and human control;
- procedures for responding to violations, including mechanisms for public reporting, prosecution and engagement with civil society.

The White Paper is an official document with open access, which is updated at least every two years and is subject to public discussion with the involvement of representatives of the public, business, the IT sector and international partners.

CHAPTER 33. THE PRINCIPLE OF DIGITAL SOVEREIGNTY AND NATIONAL CONTROL OVER DATA

The digital sovereignty of the State is defined as the totality of exclusive, inalienable and legally protected public law powers of the State to form, implement and control all policies, regulations, technological architectures and practices that regulate the circulation of data on the territory of the country and in its cyberspace, as well as in cases where the data are related to citizens or residents of the State.

Digital sovereignty is a legally binding principle that defines the limits of the legitimate functioning of artificial intelligence systems and other digital services on the territory of the State. No AI system or digital infrastructure that affects the management of state or socially significant processes is subject to extraterritorial treatment in relation to the norms of national law. Data sovereignty means that data are governed by the laws of the State in whose territory they are created, stored and processed, and are subject to processing and transfer exclusively in accordance with the laws of that State, including when they are transferred outside its jurisdiction.

Digital sovereignty includes the right to establish legal, technical, cryptographic, ethical and institutional frameworks for the processing, analysis, storage, transfer, aggregation, cross-border processing, interoperability, reuse, localization and destruction of all types of data — regardless of their personalized status — that are directly or indirectly related to the jurisdiction of the State, its natural or legal persons, digital infrastructure or information sovereignty.

Institutional control is carried out by concluding contracts with operators and owners of digital systems, maintaining technical documentation, conducting audits and exercising jurisdictional supervision over their activities. Mandatory local storage of data on the territory of the State and restriction of its transfer outside the State is a key tool for the implementation of digital sovereignty, which is used to prevent the loss of control over data and related strategic resources.

To implement digital sovereignty, the state establishes: localization of servers and data on its territory in order to ensure control and access; state control over the logic of the systems through API openness, oversight of updates, and independent audits of digital systems; measures for disaster recovery and ensuring autonomous management in case of cyber incidents or loss of control by a foreign operator. The supreme executive body of the State (or another central executive body authorized by law) approves and maintains the Register of Sovereign Critical IT Systems and establishes for them special licensing conditions, requirements for data placement and encryption, the procedure for auditing AI algorithms and the mandatory return (repatriation) of data to the territory of the State Law.

All digital platforms, operators, providers, artificial intelligence systems and analytical modules that operate with data of national origin or access national data sets are obliged to:

provide physical storage of primary data in certified national data centres;

notify the state regulator of any data transfer outside the country, indicating the purposes, recipients, protection channels and deadlines — no later than 72 hours from the date of the decision on the transfer;

implement monitoring interfaces for state control with transparency of access and protection of trade secrets;

carry out mandatory localization of critical data, the list of which is determined by law;

comply with the policy of assessing countries by the level of legal and technical protection of information in case of cross-border data processing.

The authorized central executive body in the field of digital technologies and artificial intelligence establishes and approves technical, organizational and cryptographic requirements for the architecture of data handling; carries out a systematic audit of digital platforms for compliance with these requirements; maintains and updates the Register of Critical Data; determines the modes of their storage and processing depending on the categories of influence on national interests; and is empowered to issue binding orders to restrict or terminate the activities of digital platforms in case of violation of the requirements of this article.

Violation of the principle of digital sovereignty entails, depending on the nature and degree of the violation, the application of a set of legal measures, in particular: administrative liability, financial

sanctions, temporary suspension of market access, criminal prosecution in case of a threat to state security, public entry into the Register of Digital Sovereignty Violators with the determination of the term and nature of restrictions on public-private turnover, as well as other measures provided for by law.

The Government of the State is obliged to ensure the functioning of a sustainable independent digital infrastructure, which includes:

public cloud environment using open source technologies;

geo-distributed network of certified data centres;

unified interoperable systems for logging access to data;

mechanisms of the digital gateway of sovereignty (tools for filtering and controlling cross-border requests for data).

This infrastructure must guarantee technological independence, comply with the principles of cyber defence, energy autonomy and ensuring human rights in the digital dimension, taking into account the international obligations of the State.

33.1 Digital sovereignty is the ability of the state to establish and protect its own rules for the functioning of digital systems. For this reason, many countries are implementing data localization policies, fearing that their storage and processing outside of national jurisdiction could lead to a loss of control and jeopardize state sovereignty [314, 315].

The concept of "*data sovereignty*" means that data is subject to the laws of the country where it is hosted, and the state has the right to establish rules for its collection, storage, processing, transfer and protection[316]. This principle covers not only technical aspects, but also the legal subordination of all information processes. Digital sovereignty in the modern legal order is considered not as a declarative political value, but as a legally binding principle. Accordingly, no AI system or digital infrastructure can be considered "extraterritorial" in relation to the jurisdiction of the State if it directly or indirectly affects the management of state functions or socially significant processes [317].

Digital sovereignty of a State means the sovereign, inalienable and exclusive right of the State, its citizens and residents to formulate, implement and control policies, legal regulations, technical protocols, digital architectures and ethical restrictions governing the processing, analysis, storage, transfer, aggregation, interoperability, monetization, reuse and destruction of data that:

are generated on the territory of the State or in its digital space;

relate to natural and legal persons under the jurisdiction of the State;

are processed on digital platforms, servers, cloud environments or infrastructures capable of directly or indirectly affecting the sovereign rights of the State;

are the result of the activities of artificial intelligence systems, digital agents, sensor systems, machine learning platforms or other forms of digital analytics operating within national cyberspace or using national data.

Digital sovereignty encompasses both personal and non-personal data, including open, biometric, behavioural, analytical, geospatial, energy, financial, educational and other types of data, which are recognized as strategic or critical for sustainable development, security, technological independence and democratic self-determination of the State, in accordance with the list determined by law.

All artificial intelligence platforms, regardless of the country of origin, form of ownership, technical architecture or place of legal registration, which operate on the territory of the State and receive, accumulate or process data on natural or legal persons under the jurisdiction of the State, are obliged to comply with the following mandatory requirements:

ensure the storage of all primary (raw) data sets in certified national data centres located on the territory of the State and meeting state requirements for security, energy autonomy and protection against physical and cyber threats;

notify the regulator of each case of data transfer outside the country no later than within 72 hours from the date of the transfer decision — indicating the justified purpose, category of data, identification data and legal status of the recipient, technical security protocols and storage periods abroad;

integrate API access for the state digital sovereignty control body into its own access management and monitoring systems, which allows you to track, document and analyse access to data arrays, their movement, copying, aggregation and connection of third parties;

adhere to the policy of mandatory localization of critical data (including biometric, behavioural, financial, medical, educational, geolocation), as well as refrain from transferring data to countries officially recognized as not guaranteeing an adequate level of legal, technical and ethical protection of information.

The principle of digital sovereignty applies in cases where [318]:

1) the state register is located in a cloud storage controlled by a foreign company;

2) administration of the algorithm that regulates the movement of transport or the functioning of the energy system is carried out outside the State;

3) AI systems decisions on social benefits or lawsuits are based on servers or program code that are inaccessible to government agencies.

For public institutions, the risk of foreign governments' access to data stored abroad poses a significant threat to national security and necessitates the introduction of restrictions on the activities of companies that provide such services.

33.2 This principle forms the basis of sovereign technological governance and determines the subordination of digital architecture to the requirements of national law. States establish legal regulation regarding the storage, processing and transfer of data to guarantee security and ensure sovereign control [319]. One of the key mechanisms for restoring digital sovereignty, especially for countries that do not have a dominant position in the field of technology, is mandatory local data storage and restrictions their relocation outside the national jurisdiction [320].

33.3 For the implementation of digital sovereignty, it is provided:

mandatory territorial localization of servers — data created within the state must be stored on its territory, which ensures control over access and processing in accordance with national legislation;

state control over the logic of the functioning of systems — through API openness, supervision of updates, and independent audits;

mechanisms for emergency recovery of management autonomy, in cases of cyber incidents or loss of control by a foreign operator.

33.4 The ban on digital colonization reflects the current global challenges associated with excessive dependence on the activities of international platforms and services. Such dependence can lead to the subordination of digital infrastructure to the interests of foreign states or multinational corporations. Studies show that digital dominance, concentrated in a few global players that control technology, data and network resources, creates "neo-colonial" dependence and poses a threat to the sovereignty of states that do not have sufficient technological potential. In this context, the legal principle of a direct ban on digital colonization and any forms of indirect interference by foreign actors in critical digital infrastructure is enshrined. Critical IT systems and artificial intelligence solutions that affect state or socially significant processes are subject to exclusive control within the national jurisdiction.

33.5 Any direct or indirect interference by foreign actors in critical digital infrastructure and AI systems is prohibited if such interference could lead to dependence on external technology companies. The dominance of individual states or corporations in the digital sphere creates a risk of "neo-colonial" dependence and poses a threat to the digital sovereignty of other countries. The principle of digital sovereignty applies, in particular, in cases where: the state register or other key data is hosted in a cloud service under the control of a foreign company; the algorithm that regulates the transport or energy system is administered outside the State; The decision of the AI system on social benefits or lawsuits is based on servers or code that are inaccessible to government agencies. The risk of offshore data storage, especially for public authorities and agencies, necessitates the restriction of the activities of providers who exercise control over data outside the State.

Prohibited:

any hidden, opaque or unauthorized collection, aggregation, analysis, correlation or combination of personal and non-personalized data without the express, informed and voluntary consent of the data subject, obtained in accordance with the requirements of data protection legislation;

transfer of data for possession, use or temporary use to other persons or systems, regardless of jurisdiction, including their resale — without direct notification of the data subject, indication of the purpose, term, security conditions and ensuring the possibility of revocation of the authorization;

delegating the functions of processing, analysing or storing personal or critical data to third parties located in the jurisdictions of foreign countries, without going through a mandatory procedure for a comprehensive assessment of security, legal, technological and geopolitical risks;

the use of digital platforms, algorithms or infrastructure on the territory of the State that have not passed the procedure for assessing compliance with the Digital Sovereignty Policy, are not included in the National Register of AI Entities and are not integrated with state mechanisms for transparency, monitoring and control of data circulation [321].

33.6 Authorized central executive body in the field of digital sovereignty:

establishes and approves a set of technical, organizational and cryptographic requirements for the architecture of storage, routing, processing, exchange, encryption and backup of data, taking into account the category of data, the level of their criticality and requirements for interagency and cross-border interaction;

regularly audits foreign and domestic platforms — including cloud services, data centres, artificial intelligence modules, application programming interfaces (APIs) and other elements of digital infrastructure — for their compliance with the requirements of the digital sovereignty policy, national interests, as well as the requirements of cybersecurity and personal data protection;

maintains and updates the Register of Critical Data, classifying information according to the degree of impact on national security, economic stability, defence capability, infrastructure, healthcare, education, energy, science, culture and information space, and determines a special regime for their storage, processing, access, control and recovery;

empowered to issue binding orders on the restriction, suspension or complete termination of the activities of a digital platform that violates the requirements of this Article, as well as to initiate, in cooperation with national and international telecom operators, the blocking of technical protocols, interfaces or access to the network in case of detection of violations that pose a threat to digital sovereignty.

33.7 Violation of the principle of digital sovereignty entails the application of sanctions, administrative, financial and criminal measures of responsibility, determined depending on the scale, nature, recurrence and severity of the violation [322]. In particular, this may include:

bringing to administrative responsibility with the imposition of fines, suspension of the license or restriction of the functionality of the relevant service;

criminal prosecution if the violation is intentional, repeated or systemic and is committed with the aim of harming state sovereignty, national security, citizens' rights or critical data infrastructure;

temporary suspension or complete blocking of the operation of the service, digital platform or data transmission channel on the territory of the State until the violation is completely eliminated and compliance with the requirements of digital sovereignty is confirmed;

inclusion of an entity in the National List of Sanctions in the Field of Digital Activities, if it violates the digital rights of citizens of the State or state policy in the field of data, which entails restrictions on access to the market, banking operations, public financing, public tenders, research cooperation and investments;

publication of information about the subject in a special open Register of Digital Sovereignty Violators, which is maintained by the authorized body and used by public and private counterparties to assess the risks of interaction with such an entity.

All measures are accompanied by a legal assessment, ensuring the right to appeal and establishing a procedure for reinstatement in case of voluntary elimination of the violation and passing a repeated compliance audit.

33.8 The Government of the State is obliged to organize the development, financing, legal support, guarantee the operational resilience and strategic renewal of the independent state digital infrastructure for secure processing, long-term storage, analytics and access to critical data. Such infrastructure must meet the requirements of national digital sovereignty, information security and technological independence. Its key components are:

sovereign government cloud, built on the basis of open source or controlled technologies, with scalability, state-level encryption and isolation from external technology platforms;

geo-distributed network of data centres with mandatory domestic jurisdiction, cyber protection certification, autonomous power supply and physical protection infrastructure in accordance with international standards (not lower than Tier III);

national interoperable platforms for exchanging, monitoring, and logging AI systems' access to data sets that provide end-to-end recording of all queries, processing, modifications, and analytical operations with the possibility of independent audit;

Digital Sovereignty Gateway mechanisms, which provide control over interaction with transnational digital structures through mandatory identification, filtering and rejection of requests that do not comply with the policy of national data control.

CHAPTER 34. THE PRINCIPLE OF DIGITAL DEMILITARIZATION AND THE PROHIBITION OF ARTIFICIAL INTELLIGENCE AGAINST VICTIMS OF MASS DESTRUCTION

The state recognizes the integration of artificial intelligence into the processes of creating, managing, testing or using weapons of mass destruction as unacceptable from an ethical, humanitarian and security point of view. Such integration contradicts the fundamental principles of international law, undermines human control over the conduct of war and poses a threat to the existence of civilization in conditions of technological self-destruction.

Weapons of mass destruction include any type of weapon or technology capable of causing catastrophic damage to the general population, the environment, critical infrastructure or the institutional stability of states. In particular, such means include nuclear, chemical, biological, hypersonic, autonomous, cyberphysical, and combined forms of combat potential, in cases where their architecture contains elements of artificial intelligence.

It is prohibited to use artificial intelligence at any stage of the life cycle of these tools — from research and design to production, modernization, testing, management, and combat use. In particular, the following are unacceptable: automated target designation; simulation optimization of combat algorithms; algorithmic control of damage trajectories; autonomous decision-making on the use of force; delegating to machine learning systems the function of assessing strategic threats or selecting attack targets.

Any natural or legal person, public institution or private organization engaged in the field of military research, production, modernization or operation of defence technologies is required to undergo mandatory ethical, legal and security expertise. Such an examination is carried out in order to prevent the inclusion of artificial intelligence components in decisions that can affect the control, launch or escalation of combat potential with potentially massive consequences.

The state recognizes the principle of digital demilitarization as a basic component of national and global security and initiates the creation of an international digital non-proliferation regime, which should become an analogue of nuclear control in the digital age. Within the framework of this regime, it is envisaged to: a moratorium on the use of AI in means of mass destruction; international certification of dual-use systems; independent international verification of combat algorithms; open audit of their architectures; implementation of sanctions and prosecution mechanisms for attempts to circumvent established restrictions.

Any detected violation of this article — in particular, the facts of the use or development of AI components in the context of weapons of mass destruction — is recognized as an act of global danger. Such actions qualify as crimes against peace, humanity and international law within the meaning of international criminal law, and entail international criminal prosecution, the application of sanctions, as well as authorized measures to block, neutralize or destroy the relevant digital or physical infrastructure that poses a threat.

The member states of the international community call on all countries of the world, international organizations, the scientific community and the high-tech industry to support the development of an international convention on the prohibition of the use of artificial intelligence in the context of means of mass destruction. Such a convention should enshrine in international law the fundamental principle:

34.1 The state recognizes the use of artificial intelligence technologies as unacceptable and incompatible with the principles of international humanitarian law, ethics and national security (AI) in the creation, development, testing, modernization, maintenance, as well as in the tactical or strategic management of weapons of mass destruction (WMD) [323].

Such WMDs include, in particular, but are not limited to:

- nuclear weapons — including guidance systems, trigger mechanisms, countdown logic, automated response scenarios [324, 325, 326];
- biological weapons — including algorithmic simulations of pathogenesis, synthesis of artificial pathogens, processing of biomedical data to optimize virulence [327];

- chemical weapons — including automated creation of combat compounds, real-time forecasting of the spread of toxic substances using models AI;
- cyber weapons of strategic destruction — capable of disabling critical infrastructure, causing mass man-made disasters or human casualties because of cyberphysical effects [328];
- hypersonic, orbital, underwater, autonomous or multi-vector WMD delivery platforms operating under conditions of minimal or complete autonomy of the AI system, excluding humans from the decision-making cycle (human-out-of-the-loop) [329].

Any involvement of artificial intelligence in the processes of modelling, evaluation, performance forecasting, management or decision-making in relation to WMD is recognized as a fundamental violation of the principles of democratic civilian control over military technologies and creates unacceptable risks for the survival of humanity.

34.2 Prohibited:

- any use of artificial intelligence systems to control, launch, coordinate, target, or take actions to activate the launch mechanisms of weapons of mass destruction, including algorithmic control of detonation, calculation of trajectories, course adjustments, and determination of time or objects of destruction in any manner [330];
- integration of AI into aiming systems that operate autonomously or semi-autonomously, without the mandatory presence of a human operator, who has a real opportunity to intercept, change or cancel the fire command at any time, with the obligatory recording of this intervention in decision logs [331];
- delegating the processes of threat assessment, target selection and decision-making on the use of lethal force, including at the strategic level (mass destruction, escalation of conflicts, destruction of critical infrastructure facilities), to any digital system, algorithm or neural network without guarantees of continuous and legally significant human control [332, 333];
- training, training, additional training or testing of artificial intelligence models on data sets, simulation models or combat scenarios that are directly or indirectly related to the analysis of the effectiveness, scale or mechanics of weapons of mass destruction, in particular through the integration of military databases, information from satellite intelligence, open scientific data and simulations of physical, biological or chemical processes of destruction [334].

34.3 Any organizations, institutions or business entities — regardless of ownership, organizational and legal status or subordination — that carry out activities in the field of research, design, development, testing, production, supply, maintenance or modernization of weapons, military technologies or critical infrastructure and dual-use technologies are required to undergo a mandatory multi-level examination to prevent the integration of components artificial intelligence in procedures that:

- directly or indirectly affect the control over means of mass destruction;
- create the risk of delegating lethal decisions to digital systems;
- are used as elements of autonomous combat platforms;
- generate, interpret, or transmit combat commands without confirmed human intervention.

This examination covers the ethical, legal, technical and security components of the assessment, is carried out by notified bodies with the participation of independent international experts and is a prerequisite for obtaining authorization for any stage of the life cycle of a defence product.

34.4 The state officially recognizes the principle «Digital Non-Proliferation» (digital non-proliferation in the field of weapons) as a key element of the global strategic security regime aimed at preventing the use of artificial intelligence technologies in the field of development, optimization, proliferation and use of weapons of mass destruction.

In this context, the State initiates the creation of an International Agreement on the Prohibition of the Use of AI in the Field of Weapons of Mass Destruction, which should provide for:

- recognition of AI as a technology subject to a special regime of international control when combined with WMD;
- a mechanism for independent certification and verification of military digital systems for the absence of uncontrolled AI components in them;

− introduction of a moratorium on research and testing of AI algorithms in the context of strategic weapons until the development of universal norms of international law;

− the establishment of a permanent oversight body under the auspices of the United Nations or a global multilateral alliance mandated for access, verification and response;

− criminalization in international law of attempts to circumvent, conceal or imitate AI components in strategic weapons systems.

The state undertakes to promote this initiative at the level of of the United Nations General Assembly, as well as within regional organizations and recognizes this policy as an integral part of the international digital legal order and the global security system.

Detection of any violation of the provisions of this Article, in particular the facts of the development, testing, implementation, use, transfer, licensing or export of weapons, systems or components that combine the functions of means of mass destruction with elements of artificial intelligence (both autonomous and auxiliary), is recognized as an internationally wrongful act, which creates an extremely high level of risk to the foundations of national sovereignty of states, the viability of its democratic institutions, the physical security of citizens, the state of the environment, global stability, the prevention of armed conflicts and the preservation of human civilization.

Such actions are classified as a threat of an international nature that goes beyond the limits of domestic jurisdiction and require a response both within the framework of the domestic law of states and through the mechanisms of universal jurisdiction, international criminal, humanitarian and security law.

Such violations entail:

- criminal liability with the qualification of relevant actions as crimes against peace, crimes against humanity and war crimes in accordance with the Rome Statute of the International Criminal Court and the national criminal codes of the member states of the international community;
- the imposition of international sanctions, including economic, technological, diplomatic and defence sanctions, against States, state and non-state organizations, individuals, companies and research institutions that have participated in or facilitated such actions;
- inclusion of violators in special sanctions lists, which provides for the blocking of assets, restriction of access to financial systems, as well as the prohibition of scientific and technological cooperation;
- public qualification of such actions as threats to all humanity with the initiative to bring perpetrators to individual and institutional responsibility within international courts or specially created hybrid tribunals;
- authorized neutralization or elimination of infrastructure, digital systems, research facilities or cyber environments used to implement violations — with documented international control and in coordination with the UN Security Council or other authorized international body.

CHAPTER 35. PROTECTING THE DIGITAL ENVIRONMENT AND THE RIGHT OF FUTURE GENERATIONS TO A SAFE DIGITAL ENVIRONMENT

The digital ecosystem is recognized as an integral part of the social, cultural, and mental landscape of modern society. It encompasses all levels of human-digital interaction — from infrastructural to cognitive — and is shaped, used and regulated in accordance with the principles of digital sustainability, intergenerational responsibility, safe technological progress and preservation of the living environment for future generations.

Digital ecology in this Act is a multidimensional system in which technical, ethical, cultural and social factors coexist harmoniously without prejudice to the mental and physical health of a person, his cognitive and informational autonomy. All digital systems, platforms, infrastructures, artificial intelligence systems and autonomous agents are subject to development, implementation and use exclusively in a way that ensures the preservation of the ecological balance of the digital space, guarantees the humanistic principles of social development and prevents the emergence of new forms of digital inequality or intellectual colonialism.

It is prohibited to use algorithms that cause psychological pressure, impose patterns of behaviour or restrict freedom of choice, form addiction or create self-replicating digital environments that lead to the loss of a person's ability to think critically and autonomously. It is not allowed to use digital systems that affect biodiversity, energy consumption, climate change or generate waste without a Life Cycle Assessment in accordance with international environmental standards.

The state provides a national policy of digital ecology, which provides: the introduction of ethical standards for the design and use of digital technologies; conducting mandatory environmental expertise of digital infrastructures; protection of the rights of the most vulnerable groups of the population — children, the elderly, persons with disabilities, refugees, representatives of national minorities — from the influence of dangerous, opaque or uncontrolled digital systems.

Everyone has an inalienable right to a fair, transparent, humane and environmentally friendly digital environment that does not cause digital exhaustion, emotional discomfort or manipulative influence. Such an environment should be free from intrusive content and excessive information flows that can create chronic anxiety or distract cognitive attention without the explicit consent of the user. A person has the right to digital hygiene, which includes access to tools for monitoring their own activity, the right to depersonalize data, control over their own digital identity, and the ability to proactively restore their autonomy in the digital space.

The state initiates the creation of the International Charter of Digital Ecology and the Rights of Future Generations as a new type of normative and ethical act that enshrines the right of future generations, including the unborn, to a safe, sustainable, humane and dignified digital future. The Charter proclaims the digital ecosystem as a global public good, obliges states, companies and scientists to act within the acceptable "green digital footprint", introduces standards of digital responsibility and establishes the obligation of each generation to ensure that the digital environment is transferred in a better condition than received.

Digital ecology encompasses not only aspects of energy consumption and potential environmental damage, but also the fundamental principles of dignity, freedom, psychological safety, cultural continuity, and intergenerational justice. A person acts not only as a user of digital technologies, but also as their co-creator, keeper and ethical subject of the digital environment. The state recognizes this role and undertakes to integrate the principles of digital ecology into all areas of state policy, the education system, legal regulation mechanisms and the practice of international cooperation.

35.1 The state recognizes the digital ecosystem not only as an infrastructural, technological or economic component of modern life, but as an element of the anthroposocial space that directly affects mental health, cultural identity, civic autonomy and the right of every person to sustainable development [335].

The digital environment is recognized as a form of the new common heritage of humanity, to which the principles of responsibility, non-discrimination, sustainability, ethical design and the right to digital ecological balance should be applied [336].

All artificial intelligence systems, autonomous digital agents, algorithmic infrastructures, computing platforms, and simulation environments shall be designed, deployed, and operated [337] in such a way that:

does not disturb the balance between technological progress and human dignity;

does not replace the institutions of intergenerational responsibility with algorithmic calculations;

does not create new digital inequalities or digital exclusion zones;

contributes to the preservation of cultural diversity, environmental safety and digital integrity as basic conditions for the life of future generations.

35.2 Prohibited:

systematic or large-scale use of AI algorithms to create psychological, cultural, or informational pressure on a person, including mind manipulation, imposition of decisions, or the formation of unauthorized dependencies on digital influence systems;

uncontrolled and exponential scaling of digital agents that can self-learn, self-replicate or autonomous decision-making without clearly defined boundaries of legal regulation;

creation of digital systems capable of influencing biodiversity, climate change, energy consumption or the life cycle of hardware without proper Life Cycle Assessment.

35.3 The State undertakes to ensure [338]:

development of ethical and technological standards for digital sustainable development;

conducting mandatory environmental expertise of digital infrastructures and technological solutions, considering their impact on energy, the environment, social practices and future generations;

protection of the digital rights of children, the elderly and socially vulnerable groups in the context of the application of algorithms that make decisions in the field of education, labour, healthcare, migration or access to services.

35.4 Every person, regardless of age, gender, nationality, social status or level of digital competence, has an inalienable right to [339]:

transparent, fair, environmentally sustainable and psycho-emotionally safe digital environment, in which there is no hidden algorithmic discrimination, the imposition of commercial or political narratives, as well as excessive intensity of digital interaction that harms psycho-emotional well-being;

protection against digital noise, information overload and excessive content that distracts attention, provokes anxiety or is addictive, including advertising, content that deliberately imitates relevance and background digital streams;

implementation of individual digital hygiene, in particular:

- access to audit and monitoring tools for your own digital activity;
- the right to "digital oblivion" with the possibility of deleting accounts, anonymization of data and termination of automatic profiling;
- the ability to restrict or change your digital identity, temporarily opt out of data processing, tracking, or personalized influences;
- the right to view and re-verify digital traces (search history, actions, preferences, etc.) and proactively restore their own digital autonomy.

The state initiates the creation of the International Charter for Digital Ecology and the Rights of Future Generations, a fundamental international act of ethical, legal and strategic nature, which should enshrine the principles of interaction between humanity, digital technologies, the natural environment and unborn generations as participants in a global digital treaty (contract).

The charter should include:

declaring the digital ecosystem a global public good that requires an international legal protection regime;

consolidating the obligation of states, companies, scientists and technological developers to act responsibly within the framework of the "ecological digital footprint", considering the long-term consequences of the introduction of technologies for nature, people and culture;

setting standards for digital environmental reporting for digital platforms, including resource use, heat emissions, mental health impacts, and changing behavioural patterns;

recognition of the right of future generations to a safe, dignified, ethically controlled and non-violent digital environment that does not threaten the survival of humanity and the sustainability of the biosphere.

The basis of this initiative is the principle: "A person is considered not only as a user of the digital environment, but as its responsible co-creator, guardian and guarantor of the continuity of technological development based on respect for life."

CHAPTER 36. THE PRINCIPLE OF DIGITAL SUPERIORITY AND INNOVATIVE DEVELOPMENT

The state forms and implements digital policy aimed at unlocking the potential of artificial intelligence in the fields of innovation, science, defence, healthcare, education and public administration, on the basis of security, ethics and sustainable development.

In order to ensure digital advantage and stimulate innovative development, the state:

implements mechanisms of regulatory experiments (the so-called. "sandboxes") to test innovative solutions in controlled conditions without prejudice to human rights and freedoms;

develops and supports public-private partnerships in the field of research, development and commercialization of the results of technological developments in the field of artificial intelligence;

creates and expands national data centres, cloud computing platforms and open data infrastructure suitable for training and testing artificial intelligence models;

contributes to the formation and operation of digital clusters, incubators and research laboratories focused on the development of artificial intelligence.

A system of standards for open application interfaces (APIs) and interdepartmental interoperability of digital services is being introduced, which creates conditions for accelerating the development of solutions based on artificial intelligence, increasing their efficiency and adhering to ethical standards for their application.

36.1 The state forms and implements a single comprehensive national digital policy, which provides for a system of strategic, regulatory, institutional and financial measures aimed at achieving long-term and sustainable leadership in creating, implementing and scaling the application of artificial intelligence technologies [340]. This policy covers key innovation areas, including basic and applied scientific research, the security and defence sector, healthcare, education, industrial production, and public administration [341, 342, 343].

Policy development and implementation is based on the following principles:

the principle of security — ensuring cyber protection, information resilience and minimizing risks to national security, as well as human rights and freedoms;

the principle of ethics — ensuring the integration of ethical norms into all stages of the life cycle of AI systems, with the prevention of discrimination, bias and violation of the right to privacy;

the principle of sustainable development — ensuring a harmonious combination of economic, social and environmental interests, in particular using energy-efficient technologies in the creation and operation of artificial intelligence systems;

the principle of compliance with international standards is the harmonization of national norms with the requirements and recommendations of the EU, OECD, ISO, IEEE and other international organizations, as well as active participation in the formation of global standards and protocols.

The policy also provides for the formation of an innovation ecosystem that provides synergy between government agencies, scientific institutions, business and civil society, creating conditions for attracting investments, developing human capital and integrating the State into the international digital space.

In order to accelerate the implementation of innovations and verify their safety and compliance with legal and ethical requirements, the state creates and maintains regulatory sandboxes (hereinafter referred to as "sandboxes") — special legal regimes for testing AI technologies under controlled conditions.

Such sandboxes:

- function under the jurisdictional, organizational and technical supervision of authorized state bodies;
- provide for certain terms of testing and mandatory control over the results;
- guarantee the protection of the rights and freedoms of participants in experiments;
- Startups, scientific laboratories, public and private companies are allowed to participate under simplified admission procedures.

The results of sandbox activities are subject to mandatory publication and analysis in order to improve the regulatory framework.

The state promotes the development and operation of public-private partnerships in the field of research, development and commercialization of AI technologies.

This includes:

- joint financing of research and development works;
- creation of scientific and industrial consortia;
- exchange of research results and technological developments (know-how) between public institutions and the private sector;
- mechanisms for the distribution of intellectual property rights that ensure a balance between the interests of investors, developers and the state.

Such teams ensure a reduction in the time from laboratory tests to their practical application and commercial implementation.

The state creates, builds, and expands national data centres, cloud platforms, and open data infrastructure to train and test AI models.

This implies:

- the use of standardized data formats to ensure their compatibility and quality;
- providing free or licensed access for scientific institutions, startups and enterprises;
- data protection in accordance with legislation in the field of cybersecurity and personal data protection;
- optimization of energy consumption and the use of renewable energy sources during the operation of data centres.

36.2 The state promotes the creation of digital clusters, incubators and research laboratories focused on the development and implementation of AI solutions [344].

Such structures provide:

- concentration of highly qualified specialists and the formation of an expert environment;
- exchange of knowledge and experience between ecosystem participants;
- creation of infrastructure for rapid prototyping and testing of products;
- access to investment resources, grants, and international support programs.

36.3 A system of open application interfaces (APIs) and interagency interoperability of digital services [345] is being introduced, which:

- ensures the integration of digital services of public authorities and the private sector;

prevents duplication of functions and data;

contributes to the acceleration of the development of new state and commercial services;

creates a single national ecosystem for information exchange based on standardized protocols.

36.4 All innovations implemented within the framework of this principle undergo an independent ethical examination in order to verify the compliance of projects with these principles [346, 347]:

- transparency and explainability of algorithms;
- non-discrimination and prevention of bias;
- protection of privacy and data security;
- accountability of system developers and operators.

The results of the examination are subject to mandatory publication in the public domain, and the recommendations of experts are subject to mandatory consideration when making decisions on scaling or commercialization of technologies.

CHAPTER 37. THE PRINCIPLE OF THE RIGHT TO NON-ALGORITHMIC EXISTENCE

The right to non-algorithmic existence is inalienable and belongs to every person. It guarantees the ability to refuse interaction with artificial intelligence systems in the absence of its voluntary and informed consent, except for cases expressly established by law. Such exceptions are allowed solely for the purpose of ensuring national security, defence of the state, countering terrorist threats, protecting public health and ensuring public safety, as well as in other cases when automated systems are the only possible means of performing tasks determined by law, subject to the obligatory condition of further human control.

Each person has the right to choose an alternative in the form of direct human contact or obtaining a solution without the use of automated technologies; the right to avoid automated classification, scoring and personalized influence, as well as the right to receive clear and understandable information about the degree and nature of human control in the system.

Entities that implement or operate artificial intelligence systems are obliged to ensure that the user can opt out of algorithmic interaction without causing any harm to him without negative consequences or restriction of access to the main service or its functionality, unless otherwise expressly established by law. Technical and organizational mechanisms for disabling the algorithmic mode should be easy to use, ensure an immediate transition to the mode of operation under human control and record the fact of the user's exercise of this right.

37.1 Every person has an inalienable right to refuse to interact with an artificial intelligence system, regardless of the sphere of social relations, if such interaction occurs in the absence of his/her voluntary and informed consent [348].

Exceptions to this right are allowed only in cases expressly provided for by law, in particular:

to ensure national security, defence of the state and countering terrorist threats;

in the field of epidemiological control, health care and public safety, when the use of AI systems is necessary to prevent the spread of dangerous diseases;

in cases where automated systems are the only possible technical means of performing tasks determined by law, subject to further human control.

Each exception should be clearly defined in the law with a clear establishment of the limits of application, control procedure and effective mechanisms for protecting the rights of the individual.

37.2 Every person has a guaranteed right to:

1. An alternative in the form of human contact or a human decision is the ability to choose the consideration of his/her case, request or service without the involvement of automated systems through direct interaction with an authorized representative (operator, official, consultant).

2. The right to avoid automated classification, scoring, or personalized influence — the right not to be subjected to automatic determination of belonging to certain groups, assessment of risks, creditworthiness, social behaviour or characteristics of a person, as well as the right to refuse to receive individually targeted messages, recommendations or suggestions generated exclusively by algorithms [349].

3. The right to be fully informed about the degree of human control in the system — to receive clear, understandable and timely information about whether the results of the AI system are monitored by a person, to what extent and at what stage of decision-making [350].

These rights are exercised without any discrimination or unreasonable restrictions, and their enforcement is the responsibility of all actors implementing or operating AI systems.

37.3 Operators and other entities that implement or operate AI systems are required to:

provide the user with the opportunity to refuse algorithmic interaction at any stage, unless such interaction is provided for by law as mandatory;

to guarantee that the user's situation does not deteriorate in the event of the exercise of the right to exit, by refusing to provide the main service or non-fulfilment of obligations, unless such actions are due to critical security requirements or are not expressly established by law;

to ensure the technical and organizational possibility of immediately disabling the algorithmic mode and switching to the mode of operation under human control;

record and store information about the user's exercise of the right to exit for the purpose of monitoring, auditing and preventing human rights violations.

CHAPTER 38. THE PRINCIPLE OF TERRITORIAL SOVEREIGNTY IN THE FIELD OF AI

Only those artificial intelligence systems that have an official representation, are included in the National Register of AI Systems and operate in full compliance with the requirements of the legislation in the field of human rights, state sovereignty and national security can operate on the territory of the State.

It is prohibited:

the use of transnational AI systems in critical areas without national control;

transfer of strategically important data outside the State without conducting an independent examination and obtaining special permission from authorized bodies;

import and use of AI models, the training of which was carried out on materials that distort historical facts, humiliate national dignity or pose a threat to national security.

In order to protect the national digital space, a state Digital Border Policy Mechanism is being created, which certifies and audits AI systems, monitors the sources of their training, detects facts of uncontrolled data movement and unauthorized access to state resources, can impose temporary restrictions or blocking of systems that pose a threat, and ensures international cooperation in the field of digital security.

All provisions of this Article are aimed at ensuring territorial integrity, digital sovereignty, national security and protection of the rights and freedoms of citizens of the State.

Any artificial intelligence system operating in the territory of the State is obliged to have a clearly defined legal status, which is fixed through state registration, licensing and permanent state control [351]. Registration in the National Register of AI Systems performs not only the function of identifying and accounting for technological solutions but also ensures the transparency of information about their owners, the responsibility of operators, as well as the ability to trace the supply chain and the origin of software.

Official representation on the territory of the State is a prerequisite that provides a legal mechanism for the protection of users' rights, the application of sanctions and the exercise of judicial control. The existence of such an institution provides the state with the opportunity to bring to justice subjects in cases of harm to individuals or legal entities or society.

Compliance with national legislation implies not only formal consistency, but also compliance with substantive standards: respect for human rights and freedoms, protection of personal data, guarantee of information and national security, as well as compliance with the requirements for the transparency of algorithms and non-discrimination of their decisions.

Thus, the principle of digital jurisdiction is enshrined in law, which extends state sovereignty to all digital technologies and services operating on the territory of the State. This means that any activity of AI systems must be carried out within the national legal space of the State, considering the principles of the constitutional order, the requirements of legislation in the field of human rights, information security and data protection.

Supervision of the activities of such systems is carried out by authorized state bodies that have the authority to conduct audits, issue orders to eliminate violations, apply sanctions and, if necessary, limit or stop the operation of systems that pose a danger.

This principle determines that the digital space, despite its cross-border nature, is subject to the territorial jurisdiction of the state and is an integral part of its national sovereignty.

Prohibitions and restrictions on the use of critical data and models. It is prohibited to use transnational AI systems in critical areas (defence, energy, transport, healthcare, finance) without proper national control and oversight, as this poses an immediate threat of interference in the functioning of strategic infrastructures and can lead to paralysis of public administration systems or undermining the defence capability of the state.

The transfer of strategic data outside the State without conducting an independent examination and obtaining special permission is considered as a direct and potentially critical threat to state sovereignty.

Such a transfer creates a risk of establishing external control over the country's key information resources and can be used to exert economic, political or military pressure.

Strategic data includes, but is not limited to:

- defence orders and military developments;
- geospatial data on critical objects of energy and transport infrastructure;
- detailed financial and banking transactions;
- large arrays of biometric and medical data;
- results of scientific and technical research of dual use;
- data containing information about cyber defence systems and algorithms for managing state networks.

Any transfer or export of such data outside the State must undergo a mandatory multi-level examination, which includes:

- technical verification of information security;
- legal assessment of compliance with international treaties and national legislation;
- risk assessment carried out by specially authorized bodies in the field of national security.

The examination should determine the degree of risk to the state, potential scenarios for the use of this data for hostile purposes and provide mechanisms for compensation for damage and control over their further use. Only after a positive conclusion of all three levels of expertise and the granting of a special permission from the State can a decision be made on the limited and controlled transfer of information outside the State.

The import of artificial intelligence models, the training of which was carried out on materials that distort historical facts, deny or justify the facts of armed aggression against the State, humiliate national dignity or contain elements of enemy propaganda, is considered as the use of information weapons. Such models can spread distorted narratives, undermine trust in state institutions and form dependence on external information sources.

The prohibitions enshrined in this paragraph implement the concept of digital memory security, which provides for the systematic protection of society from the manipulation of history and collective identity through the use of artificial intelligence technologies. In addition, they ensure the information stability of the state from threats, which is aimed at preserving the historical truth, cultural heritage and independence of political processes.

The Digital Border Policy Mechanism functions as a "digital border" and integrates legal, technical and organizational tools, aimed at preventing uncontrolled penetration of AI systems into the territory of the State. It establishes mandatory certification of AI systems before their use, audit of training sources, verification of the transparency of algorithms and their compliance with international standards and national legislation.

Its functions include: continuous monitoring of cross-border data flow; analysis of large volumes of information traffic to identify hidden channels of information leakage; application of procedures for temporary suspension or blocking of dangerous systems; control over the sources of supply of software and hardware components; formation of a register of risk technologies.

The mechanism provides for the creation of special structures — units of the Cyber Border Service, which control access points to global networks, apply data inspection, monitoring and filtering technologies, are functionally combined with the national cyber defence system and interact with international partners in the field of digital security. It is also empowered to promptly respond to identified threats, including the use of measures isolation of individual network segments or forced withdrawal from circulation of dangerous systems.

As a result of the implementation of this mechanism, a system of digital protection of the state is formed, which provides both reactive protection against already identified risks and preventive containment of dangerous technologies. It is a component of a systemic strategy for controlling global information flows, contributes to the formation of trust between the state, citizens and business, and guarantees the preservation

of the security of the digital environment of the State and its subordination to the principles of national sovereignty.

The consolidation of the digital space as an equivalent component of state sovereignty reflects the formation of a new paradigm of state policy, in which the digital sphere is recognized as equivalent to the classical territorial dimensions — land, sea and air territory, as well as national information infrastructure and critical digital services. Any artificial intelligence systems operating on the territory of the State or purposefully providing services to users on its territory (extraterritorial effect) are subject to national legislation and the jurisdiction of national courts and regulatory authorities.

On this basis, the concept of electronic jurisdiction is formed: the state determines the rules for access to the digital market, establishes requirements for registration, certification and audit of AI systems, introduces modes for localization and storage of strategic data, and ensures the enforcement of decisions regarding foreign entities operating on its digital territory. Electronic jurisdiction combines the territorial principle (place of provision of services or location of infrastructure) with the principle of target orientation (user orientation).

The strategic direction of state policy in the field of digital security is based on a triune logic:

- preservation of national identity and historical truth;
- assertion of information sovereignty and increasing resilience to hybrid threats and influences;
- stimulating innovation and development of the national AI ecosystem.

To implement this direction, the state: establishes requirements for transparency and accountability of algorithms; introduces a mandatory Human Rights & Security Impact Assessment (AI/Human Rights & Security Impact Assessment); ensures interoperability and implementation of open standards in areas of public importance; supports the localization of critical computing infrastructure and government data centres.

In order to prevent digital colonialism, mechanisms are being introduced to ensure non-discriminatory market access and limit the abuse of the dominant position of "digital gatekeepers", in particular: requirements for interoperability and data portability; obligations regarding the transparency of tariffs and conditions of access to the API; safeguards against the imposition of unprofitable contractual practices; mechanisms of forced disclosure of technical information for conducting a state audit, subject to the preservation of trade secrets. In the field of public services, the state can set minimum levels of service (public service obligations) and requirements for open interfaces.

The institutional architecture for the implementation of this principle includes: the Authorized Central Authority for AI; National Centre for Digital Inclusion and Trust; the mechanism of digital border policy; industry regulators and the Independent Digital Rights Ombudsman. Their responsibilities include supervision, auditing, crisis response coordination, record-keeping, and public reporting. At the same time, procedural fairness is guaranteed, which includes: the right to be notified, the right to participate in the procedure, the right to receive a reasoned decision, the right to appeal and the right to judicial review.

For practical implementation, milestones and key performance indicators (KPIs) are established, in particular: the share of registered AI systems in the market; percentage of completed audits; average response time to incidents; the level of localization of computing resources; Share of content with confirmed provenance attribution; number of international agreements on mutual recognition of certification. Transitional periods and a "grandfathering" mechanism are envisaged (preservation of rights for existing systems subject to their gradual alignment), as well as the application of proportional sanctions for detected violations.

In the international dimension, it is envisaged to integrate cybersecurity and trusted data into democratic alliances, conclude agreements on mutual assistance in digital investigations, mutual recognition of technical standards and certificates, as well as participate in the development of global norms for responsible AI. All measures are carried out with respect for the principles of necessity and proportionality, respect for freedom of expression and privacy, as well as considering the balance between security and innovative development of the economy.

CHAPTER 39. THE PRINCIPLE OF NON-SIMULACRITY OF A PERSON

The human personality is not reduced to an algorithmic profile, a scoring index, or a behavioural or psychometric model. Any digital representation (in particular, a "digital twin", an avatar or a shadow profile) is an exclusively auxiliary analytical construct and cannot replace the will, autonomy and right to self-determination of a person.

Decisions that have a significant impact on dignity, autonomy, legal personality, access to basic services, or reputation are not made exclusively by automated AI systems. The following are provided: a) meaningful human supervision; b) the right to human, non-algorithmic review; c) personalized and understandable justification of the decision.

"Total scoring" and social ratings are prohibited the formation of integral indices of "value" or "trust" of a person and cross-domain cross-linking of data that have a decisive impact on rights and opportunities. Only narrowly targeted scoring practices within a specific legitimate goal, without transferring the results to other areas, if transparency, independent auditing and human review are ensured.

The creation and use of digital twins and facial simulations is only permitted with the explicit, informed and revocable consent, for clearly limited purposes, with permanent marking and technical safeguards against impersonalization. Special protection regimes are established for minors, vulnerable, deceased persons and public figures. It is prohibited to make political or commercial "endorsements" on their behalf, as well as actions that may be perceived as legally significant statements.

Suppliers and operators act on the principle of dignity-by-design: they integrate human dignity into the design, data processing and operation of systems; provide explainability and contextualization of results; avoid the use of proxy features that may lead to discrimination; guarantee "human exception" as a right to intervene; keep journals of decisions and publish descriptions of the logic of high-risk processes; manage data in accordance with the principles of minimization and security; maintain incident response plans; control the supply chain; organize systematic staff training and report on key performance indicators (KPIs).

Procedural rights are guaranteed: preliminary information about the use of AI systems; a clear explanation of the key factors of the decision; the ability to present context and evidence; human (non-algorithmic) review available within a specified timeframe; the right to a representative; suspension of the execution of the disputed decision in sensitive areas; access to data and relevant logs (audit-trails); the right to compensation for damages; protection from repression; proper fixation and accountability of decisions; escalation to the ombudsman or regulator.

In the areas of justice, medicine, social assistance, education, labour, finance, public services and electoral processes, there are increased standards: mandatory human participation, double checks for critical decisions, prohibition of predictive policing as a self-sufficient basis for interference, ensuring transparency and external audits; Automatic denial of basic services is not allowed.

High-risk applications undergo the Personhood Impact Assessment (PIA) before deployment and periodically after: Target and data analysis is performed; assess risks to dignity, autonomy, non-discrimination and privacy; a test of necessity and proportionality is carried out; errors and uncertainties are checked; human oversight and appeal procedures are being designed; an incident management plan and an immediate stop mechanism (kill switch) are developed; KPI monitoring is implemented and stakeholder participation is ensured.

Exceptions for science, art, satire and education are allowed only if: a) obvious marking of artificial nature; b) unambiguous separation from factual statements; c) the absence of substitution of identity or significant harm to dignity or privacy. Such exceptions do not apply to materials that can be perceived as legally significant statements or means of authentication.

For violations, the following are established: a) administrative sanctions, including a share of global turnover; b) civil remedies (compensation, annulment of decisions, correction of records, class actions); c) criminal liability in case of intentional and systemic violations. The burden of proving compliance rests with the operator or supplier; mechanisms for cross-border enforcement of judgments are provided.

The provisions of this Law are interpreted in favour of human dignity (in dubio pro dignitate; pro persona). Disclaimers of minimum guarantees are null and void. The principle of divisibility of provisions is in force and transition periods are provided for existing systems. In case of doubts about the permissibility of total scoring, simulations or decisions without human supervision, the ban is preferred if the operator does not prove the presence of effective safeguards and the proportionality of the intervention.

General rule (dignity over model). A human person cannot be reduced to an algorithmic profile, scoring index, behavioural or psychometric model - regardless of accuracy, data volume, methodology, context or stated goals[352]. Any digital representation of a person (including a "digital twin", simulation, avatar, shadow profile, personalized bot or virtual assistant) is only an analytical abstraction and does not reflect the completeness of its subjectivity, intentions, life context and emotional-value sphere [353]. Such representation cannot substitute for will, autonomy and the right to self-determination and does not create any presumption of truth, "character" or "identity" with a real person [354]. The use of digital representations is allowed solely as ancillary evidence and cannot be a self-sufficient basis for definitive conclusions about a person's behaviour, motives, or "trustworthiness."

Algorithmic indicators, proxy signs, and digital profiles cannot be recognized as decisive evidence in matters relating to dignity and fundamental human rights. In the event of a conflict between a digital profile and an individual's explanations or evidence, individualized human context-based assessment takes precedence. It is prohibited to give digital simulations the status of "real will" or "character" of a person, to form integral judgments about "value" or "trust" on their basis, as well as to establish any obligations for a person without his conscious, free and specific participation. The use of simulated conclusions outside the declared boundaries of the model, without a transparent description of errors, uncertainties and risks of confusion with reality, is not allowed.

Prohibition of decisions affecting dignity and autonomy without human participation [355]. No decision that significantly affects a person's dignity, autonomy, self-determination, legal personality, freedom of movement, access to rights and services, property and labor rights, medical care, family status, education, migration and political rights, or a person's reputation can be made exclusively by the AI system — even if it is certified or approved by the state.

For the purposes of this paragraph, a decision with a significant impact is a decision that: rise to legal consequences or otherwise significantly changes the position of a person; creates a high risk of harm, stigmatization or long-term disability; causes denial or admission to vital services or access to critical infrastructure; forms "labels" of unreliability or risk, which determine further decisions of other subjects regarding this person.

In such cases, the following must be provided:

− meaningful human oversight, which means that the authorized person: is properly trained and independent; has sufficient time, full access to materials and technical tools to understand the logic of the model, its limits, errors, uncertainties, data used and alternative scenarios; has real competence and authority to change or cancel an automated conclusion; documents the grounds for the decision and takes into account the explanations and evidence of the person; applies safeguards against "automation bias" and "proxy discrimination";

− the right to a non-algorithmic viewing channel, which guarantees a person the opportunity to apply for a "live" consideration with the participation of a responsible official or commission, submit additional data, context and objections, as well as receive a reasoned decision within a reasonable time. Such a channel should be accessible, free of charge, inclusive (considering the needs of persons with disabilities) and understandable in the language of communication of the person;

− the right to a personalized and contextually sensitive justification, which includes a clear explanation of the key factors that influenced the decision; description of the boundaries of the model and its uncertainties; indication of available means of appeal; ensuring contact with the responsible person who carries out the review. Requirements for the organization of human supervision and review:

1. responsible persons (case handlers) are appointed, endowed with competence and authority to make final decisions;

2. event logs, saving prompts, model and parameter versions, as well as fixing the grounds for confirming or rejecting an automated conclusion are provided;

3. qualification requirements and regular training of personnel on ethics, non-discrimination, explainability and work with vulnerable groups are introduced;

4. Supervising quality metrics are established (in particular, cancellation rate, number of inspections, average response time, appeal rate, errors detected) to prevent formal supervision;

5. Rubberstamping, automatic timeouts equated to approval, and concealment of significant features or errors of the model from the person reviewing are prohibited;

6. In sensitive contexts (justice, health, social assistance, education, migration, labour, finance, public services), an increased standard applies: mandatory personal interaction with the person by the examiner upon request, the possibility of an oral hearing and the right to the assistance of a lawyer or representative.

Emergencies and exceptions. Temporary primary reliance on automated conclusion is allowed only in cases of urgent threat to human life or health or to prevent immediate significant harm, if human supervision is objectively impossible at the time of decision-making. In this case,:

a) the decision is immediately subject to mandatory human confirmation in the shortest reasonably possible time;

b) a person is guaranteed the right to priority review and restoration of violated rights;

c) All actions are subject to detailed recording for further audit.

Any internal policies or technical configurations that directly or indirectly neutralize meaningful human oversight, limit the right to non-algorithmic review, or replace personalized justification with uniform or templated responses without regard to context, are recognized as violating this clause and the principle of non-simulacrity of the person.

Prohibition of Reduction "Total Scoring" and Social Rating. It is prohibited to create, maintain, sell, buy or use integral indices or ratings that form a generalized assessment of "value", "trust", "social reliability" or other universal indicator of a person (social rating), as well as decision-making based on them or according to their determining influence in the areas of labour, credit, insurance, education, housing, access to public and medical services, justice, migration, public space and e-governance.

Total scoring includes practices that:

aggregate heterogeneous data from several domains (social networks, consumer habits, geolocations, finance, search history, biometrics, etc.) to derive a single integral assessment;

Cross-platform ID stitching, including data brokers or shadow profiles.

transfer conclusions from one context to another (function creep), creating long-term labels of "risk" or "trust" that influence further decisions of third parties;

use proxy signs that reproduce discrimination on protected grounds or socioeconomic status;

create "blacklists" or "whitelists" of individuals or groups based on behavioural or social characteristics without individual verification.

Prohibited practices include, but are not limited to: integral "trustworthiness indices" or "general trust"; reputational points that determine access to basic services or public spaces; tools that rank citizens by "public value"; algorithmic decisions that automatically reduce or increase the rights or opportunities of a person on the basis of his/her out-of-the-box behaviour.

Permissible narrowly targeted scoring is possible only if the following conditions are simultaneously present:

a. purpose is specific, legal, and clearly limited to the domain (e.g., credit risk assessment — exclusively for a specific credit product);

b. scoring results are not transferred to other goals or areas;

c. The use of protected features and their proxies is excluded;

d. Transparency of the methodology is ensured at a level that does not harm intellectual property and security, but allows for regulatory audit and clear explanation of identity;

e. There is a non-algorithmic viewing channel and a real possibility of human adjustment;

f. a necessity and proportionality assessment was carried out, including a social impact test, and a risk mitigation plan was put in place.

Suppliers and operators are obliged to: document the sources of data and the legal basis for their use; introduce mechanisms for refusing cross-domain crosslinking of identifiers; to prevent the covert acquisition or use of reputation indices from data brokers; ensure that a person has the right to access, correct or delete data, submit objections, appeal and receive an explanation; publish high-level logic descriptions for high-risk applications; keep decision logs and implement bias monitoring.

Responsibility. Systems and practices of total scoring are recognized as violating human dignity and the principle of non-simulacracy of a person. Their application provides for termination orders, fines (including a share of global turnover), restriction of access to public procurement or critical infrastructure, and in case of intentional and systematic violations, suspension of system operation and personal responsibility of officials.

Digital twins, simulations and personal avatars. A "digital twin", "simulation" or "avatar" means a model that imitates the appearance, voice, manner of expression, thinking style or behaviour of a specific individual, including "voice clones", "visual clones", memorial or posthumous bots, personalized chat assistants that create the impression of identity with a real person or claim the ability to predict their "true" will.

The creation, training, distribution and use of such models is allowed only if all the following requirements are simultaneously met:

explicit, informed, specific and revocable consent of the subject (or his/her legal representative) with the fixation of the purpose, boundaries, term and distribution channels; the revocation mechanism should ensure the cessation of generation or publication, deletion or anonymization of data and provide for an obligation to notify all recipients of the revocation;

limitation of goals — exclusively defined areas (education, science, museum or archival practices, adaptive interfaces, therapeutic services with the participation of a specialist); prohibited purposes — political campaigning, commercial endorsements, use of a person's image or name in advertising or marketing, medical, legal or financial decisions with consequences for a person, as well as any acts that may be perceived as legally significant statements;

transparent, permanent and machine-readable marking (visual/audio signals, metadata, non-removable watermarks/provenance), which makes it impossible to mix with real statements or actions and allows automated detection;

technical and organizational safeguards against manipulation, impersonalization and identity spoofing: KYC/KYB for access to high-risk functions, restriction of mass transactions, speed limit control, event and prompt logs, kill switch mechanisms, malicious tip filters, detection and blocking of bypasses;

for minors — double consent (of the child and parents/guardians), the principle of the best interests of the child, the prohibition of monetization and political/advertising use, increased standards of privacy, geofencing and short storage periods; for persons with reduced legal capacity — the participation of a legal representative or a medical specialist, the prohibition of influencing their autonomy;

for deceased persons — use is allowed only with the consent of successors or heirs in non-property rights with a clear marking "posthumous"; Political or commercial endorsements on their behalf and degrading use are prohibited;

for public figures and officials — mandatory warning disclaimers; during the election or referendum period — additional restrictions, in particular, the prohibition to create the impression of official statements or instructions and the prohibition of targeted political manipulation;

The supplier or operator is obliged to publish the model passport (goals, limits, data types, sources/licenses, known errors, moderation protocols), keep usage logs, ensure access, correction/deletion, portability, as well as the functioning of complaint and appeal channels;

data for training and functioning must comply with the principles of minimization and security: non-targeted scraping of biometrics is prohibited; biometric templates are subject to encryption and access control; storage period — minimum necessary;

imitation of passing authentication checks (KBA/2FA/biometrics), access to financial accounts, concluding transactions or forming a "statement of will" on behalf of a person is prohibited;

it is prohibited to create or distribute content that misleads relatives or professional counterparties as "real" communication;

sexualized or degrading simulations without unconditional consent are prohibited;

It is prohibited to use for harassment, harassment, fraud, blackmail or defamatory campaigns.

Surveillance and response.

Incidents of impersonalization or manipulation are to be reported to the regulator within 24 to 72 hours, depending on the severity, and the content is subject to prompt removal.

In the event of a withdrawal of consent or an established violation, the provider is obliged to block the public endpoints of the model, terminate access to third parties, initiate the revocation of copies and notify users.

High-risk simulations are subject to independent audit at least once a year.

These requirements do not restrict the freedom of artistic expression, scientific research or satire, provided that the artificial nature of the material is clearly and noticeably marked, its unambiguous separation from factual statements, as well as there is no risk of substitution of a person or causing significant damage to his dignity, privacy or reputation.

39.1 Responsibilities of suppliers and operators (dignity by design). Suppliers and operators of AI systems are required to organize, design, and operate systems so that human dignity and subjectivity are embedded at all stages of the life cycle. Minimum requirements include:

design restrictions against reducing a person to a single indicator: prohibition of integral "blocking" points; multi-criteria assessments with clear boundaries of application; mandatory "human exceptions" and individualized context in final decision-making; prohibition of automatic transfer of conclusions to other domains;

mechanisms of contextualization and explainability: clear explanation of model boundaries, uncertainties and errors; display of key decision factors; the ability for an individual to add context and evidence; available interfaces to explain solutions; warning about the inadmissibility of out-of-domain use of model results;

prohibition of proxy signs and discriminatory practices: registers of prohibited/proxy poles; regular monitoring of biases before, during and after deployment (fairness metrics, comparison of errors between groups); plans to eliminate biases; prevention of the reverse withdrawal of protected features;

procedures for individualized review and "human exception": SLA (Service Level Agreement) for consideration and appeal; an independent case handler with the authority to change or cancel an automated conclusion; documentation of grounds; online and offline channels are available (taking into account the needs of people with disabilities); prohibition of fines or hidden barriers to appeal;

traceability and decision logs: fixing versions of models, parameters, thresholds and used features; preservation of prompts and context; protection and defined storage periods of logs; a full-fledged audit trail with access to the regulator and internal quality control;

public descriptions of the logic of high-risk decisions: model cards/datasheets indicating the purpose, restrictions, known risks, and typical errors; user references; transparency about the error ranges and conditions in which the model should not be used;

data management: minimization and quality of data; lawful grounds for processing; prohibition of shadow profiles and non-targeted stitching of domains; mechanisms of access, correction, deletion; checking the sources of datasets and licenses; retention policy with short terms and anti-re-identification;

security and incident management: regular red teaming and resilience testing; kill switch/rollback; incident response plan; notification of regulators and stakeholders within 24-72 hours, depending on the severity; patch management; mechanisms for stopping automatic actions in case of a risk to dignity or human rights;

supply chain: due diligence of counterparties; inclusion of "dignity by design" obligations in contracts; the right to audit; dependency registry (SBOM/MBOM); policy of interaction with data brokers; the requirement for suppliers to adhere to the same standards;

competence and culture: mandatory training of personnel on ethics, human rights, non-discrimination, explainability and work with vulnerable groups; training against automation bias; periodic re-certification of specialists who carry out human supervision;

reporting and KPIs: regular internal and public reports (the share of human "overrides", error levels by groups, the share of appeals and their results, response time); independent external audits at least once a year; Continuous improvement plans with clear deadlines and responsible persons.

39.2 Procedural guarantees for a person. Each person has inalienable procedural rights in interaction with algorithmic systems. In the minimum composition are guaranteed:

Right to prior information: before or at the time of application of AI, the person is provided with clear notice about: purpose and legal basis, model class/type, categories and data sources, expected impact, availability of human supervision, contact of the responsible person and available appeal channels.

The right to a meaningful explanation: the person receives an explanation of the key factors and thresholds that influenced the conclusion; information about uncertainties, errors and limitations of the model; indication of the use of potentially sensitive or proxy signs; if possible, counterfactual or applied explanations and options for changing the result.

Right to submit context and evidence: the person may provide additional data, explanations, metadata, objections, request consideration of alternative sources, correct errors, and request a temporary suspension of the decision pending review in sensitive areas.

Right to non-algorithmic review: the decision is reviewed by a duly trained and independent official or commission with the authority to modify or cancel the automated conclusion; Oral hearings (upon request), recording and personalized reasoned decision are provided.

Terms and SLA: confirmation of acceptance of the application — no later than 72 hours; full consideration — within 15 working days (in the areas of basic services — 5 working days), with a possible justified extension with notification of the person.

Right to a representative: a person can engage a lawyer or representative; Vulnerable groups are provided with translation, accessibility and reasonable accommodations, including for persons with disabilities and linguistic minorities.

Right to suspension of execution: In critical contexts (healthcare, social assistance, justice, access to public services, vital services), the execution of the disputed decision is suspended until the review is completed, unless otherwise due to an urgent security need.

Data and log rights: access to personal data and relevant attributes/scores, source categories, key logs that influenced decisions; rights to rectification, deletion of excessive or erroneous data, restriction of processing, refusal of cross-domain crosslinking.

Right to compensation and restoration: In the event of an erroneous or unlawful decision, a person is entitled to compensation for property and non-property damage, restoration of access to services, correction of records and notification of third parties about the correction.

Protection from repression: any pressure, punishment or deterioration of a person's position for exercising the right to appeal is prohibited; The protection also applies to whistleblowers.

Quality standards of the procedure: individualized, reasoned decision with reference to facts and norms; the full audit trail is kept for at least 3 years in compliance with confidentiality; Statistical reporting on appeals and "overrides" is published in aggregate form.

Escalation and supervision: a person has the right to apply to the Digital Rights Ombudsman or regulator; The provider is obliged to provide contacts, complaint templates and free channels (online/offline, "hotline").

39.3 Particularly sensitive contexts (extended). In the areas of justice, prosecution, medicine, psychiatry, social assistance, migration and asylum, education, labour, financial access, public services and electoral processes, it is prohibited to rely solely on algorithmic conclusions on "risk", "reliability", "profile

compliance" or similar integral indicators without professional human consideration and individual assessment of the circumstances.

General requirements of the increased standard.

(a) Prior application authorization or certification;
(b) Mandatory human-in/on-the-loop architecture;
(c) Dual control for critical decisions (second opinion);
(d) the right to an oral hearing and the assistance of a representative;
(e) detailed logging and preservation of materials;
(f) prohibition of the use of proxy signs and transfer of conclusions between domains;
(g) lower thresholds for regulator intervention in incidents;
(h) Regular external audits and reporting.

Justice and Prosecution: It is prohibited to make decisions on detention, bail, parole, determination of a measure of restraint or punishment solely on the basis of risk scoring or predictive systems. The tools of "crime forecasting" and "distribution of patrols" cannot serve as a basis for individual intervention without independent evidence. The defence party has the right to access relevant parameters, data, error metrics and the possibility of expert refutation.

Health and Psychiatry: Clinical decisions cannot be based solely on AI findings; The doctor is obliged to check the relevance of the conclusion, alternatives and individual factors of the patient. Automated denial of vital services (resuscitation, emergency care, palliative support) without immediate human confirmation is prohibited. Patient data is subject to increased protection; Models should have clinical application logs and decision recall mechanisms.

Social assistance, education, work: payments, benefits, enrolment or expulsion from training, disciplinary sanctions, hiring, dismissal or transfer cannot be based solely on algorithmic conclusions. Algorithmic monitoring proctoring is identified as high-risk: transparency is mandatory, the availability of alternative forms of assessment, and the prohibition of automatic conclusions about "fraud" or "cheating". In the field of work, covert supervision or profiling of performance with sanctioning consequences without human review and without the informed consent of the person is prohibited.

Financial access and public services: the refusal to open bank accounts, loans, insurance, housing programs or public services cannot be automatic; Short terms of human review, temporary suspension of the execution of a negative decision and the opportunity to submit alternative evidence by a person are provided.

Electoral processes and public participation: The use of AI for individual voter suppression or manipulative targeting based on inference of political views, ethnicity, or health status is prohibited. Any moderation or recommendation tools during election periods are subject to increased oversight and transparency.

39.4 Personhood Impact Assessment (PIA). For high-risk applications, preliminary (ex ante) and periodic applications are mandatory (ex post) PIA.

PIA is held:

(i) before deploying high-risk systems;
(j) in case of significant updates to the model, data or goals;
(k) when changing the domain or audience (including minors or vulnerable groups);
(l) after a serious incident or a sharp increase in the number of complaints;
(m) in the case of the use of sensitive or biometric data or proxy features.

PIA's content includes:

(a) Description of the purposes and legal basis;
(b) data map (sources, licenses, quality, minimization);
(c) analysis of risks to dignity, autonomy, non-discrimination and privacy;
(d) test of necessity and proportionality, assessment of the availability of less burdensome alternatives;
(e) estimation of errors, uncertainties and calibration;

(f) checking for proxy signs and cross-domain transfers;
(g) design of human oversight and appeals procedures;
(h) Incident Response Plan and Mechanism «kill switch»;
(i) Red Teaming and EVALS Results (Evaluations);
(j) Staff Training Plans;
(k) monitoring metrics and KPIs;
(l) Communication plan with users.

Procedure and accountability. PIA is signed by an authorized AI Officer (AIO) and Data Protection Officer (DPO) or Chief Information Security Officer (CISO), registered in a state or internal registry of high-risk applications, and provided to the regulator upon request. The final summary is published without disclosing trade secrets. The review shall be carried out at least once every 12 months or upon the occurrence of triggers determined by this law.

Procedure and accountability. PIA is signed by an authorized AI Officer (AIO) and Data Protection Officer (DPO) or Chief Information Security Officer (CISO), registered in a state or internal registry of high-risk applications, and provided to the regulator upon request. The final summary is published without disclosing trade secrets. The review shall be carried out at least once every 12 months or upon the occurrence of triggers determined by this law.

1) Participation of stakeholders. Consultations with representatives of target groups, including non-governmental organizations, are mandatory, human rights experts and associations of people with disabilities. A mechanism for submitting comments and providing answers to them is introduced with fixation in the report.

39.5 Exceptions: science, art, satire (expanded). Research, artistic, satirical and educational simulations are allowed under conditions of strict transparency and non-interference with the rights of specific individuals.

Eligibility conditions.

obvious marking of an artificial nature (visual/audio and metadata/watermarks);

(a) Unequivocal separation from factual statements;

(b) lack of intent or consequence of identity substitution, manipulation of will, or causing substantial damage to reputation or privacy;

(c) application of the principles of data minimization and respect for intellectual property rights;

(d) refusal of targeted political influence and commercial endorsements on behalf of the individual;

(e) at the request of the subject — providing additional context or disclaimers about the artistic nature of the work.

Exception limits. Exceptions do not apply to material that could be perceived as legally significant statements, medical, legal or financial advice on behalf of an individual, content to simulate authentication or access to accounts, or systematic campaigns aimed at defamation or harassment.

Additional Warranties. For minors and vulnerable persons, enhanced restrictions apply; for public figures during election periods — extended disclaimers and targeting restrictions. The right to quickly respond to complaints and a mechanism for voluntary mediation settlement are provided.

39.6 Liability and remedies.

Administrative sanctions. Prorated fines (including as a share of global turnover), daily fines for non-compliance with regulations, temporary restriction of functions or geofencing, suspension of operation, revocation of permit/registration, exclusion from public procurement, orders for correction and independent monitoring (monitoring).

Civil law protection. The right to compensation for property and non-property damage, injunction, cancellation or invalidation of an automated decision, the obligation to correct records and notify third parties; reimbursement of legal aid costs. The possibility of class action for mass violations is allowed.

Criminal law aspects. Deliberate systematic impersonalization, fraud, extortion, proven campaigns of harassment or discrediting using simulations are grounds for criminal liability of responsible persons in accordance with the law.

Burden sharing and compliance. It is the operator's/supplier's responsibility to prove compliance with the requirements of this article (availability of PIA, logs, human supervision, labelling, etc.). Conscientious and immediate elimination of the violation and cooperation with the regulator may be taken into account as mitigating circumstances, but do not exempt from liability for the damage caused.

Cross-border enforcement. It provides for tools for notifying foreign suppliers, requirements of a local representative office, geo-blocking of dangerous services, interstate legal assistance and mutual recognition of sanctions within the framework of international agreements.

Interpretation (in dubio pro dignitate) — extended. In case of doubt, the provisions of this article shall be interpreted in favour of the preservation of human dignity, autonomy and the right to self-determination; In conflicts between innovation and dignity, dignity takes precedence (pro persona principle).

Correlation with other norms. If other acts allow for more lenient standards, the stricter standards of this article (the principle of lex specialist and the priority of protection of dignity) apply. The norms of this article do not cancel the guarantees established by the laws on the protection of personal data, non-discrimination, consumer rights, health care, etc., but are in force additionally (cumulatively) and do not reduce the level of protection.

Ineffectiveness of waiver of rights. Any contractual terms or "consents" derogating from the minimum guarantees of this article are null and void. Coercion or misrepresentation regarding the waiver of the rights to human review and explanation is not allowed.

Severability and transitional provisions. The recognition of certain provisions as invalid does not affect the validity of others. Transition periods are established for existing systems with phased proof of compliance; it is prohibited to launch new systems that contradict the basic requirements.

Corrective interpretation effect. Any doubts about the permissibility of "total scoring" practices, simulations or decisions without human supervision are resolved in favour of their prohibition, if the supplier/operator has not proven the presence of effective safeguards and proportionality.

International dimension. Within the framework of international cooperation and conflicts of jurisdictions, an interpretation that better protects dignity and human rights is preferred; while ensuring respect for national sovereignty and the principles of e-jurisdiction.

CHAPTER 40. THE PRINCIPLE OF CONTROLLED SIMULACRUM MONETIZATION

The subject of regulation is the establishment of rules for the creation, distribution and monetization of facial simulations (simulacra), as well as the use of language, behavioural and communicative patterns; determination of the obligations of subjects, supervision procedures and grounds for administrative liability.

It is prohibited: to generate messages on behalf of a person without his direct, informed and revocable consent; sell, transfer or otherwise commercialize personality patterns (linguistic, behavioural, communicative) without separate consent; simulate the participation of citizens in democratic procedures, create artificial support or opposition; use simulacra to authenticate, conclude transactions or access government and financial services; form or apply integral indices of "social value" or "trust" based on simulacra and patterns.

Subjects of political communications are subject to mandatory registration in the Register of Political Authenticity and Transparency maintained by the authorized electoral body. Each instance of political AI content must contain a registry ID, a visible disclaimer, and steel, machine-readable metadata (including digital watermarks). Platforms are required to ensure the preservation of labelling and provide open programmatic access to posting, targeting, and moderation logs.

During election periods, an increased regime is applied, which provides: prompt suspension of the distribution of suspicious or unlabelled content; prohibition of microtargeting on sensitive grounds; mandatory preliminary and final audits; round-the-clock interaction of platforms with electoral authorities; restriction of algorithmic reinforcement of political content; public disclosure of ad data as close to real-time as possible.

Monetization of a simulacrum is allowed only if the person expressly, specifically and irrevocably agrees with granular (detailed) parameters of use. A person is guaranteed: the right to preview and veto; defined terms of response to revocation with suspension of distribution and cascading removal of copies with notification of recipients; transparent financial reporting; prohibition of sublicensing without separate consent. For minors and vulnerable persons, there are enhanced guarantees, and posthumous simulations are allowed only for memorial or scientific purposes with the consent of the successors.

Platforms, marketplaces, and other intermediaries are required to identify customers; limit mass operations, counteract botnets and practices of artificially creating the appearance of mass support (astroturfing); ensure the labelling and provenance of each instance of content; keep event logs; provide reporting and undergo external audits. Data brokers are prohibited from trading personality patterns; Only personal licensing is allowed without cross-domain cross-linking of identifiers.

The person is guaranteed the right to access data and logs, a clear explanation of the grounds for decisions and uses, the ability to withdraw consent with an immediate stop of distribution and a cascading removal of copies with notice to the recipients, the right to fair remuneration and correction of financial settlements, the right to terminate the license in the event of violations or a risk to dignity or safety, and available channels for complaints and appeals.

An administrative offense is: failure to place in the register or submission of false data; absence or forgery of markings and watermarks; simulation of citizen participation in democratic processes; selling or transferring patterns without consent; using a simulacrum for authentication or transactions; microtargeting by sensitive characteristics; failure to comply with the terms of recall; failure to provide or falsify API logs; violation of the increased electoral regime; illegal monetization in areas allowed only for non-commercial use.

For committing administrative offenses, the following are applied: fine; an order to eliminate violations; temporary suspension of distribution or operation; disabling access and application programming interfaces (APIs); geo-blocking of services; exclusion from public procurement; recovery of illegally obtained income and mandatory refutation and removal of content. Sanctions are strengthened for repetition and significant social impact.

The obligation to prove the proper labelling, the validity and legality of the consent, compliance with the electoral regime and the completeness and reliability of the logs rests with the supplier, operator and platform. Logs and metadata are stored within the specified time frames. During the elections, the proceedings on complaints are carried out as a priority, the submission of complaints is free of charge, in sensitive areas the enforcement of disputed decisions is suspended until the review is completed.

The provisions of this section are interpreted in favour of protecting the dignity, privacy and authenticity of political communication. Any waivers of minimum warranties are null and void. The transitional provisions do not apply to new services that conflict with the basic requirements.

Basic prohibitions (inviolability of the will and participation of the citizen).

Any AI system is prohibited:

Generate political, social or commercial messages on behalf of the individual without their direct, informed, specific and revocable involvement; without effective control over the content in real time (prior approval, editing, veto power); without logging and saving consent logs; without constant visible and machine-readable labelling of such content as automated.

Sell, transfer, rent, or make available a person's speech, behavioural, or communicative patterns to third parties as a good or service—including any derived feature sets (acoustic/voice prints, biometric patterns, linguistic and style profiles, reaction patterns, and preferences), their aggregates, inferences, and embeddings (vector representations); carry out their cross-domain crosslinking or de-anonymization with the possibility of reverse identification; commercialize through data brokers or provide via API/SDK — without the separate, explicit, specific, and revocable consent of the subject, without explicit restrictions on the purposes and terms of use, without transparent labelling and logging of accesses, and without guarantees of the impossibility of reconstructing the person.

Simulate citizen participation in debates, public consultations, electronic voting, petitions, or other forms of democracy, including by automated creation or submission of comments, statements, votes, signatures, survey responses, applications to participate, or speak; as well as the mass creation or management of accounts ("sock puppets"); creating, buying, or coordinating "artificial support/counteraction" (astroturfing) through botnets, clones, mass digital twins, synthetic audiences, or artificial promotion ("cheating") of ratings/trends/visibility of content; using deepfake voices/images to participate remotely instead of real persons or to mislead moderators/observers; substitution of quorum or requirements for the minimum number of signatures/comments/votes in participation processes; as well as creating, providing or integrating platforms, tools or APIs/SDKs to perform any of these actions.

Use simulacra for verification/authentication (KBA/2FA/biometrics, including face/voice, fingerprint, iris, gestural and behavioural biometrics) or to bypass liveness checks and anti-counterfeiting systems; for concluding and executing electronic transactions; electronic signatures (including advanced/qualified); issuance/acceptance of powers of attorney; notarial acts; submission of procedural documents, applications, complaints, voting, access to state registers and services, opening or managing bank and payment accounts — on behalf of a person. It is also prohibited to create, provide, or integrate tools (APIs/SDKs/plugins, verification bypass scripts) that enable impersonalization for these actions.

Apply simulacra or patterns to form or use integral assessments of "social value", "trust", "trustworthiness" or generalized indices of "risk" or "reputation", including through cross-domain data cross-linking, the use of proxy signs, inferences and embeddings. It is prohibited to use such assessments as a determining or essential basis for access to work, credit, insurance, education, housing, health and public services, justice, migration procedures or the use of public spaces; create, sell, license or make available via API/SDK similar indices without the separate, explicit and revocable consent of the subject and without clear boundaries of purpose, transparency of methodology and independent audit for non-discrimination.

40.1 Register of Political Authenticity and Transparency (RPAT)

The Register of Political Authenticity and Transparency is being created (RPAT) — a public, mandatory system of pre-registration and post-market surveillance and accounting for all suppliers/operators, advertising networks and platforms that create or distribute AI-content in the political

space. The register stores and publishes (in the form of open data) information about: identification and beneficiaries of the subject; contact of the responsible person; a list of models/versions and marking methods (provenance/watermarks); Creative and campaign libraries, targeting options, placement periods, and budgets/costs. links to the Log APIs. Each instance of political AI content must contain a RPAP identifier in metadata and visible markings. Failure to place in the register, lack of an identifier or submission of inaccurate data entails the immediate suspension of the distribution of such content, administrative sanctions and the obligation to correct records; for foreign entities, a local representative in the State is additionally required.

All AI-content used in political communications must be labelled as automated/inauthentic and accompanied by persistent provenance-attribution. Requirements:

(i) visible disclaimers in the audience's language, placed next to the content/on the screen throughout the show (for video — a permanent badge; for images — a prominent inscription in the frame; for text — a prefix mark; for audio/calls/radio/podcasts — a verbal message at the beginning and a periodic reminder at least every 30 seconds);

(ii) machine-readable metadata according to the C2PA/Content Credentials standard or equivalent: RPAP ID, unique content ID, manufacturer/platform, model and version, date/time (UTC), campaign objectives/content category, basic generation parameters, "election period" label (if applicable), link to registry entry, prompt/source hash digests (no disclosure of sensitive data);

(iii) stable (robust) watermarks/signals (provenance) embedded in media and protected from transcoding, cropping, changing resolution/bitrate/frame rate, replaying/recording (for audio) and taking screenshots; for purely textual materials — cryptographic signature/micro-markup;

(iv) preservation of labelling: platforms do not remove or change metadata/tags when uploading, editing, or redistributing, ensure their migration, and provide APIs for verification;

(v) Circumvention prohibition: attempts to hide, remove, forge or weaken markings/watermarks qualify as a separate offense with increased liability.

The platforms provide standardized, documented API access that real-time and retrospectively provides:

(i) labelling/provenance data (RPAP ID, unique content ID, model/version, UTC event time, source/prompt hash digests, link to registry entry);

(ii) placement logs (time/place/format, account/advertiser with KYB ID, budget/cost, campaign period, creative/version, media hash, URL/placement ID);

(iii) Creative library and targeting options (geography, languages, audience segments without sensitive characteristics, frequency limits, exclusions, optimization types);

delivery and moderation metrics (impressions/reach/clicks/CTR/complaints/removals, response time, reasons/decision codes, appeal history). Access is protected by OAuth2 with granular rights; public keys are provided to verify signatures and schemas; webhooks (automatic notification systems) are provided to inform about withdrawals/appeals. The API has an SLA of at least 99.5% during election periods. Logs and metadata are stored for at least 4 years, with minimization of personal data (aggregated/pseudonymised indicators); access logs are kept. The regulator and electoral authorities receive free priority access; For researchers, open endpoints with controlled speed are envisaged. Formats — JSON/CSV and signed C2PA packages; Documentation and schemes are public.

During election periods, **an increased regime** is applied, including:

(i) accelerated removal of unlabelled or counterfeit AI content — TTD $\leq$ 60 minutes for confirmed violations and $\leq$ 2 hours for controversial cases; mandatory preventive suspension of distribution within 15 minutes after an officially verified signal from the election authorities;

(ii) prohibition of microtargeting based on sensitive characteristics or their inferences (race/ethnicity, political views, religion, health, trade union membership, sexual orientation, biometrics, precise geolocation, children's data); Only macro-targeting with a geography no smaller than the region/city and audiences $\geq$ 100 thousand is allowed.;

(iii) mandatory independent audit before the start of political campaigns and post-audit with a public summary;

(iv) 24/7 "election war room": moderators on duty, quick lines of communication with the CEC/NGO/media, internal playbooks, webhooks (automatic notification systems) for emergency withdrawals;

(v) freezing experiments — prohibiting the launch of new formats, models, or algorithmic changes without the approval of the regulator;

(vi) Recommendation system restrictions: lowering the ranking of unlabelled or suspicious materials, speed limits, prohibiting boost/paid boosts for political AI content;

(vii) advanced disclosure — public library of political announcements in near real time (≤ 1 hour), budgets/targeting parameters;

(viii) enhanced deepfake detection (detector ensembles, manual verification of controversial cases);

(ix) sanctions for marking circumvention and for repeated violations: temporary blocking of advertising accounts/domains, geofencing, fines, disabling APIs.

40.2 Consent, Licensing, and Fair Remuneration.

1. Monetization of a simulacrum is possible only on the basis of the explicit, informed, specific and revocable consent of the person with the definition: goals, duration, territory, distribution channels, technical boundaries and recall mechanisms (including "kill switch").»). Additionally, the following are mandatory:

a) proof of the identity of the subject (KYC/strong identity) and verification of legal capacity;

b) granularity of consent — a list of allowed formats (images/video/audio/text/3D), platforms and audiences, frequency limits, time windows, geographic zones, prohibited domains of use;

c) the right to preview and veto each instance of commercial use or pre-agreed scenarios with the possibility of revocation at any time;

d) SLA withdrawal of consent: suspension of publication/distribution — immediately, but no later than 24 hours (in sensitive contexts — 2 hours), cascading notification of partners and disabling access/APIs, complete deletion of copies — within 7 days;

e) Recording consent and use (audit trail): date/time, contract version, user ID, technical parameters, links to media/creatives;

f) prohibition of "bound consents", dark patterns and any fines/obstacles to refusal;

g) re-authorization for each significant change in goals, technology, audience, or reward conditions;

h) transparent financial reporting and access of the subject to an interactive dashboard with analytics of usage and payments;

i) prohibition of sublicensing or transfer of rights to third parties without separate revocable consent;

j) accessibility — provision in clear language, adapted formats and with reasonable accommodations for persons with disabilities.

2. The agreement (license) fixes the **terms of remuneration** (royalties, lump sum payments, micro-remuneration), a person's access to reporting, the right to audit, the limits of use and the prohibition of transferring rights to third parties without the separate consent of the subject.

3. Minors and vulnerable persons. Double consent is applied — of the person himself (if he/she has reached the age of informed consent) and his/her legal representative — with verification of age and status; The principle of the best interests of the child applies. Political, commercial and advertising monetization of the simulacrum is prohibited; Microtargeting and profiling on behavioural or biometric grounds are prohibited. The default enhanced privacy mode is set (data minimization, disabled trackers/SDKs, cross-domain stitching restrictions). Storage periods are shortened — only to the extent necessary for the stated purpose, but not more than 90 days, without "deferred" copies and reserves. The rights of immediate withdrawal of consent and deletion with cascading deletion of copies are guaranteed; a parental/guardian dashboard of access control is provided; geofencing and blocking of dangerous functions are applied. PIA

and external audits are required at least once a year. It is prohibited to transfer patterns or data to third parties and brokers, except as expressly permitted by law and only in the best interests of the child. For adult vulnerable persons, it is envisaged to involve a legal representative or authorized person (if any), as well as a specialist, if necessary; information is provided in accessible formats and using reasonable devices.

4. Posthumous simulations are allowed only in the presence of a previously expressed will of the deceased person (will or separate consent) or, in its absence, with the written consent of the successors/holders of personal non-property rights. The scope and purpose of application should be clearly limited (museum, archival, scientific, memorial practices) with a stable identifiable labelling of "posthumous AI" and a ban on identity substitution. Assignees have the right to withdraw consent at any time, request termination of publication or distribution, cascade removal of copies, and reporting on usage and revenue. Political or commercial "endorsements" on behalf of the deceased, use in advertising, lobbying, fundraising, commercial campaigns, as well as imitation of legally significant statements are prohibited. The protection of honour, dignity and reputation, respect for cultural and religious burial practices are mandatory; limited data retention periods; prohibition of transferring patterns or data to third parties; Annual independent audit and application of the "kill switch" at the request of successors.

40.3 Political and advertising communications.

1. It is prohibited to create, order or distribute simulacra (audio, video, images, texts, calls, streams) that imitate official statements or messages, political appeals, endorsements, fundraising appeals, participation in debates, interviews or other political communication on behalf of officials or candidates without their confirmed participation, which is evidenced by cryptographic mechanisms (including Content Credentials/C2PA), verified channels or personal fixation, and without prior consent, explicit, specific and revocable consent. This prohibition applies regardless of the presence of the "AI/simulation" label, satirical disclaimers or art form, and applies to deepfake voices and images, synthetic avatars, bot accounts, as well as paid and free advertising formats, including the "silence" period (silent period).

2. Extended disclaimers and transparency are mandatory in political advertising:

(i) political ads created or distributed using AI or automated systems must be accompanied by clearly visible and/or audible "AI/automated political content" labels throughout the screening: for videos, a permanent contrast badge; for images — superimposed inscription; for text, a prefix in the title or at the beginning of the message; for audio/calls, a verbal message at the beginning and a periodic reminder. Such marks must indicate: the advertiser and the source of funding; a unique identifier in the Register of Political Audiovisual Advertising (RPAP); the model and version of the AI system with which the content was created; an active link (or QR code) to an entry in the register; mark "election period" (if applicable);

(ii)It is prohibited to use micro-targeting based on sensitive characteristics and their inferences, including race, ethnic origin, political opinions, religious beliefs, health status, trade union membership, sexual orientation, biometrics, precise geolocation and children's data. Only aggregated targeting with a minimum audience of at least 100,000 people and by territorial unit of the level not lower than the region (region) or city with a population of more than 500,000 people, with openly published frequency limits of impressions subject to verification by the regulatory authority, is allowed;

(iii) A public library of political ads with an open API is created and maintained, which provides access to all creatives and their versions, targeting parameters, budgets and costs, delivery and moderation metrics, RPAP IDs and funding sources in real time, but no later than 1 hour from the date of publication. Data is stored for at least 4 years without the right to early deletion or restriction of access and must be searchable and downloaded in open machine-readable formats (JSON, CSV, XML or other internationally recognized standards).

3. During the elections, there is a "quiet period" — at least 48 hours before the start of voting and until its completion, during which the launch and distribution of new experimental political simulacra (deepfake videos, generative calls, avatars, chatbots) is prohibited. Digital platforms are obliged to ensure priority moderation of political content and a quick response to complaints from users and regulators at this time.

40.4 Exceptions (science, art, satire, education).

1. Only non-commercial simulations are allowed, provided:

(i) permanent visible labelling and machine-readable provenance attribution (C2PA/Content Credentials) specifying the author, platform, and generation method;

(ii) clearly and unambiguously separating simulations from factual statements and prohibiting the substitution of identity or creating imitations of authentic statements capable of forming an idea of support or approval by a particular individual or legal entity;

(iii) the absence of significant damage to dignity, privacy and reputation, with special safeguards for minors and vulnerable persons;

(iv) compliance with intellectual property rights (right to own image, right to own voice, copyright and related rights) and data minimization principles (no non-targeted biometric scraping, short retention periods, pseudonymization/anonymization);

(v) availability of a mechanism for rapid response to the subject's requests (takedown, correction, additional context) and fixation of sources and procedures of processing;

(vi) prohibition of any direct or indirect monetization (advertising, paywall, sponsorship, donations, profiling for targeting) and political targeting/microtargeting.

SECTION V. RIGHTS, FREEDOMS AND DIGITAL DIGNITY OF THE INDIVIDUAL

CHAPTER 41. THE RIGHT TO DIGITAL DIGNITY AND NON-ALGORITHMIC EXISTENCE

Every person has an inalienable, natural, non-quantifiable, encoding or algorithmic right to dignity in the digital environment. This right implies the recognition of the uniqueness of a person and the prohibition of reducing human existence to mathematical or engineering abstractions, typical behavioural patterns or predictable indices. Digital dignity is based on the principle that personality transcends any model, algorithm or simulation and cannot be reduced to a digital control object or a prediction tool.

The right to digital dignity includes the freedom not to be classified, profiled or algorithmically computed without ethical, legal and procedural grounds defined by law, as well as without a real right to object, appeal, cancellation or explanation. No digital system, including the most complex architectures of artificial intelligence, can cancel, limit or replace an individual's free will, his capacity for paradoxes, emotions, errors, inconsistencies and moral choices that form the basis of humanity.

It is prohibited to use automated or semi-automated decision-making systems that determine access to education, medical care, employment, housing, financial services, social protection or the legal status of a person, without the mandatory participation of a person at the final stage of decision-making that gives rise to legal or factual consequences. All algorithmic tools used in these areas should be open to review by authorized bodies and independent auditors, accountable and explainable both in terms of the logic of their functioning and in terms of the actual consequences of the decisions made. It is prohibited to create and use systems that provide for the uncontested rating of persons, algorithmic behavioral assessment, or force participation in digital trust systems without the right of voluntary consent and the possibility of refusal without negative consequences.

Any creation or use of algorithms that form discriminatory digital profiles based on racial, gender, age, social, religious or other characteristics, which may become the basis for discrimination, is recognized as a violation of the dignity of a person and is the basis for legal liability in the manner determined by law. A person is not an object of an algorithmic pipeline — he is a carrier of an undimensional identity.

The State guarantees every citizen the right to have decisions made by automated systems concerning his rights or obligations reviewed by a competent natural person. Everyone has the right to be heard, to provide counterarguments, to receive an explanation of the logic of the decision that is understandable to the average user and to initiate its revision. No decision that has legal consequences for a person can be made without the possibility of appealing it with mandatory consideration by a competent individual.

Everyone has a legally enshrined right to access information about data sources, logic, algorithms used, and the potential consequences of decisions made by automated systems. A person has the right, in accordance with the procedure established by law, to initiate the correction, blocking or deletion of a digital profile or data that forms it, if they are inaccurate, biased or collected without his/her consent. The state guarantees protection against refusal to provide services or exercise rights if such a decision is made exclusively by an algorithmic system without decisive human participation.

The state initiates the creation of the European Platform "For Digital Dignity" as an international intersectoral consortium aimed at developing an ethical standard "human-in-design". This standard should ensure the priority of preserving human meaning, freedom of expression and dignity at all stages of design, development and use of digital systems. The platform's activities are based on the principle of full transparency of the AI architecture, the explainability of algorithmic decisions, inclusivity considering the cultural, age, and social differences of users, as well as the prohibition of design aimed at creating addiction or cognitive subjugation.

The right to non-algorithmic existence is recognized as an integral guarantee of the implementation of human rights and freedoms in the digital age. This right ensures that a person could live, interact, work,

study, consume and create without coercion to digital participation or algorithmic influence, unless otherwise expressly required by law in the interests of safety, health or public order. A person has the right to remain offline, not subject to automated profiling, algorithmic identification or evaluation — without limiting their legal status, access to basic services, or social recognition.

The state undertakes to ensure the existence of alternative offline and non-algorithmic mechanisms for access to basic services and rights, to guarantee the protection of persons who consciously choose digital minimalism, as well as to ensure legal recognition of personal choice regarding the scope and depth of digital presence. The right to dignity in the digital world means the right of a person to preserve their identity, autonomy and human subjectivity despite digital logic.

CHAPTER 42. DIGITAL SOVEREIGNTY OF THE INDIVIDUAL AND CONTROL OVER THEIR OWN INFORMATION CLOUD

Everyone has an inalienable, natural, legally guaranteed right to digital sovereignty — the right to full and exclusive control over all aspects of their digital presence, actions, and footprints in the virtual environment. This right includes the ability to independently determine the procedure for the formation, modification, use, transfer and destruction of all forms of personal, biometric, behavioural, cognitive, medical, genetic, neural data and any other data that directly or indirectly allow to identify a person or create his/her personal, psycho-emotional, cognitive, informational or social profile. A person has the exclusive right to determine the conditions for processing, dissemination, storage, transfer, withdrawal of consent and access to any information concerning him/her, with a guarantee of the exercise of the right to be forgotten (erasure right) and the prohibition of any form of digital stigma, discrimination or biased profiling.

The right to digital sovereignty implies a guarantee that no state, corporation, digital platform or algorithmic system may collect, process or use a person's data as an object of economic, commercial or analytical exploitation without their explicit, specific, free and informed consent, given in writing or other recorded form. All forms of inclusion of a person in digital registers, social graphs, analytical platforms, blockchain networks or other digital identification and accounting systems are carried out solely based on explicit consent, with a guarantee of the right of a person to renounce his digital presence without losing or restricting access to fundamental rights, basic state and social services.

Each person is recognized as the exclusive owner of his/her "information cloud" — a multidimensional array of digital information that includes personal, behavioural, cognitive, biometric, neural data and other data that is formed because of interaction with information and communication systems, sensor devices, digital platforms, networks or biodigital technologies. The information cloud includes data and information structures created, generated or synthesized, from browsing stories to neurobehavioral profiles. Its elements, including emotional feedback, neurotrophies, metadata, algorithmic portraits, biometric casts, personified reactions, and any other digital representations of an individual, may not be alienated, aggregated, used, or transmitted without the explicit, specific, free, and informed consent of that person.

A person's information cloud cannot be the subject of commercial use without the previously provided, explicit, specific and informed consent of the owner. It cannot be used to train artificial intelligence systems without separate and explicit consent. Information cloud data may not be stored longer than necessary to achieve a predetermined legitimate purpose of processing and are subject to seizure or anonymization at the request of the owner, except as expressly provided for by law. Each person has the right to destroy or edit any element of their information cloud, regardless of the form of preservation, technological platform or source of generation.

The state guarantees the creation and functioning of the national infrastructure of digital sovereignty as a system of legal, organizational and technological means, including, in particular: National Register of Individual Information Clouds (on a voluntary basis and in compliance with the principle of informed consent), blockchain protocols for fixing requests and interventions with guaranteed immutability and verifiability of records, multi-level interfaces for managing digital data to ensure access, editing, withdrawal of consent and deletion. All infrastructure must comply with the principles of technological neutrality, interoperability and interoperability of systems, openness and transparency, as well as the priority of protecting the individual as an active digital actor.

Any unauthorized interference with a person's information cloud (the totality of his personal, behavioural, cognitive, medical and other data in the digital environment) without his consent is considered an act of digital aggression — a violation of information self-determination, which entails liability provided for by national law (civil, administrative or criminal, depending on the degree of offense). Such interference is recognized as an infringement of a person's constitutional rights — privacy, mental and cognitive integrity, digital dignity — and may be the subject of proceedings in national courts, international judicial

and quasi-judicial institutions, human rights bodies, as well as in specialized digital tribunals (if established).

Digital sovereignty is an integral part of human legal personality in the information age and guarantees the right to information autonomy, dignity, cognitive and psychological security, as well as freedom from coercion to digital dependence, algorithmic control or manipulative technologies.

42.1 Everyone has an inalienable, natural and legally guaranteed right to digital sovereignty — full and exclusive control over all aspects of their presence, actions and traces in the digital space[356,357]. This right covers both basic identification and contact data, as well as all other types of information that directly or indirectly allow to identify a person, characterize his behaviour, psycho-emotional states, preferences, value and worldview characteristics, neurocognitive and behavioural characteristics and interactions in digital environments, as well as any other data that can form a digital profile of a person [358].

The right to digital sovereignty [359] includes:

the freedom to independently determine the rules for collecting, processing, aggregating, transmitting, sharing, modifying, storing, copying and permanently destroying all forms of personal, biometric, behavioural, social, cognitive, medical, genetic, neural data and any other data that directly or indirectly allow the identification of an individual;

full legal, ethical and technical control over the sources, mechanisms, channels and purposes of access to digital information that directly or indirectly concerns a person;

the ability to withdraw or modify any previously given consent to data processing without adverse legal consequences or restriction of access to basic social services;

the possibility of exercising the right to be forgotten in the digital space through the mandatory destruction or complete de-identification of information that has lost its relevance, does not have a legitimate purpose of processing or creates a risk of digital stigma of the person;

protection against forced inclusion in digital registries, directories, profiles, databases, blockchain networks or social identity systems without expressed, explicit, free and informed consent.

Digital sovereignty ensures that no state, corporation, digital platform or algorithmic system can collect, process or use a person's data, turning them into an object of exclusively automated processing, without their personally, voluntarily and in the proper form of granted, explicit, specific, free and informed consent, recorded in writing or in another form recognized by law. A person in the digital space has a guaranteed right to preserve his dignity, autonomy, privacy, as well as the right to remain outside the digital environment (the right to offline existence) without losing his legal status, access to fundamental rights and basic state or social services.

Each person is the exclusive owner and subject of his/her "information cloud" — a complex and multidimensional set of digital information formed as a result of the use of information and communication systems, digital services, algorithmic platforms, sensor devices, bio digital technologies and other technological environments [360]. The information cloud includes both data directly created by a person and derived or automatically generated data and information structures that arise in the process of analytical, neural network, behavioral-analytical, and predictive processing.

A person's information cloud includes (but is not limited to), in particular:

1) digital interaction history (including searches, views, likes, purchases, directions, subscriptions, reactions)

2) algorithmic profiles formed on the basis of observation, classification or inductive modelling of behaviour;

3) metadata accompanying any activity (time, place, device, user ID);

4) digital biography — a chronological structure of changes in digital roles, participation in network ecosystems, and digital footprints;

5) codes of behaviour — repetitive patterns of reactions, preferences, emotional and sensory rhythms;

6) emotional responses detected using affective recognition technologies, including voice, visual, or neurophysiological markers;

7) neuroprofiles — personalized maps of the nervous system's reactions to external digital stimuli received through brain-computer interfaces, activity trackers, implants, or neuroscanning methods;

8) any other data that can be used to create a cognitive, psychological, cultural, medical or identification portrait of an individual, regardless of the method of collection, form of preservation or technological structure.

The ownership of a person's information cloud is absolute and inviolable and cannot be restricted otherwise than on the basis and in the manner expressly established by law, without personally given, explicit, specific, free and informed consent, expressed in writing or in another form duly recorded in accordance with the law.

42.2 A person's information cloud cannot be:

the subject of commercial use without prior granted, explicit, specific, free and informed consent, duly recorded in accordance with the law;

transferred to third parties — public or private entities — including in automatic or intersystem mode without the specific consent of the owner;

used to train artificial intelligence models without the separately provided, explicit and informed consent of the person;

be stored longer than determined by the person, without ensuring the possibility of its destruction or seizure.

42.3 The state is obliged to provide a comprehensive, multi-level technical and legal infrastructure that allows individuals to freely, effectively and reliably exercise digital sovereignty in accordance with the principles of dignity, autonomy and information security.

Such infrastructure includes, in particular:

the National Register of Individual Information Clouds is a secure state information system that, with the consent of a person, records the owners of information clouds, access models, permission levels, history of changes and requests, providing individual access management, including the possibility of temporarily blocking, delegating or revoking consent.

multi-level digital data management interfaces that guarantee a person the technical ability to independently view, control, edit, upload, export, as well as fully or partially destroy personal data or their individual segments without the involvement of third parties, except as expressly established by law;

blockchain-based protocols for fixing access and changes — technologies for immutably recording any interference with the information cloud, including remote request, viewing, copying, analysis, or aggregation of data. Each transaction must contain the requester's ID, the purpose of access, the technical platform, the time and level of intervention and be available to the owner for review in the format of a legally recognized legal journal (log).

This infrastructure must comply with the principles of openness, interoperability, transparency, technological neutrality and the priority of protecting the rights and legitimate interests of a person as an active subject of digital legal relations.

42.4 Any unauthorized interference, including access attempts, covert analysis, automated copying, reading, transmission, monitoring or manipulation of a person's information cloud without personally given, explicit, specific, free and informed consent, duly recorded in accordance with the law, with the right to withdraw it, is recognized as an act of digital aggression — a type of digital violence that entails legal and social consequences for both the individual and society as a whole.

Unauthorized interference with a person's information cloud:

constitutes a direct violation of the constitutional right of a person to privacy, information autonomy and inviolability of his/her cognitive, psycho-emotional and behavioural profile;

qualifies in accordance with national legislation as a digital offense with potential civil, administrative and criminal liability, including fines, restriction of access to information resources, blocking of offending entities and their assets;

creates a basis for recourse to national and international judicial and quasi-judicial institutions, in particular specialized human rights bodies, digital tribunals or arbitration institutions in the field of information rights;

may be qualified as an act of information encroachment in accordance with the norms of international law, in particular as a violation of the articles of the International Covenant on Civil and Political Rights, the European Convention on Human Rights, the UNESCO Principles on Digital Ethics and other international legal obligations of the state.

Interference without a person's consent in the information cloud is not only a technical action, but a violation of the moral and legal contour of the individual, which undermines trust in the digital order, creates risks of reputational damage, psychological pressure and loss of digital identity in the age of algorithms.

CHAPTER 43. THE RIGHT TO CYBER NEUTRALITY AND NON-ALIENATION OF DIGITAL SUBJECTIVITY

Everyone has the right to cyber neutrality — the freedom to remain autonomous, non-integrated, untied to a specific digital ecosystem, technology platform, operating environment, interface, or protocol. This right guarantees the prohibition of coercion to integrate into a certain digital environment as a condition for access to services, social inclusion or the exercise of basic rights.

Every person has an inalienable right to the non-alienation of their own digital subjectivity — that is, a guarantee that their digital identity, visual, behavioural, cognitive, intellectual and AI projections cannot be alienated, reproduced or externally exploited without personally given, explicit, specific, free and informed consent. This includes prohibiting the automatic transfer, delegation, consolidation, or sale of digital identities to third parties; prohibition of automated creation, reproduction or replication of digital avatars, models or reconstructions of a person without his/her consent; the right to control digital manifestations in augmented, virtual, mixed or immersive reality; prohibiting the use of cognitive, emotional, or mental portraits for predictive modelling, manipulation, or influence without the person's prior given, explicit, and informed consent.

Any interference with the integrity of digital subjectivity without personally given, explicit, specific, free and informed consent is recognized as an act of digital alienation — a hybrid form of encroachment that violates the integrity of the individual, distorts his identity or replaces his presence with an artificial construction. Such a violation entails legal liability in the form of civil, administrative or criminal, depending on the level of damage, intent and technological complexity of the invasion. Everyone has the right to protection from digital exclusion, including by applying to the Digital Ombudsman, specialized tribunals or international human rights protection mechanisms.

In a post-algorithmic (post-AI) society, where digital technologies penetrate deeply into all spheres of life, the state recognizes the fundamental human right to remain non-imitative, unidentified, non-profiled and non-avatarized. This means the right not to be subjected to algorithmic categorization or typology; Not be included in mass personalization or targeting systems. do not create or recognize digital substitute avatars; not be included in predictive modelling systems without consent. The state undertakes to provide legislative, technical and educational guarantees for the implementation of this right.

In order to counteract digital coercion, the state creates a multi-level technical, legal and organizational infrastructure to support the freedom of choice of the individual: provides guaranteed offline access to basic public services, enshrines the principle of technological neutrality, prohibits practices, requiring mandatory full digital identification as a condition for access to rights or services, develops interfaces, that exclude the coercion of digital participation. Control over compliance with these mechanisms is carried out by independent bodies with the mandatory participation of civil society. It ensures the freedom of a person to participate or not to participate in the digital environment without losing dignity, social status or access to fundamental rights.

Everyone has the right to cyber neutrality — the freedom to remain autonomous, non-integrated, and untied to a specific digital ecosystem, technology platform, operating environment, interface, or protocol [361,362]. This right guarantees the prohibition of coercion to integrate into a certain digital environment as a condition for access to services, social inclusion or the exercise of basic rights.

Individuals have an inalienable right to the inalienability (non-alienation) of their own digital subjectivity — that is, a guarantee that their digital identity, including visual, behavioural, cognitive and artificial-intellectual projections, may not be subject to alienation, copying, simulated reproduction or any external use without a clearly expressed, conscious, informed and voluntary consent of the person himself or herself [363]. This right includes:

prohibition of any form of alienation of digital identity, including its automatic transfer, delegation, consolidation or sale to third parties, including government agencies, corporations, platforms or other digital entities — without a special, legally recorded expression of the will of the person, except as expressly provided for by law;

prohibition of automated or semi-automated creation, generation or replication of digital avatars, behavioural models, speech, voice, mimic, gesture, emotional or visual reconstructions of a person using AI or other simulation technologies — without the prior informed, explicit consent of the person and the definition of clear boundaries and conditions of use;

the right to full and exclusive control over digital visual, behavioural and avatar representations of one's own person in augmented, virtual, mixed or immersive reality, including the right to demand at any time the revocation of such images, their blocking, editing, deactivation, termination of broadcast or complete technical destruction with the cascading deletion of all copies;

prohibiting the use of a cognitive portrait, psychometric models, digital twin, reconstructed mental model, or emotional simulacrum of an individual to participate in digital prediction, modelling, social impact, experimentation, or personalized manipulation systems against other individuals or the individual — without their explicit consent, provided separately for each specific purpose and technological environment.

Any simulation of an identity without its express and informed consent is equated to unauthorized digital replication — reproduction of identity, which creates the risk of digital substitution of a person, unauthorized intrusion into his/her identity and levelling of individual uniqueness, which poses a fundamental threat to human dignity in the digital age.

43.1 Any interference with the integrity of a person's digital subjectivity — including creating, editing, duplicating, modifying, distorting, merging, simulating, or replacing its digital manifestations (visual, voice, behavioural, emotional, cognitive, neurobehavioral, or otherwise) — without a clearly expressed, conscious, specified and duly recorded consent of a person is recognized as an act of digital alienation [364].

Such digital alienation is considered as a form of hybrid information encroachment with a high level of public danger, which has the potential to violate the integrity of the individual, distort his digital identity, depersonalize or replace his presence in the digital space with an artificial construction. This violation qualifies as a dangerous interference with fundamental human rights, in particular the right to privacy, autonomy, honour, reputation, informational self-determination and dignity.

Digital alienation entails legal liability, in particular:

− civil liability in the form of compensation for property and non-property damage, compensation for moral damage and restoration of business reputation;

− administrative sanctions — fines, temporary or permanent blocking of access to digital resources, prohibition on further use of illegally obtained or alienated digital data;

− criminal liability — in cases of systematic or intentional creation of digital simulacra (falsified avatars, cognitive models or other recreations of identity) that have signs of fraud, manipulation or digital security violations.

Everyone has the right to judicial and extrajudicial protection against digital exclusion at the national and international levels, including through independent digital ombudsman mechanisms, specialised digital tribunals or international digital identity protection bodies.

43.2 The state recognizes that in a post-algorithmic society, where digital technologies penetrate all spheres of private and public existence, the right to remain a non-imitative, non-profiled and non-avatarized person is a fundamental element of human dignity, autonomy and subjectivity [365].

This right includes human freedom this right includes human freedom:

have the right not to be subjected to algorithmic categorization or imposed automated typology;

have the right not to be included in the systems of mass classification, content personalization, social targeting or information optimization, which limit the multidimensionality of its manifestations;

have the right not to create, maintain or recognize digital avatars that are formed based on algorithmic constructs and function as identity substitutes (digital constructs that mimic or substitute for a person's individual identity) in virtual, augmented or mixed environments;

have the right to refuse to participate in predictive modelling systems that determine or significantly affect the rights, freedoms or life opportunities of a person without his/her explicit and recorded consent.

The state undertakes:

to create legislative guarantees for the implementation of this right in all spheres of public life, in particular in education, medicine, employment and culture;

develop and implement technical and organizational standards that prevent forced avatarization, profiling, or unauthorized digital replication of an identity without explicit and recorded consent;

support educational, ethical and cultural initiatives aimed at preserving non-imitative humanity (authenticity of the human person) as a basis for digital collaboration, creativity and empathy in the future technological world.

43.3 The state is creating a multi-level system of mechanisms for preventing and countering digital coercion, aimed at ensuring the freedom of choice of a person in the field of his/her participation in digital ecosystems, avoiding imposed digital totality and guaranteeing the preservation of physical, offline and alternative forms of access to basic rights.

Such mechanisms include:

guarantees of real offline accessibility to basic public, administrative, social and vital services (birth registration, medical care, education, justice, pensions, voting), regardless of a person's participation in digital systems;

regulatory consolidation of the principle of technological neutrality — the obligation of the state and the requirement for private providers to provide equal access to services through different technological platforms, without being forcibly tied to a single environment or digital tool;

prohibition of practices that require a person to transition to a "total digital identity" as the only prerequisite for participation in public life (for example, the prohibition of denial of service due to the absence of a digital signature, identifier or digital passport);

development and implementation of "anti-coercive interfaces" — technological protocols, algorithms or functional mechanisms designed to: identify signs of digital coercion (lack of alternatives, restriction of offline choice, discriminatory access conditions), inform a person about his/her rights and provide him with access to alternative scenarios of actions.

The state guarantees independent monitoring of such mechanisms with the involvement of representatives of civil society, digital ombudsmen and technical auditors, as well as the development and application of legal instruments of protection against digital coercion as a form of violation of human dignity, personal autonomy and the right to informational self-determination.

CHAPTER 44. DIGITAL EMPATHY AND GUARANTEES OF INDEPENDENT COGNITIVE SELF-REGULATION

Everyone has an inalienable right to digital empathy — that is, to an ethical, inclusive and non-discriminatory attitude towards oneself on the part of digital systems, platforms, devices, agents and algorithms involved in communication, training, modelling, recommendations or other forms of interaction. The principle of digital empathy imposes on operators and developers of digital technologies the obligation to ensure their functioning in a way that: does not disturb the psychological balance of the user; respects its cognitive and physiological limits; avoids covert influence, pressure and manipulation; supports intellectual freedom, self-reflection and comfort of the face.

This principle is mandatory in the fields of education, healthcare, law, social support and cultural and information exchange. Its violation is considered as a failure to comply with ethical and legal standards for the design and use of digital technologies.

Everyone has a fundamental right to independent cognitive self-regulation — the ability to independently control their thought processes, maintain worldview autonomy, protect psycho-emotional integrity and maintain mental balance in the digital environment. This right provides protection against information overload, algorithmic manipulations, obsessive recommendations, the introduction of destructive thinking patterns and unauthorized personalization.

A person has the right to form, store or change their beliefs, thoughts and emotional assessments without undue influence, distortion or forced modelling by digital platforms, as well as the right to a cognitive pause — temporary liberation from the information space, digital pressure and external stimuli, which is realized through modes of mental rest, information silence or psychodigital detoxification. This right imposes on digital technology developers the obligation to create tools for self-control of loads, adaptive interfaces and means of supporting individual rhythms of mental interaction.

All digital products that interact with users in the fields of education, medicine, information or management must comply with the principles of empathic design. This includes considering the user's psycho-emotional state (fatigue, stress, mood), ensuring the right to informational silence, the absence of coercion to constant reaction, as well as the possibility of individual settings of the pace, style and intensity of digital interaction.

The use of artificial intelligence systems or other technologies capable of analyzing, simulating or predicting the emotional state, cognitive patterns or patterns of thinking or psycho-emotional reactions of a person is allowed only if a special written consent of the user is obtained, a predetermined and agreed scenario of interaction is obtained, it is possible to immediately terminate such interaction by the user without the need to provide explanations, as well as the introduction of technical and legal restrictions, that make it impossible to exploit an individual's vulnerability for manipulative purposes in marketing, political, therapeutic or any other context. Such systems are subject to an independent ethical audit and must be clearly marked in the interface as carrying out analysis or modeling of the emotional state.

The state undertakes to ensure the integration of a multi-level program of digital ethics into the national education system to form a culture of respect for the emotional, cognitive and identification integrity of the individual. Such a program includes courses for students, teachers and parents on emotional literacy, critical analysis of algorithms, neuropsychological safety, the right to offline access and respite from the digital environment, the development of empathic teaching, as well as support for experimental educational programs aimed at educating an independent, humane and conscious generation in the digital age.

44.1 Everyone has an inalienable right to digital empathy — that is, guaranteed ethical, humane and non-discriminatory treatment of them by digital systems, platforms, devices, virtual agents, algorithms and other digital technologies involved in the processes of communication, information exchange, training, recommendations, analysis or modeling.

The principle of digital empathy means that digital technologies must be designed and configured in such a way that:

do not disturb the psychological balance of the user;

respect the cognitive boundaries of the user, his rhythm of thinking and perception;

refrain from any form of covert influence, pressure, manipulation or artificial induction of emotional vulnerability;

stimulate the development of intellectual freedom and the ability to self-reflection;

ensure inclusivity for users with different emotional, age, cognitive, or psychophysiological characteristics.

The application of the principle of digital empathy is mandatory for systems used in the educational, medical, legal, social and informational and cultural spheres. Violation of this principle is considered as a violation of the ethics of the design and use of digital technologies.

44.2 Every person has an inalienable fundamental right to independent cognitive self-regulation — the ability to maintain control over their own thought processes, maintain worldview autonomy and psychological comfort in the digital environment. This right encompasses the protection of an individual from hidden external information influence, digital overload, unauthorized personalization and obsessive thought patterns formed by algorithmically controlled systems.

The right to cognitive self-regulation includes, but is not limited to::

- the freedom to shape, change or store one's own thoughts, beliefs, assessments and views without coercion, distortion or covert modelling by digital technologies, including personalization algorithms, rating systems, recommendation platforms or neurosemantic (artificially created) environments;

protection against information overflow, digital overwhelm, neural noise (noisy attention), as well as the risks of forming the effect of fragmented consciousness through excessive multitasking and intrusive content flows;

The right to a cognitive pause is a temporary disconnection from the information space, the cessation of the action of digital stimuli and a return to a state of mental silence. Such a pause is implemented through special modes of mental rest (mind-rest), informational silence (info-silence), psychodigital detoxification, as well as by creating a space without screens, notifications, tracking and digital identification.

In order to ensure the right to cognitive self-regulation, digital technology providers are obliged to provide users with tools for self-control of mental load, means of self-monitoring and displaying its level, as well as the possibility of flexible personalization of interfaces and modes of interaction, considering the individual cognitive and psychophysiological characteristics of the person.

44.3 Digital products that interact with humans, including in the fields of education, healthcare, social support, information environment, and public administration — should provide empathic design mechanisms, in particular: ensuring that there is no coercion to have a permanent presence in the digital environment, as well as to react or respond continuously; taking into account states of fatigue, mood swings and stressful loads; the ability to personally customize the pace, tone, style and intensity of digital interaction.

44.4 The use of artificial intelligence or other digital systems capable of identifying, analysing, tracking or predicting the emotional state, mood, cognitive patterns and patterns (patterns), non-verbal reactions or psycho-emotional markers of a person is allowed only if the following requirements are met:

−on the basis of a special written consent of the person, containing a specific description of the purposes, duration, technologies, subjects of access to data and a guarantee of the possibility of revoking this consent at any time without the need to explain the reasons;

−within a predetermined and agreed interaction scenario that outlines acceptable analysis algorithms, data sources, interpretation boundaries and system response interface; any deviation from such a scenario is considered a violation of the principle of predictability of interaction;

− with the obligatory provision of an unhindered right of a person to terminate interaction at any time, including automatically, through the use of empathic brake lights (digital or physical) that do not require additional explanations or confirmation of reasons;

− in the presence of technical and legal barriers that make it impossible to exploit emotional vulnerability, psycho-emotional exhaustion or manipulative influence by systems — in particular in marketing, commercial, political, ideological, educational or therapeutic or any other contexts where it is possible to exploit a person's vulnerability for manipulative purposes.

Such systems are subject to independent ethical audits, and their use must be transparent, controlled and clearly labelled for the user.

44.5 The state ensures the introduction of a multi-level program of digital ethics into the national education system, aimed at creating a culture of respect for the emotional, cognitive and identification integrity of a person.

Such a program includes:

−training courses and educational modules for schoolchildren, students, teachers and parents dedicated to the topics of emotional literacy in the digital environment, the risks of algorithmic manipulation, respect for mental boundaries and the right to be offline;

− creation of educational scenarios that take into account the individual neuropsychological characteristics of students and their rhythm of learning, slow or asynchronous, and provide the possibility of offline participation (non-digital form) in the educational process;

− development of ethical competence of digital educators — the ability to consciously and responsibly integrate digital tools into teaching, considering the principles of respect for the dignity, safety and freedom of students;

− inclusion of the principles of intellectual self-preservation, psycho-emotional safety, virtual ethics and digital empathy in educational standards;

− support for experimental schools, university courses and interdisciplinary projects that develop and test new models of empathetic digital education.

The digital ethics program is aimed at forming a generation capable of autonomous thinking, critical evaluation of algorithms and preservation of humanity in the technological environment.

CHAPTER 45. THE RIGHT TO DIGITAL PHYSICALITY AND PSYCHOPHYSICAL INTEGRITY IN VIRTUAL ENVIRONMENTS

Every person has an inalienable right to digital physicality — legally and ethically recognized right to integrity, integrity, autonomy and safe existence of its digital body, avatar or other personified embodiment in a virtual, augmented, blended, or immersive environment. Digital corporeality encompasses not only the appearance of the avatar, but also all the sensory, emotional, physiological, and psychological aspects of interaction that occur within the digital experience.

This right includes the legal recognition of the digital avatar as an individualized projection of the individual, requiring the same level of respect and protection as the physical body in physical space. A person has the right to protect his/her virtual space, determine the boundaries of the personal area, protect against unwanted sensory influences, simulated touches, aggressive visualization or other forms of digital interference. It is prohibited to use tools that distort, depersonalize or alter the digital corporeality without the voluntary and explicit consent.

Any violation of digital physicality is recognized as a digital encroachment and entails legal liability established by law, in accordance with the principle of the inviolability of the person, which applies to virtual and other digital environments.

Any form of digital violence, interference or psychological pressure that violates personal boundaries, causes fear, humiliation, psycho-emotional disorientation, anxiety or causes emotional or cognitive damage is prohibited. Such actions include simulating attacks on a virtual body, invading a personal area without consent, virtual harassment, demonstrating hostile intent, creating sensory overload, or intentionally using harmful or aggressive content.

A person has the right to control his/her digital physicality: determine access limits, configure sensory interaction parameters, activate protection modes ("digital privacy"), use technical means to block unwanted interaction, and apply interface signals or other mechanisms for immediate termination of contact in real time. Such tools should be universally accessible, context-responsive, and independent of user status or technical limitations of the platform.

Each virtual environment in which social, professional, educational or cultural interaction takes place must guarantee the principles of psychophysiological safety, emotional sensitivity and adaptability, predictability of interaction between users and AI agents, individual configuration of sensory influence, and the prevention of accidental, aggressive or discriminatory scenarios.

Platforms operating in immersive environments are required to implement comprehensive protocols for protecting digital physicality. This includes: providing anonymity and pseudonym regimes; warning of potential sensory or cognitive load; provision of emergency exit mechanisms from the simulation; creation of digital violence response services; support for transparent and effective channels for appeals, complaints and protection of rights in the digital space.

The right to digital physicality is recognized as fundamental in the conditions of virtual coexistence and is one of the main guarantees of psychological well-being, freedom of expression and digital dignity of a person in immersive environments with an increased level of presence.

45.1 Every person has an inalienable right to digital physicality — legally and ethically recognized right to the integrity, integrity, control over one's own digital body and the safe existence of her avatar or any other personified embodiment in the virtual, augmented, blended, or immersive environment. Digital physicality encompasses not only the image or shape of the avatar, but also all sensory, emotional and psychophysiological aspects of user interaction in the digital environment.

This right includes:

recognition of the legal status of the avatar as a digital projection of the personality, which requires the same protection as physical corporeality in real space;

protection of virtual space defined by a person as his/her own, in particular "personal zones" in virtual rooms, environments and immersive scenes;

prohibition of imitation touches and aggressive visual, audio, tactile and other sensory influences that cause discomfort, fear, shame, anxiety or a harmful physiological reaction;

the right to form the boundaries of the avatar, determine its reactivity, visibility, physical characteristics, clothing, movements and feedback;

prohibiting the forced use of tools that distort or depersonalize a person's digital physicality (including automatic patterns that alter gender, emotions, appearance, or voice without consent);

the right to opt out of virtual bodily interaction, including the ability to be present without an impersonation, avatar, or simulation shell.

Any violation of digital physicality is recognized as a digital encroachment and entails legal liability established by law, in accordance with the principles of the inviolability of the individual in the virtual space.

45.2 Any form of digital violence, interference or manipulative psychological influence in a virtual, augmented, mixed or immersive environment that violates the boundaries of the individual, causes fear, disorientation, anxiety or humiliation is prohibited.

Digital violence includes:

simulation of actions that simulate encroachment on the body (touch, blows, grabs, shock effects);

intrusion into personal areas without the user's consent — approaching the avatar, obsessive stalking, attempts at forced contact, virtual threats or demonstration of hostile intentions;

visual, audio, tactile or other sensory influences that intentionally cause panic, phobias, cognitive blocks, decision paralysis, as well as simulations of excessive noise, lighting or spatial disorientation;

posting content in the environment that causes stress, disgust, humiliation or stigmatization, especially due to the racial, gender, sexual, age, physiological or cultural characteristics of the person.

Such actions are recognized as digital encroachment, which is prohibited regardless of the format of the gaming, social or commercial virtual environment and entails liability provided for by law.

45.3 All users have the right to active control (control) of their own digital corporeality, which includes:

use of technical and software means of personal protection of their avatar or digital image from any unauthorized or undesirable actions, including visual, auditory, tactile and other sensory influences;

determining the individual boundaries of the virtual body, including adjusting the distance from which other subjects (users or AI agents) are allowed to approach, interaction parameters, acceptable touch or simulation interaction formats, as well as the types of avatar reactions to external stimuli;

establishment and activation of the "digital privacy" mode — a technological state that automatically blocks any contact attempts, unauthorized influences, virtual gestures, touches, as well as attempts at visual or other sensory penetration into personal space that exceed the established threshold of permitted interaction;

access to real-time rapid warning, alarm and immediate termination systems, including the "Emergency Digital Exit" function, as well as speech, tactile or non-verbal signals configured by the interface according to personal needs and context of the situation.

These rights constitute the basic level of digital security of the individual in the virtual environment and must be implemented regardless of the type of platform, the format of the environment, or the user's status.

45.4 Virtual environments in which social, educational, cultural, therapeutic or professional interactions take place are obliged to adhere to a set of principles that guarantee human, safe and individualized interaction:

- psychophysiological safety — prevention of elements that can cause stress, sensory overload, spatial orientation disorders or anxiety; ensuring physiological comfort and predictability of sensory reactions;
- emotional adaptability — the ability of the environment to adequately and timely respond to the emotional state of a person, providing soft transitions, emotionally sensitive behaviour of virtual agents, as well as the possibility of a temporary pause in interaction (detachment) without loss of progress or participation in interaction;

- individual configuration of the contact level — the ability of each user to independently adjust the intensity, nature and volume of interaction with other subjects and systems, including visual, sound, tactile and cognitive parameters;
- predictability of actions of other users and AI agents — clear labelling of behavioural scenarios, the absence of accidental or provocative actions, the transparency of agent management algorithms, and the predictability of their reactions in accordance with the standards of digital ethics and dignity.

45.5 Each platform that provides services in immersive environments (virtual reality, augmented reality, mixed environment, Metaverse) is required to implement and maintain a comprehensive set of security protocols that provide:

privacy modes that allow you to act incognito or anonymously without forcing the disclosure of personal, biometric or behavioural identifiers, including by controlling visual display, voice, avatar and spatial localization;

mandatory warnings about potential sensory, cognitive, or psycho-emotional stresses that may occur when using content or scenarios (e.g., virtual scenes with bright flashes, rapid movements, emotionally charged dialogue, or socially sensitive elements);

Mechanisms for instant exit from the simulation are available and effective ("emergency shutdown", "panic button", "emergency exit"), which provide the user with the opportunity to leave the virtual environment without delays and technical obstacles in case of distress, disorientation or the threat of digital overload;

establishing and maintaining an independent response service for cases of digital violence, online harassment, toxic behaviour or cognitive overload, including moderators, digital ombudsmen, mental health professionals and technology auditors, and functioning on the basis of transparency, confidentiality, responsiveness and ensuring the individual's right to protection.

CHAPTER 46. PROTECTION OF THE RIGHT TO DIGITAL SECRECY AND UNOBSERVABILITY

Every person has an inalienable right to digital secrecy, which means the ability to exist, move, interact, view and communicate in the digital environment without being forced to, automated or hidden tracking, identification, monitoring, profiling, or collecting personal, behavioral, or analytical data. This right includes freedom from any digital identification or traceability that has not been initiated by the person himself or herself or is not directly justified by legal regulations.

Digital secrecy implies full and exclusive control over one's own digital presence, including the right to choose the conditions, method, time and degree of detection in the digital space; the right to withdraw or prevent entry into databases, search indices or social graphs; the right to prohibit unauthorized aggregation or re-analysis of their digital actions. Digital surveillance technologies are allowed only if they are legal, proportionate, clearly defined and openly regulated.

Digital untraceability guarantees a person invisibility or a controlled minimum level of detection in the digital space, in particular, protection from automated analysis, data accumulation or tracking *without their direct, specified and informed consent. A person has the right to operate in zero visibility mode, in which his/her activity is not recorded in logs, is not subject to indexing or analytical processing, does not leave a digital trace and does not form a profile without of the user's direct will.*

Any actions involving the use of covert data collection technologies, including facial recognition, biometric identification, microexpression analysis, spatial tracking, or neurophysiological reading, are allowed only with the specific, voluntary and informed consent of the person, clear marking of data collection areas, the possibility of blocking them, and conducting an independent audit of such systems.

Direct or indirect coercion to digital identification, mandatory registration or verification in the digital environment is prohibited if such a procedure is not objectively necessary to ensure public order, law and order, national security or vital interests of society. This applies to verification requirements for access to basic services, public information, open platforms or platforms whose activities are not related to high-risk functions.

Digital authentication should be based on the principles of voluntariness, proportionality, the availability of alternative methods and the right to anonymous or minimally identifiable access, except as expressly defined by security or justice legislation.

The state is obliged to ensure the development and functioning of a full-fledged infrastructure of anonymity as a component of the digital sovereignty of the individual. This includes creating virtual spaces without coercive identification mechanisms; providing modes of anonymous access to basic government services; guaranteeing legal protection of persons exercising their right to digital secrecy; and the development and implementation of design standards that make it impossible to collect data covertly.

The state pays special attention to educational and outreach programs that raise awareness of digital secrecy, human rights to privacy, identification risks, and self-defense tools in the digital world. Such measures should be integrated into national digital policy, human rights practice and information ethics.

46.1 Everyone has an inalienable right to digital secrecy, i.e. the ability to exist, move, communicate and interact in the digital environment without coercive, automated or covert tracking, identification, monitoring, profiling or collection of personal, behavioural or analytical data. This right includes freedom from any forms of forced digital detection that have not been directly initiated by the person or are not objectively necessary to carry out legally defined procedures.

Digital stealth also encompasses:

control over one's own digital presence, which means the ability to determine when, under what conditions and in what form an individual appears in the virtual environment;

the right to extract data or prevent their inclusion in databases, search indexes, social graphs and other forms of digital aggregation without explicit and specified consent;

guarantee of the use of digital surveillance technologies only if they are legal, proportionate, reasonable and clearly regulated.

Any interference with digital secrecy is considered as a significant restriction of a person's information autonomy and is subject to legal assessment for violation of his/her rights.

46.2 Digital unobservability encompasses a system of legal, technical and behavioural guarantees that ensure invisibility or a controlled minimum level of detection of a person in the digital space. This means the individual's right not to have unwanted digital analysis, tracking, or accumulation of data, the collection of which has not been expressly authorized by the person himself or herself.

This right includes:

freedom not to be subject to algorithmic analysis, automated categorization, recognition, contextual reading or digital profile construction without the direct, specified and informed consent of the user;

the ability to function in the "zero-visibility" mode, in which all a person's activities are protected from analytics, tracking, fixing a digital footprint or saving in system logs;

the right not to leave a digital footprint when performing private, household or non-formal actions (viewing materials, personal navigation, autonomous presence in the environment), if they are not related to a public, official or legally significant function;

protection against any form of passive or covert data collection, covering behavioural patterns, biometric signals, social interactions, neurobehavioral responses, and other types of data that can form a digital profile without the active participation of the individual.

Digital unobservability is recognized as a key element of information self-preservation and is a component of human dignity, autonomy and intellectual sovereignty in the information age.

46.3 The use of technologies for covert or contactless collection of personal and behavioural data (in particular, facial recognition, spatial tracking, biometric identification, analysis of microexpressions or neurophysiological reading) is allowed only if:

the presence of special, specified and informed consent of the user;

clear and public marking of devices and collection areas;

ensuring the possibility of refusal or blocking by the user at any time;

conducting an independent technical and ethical audit of the transparency of the system.

46.4 Any form of direct or indirect coercion to undergo digital identification, mandatory registration or verification of a person in the digital environment is prohibited, unless such a procedure is necessary to meet critical security requirements (transactional, legal or operational), exercise public authority or protect vital public interests.

This prohibition applies to:

mandatory verification requirement for access to digital content, platforms or services that do not perform a public, state or high-risk function;

imposing on users the use of biometric or behavioural identification algorithms in informational, educational, cultural, social or entertainment environments;

creating conditions under which a person is actually deprived of the opportunity to exercise his/her right to digital secrecy without restricting other rights or access to basic services.

All digital authentication mechanisms should be based on the principles of voluntariness, proportionality, openness of alternatives and guaranteed choice in favour of anonymous or minimally identifiable access, unless otherwise contrary to security or justice requirements.

46.5 The state guarantees the development of a comprehensive infrastructure of anonymity as a component of the digital sovereignty of the individual. This includes not only the provision of technical means, but also legal, ethical and educational mechanisms to support digital secrecy, in particular:

creation and maintenance of virtual public spaces (forums, services, educational platforms), which technically and normatively lack mechanisms for collecting, analysing, processing or storing user identification data;

introducing anonymous use regimes for government digital services (e.g., reviewing regulations, accessing public information, participating in public hearings, etc.) in cases where these services do not require mandatory personalization or identity verification;

creation of a legal protection mechanism, including judicial, for persons who choose digital secrecy as a legitimate form of exercising their autonomy, privacy or protest excessive digital control;

development and implementation of special digital design standards that exclude the possibility of covert data collection in the basic interfaces of digital interaction and guarantee the right to anonymous stay without loss of functionality;

support for educational and outreach programs on digital stealth practices, total tracking risks, and personal privacy protection.

Such events are part of the state's digital policy and are integrated into the strategies for the development of digital human rights, information ethics and inclusive infrastructure.

CHAPTER 47. THE RIGHT TO DIGITAL SELF-DEVELOPMENT AND IDENTIFICATION SELF-PRESENTATION

Everyone has an inalienable right to independently form, define and change their digital identity — both public and private — in any digital environment, including information platforms, social networks, communication services, virtual spaces, and Metaverses. This right covers the freedom to choose a name, avatar image, visual style and design, symbolic or aesthetic elements, manner of communication, emotional colouring, as well as the scope of disclosure of personal or ideological characteristics.

An individual has the right to create, maintain, and modify multiple, dynamic, context-sensitive, or fragmented forms of their digital identity, including pseudonymous, gaming, artistic, protest, or alternative. Changing (including updating or reconfiguring) a digital identity may not be grounds for restricting access to services, lowering algorithmic rating or other digital index, discriminating against or violating the rights of an individual, except when such identity is used to fraudulently or violate the rights of others.

Identification self-presentation in the digital environment is recognized as a form of freedom of expression and covers the appearance of the avatar, the choice of visual style and design, the rhythm of publications, vocabulary, voice (including synthetic or stylized), as well as narrative and emotional patterns of interaction. An individual has the right to change the way they have a digital presence according to their context, platform, role, emotional state, or social behaviour strategy.

It is prohibited to force a person to combine all digital identities, accounts or images into a single profile; until a universal or global permanent identifier is established; to pass the verification of personal, psychological or cultural characteristics, which narrows the possibility of self-establishment to standardized, binary or predetermined categories. It is also prohibited to automatic, coercive or carried out without proper information and consent profiling, which reduces the multiplicity of manifestations of a person to one generalized pattern.

The state, digital platforms and operators of AI systems are obliged to provide an appropriate, flexible and technologically supported environment for free digital expression of a person. In particular, the ability to edit, mask, archive, save, restore or revert to previous images without sanctions or the need for additional justification should be guaranteed.

Digital interfaces should provide a person with mechanisms for managing self-presentation: changing the name, style, visual image; choice of the level of publicity; blocking algorithmic profiling; managing digital statuses and roles. It is not allowed to use any element of style, colour, vocabulary or frequency of publications for automated personality analysis without the prior, explicit consent of the person.

The right to digital self-determination is a fundamental human right in the digital age, intrinsically linked to respect for dignity, guarantee of freedom of expression, ensuring psychological integrity and recognition of cultural pluralism in the global information environment.

Every person has an inalienable right to independently define, form and transform their digital identity — both public and private — in any digital environment, including information platforms, virtual spaces, communication services and social networks.

This right includes freedom of choice:

own name, pseudonym, avatar image, graphic symbols, colour palette, emblems, digital voice (including synthetic or modified) or visual embodiment;

the volume and depth of disclosure of personal information, character traits, professional role, status, worldview positions or emotional openness;

creating and maintaining multiple, dynamic, context-sensitive, pseudonymous, fragmentary, or anonymous profiles that reflect different aspects of the personality without limiting or coercion unified personification.

Digital self-development implies the recognition of the right to a variety of embodiments of one's own self — including playful, professional, artistic, ironic or protest forms — provided that such forms of self-expression do not violate the rights of others and are not used for destructive behaviour.

A change in digital identity cannot be grounds for discrimination, restriction of access to services, or reduction of the level of digital trust or reputation. Such a restriction is allowed only in cases where the user acts in bad faith and resorts to deliberate disguise for the purpose of misleading or violating the law.

47.1 Identification self-presentation in the digital environment covers the full range of external and interface-communication self-expression of a person, which is formed, transformed or implemented in virtual, social, information and communication platforms. It includes:

choice of name (real or pseudonym), avatar (3D model, image or emoji avatar), visual design, voice embodiment (synthesized or stylized), as well as cognitive (neuronal), narrative, emotional or stylistic profile of interaction;

determination of the parameters of individual communication — content, frequency, duration, tone, form, rhythm and depth of publications, replicas, reactions, digital traces (including timestamps, typical interactions), as well as linguistic, lexical or aesthetic models (patterns);

creation and support of game, alternative, symbolic, allegorical, artistic, protest or imaginative embodiments of oneself in the digital space with varying degrees of intensity and realism — in particular, role-playing images in the Metaverses, creative stylized images in social networks, or recreated (reconstructed) visual identities in VR;

the ability to freely transition between different digital images, modes, statuses and roles depending on the platform, context, situation, emotional state or personal presence strategy.

47.2 It is prohibited to force a person to:

merging all their digital identities, accounts, avatars, or roles into a single unified profile, which removes their multiplicity or pseudonymity;

the consolidation of a single permanent or immutable image, avatar, name, or profile that restricts an individual's right to freely change or update their digital self-presentation according to the emotional, psychological, or social context;

passing identification or psychological (cultural and value) verification, which limits the right of a person to self-determination and change of identity over time.

It is also prohibited to use profiling tools that automatically reduce the multiplicity of digital manifestations of a person to a single standardized cognitive or social model without their knowledge and consent.

The state, digital platforms, and AI system operators are obliged to create an appropriate, flexible, and non-discriminatory environment for the realization of an individual's right to digital expression in all its forms.

In particular, they are obliged to:

recognize the right of a person to change their digital identity, update their self-presentation or transition to a new avatar role, regardless of the time, frequency or form of such changes, without the obligation to verify them or provide explanations;

refrain from imposing any sanctions, restrictions, blocks, reduction of algorithmic rating or digital reputation, or deprivation of access to the services on the basis that the user has chosen several parallel or interrelated forms of self-expression within the same or different platforms;

- provide flexibility and multi-level controllability of digital interfaces, providing a person with the technical capability:
- edit digital images according to a change in identity or social context;
- hide (mask) part or all of the attributes of your digital appearance;
- back up or archive previously used avatars;
- revert to previous versions of the digital image or perform a full restart of the digital image;
- not to carry out automatic or forced profiling of persons on the grounds of self-expression style, vocabulary, color scheme, frequency of publications, avatar style or other markers without the prior express informed consent of the person and with the possibility to withdraw such consent at any time.

The right to digital self-formation is recognized as a fundamental principle of respect for the individual in the information society and is of particular importance for ensuring freedom of expression, pluralism of opinion and psychological safety in the digital age.

CHAPTER 48. THE RIGHT TO PSEUDONYMITY, ANONYMITY AND DIGITAL CONCEALMENT OF DATA AND TRACES

Every person has an inalienable right to use a pseudonym, a digital nickname, avatar or other form of alternative identification that does not reveal their real name, natural person, or official status in the digital environment. This right is guaranteed in all forms of online interaction — in social networks, gaming environments, professional platforms, forums, service providers, educational systems, and any other digital spaces.

Pseudonymity is recognized as a form of self-expression, privacy protection, psychological safety, and digital self-realization. It can be permanent, temporary, situational or multiple at the user's choice and does not require mandatory registration, confirmation or binding to an official, except in cases expressly provided for by law and limited by the principle of necessity and proportionality in the interests of national security, prevention of cybercrimes, ensuring justice or protection of the rights of others.

Anonymity in the digital space is recognized as a recognized (legitimate) form of participation in public life, expression of views, obtaining information, protest, self-education and protection of vulnerable groups. It is prohibited to establish technical, organizational or administrative barriers to anonymous access to open resources, limit the functionality or quality of the service on the basis of anonymity, as well as discriminate or stigmatize persons who have chosen a non-personalized form of digital presence.

Each person has the right to form and control their own digital presence, including the creation of several digital profiles with different levels of personification, style, functional purpose and thematic focus. A person has the right to independently determine the volume of digital traces (activity data) that he/she leaves, in particular, to request their deletion, to carry out de-indexing or cleaning, as well as to use modern means of digital invisibility — encryption, virtual private networks, synthetic voices, digital simulation technologies, pseudo-identifiers or other tools of concealment (masking), if such use does not pursue an illegal purpose.

Digital platforms that provide public access to services, information or interactions, regardless of jurisdiction, form of ownership or scale, are obliged to ensure the right of users to voluntary identification, guarantee access without mandatory account registration, refrain from collecting personal data and other identifying information without the direct consent of the user, and ensure full control of users over their own data, including the right to delete it, Export, view, and temporarily block (pause processing).

The use of pseudonyms or anonymous presence cannot be grounds for discrimination, restriction of participation in public processes, reduction of the significance of statements or prohibition of use of services. It is prohibited to automatically restrict or exclude such users from discussions, consultations, petitions, crowdsourcing initiatives, competitions or digital e-government platforms, unless the identification is expressly necessary to ensure legality or the public interest.

48.1 Every person has an inalienable right to use a pseudonym, digital nickname, avatar or other form of alternative identification that does not reveal their real personal data or official status in the digital environment [366]. This right is guaranteed in any form of online interaction — in social networks, gaming platforms, professional environments, forums, service providers, educational systems, etc [367,368].

Pseudonymity is recognized as a form of self-expression, privacy protection, psychological safety, and digital self-realization. It can be permanent, temporary, situational or multiple at the user's choice and does not require mandatory registration, confirmation or binding to an official [369].

Restrictions on the use of a pseudonym may be applied only in cases expressly provided for by law and limited by the principle of necessity and proportionality [370], for the purpose of:

protection of national security and public order;

prevention of fraudulent activities or other cybercrimes;

ensuring justice and law enforcement;

protection of the rights and freedoms of others, in cases of defamation, discrediting or concealment of aggressive or illegal behaviour.

48.2 Anonymity in the digital space is recognized as a recognized form of self-expression, participation in public life, exchange of information, self-education, protest and protection of vulnerable groups [371].

Prohibited:

establishment of administrative, organizational or technical barriers for anonymous access to open information resources;

limitation of quality (reduction of the level of service) or functionality of the service for anonymous users in the absence of a reasonable technical need;

discrimination or stigmatization of individuals who have chosen an anonymous or pseudonymous form of digital presence.

48.3 A person has the right to freely form and control their own digital presence, in particular:

create, maintain and use multiple digital profiles with varying degrees of openness, personification, theme, visual style, functional orientation or identity depending on the context, purpose and personal preferences;

independently determine the volume and nature of digital traces that remain in the process of interaction with information environments to be able to completely or partially delete them, configure periodic cleaning, deindexing from search engines, automatically block the storage of session data or prevent tracking by trackers;

use modern digital stealth technologies, including concealment or imitation of digital activity, metadata encryption, VPN, TOR, pseudo-identifier generation, use of avatars, holographic shells, digital animation, stylistic filters or synthetic voices, if this does not contradict applicable law and does not pursue the purpose of committing illegal actions.

48.4 All digital platforms that provide public access to information, communication or services, regardless of jurisdiction, ownership models or audience scale, are obliged to guarantee compliance with the principles of anonymity, self-control over the digital footprint and voluntary identification [372].

They must:

provide access to the basic functionality of the service (viewing, navigation, searching, downloading) without the obligatory registration of an account and creating an account or logging in through third-party identification platforms;

Covertly or automatically collect identifying, behavioural, or technical data (including IP addresses, geolocation, device configuration, browser fingerprints, cursor movements) using analytics scripts, plug-ins, sensors, or algorithmic trackers without the express informed consent of the user;

provide users with full control over their data, including the right to view, edit, export, suspend, permanently delete and manage information stored or generated during the use of the Service, including analytical data, session logs and hidden profiles;

undergo an independent external audit of systems and practices to comply with the principles of anonymity, voluntary verification and transparency of profiling, and publish the results of such audit at least once a year in the public register of privacy policies and practices.

48.5 The use of anonymity or pseudonym cannot be the basis for any form of discrimination, stigmatization or restriction of a person's access to digital rights, services and forms of participation in public life.

It is prohibited:

reduce the credibility or weight of statements in digital discussions solely on the grounds that the user has not disclosed their identity, or deny them the right to publicly comment, post or respond;

automatically restrict or exclude anonymous or pseudonymous users from participating in contests, consultations, discussions, crowdsourcing initiatives, petitions, surveys, or other forms of public activity;

restrict access to tools for participation in the formation of digital policy, the implementation of digital justice, the use of e-governance institutions or administrative services only due to the lack of personalized verification, if such verification is not necessary to ensure security, legality or protection of the public interest.

CHAPTER 49. THE RIGHT TO DIGITAL OBLIVION AND THE RESTART OF DIGITAL HISTORY

Every person has an inalienable right to digital oblivion — that is, for full or partial deletion, masking, restriction of access, deindexing, depersonalization or transfer to archival mode of digital information, that has lost relevance or legal or social relevance, violates privacy, damages the dignity or reputation of a person or is discriminatory in form or consequences.

This right applies to both personalized materials and any digital data that directly or indirectly identifies an individual, in particular: open information (profiles, comments, content), hidden digital records (logs, caches, technical mirrors), shadow profiles and analytical indices, digital traces and contextual markers, recommendation layers, automatically generated characteristics (summary) or interpretations, as well as any data created or collected by third parties without the informed consent of the individual.

The right to digital oblivion is a component of informational self-determination and is implemented through the principle of dynamic control over personal digital reputation, privacy and the right to change or terminate its own digital presence.

An individual has the right to request the total or partial deletion of any digital information that directly or indirectly identifies him, regardless of the source of publication, if such information has lost its relevance or public importance, violates the right to privacy, causes damage to dignity or reputation. It has the right to request the suspension of access to archived or duplicate copies, to challenge the results of search results, algorithmic ranking or personalized recommendations, as well as to seek the labelling of content as outdated or inconsistent with its current status.

Digital platforms, regardless of ownership, jurisdiction or scale of activity, are obliged to provide effective mechanisms for the implementation of this right: provide users with interfaces for self-deletion or submission of an official appeal, promptly and transparently consider such appeals, implement procedures for complete deletion (digital data cleaning protocols) or information anonymization, as well as to stop the practice of storing or redistributing data after its deletion, except in cases expressly provided for by law.

The exercise of the right to digital oblivion cannot be limited by formal or technical grounds for refusal, in particular: invoking the public interest without proper justification of its significance; technical failure of the platform, which is the result of unfair design; or prior consent without the right of revocation granted without full awareness of the extent of the processing of their data.

49.1 Everyone has a guaranteed right to digital oblivion — that is, to delete, restrict access, mask, deindex or depersonalize digital information that has lost its relevance, is false or inaccurate, violates privacy, damages dignity or reputation, restricts the rights and freedoms of a person, or promotes discrimination. This right applies to both open information and hidden digital records, shadow profiles, algorithmic labels, analytical links and data generated by third parties about the person.

49.2 The right to digital oblivion covers all types of information that directly or indirectly identifies a person and that is no longer relevant, or is processed without an appropriate legal basis, or is not subject to further access or processing. This is:

personal data, including photos, videos, texts, messages and comments, which were published or previously generated by or with the user's participation;

information that has become irrelevant due to changes in the professional, social, ideological, medical or legal status of a person, or that no longer reflects a reliable or up-to-date image of the user;

digital traces of activity on Internet resources and electronic services, including social networks, forums, chats, databases, educational platforms, search engines, delivery services, financial services, as well as other digital platforms and electronic environments that store the history of a person's interactions;

- Profiles, logs, behavioural patterns, content recommendations, analytical reports, or machine-generated descriptions of user actions automatically collected and stored by systems or algorithms, which are stored without their express informed consent.

49.3 A person has the right to exercise control over their digital presence in terms of its historicity, accessibility, and relevance. Including:

demand the complete or partial deletion of any digital information (including metadata, backups, algorithmically generated inferences) associated with it, regardless of the source of publication, if such information has lost its legal or social relevance or harms privacy, dignity or reputation;

request the suspension of access or temporary blocking of archived or duplicate copies located on third-party servers or in cached search engines;

challenge the results of search results, algorithmic ranking or content systems that lead to distortion of the image of the person, discredit, moral humiliation or associate it with irrelevant or harmful contexts;

- require that content be labelled as historically outdated, archival, irrelevant, or not reflective of the person's current status, and that a special interface be put in place for such labels in digital profiles and media archives.

Digital platforms, regardless of ownership, jurisdiction or scale of activity, are obliged to implement systematic and technically effective mechanisms for exercising the right to digital oblivion.

They have:

provide users with a clear, accessible and guaranteed effective interface for self-deletion or submission of a formal request for administrative deletion of any information from a personal profile, including activity, metadata, interactions and automatically generated records;

consider requests for deletion or restriction of data without excessive procedural obstacles, within a period not exceeding 15 working days from the date of submission, with the obligatory provision of confirmation and justification of the decision;

implement modes of full digital history update — special technical protocols that allow the user to fully update their digital presence at their own discretion (change identifiers, remove traces of previous activity, adjust personalization logic) without losing the legal continuity of the account;

- Stop the practice of storing or redistributing information after its deletion, unless otherwise expressly provided by law, and ensure that backups, caches or third-party services do not contain information accessed without the user's informed consent.

49.4 The exercise of the right to digital oblivion cannot be limited by unreasonable, technically unsubstantiated or purely formal grounds. It cannot be rejected:

based on a generalized reference to the public interest, if the relevant information does not have a reasonable public significance or does not perform a critical function in the field of freedom of speech, transparency or public control;

due to the technical failure of the platform, if such failure is the result of unfair infrastructure design, centralized storage, lack of deletion protocols or failure to fulfil the obligation to ensure the implementation of user rights;

due to past consent to data processing, if such consent has been given:

- no revocation option;
- without sufficiently informing the person about the consequences and extent of the processing;
- fs part of non-transparent or discriminatory contract terms that deprived a person of the ability to exercise control over data in a dynamic mode.

CHAPTER 50. THE RIGHT TO INFORMATION INVISIBILITY AND OFFLINE EXISTENCE STATUS

Every person has the right to information invisibility — a deliberately chosen state of lack of digital activity or inaccessibility for technological surveillance in digital environments, regardless of the level of their accessibility. This right includes the ability to exist outside of systemic digital monitoring, tracking, analysis, or logging without being forced to constantly interact with interfaces, networks, devices, or algorithms.

The right to offline existence status guarantees the ability to be outside the digital environment without losing basic civil, social, economic or cultural rights. A person cannot be obliged to be permanently present in the digital space to exercise his/her basic rights and freedoms, receive state, medical, educational or financial services, participate in elections, public life or communicate with authorities.

No technological system can interpret the absence of a digital footprint, unwillingness to register, disabling geolocation, blocking tracking, or restricting interaction with devices as suspicious, anomalous, or unlawful (deviant) behaviour. The lack of digital activity is not a reason for forming a risk profile, downgrading, monitoring, or restricting access to basic digital citizenship functions.

A person has the right to refuse the use of biometric systems, geospatial interfaces, algorithmic identifiers, automated presence recorders, electronic bracelets, smart devices or wearable sensors, unless otherwise expressly provided by law in exceptional cases. The technological autonomy of the individual takes precedence over the interests of data collection and analytical processing.

Digital systems are obliged to provide the ability to function in the mode of information invisibility ("invisible user") — without storing technical and behavioural data (including logs, metadata, IP addresses, behavioural markers), without automatic profiling and algorithmic determination of interests or reactions. A person has the right to a temporary (session) presence in the digital environment without being forced to leave a permanent digital footprint.

A digital platform or system that does not provide the possibility of information invisibility or forcibly requires permanent registration, assignment of a digital index, identification number or creation of a personal account is considered to violate the right to offline existence and entails liability in accordance with this law.

50.1 Every person has an inalienable and fundamental right to information invisibility as a modern form of privacy, autonomy and digital dignity [373,374]. This right includes a conscious and voluntary decision, guaranteed by the protection of the law, to be inaccessible to any form of digital surveillance, profiling, modelling, analytics, tracking, prediction, personalized exploitation, or use for the benefit of third parties — for commercial, governmental or technocratic purposes.

Information invisibility means the ability of a person to be in a state of non-presence in the digital environment: stay out of the reach of digital agents, algorithms, platforms and services that record, record, interpret or evaluate the behaviour, emotional state, cognitive profile or digital footprints of a person [375].

The right to information invisibility guarantees avoidance:

constant digital surveillance — both open (through activity tracking) and hidden (through backend analytics and machine learning);

analytical indexing — creation of databases or registers with behavioural, consumer, cognitive or social markers;

assignment of evaluation metrics (ratings, indices, trust score, risk indicators, etc.) to a person by systems, which are formed without a request or consent;

automated formation of predictive or strategic profiles that can influence decision-making about a person in the financial, medical, educational, labour or migration spheres.

The implementation of this right implies that digital systems cannot create a computational profile of a person — even indirect or impersonal — without clear, informed, specified and dynamically revocable consent. Such consent is not considered valid if it was given under conditions of cognitive pressure (manipulative influence), systemic asymmetry or without the technical possibility of easily refusing [376].

50.2 The status of information invisibility implies both the physical and digital absence of a person at a specific point in time, as well as the extended right not to generate, transmit, accumulate or store any digital traces in situations that are private or non-public in nature or do not require digital fixation. This right is proactive — it not only responds to interference but also guarantees the freedom to prevent the digital representation of a person in forms that do not correspond to it [377].

Within the framework of the exercise of this right, a person has the right to:

completely disabling or refusing to use permanent digital identifiers (including UUIDs, device fingerprints, cookie profiles, session keys, etc.) and replacing them with temporary or pseudonymous attributes that do not directly or indirectly identify a person;

refusal to interact with digital systems, services or devices that require a disproportionate amount of personal, biometric, behavioural, cognitive or contextual data, if such requirements are not mandatory for the operation of the basic service and exceed the limits justified by the principle of data minimization;

the right to ignore digital or automated requests, including consent forms, surveys, marketing proposals, forecasts or recommendation systems, if they are aimed at creating or clarifying a person's social, psycho-emotional, ideological, value or consumer profile without their direct initiative and control.

Thus, the status of information invisibility is not only technical, but also regulatory in nature and reflects a person's conscious position on the boundaries of disclosure and use of their own data in the digital environment. It should be guaranteed both at the level of user tools and at the level of regulation, design of digital systems and the formation of public digital policies.

50.3 Digital systems, services, platforms and devices that interact with the user are obliged to ensure compliance with the information invisibility regime chosen by the user. This means not only refraining from collecting data without consent, but also actively providing technological support for modes in which an individual can act without leaving digital traces, controlling the volume, nature and duration of their own digital representation.

Digital systems have:

to implement full-fledged modes of information invisibility, including:

"Incognito" mode - interaction with digital services, including websites, applications and digital platforms, without automatically saving browsing history, cookies, session data and local cache;

"temporary presence" mode means a one-time session of interaction with a digital service without saving digital traces after its completion, including device identifiers, IP address, geolocation, navigation route and actions;

"Contextual masking" mode is an interaction mode that allows the user to choose the level of their presence (from complete anonymity to pseudonym or partial identification) when participating in public discussions, voting, online broadcasts, conferences or other digital spaces;

ensure mandatory labelling and anonymization of all data generated in such modes, including metadata, logs, device tokens, time stamps, frequency responses, and other similar technical indicators, with subsequent automatic deletion or cryptographic destruction after the end of the session, expiration of a specified period, or at the request of the user;

introduce visually clear and standardized interfaces that demonstrate the status of its visibility in the system in a way that is understandable to a person — in particular, through color indicators, icons, indication of active collection, statuses "unindexed", "anonymous", "temporary", as well as notifications about changes in the policy of data collection or fixation;

provide each user with built-in verification tools that allow you to determine whether interaction with the service leads to the storage, processing, transmission or transit of any data (including through hidden mechanisms, embedded APIs or third-party services) and whether such interaction corresponds to the level of invisibility chosen by the user. Such tools should include a transparent log of operations (actions) of the data processing system, notification of crossing the prohibited collection limits, as well as the ability to immediately terminate the session or switch the mode.

The requirements of this provision are mandatory for all digital systems working with personal data, virtual identities or behavioural information of an individual, regardless of their jurisdiction, ownership model or algorithmic architecture.

50.4 Any action aimed at circumventing the status of a person's information invisibility — i.e. an attempt to collect, process, aggregate (generalize) or analyse personal or behavioural data contrary to a deliberately chosen mode of information invisibility — is recognized not only as a technical violation of the settings, but also as a significant encroachment on the digital sovereignty and dignity of a person [378,379].

In particular, such violations are considered to be:

- the use of hidden or undocumented backend data collection mechanisms that do not appear in the user interface, but continue to record, log or transmit information without their knowledge;
- Use of hidden behavioural analytics algorithms that collect data in the background (e.g., mouse micromovements, scrolling rhythm, click sequence) for the purpose of further shaping or modelling a user profile;
- Implementation of analytical markers or tracking codes ("Tracking Codes") that store information in cookies, through Device Fingerprinting, in local caches or third-party advertising modules, regardless of the activation of the "incognito", "do-not-track" mode or other similar mechanisms.
- activation of biometric identifiers or triggers (including microexpressions, gait patterns, voice patterns, muscle response, body temperature dynamics) that have not been expertly verified or are used without the express informed and voluntary consent of the person;
- Use of third-party APIs and/or SDKs (software development kits) embedded in services that collect or potentially collect data for the benefit of third parties without explicitly stating this fact in their privacy policies or terms of use;

All the above actions are recognized as unauthorized intrusion into the information space of a person, regardless of the presence of immediate or obvious consequences. Such actions entail legal liability, including administrative, civil law, and in the case of a systemic or mass nature, criminal liability.

50.5 The right to information invisibility cannot be reduced solely to a technical function of the interface or an option in the "privacy settings" menu, since it constitutes a basic element of a person's information autonomy and requires systemic protection. This right should be recognized as fundamental — at the level of constitutional guarantees, principles of digital ethics, public law norms and contractual standards in the field of data processing.

Exercising the right to information invisibility requires:

- integration of appropriate modes and mechanisms into the design of the digital architecture (privacy by default — privacy-by-default, silence by design — silence-by-design, invisibility-first priority), rather than adding them as optional plugins or secondary user experience options;
- ensuring that this right is necessarily considered in the process of developing technical standards, digital interaction protocols, API regulations and the architecture of interaction of artificial intelligence agents;
- imperative enshrinement of invisibility provisions in contracts for the use of digital services, licenses, privacy and data protection policies, data processing agreements and terms of technology tenders.

Any violation or ignorance of this right — even in the form of undocumented bypasses, artificially created technical unsuitability, or lack of functional alternatives — is recognized as a systemic restriction of personal freedom in the digital environment.

CHAPTER 51. PROHIBITION OF DIGITAL DISCRIMINATION AND THE CREATION OF DIGITAL CASTES

Any form of digital discrimination is prohibited, including (but not limited to) direct or indirect, overt or algorithmically hidden inequalities in relation to persons or groups of persons in the digital environment on the basis of race, gender, age, nationality, language, political, religious or other beliefs, disability, social status, profession, place of residence, genetic or psycho-emotional characteristics, as well as digital behaviour, consumer rating, credit history, the frequency of activity or other dynamic parameters generated or used by artificial intelligence, personalized analytics, automated control, or other algorithmic systems.

It is prohibited to implement or use digital systems that automatically create "castes" — i.e., hierarchized or segmented groups of users with unequal treatment, to which different levels of access, service quality, cost of services, speed of service, recommendation content or level of reliability of information determined based on pre-collected or simulated digital attributes are applied.

Every person has the right to transparency, control and the possibility of objecting to any digital profile, classification or rating applied to them in an automated form. A person has the right to know what data has been used, what algorithms have been applied, what conclusions have been drawn, as well as to request the deletion, revision and correction or cancellation of such profiles if they cause bias or restrict access to equal digital opportunities.

Digital platforms, information system operators, providers of artificial intelligence, personalized analytics services and other digital entities are obliged to adhere to the principles of algorithmic fairness, explainability, impartiality, equality of access and social responsibility in the process of data processing, ranking, provision of services or interaction with users.

It is not allowed to put a person in a less favourable position only on the basis that his digital footprint does not correspond to artificially determined parameters of the algorithm, and his profile in the system is marked as "risky", "ineffective", "unattractive", "unprofitable" or "toxic", without granting him the right to appeal, receive an explanation or initiate interference.

Any systemic digital segregation — including special tariffs, limited access rights, content visibility control, differences in voting or interaction rights — based on an artificially generated rating or algorithmic classification is recognized as a gross violation of the principle of digital equality and is subject to immediate termination.

The state ensures control and supervision over compliance with the principle of non-discrimination in algorithms, decision-making procedures by AI systems and in the use of digital profiles and indices, as well as creates independent bodies to consider complaints, conduct audits and implement digital justice standards.

51.1 Digital discrimination — is any restriction, exclusion, isolation, humiliation or prejudice against an individual based on their digital behaviour, metadata, algorithmic profiles, rating scores, socio-digital indices or other chosen or imposed form of digital identity [380,381]. Such discrimination violates the principle of equality and non-discrimination in the digital space and is incompatible with the right to information autonomy and digital dignity [382,383].

Prohibited:

creation or use of systems for automatic assessment or ranking of a person without his/her informed consent, if this leads to the restriction of his/her rights and opportunities;

restricting access to services, platforms, content, or communications due to the user's use of a pseudonym, anonymous mode, or digital footprint minimization practices;

dividing users into "trust classes", "influence groups" or "quality castes" according to the criteria of non-transparent algorithms or social rating;

Using historical data—including previous interaction sessions, emotional reactivity, post style, cognitive speed—to block or restrict access to public functions, education, medicine, employment, or other socially relevant areas and resources.

51.2 In the case where algorithmic sorting, classification or ranking is used in the interests of ensuring the security, stability and reliability of the service or improving the quality of the user experience, the following actions [384] are allowed only if the following requirements are met:

- Clear legal basis: The algorithm must have a well-defined purpose, a legitimate legal basis, and a technical specification that includes a decision-making architecture;
- Guarantee of human control: the final decision on the restriction of rights or the assessment of a person cannot be made exclusively by an automated system without the mandatory participation of an authorized person empowered to analyse the context and provide clarifications;
- technical transparency: a person has the right to be informed about how, on the basis of what data and according to what rules his/her algorithmic portrait or rating is formed, as well as the right to correct, appeal or complete refusal to participate in such a system.

51.3 Every person, regardless of the level of digital experience, technical training or formed digital profile, has the right[385,386]:

- to equal and non-discriminatory access to digital platforms and services, administrative resources, educational opportunities, media space, as well as to participation in digital democracy;
- to protect against digital stigmatization, forced labelling or discriminatory labels by artificial intelligence systems and digital intermediaries;
- refusal to participate in algorithmic ranking systems without losing or restricting access to basic functions and services;
- to participate in the formation of standards of ethical and legal regulation of digital technologies, in particular by participating in advisory boards, committees, focus groups and public discussions.

51.4 Any digital platform that influences access to vital services or public goods is obliged to:

- publish complete, up-to-date and understandable documentation on algorithmic systems, including sorting principles, moderation mechanisms, visibility parameters, factors influencing the formation of recommendations and metrics for evaluating their effectiveness;
- ensure the smooth functioning of the service, i.e. guarantee equal access of users to all its functions regardless of their identity, registration method or profile type;
- ensure the functioning of permanent tools for monitoring manifestations of digital discrimination in any form that guarantee the possibility of filing a complaint, receiving a reasoned response and initiating an independent examination;
- regularly, but at least once a year, undergo an independent audit carried out by authorized or accredited entities to check the absence of structural bias in the algorithms, with the mandatory publication of conclusions and the immediate adoption of corrective measures within the established time frame in case of detection of impact asymmetry.

CHAPTER 52. THE RIGHT TO CYBER-PHYSICAL INTEGRITY IN ARTIFICIAL INTELLIGENCE SYSTEMS

Everyone has an inalienable and state-guaranteed right to cyber-physical integrity — that is, to protect physical integrity, sensory perception, spatio-motor balance, neuropsychic stability, and bioinformational integrity from any influences arising from interactions with digital, robotic, autonomous, sensory, or hybrid systems, including systems based on artificial intelligence, the Internet of Things, or immersive technologies.

It is prohibited to exert any influence from such systems that may cause physical discomfort, sensorimotor disorientation, cognitive overload, vestibular instability, panic states or disturbances in the physical and psycho-emotional state, as well as imitation of acts of aggression, touching or forcible restriction of freedom of movement without the direct, informed consent of the person.

Each person has the right to independently determine the limits, intensity and permissibility of any physical or sensory feedback from technological systems. This right includes the right to complete or partial shutdown of vibrations, tactile pulses, temperature effects, force feedback, as well as immediate termination of interaction by means of emergency commands or emergency interfaces to ensure that the process is completed safely.

All systems containing components of physical or sensory interaction must comply with the principles of technological safety, bioethical responsibility and adaptability to the user's condition, and must also provide a "zero impact" mode — that is, a mode of complete neutralization of physical and sensory stimuli for vulnerable users or those who have not provided consent.

Developers, manufacturers, suppliers, operators and other authorized entities of such systems are obliged to conduct a comprehensive assessment of potential risks, clearly mark risk modes of interaction, provide simple and unhindered access to emergency shutdown mechanisms, and carry out systematic monitoring and documentation impact on the user's health. Any changes in the impact architecture or software or firmware updates that alter the patterns of physical or sensory interaction with the user's body are allowed only if their express and informed consent is obtained.

Any violation of the right to cyber-physical integrity — including unwanted stimulation, lack of a refusal mechanism, ignoring an exit request, or using bodily influence as a means of coercion — qualifies as a serious digital security offense and entails legal liability.

Everyone has a guaranteed inalienable and universal right to cyber-physical integrity — that is, to protect the integrity, inviolability and stability of one's own physicality, sensory perception, spatial-motor balance, neuropsychic balance and bioinformational homeostasis (stability of information and biometric processes) when interacting with any digital, robotic, sensory, autonomous or hybrid systems [387].

This right should be understood as a guarantee that no digital or algorithmically controlled technology — but not limited to artificial intelligence systems, autonomous robots, wearable devices, immersive and virtually-augmented environments, haptic feedback mechanisms, smart environments, Internet of Things (IoT) devices, neural interfaces, and other similar technologies — may initiate or exert an effect on a person in a way that results in:

violation of physical comfort, bodily safety and spatial orientation;

loss of control over motor or sensory functions, overload of sensory (receptor) systems, or impaired coherence between perception and action;

changes in the neuropsychiatric state, manifested in the form of anxiety, emotional overstrain, panic reactions, cognitive overload or dissociative states;

impairment or limitation of a person's ability to independently determine and protect the boundaries of their physical and sensory integrity, including the bodily, tactile, temperature, sound, visual and vestibular spheres.

An individual's right to cyber-physical integrity covers not only protection from harmful or unwanted influences, but also imposes an obligation to ensure that all systems of the technological environment are designed and functioned taking into account human bodily sovereignty — as a basic ethical and legal

criterion in the age of physical-digital integration. This means that the human body, its integrity, reactions, perceptual limits, and ability to self-regulate have unconditional priority over the criteria of efficiency, navigational logic or performance of technological systems[388].

52.1 Any physical, sensory or neurophysiological effect on a person from digital, cyberphysical, autonomous or robotic systems is prohibited without their prior consent, expressed in an understandable form, informed and revocable at any time. Any influence that leads or may lead to a violation of the cyber-physical integrity of a person is also prohibited. All forms of actions or interventions carried out by digital, cyber-physical, autonomous or robotic systems that:

create a risk of direct or indirect physical harm, including due to pain, seizures, muscle tremors, microtrauma or sensorimotor disorientation;

disturb spatial orientation and balance of perception through visual, auditory, temperature or tactile means that cause vestibular destabilization, a feeling of falling or loss of stability;

cause excessive cognitive or informational load, leading to frustration, visual fatigue or oversaturation, decreased reactivity or maladaptation in the real environment after interaction;

simulate a physical intrusion into the user's personal space (harassment, aggression, touching, restriction of movement, compression or capture), regardless of whether such actions are implemented as an element of gameplay, educational simulation, simulator or artistic environment;

use signals of pain, fear, disgust, shame, or other basic emotions as a means of influencing and imposing a behavioural model, an element of gamified feedback, or a mechanism of "motivation" for further interaction.

Separately, the use of neuroexposure triggers (including light and sound patterns, binural frequencies, hypersynchronized visual effects, rhythm manipulation or colour impulses) that can cause trance states or affective instability, in the absence of proper medical justification and appropriate certification, is prohibited.

The principle of preventive protection establishes that the responsibility for preventing such impacts rests with the developer, supplier and operator of the system. Influence is recognized as prohibited both in the case of actual harm and in the presence of a reasonable threat to the bodily or sensory well-being of a person.

52.2 Each person has a guaranteed right to independently determine the depth, form, intensity and duration of physical, sensory or neuro-emotional interaction with any digital, cyberphysical, autonomous or robotic system. This right applies, among other things, to domestic, educational, play, professional, rehabilitation and media immersive environments and is exercised through:

the ability to adjust or completely disable all forms of physical feedback (including tactile impulse, vibration, force, thermal stimulation, pressure, ultrasonic excitation, electromuscular modulation, biometric response, or quasineuronal feedback);

interface control over the level of intensity of such influences with a mandatory preliminary visual or auditory message, which must clearly determine the nature, expected effect and duration of the relevant mechanism;

the right to refuse any physical interaction with the digital system without restricting access to other functions or content, except where such exposure is necessary to ensure the physical safety of the person or the environment (for example, emergency braking in autonomous vehicles);

- the right to immediately terminate the interaction and carry out a full-fledged "emergency withdrawal" (hereinafter referred to as "emergency egress"), which is guaranteed through:
  - visible interface element in the form of a button or gesture;
  - voice or non-verbal command (in particular, using a neural interface);
  - automatic exit timer, which is activated when the stimulation limit level is reached.

Each system that involves physical or sensory interaction must be accompanied by an "impact passport" — an official electronic or printed document containing information about all potentially active components of the reverse effect, their permissible limits, methods of disconnection, as well as technical and ethical restrictions of use for different age, physical and psycho-emotional groups of users.

52.3 All technical systems that contain components of physical or sensory interaction (including tactile, motor, thermal, acoustic, kinetic, visual, electrical stimulator, biomechanical or neurosensory modules) are required to function in accordance with the principles of technological safety, bioethical responsibility and the priority of human bodily sovereignty.

Such systems must be designed, tested and implemented in compliance with the following criteria:

standardized Safe Exposure Certification: Every device that produces physical or biologically relevant stimulation is required to undergo independent testing for compliance with national and international regulations that determine permissible levels of strength, duration, frequency, depth of penetration, temperature, inertial impact and cognitive load.

adaptive adjustment of exposure parameters: systems are required to consider age characteristics (including childhood and geriatric status), health status, history of trauma, the presence of neuropsychiatric sensibilities (e.g., epilepsy, anxiety disorders, post-traumatic stress), as well as cultural and religious factors regarding the integrity of the body.

zero-impact mode: every system with a physical interface is obliged to provide the ability to completely disable any form of influence on the user's body, which guarantees safe participation in the digital environment in absolute neutrality mode. This mode must be activated by default for users who have not given explicit consent to the other.

built-in monitoring and self-control mechanisms: systems are required to ensure registration and self-diagnosis of the level of exposure with the possibility of automatic intensity reduction or complete shutdown in case of exceeding the limits of physiologically safe parameters.

responsible firmware update: it is forbidden to implement any changes in the logic of the system (including updating vibration patterns, amplitude of motor force or pulse frequency) without first informing the user, confirming his consent and conducting a public audit.

Each organization that manufactures or distributes such systems is legally liable for damage caused as a result of violation of these criteria, including in cases of unintentional damage, if it is the result of technical negligence or constructive thoughtlessness.

52.4 Developers, manufacturers, suppliers and operators of digital, robotic, sensory and cyber-physical systems that carry out or provide for interaction with the human corporeality have an increased responsibility for ensuring and maintaining the state of cyber-physical integrity of users. Entities are required to implement the following mandatory protocols and standards:

Conducting a preliminary comprehensive risk assessment, which includes testing the technological architecture for the likelihood of causing physical, psycho-emotional, sensory or neurophysiological disturbances. Such an assessment is carried out before the implementation phase, covers both conventional and extreme application scenarios (including stress tests, long-term sessions, emergencies) and is subject to publication in the open safety register.

Mandatory provision of clear labelling of all modes or functions that may pose a risk, including through the use of:

- graphic warning signs;
- color-coding of exposure levels;
- auditory or vibration signals informing about a change in the level of exposure;
- language descriptions and warnings tailored to the needs of users with visual, hearing, or cognitive impairments.

Mandatory implementation of emergency mechanisms for shutting down the system, which can be activated by the user at any time and without explaining the reasons, including due to:

- a physical "emergency exit" or "full shutdown button»;
- voice command;
- neurointerface contraction or other biosignal (e.g., heart rate, pulse, EEG pattern);
- automatic activation when a stress response or bodily instability is detected.

Mandatory provision of constant monitoring of the system by the user, including a log of interactions, levels of influence and history of calls to security mechanisms, with the ability to individually adjust the thresholds of admissibility and save the history of reactions in a local encrypted environment.

Mandatory periodic review and updates of the device or software architecture in case of:

- mass or repeated receipt of complaints;
- detection of negative impact patterns;
- changes in international regulations that increase the requirements for sensory, motor or cognitive safety.

An individual's right to cyber-physical integrity has an absolute priority in the system of regulation of the latest technologies.

CHAPTER 53. THE RIGHT TO CONTROL PSEUDO-IDENTITY, DIGITAL AVATARS AND REPRESENTATIVES

Each person has the exclusive, inalienable and legally protected right to control the creation, configuration, use, modification, restriction of access, deactivation or destruction of all forms of their own digital pseudo-identity, which are virtual representations that directly or indirectly reproduce their personality, elements of appearance, voice, style of thinking, behaviour, vocabulary, manners, reactions or life experiences.

An individual has the exclusive right to determine the status and legal nature of his/her digital representatives and avatars, including by establishing the limits of autonomy, functional purpose, time limits, context, scope of use, and legal personality of the respective representation. It is prohibited to use such representations in forms that contradict the will, values or dignity of a person.

It is prohibited to create or use representations of a person without his/her direct, informed and revocable consent. Covert modelling, automatic generation, use in hidden or unauthorized contexts and imitation of a person that lead or may lead to distortion of his/her image, emotional manipulation, imposition of decisions or restriction of his/her rights are prohibited. Representations are required to include a digital signature of origin, authentication and parameters of the permitted use.

Digital pseudo-personality is recognized as an object of personal non-property rights. A person has the right to registration, legal protection, confirmation of authorship, inheritance, revocation or prohibition of secondary use of representation. Violation of authorship or self-determination in a virtual environment entails legal responsibility determined by law.

A person has the right to complete destruction or deactivation of his/her digital representation at any time. This right applies both to digital forms created by it and to those created by third parties without consent or in violation of the original terms. Resumption of representation after withdrawal of consent is prohibited without repeated permission.

The right to control a digital pseudo-person is extraterritorial in nature and is subject to international protection. In case of legal conflicts or transnational use of representations, the principle of digital self-determination of the individual is preferred. The state is obliged to ensure the implementation of this right through the use of diplomatic, legal and technical instruments of international cooperation.

53.1 Each person has the exclusive right to create, edit, use, restrict or destroy their digital avatars, agents, representatives, simulacracers, digital shadows, voice clones, facial replicas, virtual copies, holograms, as well as any other forms of personified or partially personified presence in the digital environment created using artificial intelligence technologies, generative models, procedural modelling, virtual reality or mixed platforms[389,390,391].

53.2 Each person has the exclusive, inalienable and legally protected right to control the creation, configuration, use, modification, restriction of access, deactivation or destruction of all forms of their own digital pseudo-identity, which are virtual representations that directly or indirectly embody their identity, elements of appearance, voice, style of thinking, behaviour, vocabulary, manners, reactions or life experiences[392].

Digital pseudo-personality (within the meaning of this article) includes:

digital avatars that imitate or display a person's appearance in virtual environments;

voice or speech copies, including synthesized speech models created from voice recordings or speech style;

virtual agents, digital representatives, or chatbots that interact on behalf of a person in social networks, administrative services, business platforms, educational spaces, or in legal relationships;

digital shadows used in recommendation, analysis or personalization systems and based on a person's behaviour patterns, preferences or history of actions;

holograms, 3D models, avatars for Metaverses, digital reconstructions, or posthumous personifications created from an individual's public or private data.

The right of control provides that none of the listed forms can be created, used or transformed:

without the direct, informed and predetermined consent of the person — separately for each context (personal, commercial, legal, educational, etc.);

without the possibility for the individual to withdraw such consent, deactivate or completely destroy the corresponding avatar, copy or interaction interface;

without a clear designation for third parties that they are an avatar or virtual agent, and not an individual, with the provision of authentication mechanisms and a source of control over the relevant digital representation;

without ensuring the functioning of the system for logging actions, changes, updates and interactions carried out on behalf of a person, with the possibility of further auditing, verification, suspension or appeal of such actions.

The right to control a pseudo-personality includes the prohibition of creating a representation of a person in a misleading, humiliating or unacceptable form, including through the use of deepfakes, synthetic voice technologies, stylistic or behavioural falsifications, as well as simulations in toxic or unacceptable contexts, including in the areas of political agitation, disinformation, discrimination, humiliation of honor and dignity or manipulation of the consciousness of third parties [393].

53.3 A person has the exclusive right to independently determine the status, functional purpose and legal nature of his digital representatives and avatars, in particular by:

establishing a regime of control over behaviour, responses, appearance, style of representation, as well as the scope and boundaries of the permitted use of the digital image;

prohibiting the use of her avatar, voice, or style in contexts contrary to her will, values, political, ethical, or religious beliefs—including satirical, parody, commercial, fictional, erotic, or violent scenarios;

determining the form of legal personality of digital representation — whether it is exclusively visual, functional or communicative, as well as whether it can be endowed with delegated rights, in particular, the right to electronic signature, consent, participate in negotiations, trials, consultations, etc.;

limitation of the time of representation, definition of scenarios and platforms for its application, as well as technical conditions for storage and distribution, including the right to "self-destruct" a digital copy after the expiration of a set period or the occurrence of a certain event.

In the event of death or loss of legal capacity of a person, the right to manage his/her digital representations passes to a specially authorized heir (digital principal) specified in the digital will or according to the general procedure for inheriting digital rights.

53.4 No digital system, regardless of its functional purpose, jurisdiction, level of autonomy or source of funding, has the right to create, copy, reproduce or store avatars, voice models, virtual agents, behavioural patterns or any other form of pseudo-identity [394] without observing the following conditions:

obtaining a direct, conscious, specific and documented consent of a person to such actions;

clear definition of the purposes of creation, parameters of operation, storage time limits, level of autonomy and control mechanisms;

preliminary audit of ethical compliance and technical safety, ensuring that there is no harm to the honour, dignity, reputation or personal safety of the user during the creation or use of the representation.

Prohibited:

hidden (latent) modelling of personality based on indirect features (in particular, stylistic markers, manner of communication, thinking structures, media preferences or reactions to content), if the result is perceived as an image identified with a real person;

use of pre-existing representations in a way that goes beyond the originally agreed context (e.g., the use of an educational avatar in commercial advertising without updating consent);

automatic creation of personalized agents based on data from open sources (publications, social networks, blogs) without prior permission of the person to such a level of aggregation, processing and reconstruction of their digital behaviour.

A particularly strict regime for obtaining consent for the creation of post-death personifications is established — digital images of deceased persons that reproduce their appearance, voice, facial expressions, movements or behaviour. In such cases, the creation can be carried out only if there is:

a person's will expressed in advance during his lifetime (digital will);

official consent of the heirs of the first degree of kinship, subject to the principles of dignity, privacy and cultural sensitivity.

Violation of this paragraph is the basis for the mandatory immediate destruction of the relevant digital representation and bringing the offender to responsibility provided for by law (administrative, civil or criminal), depending on the severity of the violation.

In particular, the use of digital avatars, voice models or virtual (simulated) agents is prohibited in such cases:

in a judicial, administrative, diplomatic or medical process without authentication, documentary confirmation of the person's consent and provided that the functional autonomy of representation is limited exclusively to an informative or advisory role, without the possibility of decision-making;

in contexts of emotional manipulation — including advertising, political propaganda, imitation of emotional intimacy, or psychological influence through avatars stylized as a specific real person, if such interaction is imposed on third parties without their informed consent;

in practices of automated persuasion, negotiation or withdrawal of consent, when pseudo-personality is used as a tool to put pressure on the expression of the user's will, in particular in the field of online sales, media subscriptions, financial services or administrative services.

All digital representations created on the basis of a person's data must contain in the structure a "digital signature of origin" — an algorithmically embedded unique token that certifies:

1. identity of the person who gave consent;
2. modelling source;
3. date of creation;
4. type of data used;
5. acceptable application scenarios.

Such signatures must be available for external verification upon request, cannot be altered without the consent of the person, and must prevent forgery or simulation without the knowledge of the copyright holder of the digital representation.

The absence of such a signature or its inaccessibility is a sufficient reason to block the relevant representation, limit its visibility in open systems and investigate the legitimacy of its creation.

53.5 Every person has the right to legal recognition of his/her digital pseudo-identity as an object of personal non-property rights, which is endowed with a legal status that ensures his/her protection from unauthorized use, manipulation, forgery, misappropriation or automated creation or reproduction for commercial, political or manipulative purposes.

This means that:

any form of digital representation of a person that directly or indirectly identifies him (through appearance, voice, behaviour, style of thinking, emotional reactions, speech, context or cultural image) is recognized as an element of his personal non-property space and is subject to protection at the level of private law, copyright, ethical standards and the principle of digital immunity;

a person has the right to register his/her digital avatars, voice models, virtual agents, simulated copies or post-mortem representations in a special register of digital identities (state or self-regulated), for the purpose of documenting authorship, legitimacy, scope, permitted application scenarios, as well as for the protection of his/her rights;

Digital platforms engaged in the hosting, generation, transfer or commercial implementation of representations are obliged to ensure the technical implementation of the principle of "digital validation of the person" — i.e. verification of the authenticity of origin and the availability of the right to use an avatar, voice, virtual agent or any other object that performs a personified function;

copying or secondary use of a person's digital representation without his/her consent (including partial cloning) is recognized as a violation of his/her right to digital integrity, digital dignity and reputational self-determination.

The legal status of a digital representation can be:

a) limited (for personal use only));

b) delegated (with representative functions and limited communication powers);

c) author's (with the protection of creative stylistics or emotional image);

d) hereditary (in the case of creating a posthumous representation or memorial copies);

e) revoked (with the right to completely cease to exist (delete) a digital representation at the request of a person).

53.6 A person has the right to complete destruction or irreversible deactivation of any digital representation created on its basis, including by third parties, in the event of:

withdrawal of consent to use an image, voice, style or behavioural model;

the use of representation in a context contrary to the ethical, political, religious or personal beliefs of the person;

interference with the functioning of representation by outsiders or systems, which led to its distortion, distortion of meaning, manipulation of emotional expression, behavioural reactions or simulation of decisions on behalf of the person without his participation;

proven fact of the operation of digital representation in offline mode contrary to certain parameters (in particular, in cases of self-generation, functional escalation or integration into third-party systems).

Such a right to destruction (the right to digital erasure) is exercised through:

a person's request, submitted in digital form with authentication, to the platform that stores or distributes the relevant representation;

an automated procedure for withdrawing consent, provided during the creation of a representation or built into the management parameters;

decision of a supervisory or ethics body if it is proven that the continued existence of the representation leads to a violation of human rights, in particular the right to dignity, privacy and privacy, freedom of opinion or security.

After the destruction of digital representation, any recreation, restoration or reconstruction, including partial, without the new consent of the person, is recognized as a significant violation of digital autonomy and entails civil, administrative or criminal liability.

In the event of the death of a person, the right to initiate digital destruction or deactivation of his/her post-death representation belongs to:

a) to the heir identified in the digital will;

b) to an authorized person appointed by a court decision;

c) next of kin, in the absence of other mechanisms determined during the person's lifetime.

53.7 In the event that a digital representation of a person (avatar, voice model, digital agent or other form of pseudo-identity) is created, stored, processed or used outside the jurisdiction of the state of citizenship or permanent residence of a person, the provisions of international digital law and the principle of extraterritorial protection of personal non-property rights in the digital environment shall apply.

This means that:

the rights of a person to create, manage, prohibit or destroy digital representation belong to him/her regardless of the platform, state, technology or conditions for the creation of the relevant digital form;

any use of representation on transnational digital platforms, including in Metaverses, blockchain systems, decentralized autonomous organizations and generative models with open access, is carried out in accordance with generally recognized principles: digital dignity, digital autonomy, informed consent, the right to erasure and protection against misappropriation of a person's digital identity;

in the event of a conflict of rights recognized in different jurisdictions, preference shall be given to the person's right to self-determination in the digital environment, confirmed by his/her digital will recorded by means of an electronic signature, blockchain documentation or a registered digital manifest;

The state provides mechanisms for international legal assistance, legal support and technical support for the implementation of the protection of digital rights of an individual, including the filing of international complaints, recognition of decisions of supervisory authorities, blocking of illegal representations and termination of data processing outside national jurisdiction.

A state may initiate the conclusion of international treaties on the mutual recognition of rights to a digital person and its derivative forms, in the forms of an international digital passport, digital heritage, memorial image, cultural personification or virtual representation.

CHAPTER 54. RIGHT TO DIGITAL OBLIVION, DEINDEXING AND DELETION OF INFORMATION CREATED OR PROCESSED BY ARTIFICIAL INTELLIGENCE

A legal regime for the protection of an individual from excessive, non-transparent or unreasonably long storage, distribution, processing, transformation or combination of information about him/her, created, modified or systematized by artificial intelligence systems, is established, which determines the grounds for exercising the right to digital oblivion, establishes procedures for deindexing and deleting digital content, as well as imposes obligations on operators, providers and third parties to stop the use of such Information.

Every individual has an inalienable right to digital oblivion, which includes the right to request the termination of the storage, indexing, use, processing, distribution or creation (generation) of any information that directly or indirectly concerns him or her and was created, modified or systematized by AI systems without proper legal basis or contrary to the interests of the person.

Information is subject to mandatory deletion, blocking or deindexing in cases where it is factually unreliable, distorted or incomplete, in particular due to errors of AI systems; outdated, lost public significance or does not correspond to the current context of a person's life; created or used in violation of the regimes of confidentiality, privacy, protection of personal, biometric, medical or behavioural data without informed consent or other legal basis; such that has become publicly disseminated as a result of the work of algorithms for ranking, classification, recommendations, search indexing or generative analysis and has caused reputational, psychological, professional, social or physical damage; such that contains signs of offensive, discriminatory, stereotypical or misinformation content, including unauthorized deepfake images, audio and video material, digital twins, falsified avatars, false biographies or automatic labels.

A digital oblivion request can include deleting information from databases, personalization algorithms, decision logs, archives, generation modules, search engines, and virtual assistants; blocking access of third parties to such information with its preservation only for justified archival or evidentiary purposes; termination of further distribution through digital platforms, aggregators, web services and APIs with the obligation to delete cached and backup copies; deindexing in external and internal search engines; as well as preventing the use of this information as training, test or reference materials for future generations of AI systems.

AI system operators, digital service providers, database administrators and owners of algorithmic platforms are obliged to start verifying the request within no more than fifteen calendar days from the date of receipt of the request, if necessary, with the involvement of technical, legal and ethical expertise; to ensure full compliance with the applicant's requirements, provided that there are no legal grounds for further preservation of information; notify the applicant in writing about the results of the consideration or provide a reasoned refusal indicating the methods and procedure for appealing; document all requests and actions taken, including maintaining log files, audit trails, create and maintain convenient and non-discriminatory channels for submitting requests (electronic forms, hotlines, embedded interfaces, specialized APIs); and refrain from transferring or using the disputed information until the final consideration of the request and the expiration of the appeal periods.

In case of failure to comply with the requirements of a person, his/her right to apply to the Commissioner for Digital Human Rights, the relevant supervisory authority or the court in order to stop the processing of disputed information, stop the functioning of the system or a separate algorithm, delete technical traces, digital copies, transaction records and generation products, bring responsible persons to administrative, civil or criminal liability and compensate for material and moral damage. At the same time, the national digital rights authority enters information about the incident into the National Register of Digital Dignity Conflicts, organizes its investigation and ensures public information.

The provisions apply regardless of the method of creation or transformation of information, including if it is the result of direct generation by the AI system, the result of processing data sets or the result of algorithmic transformation, classification, segmentation or automated scoring.

This section defines the mechanisms for restoring digital dignity, ensuring a person's control over their own information and guaranteeing their reputational security in the algorithmic information space. The implementation of the right to digital oblivion is recognized as a prerequisite for compliance with the principles of digital ethics, information self-determination and the rule of law in the digital environment.

54.1 Any individual has an inalienable right to digital oblivion, which includes the right to request the termination of the storage, indexing, use, processing, distribution or creation, including by generation, of any information that directly or indirectly relates to them and has been created, modified, systematized or used using AI systems without legal basis or contrary to the interests of the person [395,396,397].

54.2 Any information that has at least one of the following features is subject to mandatory deletion, blocking or deindexing:

is actually unreliable, distorted, incomplete, or created based on erroneous or unusable data, in particular if such errors arose as a result of the functioning of the AI system that generated the data, classified, transformed, or combined data;

outdated, such that has lost its social significance, or does not correspond to the current context of a person's life;

is created or used in violation of the regimes of confidentiality, privacy, protection of personal, biometric, medical or behavioural data, as well as without proper informed consent of the person or other legal basis determined by law;

has become publicly disseminated because of the functioning of automated systems of ranking, classification, recommendations, search indexing or generative analysis of AI systems and has caused reputational, psychological, professional, social or physical damage to a person;

contains signs of offensive, discriminatory, stereotyped or misinformation content, including in the form of unauthorized generations of images, audio, video, digital twins, falsified avatars, false biographical information, automatic labels or tags.

54.3 A digital oblivion request may include one or more of the following requirements:

deletion of relevant information from databases, personalization algorithms, logs of decision-making systems, archives, platforms, generation modules, search engines or virtual assistants;

blocking access to this information by any subjects, except for the person who provided it or is its subject, with the possibility of storing it only for reasonable archival purposes or for evidence in court proceedings;

termination of the distribution of relevant information through digital platforms, content aggregators, web services or API access systems with the mandatory deletion of cached and backup copies;

deindexing of information in external search engines, as well as in internal search and recommendation algorithms integrated into AI systems;

prevention of further use of restricted information, such as training data, test sets or data sources, in the learning process and during the generation of results by artificial intelligence systems.

54.4 AI system operators, digital service providers, database administrators, owners and administrators of algorithmic platforms, as well as other entities that collect, store, process or otherwise act with digital information, are required to ensure:

not later than within 15 calendar days from the date of receipt of the request for the application of the right to digital oblivion, the entities specified in this paragraph are obliged to:

start checking the request with the involvement of technical, legal and ethical expertise and ensure its completion within the period established by law;

ensure compliance with the applicant's requirements if there are no legal grounds for further preservation of information;

notify the applicant in writing or electronically about the results of the request or provide a reasoned written refusal indicating the procedure and methods of appeal;

document all appeals and actions taken in response to them, with the preservation of event logs and audit records for control and verification by authorized bodies;

create convenient, accessible and non-discriminatory mechanisms for submitting digital oblivion requests, including electronic forms, support services, integrated features in user interfaces and specialized APIs;

refrain from transferring or using the relevant information until the completion of the procedure for considering the request and the expiration of the terms of appeal established by law.

54.5 In case of failure to comply with the requirements of a person regarding the exercise of his/her right to digital oblivion, such a person has the right to apply to the Commissioner for Digital Rights, the relevant supervisory authority or to the court with demands for:

termination of processing of information about which a dispute has arisen or a request for digital oblivion has been submitted;

temporary suspension of the functioning of the system or a separate algorithmic module;

removal of technical traces, deletion of digital copies, transaction records and generation products;

bringing responsible persons to civil, administrative or criminal liability;

compensation for material (property) and moral (non-property) damage.

The National Authorized Body for Digital Rights enters information about such an incident into the National Register of Digital Dignity Conflicts and ensures its investigation, proper public information and legal support.

54.6 The provisions of this article apply regardless of whether the relevant information was created as a result of generation by the AI system, formed because of processing data sets (arrays) or transformed as a result of algorithmic processing, classification, segmentation or algorithmic evaluation (scoring).

54.7 The provisions of this article are a means restoration of a person's digital dignity, a guarantee of control over personal information and ensuring its reputational security and protection from encroachments on it in the conditions of algorithmic information space.

CHAPTER 55. THE PRINCIPLE OF RESPECT FOR CULTURAL AND LINGUISTIC IDENTITY

A legal regime for the protection of cultural, linguistic, religious and regional identity of an individual in the process of its interaction with artificial intelligence systems is established. It is prohibited to use algorithms in a way that leads to the imposition of cultural stereotypes, behavioural patterns, consumer patterns and ideological guidelines that contradict the right of a person to preserve his or her national, ethnic, linguistic or religious identity.

Entities that develop, provide or operate artificial intelligence systems on the territory of the State are obliged to provide support for Ukrainian as the state language in all interfaces, messages, reference materials and feedback channels. Within the limits of technical feasibility, support is also provided for the languages of national minorities and language communities living in the respective territory.

It is prohibited to imitate accents, language constructions, humorous stylizations and cultural features in a way that humiliates national dignity, forms hostile stereotypes or leads to discrimination based on language, origin or religion. Such content is subject to immediate deletion, and, if necessary, withdrawal from circulation by the authorized body. Algorithmic settings are subject to immediate correction, considering the principle of non-discrimination

Generative model providers and operators are required to implement cultural sensitivity mechanisms and bias testing procedures; Provide warnings and labels for content with potential signs of humiliation or stereotyping; create mechanisms for free appeals and ensure prompt correction or deletion of generation results.

In the field of education, media and administrative services, the creation and functioning of a multilingual environment that ensures equal access is guaranteed. It is prohibited to reduce the quality of functionality, accuracy of translation, availability of reference materials or speed of processing requests based on the language of request. Accessible alternatives should be provided for persons who use national minority languages, sign language and other means of communication.

A person has the right to respect for their cultural identity in the process of interaction with AI systems and the right to file a complaint about cases of humiliation, stereotyping or language discrimination. Operators of AI systems and digital platforms are obliged to provide accessible channels for filing complaints and provide a reasoned written or electronic response indicating the measures taken within a specified period.

55.1 All digital information systems and artificial intelligence systems operating on the territory of the State are obliged to:

to provide support for the Ukrainian language and, in cases provided for by law, the languages of national minorities;

refrain from imitating accents, language constructions or using language forms in a humiliating or discriminatory way that violates national dignity;

to apply mechanisms for ensuring cultural sensitivity in generative models.

55.2 In the field of education, media and administrative services, the creation and maintenance of a multilingual environment is ensured, which guarantees equal access to all users.

CHAPTER 56. THE PRINCIPLE OF ALGORITHMIC DIGNITY OF WORK

The principle of the algorithmic dignity of labour of a public law nature is established: the introduction and use of artificial intelligence systems in the field of labour relations is allowed only for the purposes defined by law and under the following conditions: necessity and proportionality; data minimization; transparency of the logic of decisions; ensuring effective human supervision; with unconditional respect for the dignity, privacy and autonomy of the employee, equality opportunities and non-discrimination.

This principle applies to all forms of employment, as well as to related procedures: hiring, civil service, platform and gig work, remote and home-based work, freelancing and other civil law relations with signs of dependent work, internships and apprenticeships. In particular, the principle applies at the stages of recruitment, pre-selection, profiling, evaluation and certification, shift planning and performance management, including in cross-border digital environments.

Automated dismissal, account deactivation, reduction of access or pay, other disciplinary measures, or withholding of payments without human review and motivated written notification are prohibited; total or excessive monitoring, including biometric and so-called "emotional" monitoring, if it not expressly provided for by law and is not strictly necessary; algorithmic scoring, which is discriminatory or non-transparent, including using direct, indirect or proxy features; so-called "jurisdictional shopping" in order to circumvent labour guarantees; unjustified geolocation surveillance outside of working hours; the use of manipulative gamification mechanics and so-called "dark patterns"; as well as algorithmic obstruction of the exercise of trade union rights and collective action.

Algorithmic labour management is allowed only if there is a goal defined by law, proven necessity and proportionality; ensuring transparency of decision-making logic with clear explanations of relevant factors; minimizing the amount of data and restricting access to it according to the principle of "need to know"; effective human supervision with the right to preventive intervention, suspension or cancellation of decisions; availability of a mechanism for prompt appeal with mandatory fixation in event logs; determination of data storage periods and procedures for their deletion or anonymization.

The employer (customer) and the supplier of AI systems are obliged to conduct an algorithmic Labor Impact Assessment (AIA Labour) before putting the system into operation, which should include a description of the objectives and legal bases, characteristics of data and flows, analysis of risks and mitigation measures, results of quality and fairness tests, human oversight, appeals and incident management plans. They are required to ensure that employees or their representatives are consulted, to publish a public summary of the evaluation results without disclosing trade secrets, as well as to carry out event logging, preservation of evidentiary information and version control throughout the life cycle of the system.

The employer (customer) is obliged to ensure that the employee or candidate is informed before the start of data processing: the employee (candidate) is informed of the purpose and legal basis, processes and limits of autonomy, sources and categories of data, their storage periods, quality and fairness indicators (if any), contact of the responsible person, appeal procedures and the procedure for suspending the execution of disputed decisions. This information is reflected in internal policies, collective agreements and vacancy conditions, and significant changes are notified in advance, but no later than within a reasonable time before their application.

The employee is guaranteed the right to receive a meaningful explanation of the logic of an automated decision with a significant impact within a reasonable time, but no later than within 72 hours from the date of submission of the request; to a real human review of such a decision with its automatic suspension, defined terms of consideration and access to data and logs; to access their own data, profiles and logs (logs), to obtain their copies, to correct inaccuracies, restrict or delete them in cases determined by law; refusal of biometric or emotional monitoring without negative consequences; as well as protection from repressive measures, with the presumption of their presence in case of any negative actions of the employer within six months after filing a complaint or carrying out collective activity.

The "Bring Your Own Device" (BYOD) and geolocation monitoring policy does not apply by default; exceptions are allowed only under conditions of proven official necessity and conducting an algorithmic assessment of the impact on labour (AIA Labour), provided that a service device is provided or an isolated corporate profile is used, with a ban on access to private data, mandatory disabling of geolocation monitoring (tracking) outside the borders working hours and full logging of all data accesses. An employee's consent to monitor a private device is not recognized as an independent legal basis.

Special guarantees are established for platform and gig work: account deactivation or reduction of the level of prioritization are not allowed automatically and are possible only after a real human review and a motivated notification to the employee, with the right to appeal and automatic suspension of the execution of the disputed decision; rating, dynamic pricing, and routing algorithms should be transparent, explainable, and verifiable; Hidden sanctions ("shadow fines") and the use of discriminatory parameters are prohibited; transparency of payment terms, timeliness of payments, minimum guarantees of income, compensation in case of cancellation of an order through no fault of the employee, payment of tips in full and proper timely settlements are ensured.

Regular testing of AI systems for bias, detection of failures and false positives using representative samples and agreed fairness metrics is provided, mandatory documentation of the reasons for deviations and corrective actions taken, revalidation after each significant change in algorithms or data, independent audits for high-risk applications and version logging with the publication of a summary of changes in an accessible form for employees.

In the event of a conflict between the use of innovative technologies and human rights, unconditional preference is given to the rights and freedoms of the employee, as well as public safety; any practices that violate this principle are recognized as null and void and entail the mandatory termination of the operation of the relevant algorithmic mechanisms, the restoration of the employee's position, compensation for the damage caused by the damage and application of liability established by law.

56.1 Administrative and legal consolidation of the principle.

1. The principle of algorithmic dignity of labour is a public law principle according to which the regulation and application of artificial intelligence systems in the field of labour relations is carried out exclusively in compliance with human rights and freedoms, namely:

ensuring respect for the human dignity, privacy and autonomy of the employee, equality of opportunity, non-discrimination, occupational safety, health care and fair transparent remuneration;

to prevent automated operation, deterioration of working conditions and disproportionate or full supervision, including biometric and emotional monitoring, except when such application is strictly necessary and expressly provided for by law;

to use AI systems only to achieve the goals defined by law on the basis of the principles of necessity, proportionality and minimization of data;

to ensure the transparency of algorithmic decisions and their logic, their traceability, as well as the employee's right to receive an understandable and meaningful explanation;

guarantee mandatory human review of decisions with significant impact by a competent individual and provide effective mechanisms for appeal and appeal, with mandatory automatic suspension of the execution of such decisions until their final review;

provide for regular testing for bias, as well as for compliance with quality and safety requirements; ensure documentation of changes in algorithmic models and logging of events with ensuring evidence; conduct independent audits for high-risk applications;

ensure that the employee or candidate is informed about the use of AI systems in the decision-making process that affects his labour rights;

guarantee the right of an employee or candidate to access his/her own data, as well as to rectify or delete it in accordance with the law;

extend this principle to all forms of employment, including platform and gig work, remote work, public service and other types of employment provided for by law.

In the event of a conflict between the use of artificial intelligence systems or other digital technologies and human rights, the rights and freedoms of the employee and public safety have an unconditional priority.

2. Algorithmic management of labour activity can be applied only if:

the legitimate purpose is determined and the necessity and proportionality of data processing is proved;

data minimization and restriction of purposes and accesses on a need-to-know basis;

transparency of decision-making logic, including the right to receive an understandable and meaningful explanation of relevant factors;

effective human supervision is established with the right to preventive intervention and suspend or cancel an automated decision;

the functioning of the mechanisms of operational appeal and reconsideration with mandatory fixation in the event logs was ensured;

technical and organizational safeguards against bias, false positives and abuses (testing, auditing, version control) have been implemented;

determines the storage periods, the procedure for deleting or anonymizing data and prohibits their secondary incompatible use;

The use of biometric or emotional monitoring is prohibited, except in cases expressly provided for by law and strictly necessary to ensure security.

3. This principle applies to all forms of employment and related procedures, in particular:

1) hired labour;
2) civil service;
3) military service;
4) platform (gig-) work, including activities through digital platforms;
5) remote and home-based work;
6) freelancing and other civil law relations that have signs of dependent labour;
7) internships, apprenticeships and other forms of personnel training using algorithmic tools.

Within the framework of the principle, personnel processes are also regulated:

- Selection and pre-selection of personnel;
- Profiling and suitability analysis;
- Algorithmic assessment and attestation;
- Work Shift Planning;
- Productivity management;
- Labor monitoring and control;
- Procedures for reorganization of labour relations and dismissal.

56.2 Scope and definition.

The article applies to employers, customers of works (services), suppliers and integrators of AI systems, operators of digital platforms, as well as other persons who implement or operate AI in labour processes on the territory of the State or in relation to employees and candidates whose work is organized from the State, including in cross-border digital environments (meta environments).

Algorithmic management is the use of AI systems or other automated systems for scheduling work shifts, assigning tasks, monitoring productivity, conducting assessments, applying incentives and sanctions, as well as suspending or terminating employment relationships or restricting access to a digital platform.

Automated decision with a significant impact is any decision made without the participation of a person or with his/her nominal participation, which significantly affects the rights, obligations or opportunities of an employee or candidate,

Regarding hiring or refusal of hire, determining the terms of payment, setting work schedules, providing access to bonuses, transfers, disciplinary measures, dismissal or deactivation of the account.

56.3 Prohibited practices.

The following practices are prohibited:

1. Automated dismissal, suspension or suspension from work, deactivation or "freezing" of an account, reduction of access level, reduction of pay or working hours, as well as other disciplinary sanctions or negative personnel actions taken in a fully or predominantly automated manner, if such a decision is made without:

- real (substantive) human viewing;
- identified responsible official and his motivated signature;
- preliminary written or electronic notification of the employee with an explanation of the logic of the decision, a list of data and evidence used, determination of the terms and procedure of appeal;
- fixing and storing event logs and case materials for at least twelve months.

Until the review is completed, the execution of such an automated decision shall be suspended, except in cases where there is a direct threat to security or there is a violation of the law.

2. Total or excessive monitoring of an employee's behaviour, emotions, psychophysiological indicators or microactivity ("bossware"), including constant audio and video surveillance, keylogging, analysis of facial expressions, tone of voice, gaze, heart rate or stress levels, unless such monitoring is required by law and is not strictly necessary and proportionate.

3. Use of biometric identification and "emotion recognition" technologies to evaluate employee performance, trustworthiness, or loyalty .

4. Discriminatory or opaque algorithmic scoring candidates or employees, in particular, with the use of direct or indirect features and proxy variables (place of residence or postal code, school or university, Socioeconomic profile of the district, accessibility schedule, language or accent features, citizenship, age, disability, marital status or paternity, union membership, religious or political beliefs, biometric or external characteristics, etc.) that cause or may cause direct or indirect discrimination.

The use of "black boxes" without proper explanation and documented bias tests is prohibited; at the same time, the burden of proving the non-discriminatory nature of algorithmic scoring rests with the employer or supplier.

Evaluation of employees using algorithmic tools should be based solely on relevant and lawfully collected data and clearly defined working criteria. In case of statistically significant deviations, the employer is obliged to immediately take corrective measures and notify employees or their representatives about them.

5. "Jurisdictional shopping" and covert deployment of AI systems in order to circumvent labour guarantees — i.e. intentionally selecting or moving the base jurisdiction, place of registration, hosting or actual operation, use of affiliates, subprocessors or proxy infrastructures **(**including abroad) without proper notification of employees and the supervisory authority, without conducting an AI Labor Impact Assessment (AIA Labour) and without complying with the requirements established by law— if it results in the avoidance of labor legislation, collective agreements, tax and social obligations, trade union rights or inspection control, are considered null and void for the purposes of the legitimacy of algorithmic control.

Such actions entail:

the obligation to immediately terminate the operation of the system;

restoration of violated rights of employees;

joint and several liability of the actual beneficiary and the persons involved;

imposition of fines and orders;

suspend or revoke system admission.

6. Unjustified or persistent geolocation tracking of an employee outside of working hours, including remote access to location data, Wi-Fi/Bluetooth scans, geofencing, and other location identifiers. It is not allowed to require the installation of any means of total monitoring on the employee's private devices (MDM/spyware, keyloggers, activity trackers, audio or video recording tools), except as expressly provided for by law and only under conditions of strict necessity and proportionality. If monitoring is necessary, the employer is obliged to provide a service device with separate data profiles or a secure container, as well as clearly defined policies, access and disabling tracking policies outside of working hours. The employee's "consent" to the installation of such facilities on a private device is not considered freely granted and cannot

be a legal basis for data processing. The temporary application of geolocation is allowed only during business hours and solely for security or logistics purposes — within defined geofences, with mandatory logging of access and minimum data retention periods .

7. Automated write-off, withholding, or other negative adjustment of wages, rewards, bonuses, tips, or other benefits, as well as automated accrual of fines or penalties, if such a decision is made:

without prior written or electronic notification of the employee with an explanation of the grounds, calculation methods and data sources;

without real (substantive) human review and a motivated signature of the responsible official;

without granting the employee the right to prompt appeal and a reasonable period for filing objections;

without recording in the event logs and keeping the full package of evidence materials for at least twelve months;

if, as a result of such write-off, payments are reduced below the minimum guaranteed level or the rules of taxation or social contributions are violated;

with immediate execution until the review is completed, except in cases of an obvious technical error of double payment, when a temporary blocking of the surplus is allowed with mandatory notification of the employee.

Unlawfully withheld amounts are subject to immediate return with interest accrual (in the amount of the NBU discount rate, unless otherwise provided by the agreement or the law) and compensation for the damage caused.

8. Manipulative gamification and nudging mechanics, including temporal stimuli, tempo counters, forced ratings or badges, hidden KPIs, "dark patterns" of the interface, dynamic thresholds or quotas, or "achievement" systems that encourage exceeding safe work intensity, ignoring breaks, reducing rest, or refusing to use safety equipment.

Separately prohibited:

- tying such mechanics to payment, fines or admission to changes;
- Turn off rest reminders;
- ignoring established medical and hygienic standards and indicators of fatigue.

Only voluntary, opt-out, and penal-free forms of gamification are allowed, which are accompanied by:

- limitation of the duration of working sessions;
- mandatory pauses;
- overload warnings;
- supervision of the responsible labor protection officer.

9. Circumvention or undermining of the rights of trade unions, workers' councils, collective representation or ethical regulation initiatives, in particular:

- using AI systems to identify or profile trade union activists or their supporters;
- monitoring or intercepting communications in order to obstruct the organization;
- Algorithmic change distribution, routing, or ratings that create a chilling effect on union membership or participation in collective action;
- manipulation of access to orders, tariffs or bonuses depending on participation in trade union events;
- interference in the election of representatives, negotiation procedures or strikes;
- forced individualization of labour relations, including fictitious requalification to the status of "self-employed" or individual entrepreneurs (FOP) to avoid collective rights;
- restriction of representatives' access to necessary information, including models, policies and decision logs, refusal of consultations or their unreasonable postponement, which deprives such consultations of effectiveness.

Such practices entail the nullity of the relevant decisions, the obligation to immediately restore the situation, the prohibition of re-acting, administrative responsibility, and in case of their systemic nature, the suspension of the operation of the relevant algorithmic mechanisms.

56.4 Obligations of the employer, customer of works (services) and supplier of AI systems.

1. Conducting an algorithmic Labour Impact Assessment (AIA Labour) prior to the implementation of the system, which should include at least:

a) determination of the purposes of data processing and expected results, description of scenarios for using the system and the limits of its autonomy;

b) legal bases and compliance with labour, data protection and trade union laws;

c) description of the subjects and their roles (employer, supplier, operator, responsible official) with the definition of areas of control;

d) categories, sources, volumes and data flows, including data of third parties and cross-border transfers, their storage periods and methods of minimization or anonymization;

e) risk profiles for the employee's rights (including discriminatory, security, psychosocial), risk matrix and mitigation measures;

f) quality and reliability tests, failure and false alarm detection mechanisms, and fault tolerance strategies;

g) Methods for detecting and eliminating bias (fairness tests, proxy control), as well as target metrics and tolerance thresholds;

h) a plan of human oversight and intervention, including the right to suspend or cancel decisions;

i) appeal procedures, terms of their implementation, levels of service obligations (SLAs) and mechanisms for suspending the execution of disputed decisions;

j) logging of events, ensuring evidence and determining the terms of storage of logs;

k) update plans and version control, as well as criteria for re-AIA Labour in the event of significant changes;

l) Incident Response Plan (IRP), setting RTO/RPO parameters and identifying contact points for 24/7 support;

m) the results of consultations with employees and their representatives, as well as the comments received;

n) assessment of the residual risk and decision on its acceptability , indicating the responsible person and the date of review;

o) public resume for employees in an accessible form, without disclosure of trade secrets;

p) conditions for integration with digital environments (Metaverse) and compliance with their codes;

q) in case of high risk — notification and submission of an algorithmic assessment of the impact on labour (AIA Labour) to the supervisory authority in accordance with the established procedure; AIA Labour is drawn up in writing, approved by the head of the employer or platform and kept throughout the entire period of operation of the system and at least twelve months after its decommissioning.

2. The involvement of employees and their representatives in preliminary consultations is mandatory and is carried out before the start of operation of the AI system and includes:

a) timely written notification of the intention to implement the system, its purpose, expected impact and deadlines;

b) conducting at least two rounds of consultations with the possibility of submitting written comments;

c) provision of consolidated results of AIA Labour (without disclosure of trade secrets) and risk matrix in an accessible form;

d) familiarization with the risk mitigation plan indicating responsible persons, deadlines and key performance indicators (KPIs);

e) discussion of alternatives and compensatory measures (training, retraining, adaptation of schedules, additional safety tools);

f) recording the course of consultations in the protocols indicating the received proposals and motivated answers of the employer;

g) ensuring access of employee representatives to the necessary technical and organizational information, test logs and draft policies;

h) providing employees and their representatives with a reasonable period, but not less than fourteen days, to analyse documents and prepare a position**;**

i) the possibility of engaging independent experts on labour and data protection;

j) ensuring the publication of a consolidated summary of results for all employees.

Failure to comply with the requirements for consultations and disclosure of certain information entails the suspension of the implementation of the system and is the basis for issuing an order by the supervisory authority, bringing to administrative responsibility and recognizing the relevant decisions as null and void.

3. Providing a real (substantive) human review of all automated decisions with a significant impact, implying at least:

a) appointing a responsible official (Human in Charge) with the authority to suspend or cancel decisions and personal responsibility for signing them;

b) providing that person with full and unhindered access to data, event logs, models, and applied criteria;

c) checking the legality and proper relevance of the data used, as well as assessing their bias and proportionality;

d) mandatory recording of the motives, time and results of any intervention in the case carried out by the responsible person, with the reflection of a reasoned decision in the protocol or electronic log of events;

e) preservation of review materials (protocols, motivated decisions) and event logs (logs) for a period of at least twelve months from the date of completion of the review, with the provision of proper protection and the possibility of further audit;

f) compliance with the specified terms of consideration fixed by legislation or internal regulations, as well as the agreed indicators of the level of service (Service Level Agreement, SLA), with the provision of urgent consideration in cases where there is a risk of significant harm to the rights, freedoms or legitimate interests of a person, society or the state;

g) prevention of direct or potential conflicts of interest of persons carrying out the review, as well as the introduction of mechanisms for rotation or duplication of control over critical decisions that have a significant legal or social impact;

h) mandatory informing the person in respect of whom the decision was made about its content and the means of appeal provided for by law. Until the completion of the human review, the execution of the disputed automated decision shall be suspended, except in cases requiring immediate intervention for security reasons or to prevent violation of the law.

4. The data is processed only to the extent necessary for a specifically defined and documented purpose. The use of private communications, including personal messengers, e-mail, VoIP and private chats, is prohibited, personal social media accounts and content from social media accounts, as well as data from personal devices of family members or other third parties — without a separate clear legal basis, without prior notification to the data subject, as well as without conducting an Employee Rights Impact Assessment (AIA Labour) and a Data Protection Impact Assessment (DPIA). Minimum retention periods and deletion or anonymization policies are set (typically — no longer than necessary to achieve the purpose and terms of appeal or inspection). Access to data is provided on a need-to-know basis and with a mandatory access log. Secondary use of data that is incompatible with the original purpose of collecting, creating hidden profiles and combining them with data from open sources or brokers without proper legal basis is prohibited. The use of special categories of personal data (including biometric and emotional state data) for work purposes is allowed only in cases expressly provided for by law, with the use of enhanced technical and organizational safeguards.

5. Regular, systematic and documented testing for bias, failures and false positives is carried out both at the implementation stage and throughout the life cycle of the system and includes at least:

periodic fairness tests using representative control samples and equality metrics, in particular Statistical Parity Difference (SP/D), Disparate Impact (DI), Equal Opportunity / Equalized Odds (EO/EP), calibration, etc., with the definition of thresholds for acceptable deviations;

a) Fault tolerance testing and failure detection, in particular through stress, fuzz and chaos testing, as well as assessing the proportion of false positives and false negatives and their impact on employee rights;

b) documenting the causal factors of deviations and plans for their correction;

c) mandatory revalidation after each significant change in data, model or its hyperparameters throughout the entire life cycle of the system;

d) for high-risk systems — independent verification or audit. Model Risk Management (MRM) is carried out according to the approved Model Risk Management (MRM) procedure, which includes: impact analysis, controlled version change; A/B or "canary" deployments; possibility of operational rollback; updating of supporting documentation. Public (internal) release notes are kept, indicating the date, responsible person, description of the changes and reasons, expected effect on accuracy, fairness and safety, results of repeated tests and the decision on admission.

6. BYOD Policy (Bring Your Own Device) and geolocation: By default, a ban is set. Exceptions are only allowed in cases of strict professional necessity and after conducting AIA Labour (Employee Rights Impact Assessment) and DPIA (Data Protection Impact Assessment), with the manager's written justification and the approval of the employees' representatives. Mandatory technical restrictions are provided:

a) provision of a service device or isolated corporate profile or container with MDM (Mobile Device Management) management, which is carried out without access to the user's private data;

b) prohibition of access to microphone, camera, keylogging (keystroke logging) and geolocation outside of working hours;

c) the use of geolocation is allowed only during working hours in certain geofences, with mandatory logging of accesses and limitation of storage periods to the minimum necessary;

d) the employee has the right to refuse to use their own devices (BYOD, opt out) without any negative consequences;

e) copying, indexing, or backing up private content and metadata is prohibited;

f) periodic review of the granted permit is carried out at least once every 6 months, and its validity period is no more than 12 months;

g) a clearly stated requirement to immediately disable tracking and delete the corporate profile after the end of working hours or dismissal;

h) provides for the maintenance of a register of issued permits indicating the details, including the grounds, deadlines and responsible persons. The employee's consent to the use of his/her own devices (BYOD) or geolocation is not an independent legal basis and cannot replace the established requirements of this paragraph.

7. Logs of events and decisions taken are kept with the provision of their evidence. The storage period of such logs is at least 12 months, and in case of a dispute or inspection — until their final completion.

8. A multi-channel operational appeal is provided within the defined SLA (Service Level Agreement) and automatic suspension of the execution of the disputed automated decision in cases of risk of significant damage to the rights or legitimate interests of the employee, which includes at least:

a) the following submission channels are provided: personal account and internal portal, e-mail, telephone hotline, written offline application;

b) the application is registered with the assignment of a unique number to it and immediate confirmation of its receipt;

c) the procedure is free of charge for the employee, who has the right to access all materials, data and event logs on which the relevant decision is based;

d) the terms of consideration of appeals are determined as follows: critical cases (dismissal, deactivation, withholding of payments or security risks) are considered within no more than 24 hours; other

cases — within no more than three working days, with the possibility of extension for up to three days, provided that the employee is provided with a reasoned notice;

e) the suspension of the execution of the decision shall remain in force until the completion of its review, except in urgent cases related to an immediate threat to security or with a detected violation of the law; in such cases, the least onerous temporary measures are applied;

f) you may not modify or delete relevant data, logs, or model settings that may affect the viewing result;

g) escalation to the responsible official or independent reviewer is ensured, as well as the possibility of external appeal to the labour inspectorate or the court in case of violation of the established deadlines (SLA);

h) together with the decision, the employee is provided with a reasoned written opinion, instructions for further appeal, and if the complaint is satisfied, a decision to restore his/her rights or payments.

56.5 Rights of the employee and the candidate.

Each person (employee or candidate) has the right to:

1. Be informed, in advance, understandably and in writing (including electronically) about the use of AI systems in the processes of selection, planning, evaluation, remuneration and disciplinary procedures, as well as about their autonomy. Such notification is provided before the start of data processing or decision-making and must contain at least:

a) the purpose and legal basis for the use of the AI system;

b) a list of automated decisions that have a significant impact on the rights, obligations or position of a person (employee or candidate);

c) sources of origin, categories and volumes of data, as well as their storage periods;

d) the limits of the autonomy of the AI system, the presence of human supervision and the contact details of the responsible official;

e) key factors (signs) that can influence the decision, indicating their weight (if possible) or with a description of the logic of decision-making;

f) appeal procedures, terms of its consideration and the procedure for suspending the execution of disputed decisions;

g) links to internal model policies and registers, as well as the date and version of the model;

h) information on the assessment of algorithmic impact in the world of work (AIA Labour), the main identified risks and measures to mitigate them;

i) ensuring the language and format accessibility of information (plain language, adapted formats for persons with disabilities). The employer or the platform is obliged to notify about significant changes no later than 14 days before the date of their entry into force, except in cases of urgent corrections due to the need to ensure security.

2. Get a meaningful explanation of the logic of automated decision-making, which should include:

a) the role of the AI system and the limits of its autonomy;

b) a list of key factors (signs) indicating their relative weight, threshold values and impact on the result;

c) sources and categories of data, dates of their collection or update, legal basis for processing and retention periods;

d) model version and date, applied business rules and post-processing procedures;

e) quality metrics (accuracy, FPR/FNR), level of uncertainty or confidence, and relevant fairness metrics;

f) description of possible alternative solutions and less burdensome options for the employee;

g) contact details of the responsible official and appeal procedures;

h) providing a copy of the data used and extracts from the decision logs (within the limits of the law) in a machine-readable format;

i) in the case of restrictions related to trade secrets — providing an explanation of equivalent content sufficient for effective appeal; the period for providing information — within a reasonable time, but not more than 72 hours from the moment of the request.

3. Require a real (substantive) human review of any automated decision with a significant impact, as well as a prompt appeal involving:

a) automatic suspension of the execution of the disputed decision until the completion of its review, except for cases when it is associated with a direct threat to security or the need to comply with the requirements of the law;

b) terms of consideration: in critical cases — no more than 24 hours, in other cases — no more than three working days;

c) the right to have a representative (including a trade union representative or lawyer) and to submit additional evidence and explanations;

d) the right of access to the used data, logs and criteria or model rules — within the limits of the law;

e) obtaining a reasoned written decision with instructions for further appeal (administrative or judicial) and ensuring the restoration of rights and payments if it is satisfied;

f) prevent you from modifying or deleting relevant data, logs, or model settings while browsing.

g) fixing all procedural actions in event logs.

4. The right to access your own data, profiles, and related materials, including:

a) initial data, signs and indicators used during the assessment or screening;

b) logs (audit trail) of automated solutions and versions of applied models or rules;

c) the history of changes in productivity indicators and methods for their calculation;

d) logs of access to his data;

e) list of recipients and facts of data transfer, as well as the terms of their storage — with the provision of:

- provision of information in an understandable form and machine-readable format, free of charge, within a reasonable time, but no later than 30 days from the date of the request (shortened period — no more than 72 hours for decisions with a significant impact);

- the ability to obtain copies and extracts;

- the right to request the correction of inaccuracies, additions, restriction of processing or deletion in cases provided for by law;

- providing, in case of refusal, a reasoned explanation and specifying the procedure for appealing;

- application of restrictions or revisions for reasons of trade secret only to the extent necessary for its protection, which cannot prevent effective appeal.

5. The right to opt out of biometric or emotional monitoring , including:

a) recognition of face, fingerprints, iris, voice, gait or posture, facial expressions and analysis of emotions or stress levels - without any negative consequences for individuals;

b) exceptions are only possible in cases where it is expressly established by law and proven to be strictly necessary and proportionate for occupational safety or health, with the conduct of an algorithmic impact assessment in the world of work (AIA Labour) and an impact assessment on data protection, as well as with the implementation of non-biometric alternatives (badges, PINs, tokens) by default;

c) covert or continuous biometric or emotional monitoring, as well as its use outside working hours or in recreation areas or sanitary facilities, is prohibited;

d) Any processing of biometric data requires a separate clear legal basis, minimization of the volume and terms, local processing (where possible), logging of accesses, prior and visible notification of the person and provision of understandable information.

6. Not to be subjected to any form of repression for exercising their rights, filing complaints, participating in trade union activities or collective actions, as well as for cooperating with supervisory authorities — in particular:

a) any direct or indirect negative actions, including fully or partially automated, in particular: dismissal, suspension, deactivation or "freezing" of the account, downgrading or access to changes or

orders, deterioration of the schedule or working conditions, refusal to promote or train, deprivation of bonuses or tips, application of algorithmic penalties or adjustments, inclusion in the "black lists", as well as any form of pressure, mobbing or excessive supervision;

b) a presumption of repression is established in respect of any such actions committed within six months from the filing of a complaint or appeal or from participation in collective actions;

c) the burden of proving the absence of a causal link rests with the employer or the platform;

d) the confidentiality of the person who filed the complaint, as well as witnesses and representatives, is guaranteed, and their tracking for the purpose of identification is prohibited;

e) In the event of a violation, immediate restoration of the situation (reinstatement), compensation for material and non-property damage with the accrual of interest and the application of administrative liability measures are guaranteed, and in case of systemic violations, the suspension of the operation of the relevant algorithmic mechanisms.

56.6 Transparency and Messaging.

1. Before the start of the use of AI systems, the employer or platform is obliged to provide employees and their representatives with a written or electronic notification no later than fourteen calendar days before the launch (except for urgent security-related fixes). The notice must be in plain language, also accessible to persons with disabilities, and contain at least such information:

a) purpose of application and legal basis;

b) a list of processes and categories of automated solutions, as well as the limits of the system's autonomy;

c) sources, categories and volumes of data, as well as how to obtain them;

d) storage periods, place and conditions of data processing or transfer, including cross-border ones;

e) Model or policy version and date , quality and fairness metrics (if available), and upgrade plans

f) contact details of the responsible official (Human in Charge) and channels of appeal;

g) appeal procedures, terms of its consideration and the rule for suspending the execution of disputed decisions;

h) a link to AIA Labour's consolidated summary, internal policies and the register of models;

i) description of cybersecurity, data protection measures and their minimization;

j) terms of use of BYOD and geolocation (if any), as well as types of monitoring;

k) The procedure and frequency of updating the message.

The notification is brought to the attention of employees individually (through a personal account, e-mail or paper delivery) with the obligatory recording of the fact and time of familiarization; for candidates — before submitting data or participating in the assessment.

2. Information on the use of AI systems is necessarily displayed in the:

a) internal policies (code of conduct, privacy policy, algorithmic labor management policy);

b) collective agreements and local regulations;

c) job postings and job descriptions.

These documents must contain at least:

purpose and legal basis for the use of AI systems;

- list of processes and decisions with a significant impact;
- types and sources of data, as well as their storage periods;
- limits of autonomy of AI systems and the procedure for human supervision;
- link to AIA Labour's summary summary and data protection policy.

56.7 Features of platform (gig-) work.

1. No deactivation, downgrading or prioritization, "freezing" access to orders, changing tariffs or other restrictions on access to the platform is allowed automatically and is possible only if:

a) implementation of a real (substantive) human review with the obligatory signature of the responsible official;

b) a preliminary written or electronic motivated notification of the employee at least 24 hours before the application of the measure, indicating the grounds, evidence, applied rules or criteria and the procedure for appeal, except for urgent cases related to safety or compliance with the requirements of the law;

c) providing the employee with a real opportunity to submit explanations and additional evidence;

d) suspension of the execution of the disputed decision until the completion of its review, except in cases of direct threat to security or violation of the law;

e) ensuring full fixation of measures in event logs in compliance with the principle of proportionality: first of all, less burdensome alternatives (temporary, partial or geofenced restriction) are applied, while access to the history, balance and withdrawal of accrued funds cannot be blocked. A temporary "freeze" for a preliminary check of the incident is allowed for a period of for no more than 24 hours; in case of groundlessness, it is subject to immediate cancellation with the restoration of access, compensation for downtime or lost bonuses, and cleaning the profile from negative marks.

CHAPTER 57. THE PRINCIPLE OF DIGITAL SELF-PRESERVATION OF A PERSON

The principle of digital self-preservation of a person is established: any interaction with artificial intelligence systems must be carried out with unconditional respect for the psycho-emotional, physical, informational, biometric and existential integrity of the person; algorithmic decisions cannot create a disproportionate risk of psychological harm, the formation of dependence, the humiliation of human dignity or bodily or neural intervention without the existence of a legal basis and adequate safeguards.

It is prohibited to create, deploy or operate systems that simulate death, loss, catastrophe or other traumatic events for the purpose of emotional impact, manipulation of will or monetization of suffering; use of technologies capable of interfering with the human body, brain or nervous system without medical indication, legitimate purpose, documented necessity and proportionality and without proper informed consent; and the imposition of a human-machine symbiosis without a clearly guaranteed right of withdrawal and without negative consequences for access to services, work or education.

Suppliers and operators implement a safe design and ensure safe operation, which includes: warning of potential psychological triggers, setting age restrictions and parental controls; the application of filters of cultural and traumatic sensitivity in generative models; the limitation of the duration of sessions and the intensity of stimuli in virtual, augmented or audiovisual environments (VR, AR, etc.); "safe mode" without touch overload; the possibility of immediately terminating the system ("kill switch") at the request of the user; prohibiting hidden emotional enhancers, subliminal stimuli and manipulative interface practices ("dark patterns").

The use of biometric, physiological and neural data is allowed only in the minimum necessary volume, with local or maximally isolated processing, cryptographic protection and access delimitation; their commercialization, profiling and secondary incompatible use are prohibited. Neurotechnologies and brain-machine interfaces are used exclusively for medical reasons, under the supervision of an authorized medical professional and an ethics committee, with safety protocols, logging of events and ensuring the possibility of a complete shutdown without harm to health.

The person is guaranteed the right to refuse any form of symbiosis, biometric or neural monitoring, intense or traumatic simulations without any negative consequences; the right to access, explain, correct and delete their own biometric and neural data; the right to immediately terminate the session, switch to "safe mode" and receive an alternative, non-traumatic channel of interaction.

Before deploying high-risk solutions, a Human Well-Being Impact Assessment is carried out, including testing for psychological triggers, risk of addiction, emotional manipulation and sensory overload, with incident response protocols and mandatory staff training; the results are documented, and key findings are communicated to users in a understandable manner Language. Detected incidents are subject to immediate termination of impact, notification of the supervisory authority and affected persons, elimination of the causes and provision of adequate support.

In the event of a violation, orders to suspend or withdraw the system, remove the traumatic content, prohibit its reposting, fines, compensation for material and moral damage, as well as other measures established by law, are applied; any contracts or "consents" that derogate from the minimum guarantees of this section are null and void. The norms are interpreted in favour of preserving the psycho-emotional and existential integrity of a person, and in In case of doubt, the rights and safety of the person shall prevail.

57.1 A person has an inalienable right to preserve his psycho-emotional, physical, informational, biometric and existential integrity in the process of interaction with AI systems.

57.2 Prohibits:

creation of AI systems capable of simulating death, loss or catastrophe for the purpose of emotional impact;

the use of technologies that interfere with the human body, brain or nervous system without medical indications;

transition to symbiosis with AI systems without a clear consolidation of the human right to refuse.

CHAPTER 58. THE PRINCIPLE OF THE PROHIBITION OF AUTONOMOUS LETHAL WEAPONS

The State recognizes as unacceptable from the point of view of ethics, law, as well as national and international security any use of artificial intelligence in systems capable of independently, without immediate and effective human intervention, deciding on the use of lethal force. It is prohibited to develop, use, test, finance, export, transit, store or any other activity related to autonomous lethal weapons on the territory of the State — regardless of the field (military, law enforcement, security, cybernetic) or the subject of implementation (public, private or foreign).

Autonomous lethal weapons should be understood as systems that, using artificial intelligence components, are able to independently carry out a full cycle of actions: detection, classification, target selection, situational analysis, decision-making on damage and physical use of force. Systems operating in human-out-of-the-loop mode or providing only a formal interrupt option are subject to prohibition as incompatible with the principles of international humanitarian law and human responsibility.

The legal basis of the ban is based on the principles of distinction, proportionality and humanity; an ethical imperative that makes it impossible to delegate the right to take the life of an algorithm; as well as on the strategic risks of escalation, manipulation, destabilization of the global security order and destruction of the foundations of international humanitarian law.

The use of separate systems that combine artificial intelligence and combat functions is allowed only if the:

ensuring real and permanent human control (human-in-the-loop) capable of interrupting or cancelling a decision before its execution;

certification for compliance with the principles and norms of international humanitarian law;

conducting an independent examination with the participation of human rights defenders, ethicists, international law and representatives of civil society, with the open publication of conclusions or their summaries in case of partial secrecy.

***The authorized state body** is obliged to maintain the National Register of Dual-Use Technologies That Can Be Used in the Development of Autonomous Lethal Weapons. All entities conducting research in the field of artificial intelligence are required to declare the existence of dual-use risks, and the relevant developments are subject to approval by **the competent authority for prohibited applications of artificial intelligence**. The decisions of this body are binding and subject to publication.*

Violation of this article entails:

criminal liability in the form of imprisonment for up to fifteen years with confiscation of property;

revocation of licenses and a ban on professional activities in the field of artificial intelligence;

blocking of assets and inclusion in sanctions lists;

initiating appeals to international partners to impose global sanctions, including restrictions on patent protection, export controls, and exclusion from alliances of technological cooperation.

The principle of the prohibition of autonomous lethal weapons is a component of the digital security strategy of the State and forms the foundation of national identity as a state that integrates digital ethics into international legal practice.

58.1 On the territory of the State, it is strictly prohibited to develop, test, use, transit, store, transfer, finance or export any artificial intelligence systems capable of autonomously, without direct, immediate and permanent human intervention, deciding on the use of lethal force, hitting targets or depriving of life.

The ban applies to both systems that function fully autonomously (full autonomy) and systems that involve minimal or delayed human intervention (partial human-on-the-loop), if such intervention does not provide a real-time opportunity to prevent the execution of a lethal solution.

This rule covers all areas of application — military, law enforcement, border, private security and cyber — and applies regardless of the subject of development or operation: a public body, a private company, a foreign entity or an autonomous system.

58.2 Autonomous lethal weapons should be understood as technical systems equipped with artificial intelligence components capable of independently — that is, without direct human control or approval at the time of decision-making — to carry out a full cycle of combat activity. This includes:

identification, classification and selection of objects or persons as potential targets;

analysis of the situational context using sensor data, geospatial models or behaviour analytics;

deciding on the use of lethal force or other forms of destruction;

performing a physical action aimed at hitting an object.

Such systems do not provide for mandatory human intervention in real-time (human-out-of-the-loop) or contain only a formal control option that does not provide the ability to effectively stop, change or challenge an algorithmic decision until the moment of actual damage. Such an architecture is incompatible with legal and ethical standards of the use of force and poses a fundamental threat to the humanitarian balance in the conditions of armed conflicts.

58.3 The principle of the prohibition of autonomous lethal weapons is based on a combination of legal, ethical and strategic-security arguments that reflect the national interests of the State, universal human rights standards and international obligations in the field of humanitarian law:

the rule of law, which requires that any use of force be consistent with international humanitarian law, in particular the principles of distinction (distinction between civilians and combatants), proportionality (prohibition of excessive harm) and humanity (prohibition of cruel or inhuman means of warfare);

an ethical imperative that makes it impossible to delegate the right to deprivation of life or bodily harm to an inanimate, non-reprehensible and irresponsible algorithmic system that does not have moral subjectivity, the ability to empathize, legal responsibility or conscious assessment of the context;

a strategic rationale that includes warnings about the high probability of uncontrolled escalation of conflicts due to failures, manipulations or cyberattacks on autonomous systems; the destruction of the global trust regime; the transformation of armed conflicts into decentralized technological disasters without human responsibility; as well as the erosion of the rules of engagement, which are historically based on humanitarian principles.

58.4 Any use of systems that combine autonomous AI functions with combat platforms — including land, air, sea, space, or cyber-physical systems — is only permitted if the following mandatory conditions are met:

- ensuring real, continuous and effective human control (human-in-the-loop) over each stage of the decision-making process on the use of lethal force with the possibility of completely cancelling or changing such a decision before the action is performed;
- Specialized certification, which includes legal, humanitarian and ethical assessment of compliance with the norms of international humanitarian law, considering the risks of acting in conditions of uncertainty, limited visibility or complex civilian context;
- conducting an independent interdisciplinary examination with the participation of representatives of human rights organizations, public observers, ethics, technology and international law specialists, the results of which are subject to full public disclosure in the public domain, including conclusions, reservations and recommendations.

Any restrictions on the publicity of such materials are allowed only in cases expressly provided for by law, provided that there is a threat to state security, and must be accompanied by a public summary explaining the grounds for restricting access.

58.5 The authorized state body in the field of artificial intelligence is obliged to establish, ensure continuous updating and public maintenance of the National Register of Dual-Use Technologies. Such technologies include software, hardware and cyber-physical components that can be used for both civilian and military purposes, in particular during the development, training or operation of autonomous weapons systems.

The register must contain:

- description of the technology, its purpose and potential risks of dual use;
- information about the developer or owner of the technology;
- assessment of the possibility of its use in prohibited or uncontrolled scenarios;
- information about the stages of approval, approval, restriction or prohibition.

All entities, both public and private, engaged in the development or use of artificial intelligence technologies are required to notify the authorized supervisory authority of the presence of dual-use risks at the stages of research, prototyping, or testing.

Such developments are subject to approval by an independent competent authority (Commission) on prohibited uses of AI, which includes representatives of the scientific community, human rights organizations, military lawyers, specialists in international humanitarian law, technology ethics and civil society. This body is empowered to make decisions and provide conclusions on the admissibility, restriction or prohibition of further development of the relevant technology.

Decisions of the independent competent authority (Commission) on prohibited applications of artificial intelligence are binding, and its conclusions are subject to publication in the public register, considering the requirements of state security. All developments that fall under the criteria of dual use and can be used in potentially autonomous weapon systems are required to be approved body (Commission).

58.6 Any violation of the requirements of this article — including the development, testing, transit, sale, export, use or financing of autonomous lethal weapons — shall be recognized as a gross violation of public order, international humanitarian law, as well as national legislation in the field of security, human rights and regulation of artificial intelligence.

Such actions entail criminal liability, which includes:

− imprisonment for up to fifteen years with confiscation of property or a lifetime ban on participation in any activity related to the development of artificial intelligence technologies;

− revocation of licenses, permits and certificates of conformity;

− confiscation of assets related to the violation or obtained because of high-risk actions;

− inclusion of persons or entities in a special sanction register with restrictions on funding, access to government orders, international grants and technological infrastructure.

In addition, the authorized state body in the field of artificial intelligence has the right to apply to international partners with proposals for the introduction of coordination sanctions, including blocking patent protection, introducing export restrictions, exclusion from alliances of technological cooperation, and inclusion in export control lists (so-called "black lists").

CHAPTER 59. THE PRINCIPLE OF RESPECT FOR EDUCATIONAL SUBJECTIVITY

A public law regime for the use of AI systems in the field of education is established. Algorithmic tools are allowed to be used solely for the purpose of expanding pedagogical interaction, providing individual support to students and didactic support, but they cannot replace the activities of a teacher or teacher. All decisions that have a significant impact on the educational process or on the implementation of the rights of the student are made exclusively by a person with mandatory observance of the principles of human dignity, non-discrimination and academic integrity.

Algorithmic recommendations in the field of education are of an advisory nature only; automatic assignment of final grades, formation of academic warnings, decision-making on expulsion or disciplinary recommendations is prohibited; any content, assessment or hint generated by the AI system, are subject to mandatory marking with the mark "AI" for the teacher and the student and entering into the use log indicating the identifier and version of the model, date, time of the call, task and result; in case of discrepancies, the teacher's reasoned decision has priority; algorithmic assessment is allowed only under approved headings; assessment of "morality", "values" or other sensitive features of the student is prohibited; the applicant's right to an alternative format of interaction without the use of AI systems ("refusal without harm") is ensured; the state guarantees a regular increase in the level of AI literacy of teachers.

Social scoring of students, as well as the use of prohibited or proxy characteristics (race, ethnic origin, religion, health status, disability, socioeconomic status and other similar characteristics) is prohibited; any algorithmic ranking is subject to mandatory calibration by groups and periodic checks for non-discrimination with defining and documenting thresholds and correction plans; aggregate "rating indices" cannot be used as the only basis for allocating resources without taking into account the context and without human participation; Systems should provide local explainability of results, transparent methodology, detection and neutralization of proxy features, as well as provide for a simplified appeal procedure with automatic suspension of negative consequences until its consideration is completed.

The formation of irreversible ("fatal") educational processes that deprive the student of the right to change the educational trajectory is prohibited. It is not allowed to automatically assign educational tracks or streams (streaming), assign irreversible labels or create a "one-way door" for the student. Any algorithmic trajectory advice is provided by local explainability of the result and is accompanied by the proposal of at least two realistic alternatives. Regular review of such advice based on the results of the applicant's new educational achievements is ensured and the right to a "second chance" is guaranteed. In case of a shortage of places, the educational institution is obliged to publish clear, objective and non-discriminatory selection criteria. For underage students, increased guarantees of protection against algorithmic fixation and discriminatory influence are established.

The use of Emotion AI and invasive remote proctoring technologies in education is prohibited, except in narrow cases expressly defined by law and confirmed by a documented assessment of proportionality and the absence of less invasive alternatives. It is prohibited to carry out constant observation, collection of biometric data, mass (1: N) biometric identification and recording of the student's environment. Storage of video recordings and usage logs is allowed only for the minimum required period. All automatically generated "alarms" are subject to mandatory human review. Suppliers and operators are prohibited from making any secondary use of the collected materials.

The right of the student to refuse to use AI systems is guaranteed without any negative consequences for assessment or access to educational services. The educational institution is obliged to provide an equivalent alternative within a reasonable time. An exception is allowed only for procedures related to ensuring academic integrity or safety, provided that alternative solutions with a comparable level of control are available. All cases of refusal and alternatives provided are subject to record. A quick appeal procedure with a suspensive effect is provided.

AI systems that affect assessment, determination of educational trajectories or access to educational services are subject to preliminary ethical certification and must be developed considering the age characteristics of students. For such systems, it is mandatory to maintain a "model passport" and a "data set passport". During each interaction with the AI system, the student is provided with local explainability of the result, warnings about system limitations and, where possible, counterfactual advice on actions that can improve the outcome. The lack of explainability makes the result unsuitable for use as a basis for decision-making.

A person necessarily remains in the decision-making circuit: each algorithmic recommendation is subject to confirmation or rejection by the teacher with the obligatory journal recording of motives. It is forbidden to use the modes of automatic confirmation of results ("auto-confirmation"). AI system interfaces should contain a functional "stop button", display the version and time of the last system update, as well as the level of uncertainty of the results obtained. Any administrative or other pressure on the teacher to force the acceptance of an algorithmic result is prohibited.

Inclusivity and accessibility in the use of AI systems in the field of education are mandatory. Systems must provide support for assistive technologies, alternative formats, application of universal design principles, and compatibility at least at the level of WCAG 2.2 AA standard. Annual accessibility audits are carried out, and inclusiveness requirements are included as mandatory conditions in procurement contracts and procedures. Collection of information on the special educational needs of students is allowed only to the minimum extent necessary and solely for the purpose of providing adaptations. In cases where digital adaptation is not possible, the educational institution is obliged to provide an offline alternative without any deterioration in the conditions of access to educational services.

In the application of AI systems in education, a full cycle of quality control, fairness and robustness is ensured. The systems are subject to preliminary testing on local samples with the mandatory determination of accuracy and fairness metrics. Continuous monitoring of the operation of systems and detection of data drift are provided. All system updates are carried out in a controlled manner and do not allow the introduction of hidden ("silent") changes. All detected distortions and biases in the operation of AI systems are subject to mandatory elimination in accordance with the established procedure. In case of significant degradation of quality or fairness, the system is immediately suspended, and the educational institution and the relevant supervisory authority are subject to mandatory notification within the specified period.

The interface of the learning platform using AI systems is obliged to provide: explicit labelling of AI-generated results; contextual warnings and references to the "model passport" and "dataset passport"; availability of checklists for the teacher; double confirmation for the implementation of critical actions; the ability to cancel or roll back actions; maintaining a history of actions with the possibility of export for evidence purposes; implementation of the principle of "privacy by default"; prohibition the use of "dark patterns"; functioning of the test environment ("sandbox") for training without affecting real results.

Each installation of the AI system in the field of education is subject to entry into the state register of AI educational systems. All model challenges made during the assessment or determination of educational trajectories are subject to mandatory logging and are stored for the time limits established by law.

Participants in the educational process have the right to: be informed in advance and in an understandable form about the AI system, its purpose, categories of data processed, associated risks, available alternatives and channels for submitting complaints; to refuse to use the system without any harm to their rights or access to educational services; to human review of automated decisions with the suspension of negative consequences until the completion of such review; to access their data, their copying, correction, deletion, restriction of processing and transfer (portability); to appeal against decisions made using the system, under a simplified procedure and within a specified timeframe. Any forms of persecution or restriction of the rights of participants in the educational process in connection with the exercise of these rights are prohibited.

Data management in AI systems in education is based on the principles of minimization and targeted use. The use of data for marketing purposes, commercialization or integration of third-party trackers is

prohibited. Priority is given to local data processing, pseudonymization, the use of role-based access and the establishment of limited storage periods. Data transfers are allowed only under the condition of an appropriate level of protection and under the control of the authorized authorities. The implementation of the principle of "privacy by default" is mandatory. It is prohibited to enrich data with external profiles without a separate legal basis.

The use of biometric technologies and invasive remote proctoring in the field of education is allowed only in cases expressly determined by law, if there is a written decision and the principle of proportionality is observed. The use of Emotion AI technologies, mass (1: N) biometric search and constant monitoring of students is prohibited. Only individual (1:1) identity verification is allowed, ensuring local data processing and a minimum storage period. Students are guaranteed the right to use alternative forms of identification, the right to human viewing of automatically generated signals, as well as the right to protection from the use of materials obtained in violation of the established rules. The systems are subject to mandatory independent audits of compliance with the requirements of safety, legality and non-discrimination.

Licensing of educational materials in the field of education is mandatory. The use of materials is allowed only if the origin is confirmed and the conditions of the relevant licenses are met. Priority is given to Open Educational Resources (OER). Educational institutions and AI system providers are required to provide a license compatibility matrix, attribution of authors, and labelling of changes made. Intellectual property rights to student works belong to the authors; the use of such works for model training is allowed only with the student's separate voluntary consent in compliance with the principle of "waiver without prejudice". TDM rights (text and data mining), the use of hidden trackers and the commercialization of educational data without proper legal grounds. In the field of education, there is a "takedown" mechanism for the removal of materials used in violation of licenses, as well as annual audits of compliance with licensing requirements.

59.1 AI algorithms and systems in the field of education do not have:

1. They cannot replace the role of a teacher or lecturer and are used exclusively to expand the possibilities of pedagogical interaction, individual support of students and didactic support, namely:

a) algorithmic recommendations are exclusively advisory; final assessment and other educational decisions are made by the teacher indicating their own motivation;

b) it is prohibited to automatically generate or issue final grades, academic warnings, expulsions, or recommendations for disciplinary sanctions without mandatory human review;

c) all results generated by AI systems in educational environments are subject to mandatory explicit labelling for the teacher and the student; Auto-confirm modes ("auto-confirmation") or covert substitution of content or ratings are prohibited;

d) each case of using the AI system in the assessment or formation of recommendations is subject to mandatory logging in the learning platform indicating the identifier of the model, its version and a brief description of the task; the teacher's decision to take into account or reject the result is also subject to fixation;

e) in case of discrepancy between the result formed by the AI system and the teacher's professional judgment, the teacher's decision shall prevail, provided that he/she provides a brief motivation for such a decision; Any administrative or institutional pressure in favour of an "algorithmic" outcome is prohibited;

f) algorithmic assessment is allowed only in the presence of approved rubrics and criteria and is used only as an auxiliary tool to check the consistency of results; AI systems are not empowered to create, change or approve rubrics without the prior approval of the pedagogical council;

g) it is prohibited to use AI systems to assess the personal characteristics of an educational student, his moral or ethical qualities, value attitudes, political, religious or other worldview beliefs;

h) the teacher is obliged to provide the student with an alternative format of interaction and assessment without the use of AI systems according to the principle of "refusal without harm", if this does not contradict the requirements of academic integrity; Such refusal may not entail restrictions or negative consequences for the applicant;

i) the educational institution provides regular professional development of pedagogical staff on the use of AI systems, including the limits of their application, identification of biases, ensuring explainability and transparency of results; such training is carried out at least once every twenty-four months.

2. Algorithmic ranking of pupils and students without considering the educational context, learning conditions and individual educational needs of students is not allowed. To ensure this prohibition, the following mandatory requirements are established:

a) social scoring is prohibited, as well as the use of prohibited or proxy features that may lead to discrimination (race, ethnic origin, religion, health status, disability, socio-economic status, etc.);

b) ranking metrics are subject to mandatory calibration by groups and periodic checks for non-discrimination (minimum set: Disparate Impact Ratio, Equalized Odds, Calibration by Group) with the definition of documented thresholds and a correction plan in case of deviations;

c) It is prohibited to use aggregate "rating indices" or average scores for the allocation of educational resources (scholarships, dormitories, places in courses) without proper correction for the context (socio-economic conditions, special educational needs, language environment, status of an internally displaced person, veteran or person with a disability) and without mandatory human review;

d) the results of the ranking are subject to mandatory contextualization, considering individual curricula, temporary academic breaks or crisis circumstances (illness, war, relocation), as well as conditions of access to the Internet and devices; the absence of relevant data cannot be interpreted against the student;

e) transparency of algorithmic ranking is mandatory: for each case, local explainability is provided (key factors, weights, model limitations) and a common methodology is published; the use of "black box" systems without explanation is prohibited;

f) detection of proxy features is carried out by correlation analysis and interpretation methods (in particular, SHAP, Permutation Importance, counterfactual checks) with the obligatory removal or neutralization of such variables;

g) an audit of the fairness of algorithmic ranking is carried out at least once a semester (or academic year) with the publication of aggregated, impersonal results; the identified systemic distortions are the basis for the immediate suspension of the application of the relevant ranking mechanism;

h) it is prohibited to automatically restrict the access of students to Olympiads, advanced courses, support programs, housing or social benefits solely based on ranking results without a decision of the pedagogical or student commission;

i) a simplified procedure for individual appeal of the results of algorithmic ranking is established; At the time of the appellate review, the negative consequences of such ranking are suspended, if it does not contradict the requirements of academic integrity.

3. The formation of irreversible educational trajectories based on erroneous or biased algorithmic models is not allowed. To ensure compliance with this prohibition, the following mandatory requirements are established:

a) it is prohibited to automatically record (tracking/streaming) the educational trajectory, which restricts access to subjects, courses, scholarships or support programs without the possibility of human review and appeal;

b) any trajectory recommendations are accompanied by local explainability and at least two realistic alternatives (including those with flexible rates and additional modules);

c) it is prohibited to use irreversible labels ("weak", "risky", "unpromising") and "one-way doors"; The educational trajectory is subject to mandatory revision at least once a semester or academic year based on the results of new achievements;

d) the "right to a second chance" is guaranteed: after targeted remediation or retaking tests, the applicant has the right to a second competition or access to a higher trajectory;

e) it is prohibited to use prohibited or proxy features, socio-economic or geographical origin, as well as the school's "rating" index as a basis for restricting access; fairness checks (disparate impact, equalized odds, calibration by group) are mandatory;

f) trajectory decisions cannot be based on resource quotas or "grey" rules; in case of a shortage of places, the educational institution publishes transparent, non-discriminatory selection criteria;

g) For underage students, increased guarantees are provided: written informing of legal representatives, coordination with the pedagogical council, the mandatory possibility of transition based on the results of success;

h) Each recommendation should contain a counterfactual explanation of "what to change to gain access" (clear thresholds, required skills, additional courses);

i) the decision to determine the educational trajectory is recorded in writing, indicating the motives of the teacher or the commission; a journal of decisions with impersonal statistics is kept monitoring non-discrimination;

j) a simplified appeal procedure with a consideration period of no more than ten working days is provided; for the duration of the appeal, the applicant is granted temporary access to the selected modules, if this does not pose a threat to academic integrity.

4. The use of emotion recognition and invasive forms of remote proctoring is prohibited, except for cases expressly determined by law. In such cases, their application is allowed only based on a written decision of the management body of the educational institution, adopted based on the results of a documented proportionality assessment and analysis of available less invasive alternatives.

To comply with this prohibition, the following mandatory requirements are determined:

a) it is prohibited to use "Emotion AI" (determination of emotions, motivation or personality traits) for evaluation, selection, disciplinary decisions or determination of access to educational services;

b) Invasive proctoring tools (continuous video surveillance, room scanning, eye or face tracking, ambient audio recording) are allowed only if data is minimized and an equivalent offline or face-to-face alternative is available ("opt-out without harm");

c) It is prohibited to collect and process biometric parameters (facial expressions, heart rate, eye movements, etc.) without a separate legal basis;

d) the storage periods of video materials and logs may not exceed 30 days, unless otherwise expressly provided by law; reuse for non-academic purposes is prohibited;

e) it is not allowed to apply automatic conclusions about "fraud" or "violation of the rules" without human review and a reasoned decision;

f) About minors, increased guarantees are applied: mandatory informing of legal representatives, restriction of technical means and priority of face-to-face forms of assessment;

g) suppliers and contractors are prohibited from reusing records or metadata; violation entails liability determined by law.

5. It is prohibited to restrict access to education in connection with the student's refusal to use artificial intelligence systems, provided that such refusal does not violate the requirements of academic integrity and does not pose a threat to security.

The implementation of this prohibition is ensured through compliance with the following mandatory requirements:

a) the right to an alternative ("waiver without prejudice"): the institution is obliged to offer an equivalent offline or non-AI training or assessment procedure (face-to-face, oral test, supervised written work, etc.) without deterioration of conditions and deadlines;

b) prohibition of indirect sanctions: it is prohibited to lower grades, restrict access to courses, scholarships, dormitories, libraries or digital services in connection with the refusal to use AI; the requirement to "agree to the use of AI" as a condition of access to classes is prohibited;

c) Exceptions: The use of AI can only be a condition for measures expressly defined by law to ensure academic integrity or safety (e.g., plagiarism checking), provided that an alternative procedure with an equivalent level of scrutiny is in place;

d) availability of alternatives: the alternative procedure must be technologically and organizationally accessible; the terms of its completion may not exceed the standard terms by more than ten working days without separate justification;

e) logging and transparency: cases of refusal and provided alternatives are recorded in the institution's log without collecting redundant data; each semester, anonymized statistics are published;

f) it is prohibited to exert psychological or administrative pressure by teachers or the administration on the applicant in order to force the use of AI;

g) legal representatives must be informed about minors; reasonable accommodations are provided for persons with disabilities and other vulnerable groups;

h) informed consent: before using AI, the institution provides clear information about the purpose, scope of data, risks, and alternatives; consent can be withdrawn at any time without negative consequences;

i) appeal mechanism: a quick appeal (up to five working days) is established for the violation of this right with a suspensive effect for the duration of the hearing;

j) Liability: the violation entails the invalidation of the relevant results, the application of disciplinary measures to officials, and for the supplier, the termination of the contract and the imposition of fines in accordance with the institution's policy and the law.

59.2 Mandatory requirements for AI systems in the field of education.

1. Ethical certification and age-appropriate design. Any AI system that affects assessment, trajectory, or access to educational services is subject to prior mandatory ethical certification. The interfaces and content of such a system should be adapted to the age of students, as well as provide linguistic, cognitive and functional accessibility.

2. Transparency and explainability. For each system, a "model card" and a "datasheet for datasets" are published, including a public condensed profile and an extended version available in the "safe room" for the court, commissions or legal representatives of students if necessary. In each individual interaction, the following is necessarily ensured:

(i) local explainability of the result with indication of key factors of influence and limits of applicability;

(ii) warning about model constraints, known risks, and levels of uncertainty, including confidence intervals (if);

(iii) A counterfactual explanation in the form of a recommendation "What to change to improve the result". The "model passport" must contain at least the following information:

- Purpose, version, release date, and runtime;
- data used for training and validation, and their sources;
- known limitations and risks of bias; the results of the latest quality and fairness tests, including the Disparate Impact Ratio (DIR), Equalized Odds and Calibration by Group;
- prohibited application scenarios; update policy and incident reporting contacts).

3. The "Data Set Passport" must contain at least the following information: sources and periods of collection; geography and coverage of groups; structure and fields; cleaning and balancing procedures; identified imbalances and ways to eliminate or neutralize them; Transparency materials are provided in the state language in an understandable form (at the level of presentation accessible to the relevant age group) and in accessible formats, including adaptations for persons with disabilities. grounds for assessing or determining the educational trajectory.

4. Man in the circuit. The interface and processes should provide:

(i) explicit confirmation or rejection of each algorithmic recommendation by the teacher;

(ii) the ability to edit or cancel a recommendation before it can be applied;

(iii) mandatory recording in the journal of the reasons for acceptance or rejection;

(iv) Prohibition of any "auto-confirmation" or stealth application modes;

(v) available "stop button" and immediate return to manual procedure;

(vi) display of model version, last update time, and uncertainty level, including confidence intervals (if);

(vii) prevention of administrative pressure on the teacher in case of refusal of the algorithmic result.

5. Inclusiveness and accessibility. AI systems in the field of education are obliged to support adaptations for people with disabilities (alternative formats, special fonts, voice-overs and subtitles, keyboard navigation, etc.) and provide a universal design for all users.

a) technical requirements: compatibility with screen readers, the presence of ARIA markup, full keyboard navigation without navigation traps, correct display of focuses and statuses, the ability to zoom and reflow without loss of functionality, high contrast modes, interfaces safe for people with photosensitivity;

b) alternative content representations: subtitles and transcripts for audio and video materials, voiceovers, alternative texts for images and diagrams, support for easy-to-read formats and fonts, print and offline versions if necessary;

c) assessment and timing: the ability to extend timings or the number of attempts, "pause/continue" functions without loss of progress, non-discriminatory tasks; it is forbidden to require disclosure of disability to receive basic adaptations;

d) Interaction: support for alternative input devices (switches, sticks, on-screen keyboards), sign language or sign language interpretation if necessary, clear feedback on errors and hints for correcting them;

e) linguistic and cognitive accessibility: use of simple language for instructions, visual cues, avoidance of ambiguities, warning of complex actions and possible consequences;

f) processes: regular accessibility audits (at least once a year), testing with users with disabilities, accessibility statement and contact for requests or complaints, response times not exceeding ten working days;

g) procurement and contracts: requirement to confirm compliance with WCAG 2.2 AA level or equivalent, provision of a VPAT or similar declaration, obligation to remove barriers and prohibition of hidden (silent) degradation of accessibility;

h) privacy and data: the collection of information about special needs is carried out only to the extent necessary for the provision of adaptations, with a clear delimitation of access and prohibition of the secondary use of such data;

i) priority of proportional solutions: if digital adaptation is not possible, the institution provides an equivalent offline alternative ("refusal without harm") without deterioration of conditions.

6. Quality and fairness. Before deployment and during operation, the operator provides a full cycle of inspections of the quality, fairness and robustness of the AI system:

a) Pre-testing: multiple cross-validation and verification on local test samples (hold-out) representative of a particular institution or region; determination of the minimum sample sizes for each group; technical and platform tests (work on mobile devices, at low Internet speed, using assistive technologies).

b) Metrics: accuracy (MAE, MSE, or other quality profile metrics); fairness checks are mandatory — Disparate Impact Ratio (approximate operating range 0.8–1.25 with justification), Equalized Odds (TPR/FPR parity), Calibration by Group (Brier/ECE), as well as errors of the second kind for vulnerable groups; for ranking — NDCG, MAP with analysis by groups.

c) Monitoring in a productive environment: continuous monitoring of quality and fairness using control cards and alert thresholds; regular reports (at least once a quarter) with trend mapping.

d) Drift and updates: Identify covariative and conceptual drift of data and patterns; the update is carried out only according to the approved protocol without hidden ("silent") changes; After each release, retesting is mandatory.

e) Robustness and resilience: tests for manipulation and adversarial influences, including attempts to circumvent proctoring and prompt attacks on large language models (LLMs); checking immunity to skips, noise and overload; modelling of boundary scenarios.

f) Group justice: comparative analysis of results by gender, age, language of instruction, disability , socioeconomic conditions, IDP or veteran status, and other relevant characteristics; the use of prohibited and proxy features is prohibited.

g) Documentation and transparency: maintaining a "quality log" indicating versions, datasets, metrics applied, thresholds and results of recent checks; ensuring the availability of aggregated anonymized reports for the institution community.

h) Remediation: in case of distortions, they must be eliminated (rebalancing, retraining, post-processing), independent review and revalidation before restoring the system.

i) Suspensive safeguards: in the event of proven discriminatory effects or significant degradation of quality, the system is immediately suspended in the relevant processes, ensuring human review and notifying the students, their legal representatives and the supervisory authority within 72 hours.

7. Labelling and Interface Warranties.

The interface of the learning platform should provide explicit labelling of algorithmic results and provide the user with all the contextual information necessary for their conscious application:

a) a direct indication of "AI" for any generated content, rating, or recommendation;

b) Displaying the role of AI, confidence/uncertainty level (if any), date/version and time of the last update, applicability limits and known limitations;

c) contextual risk warnings and links to the "model passport" and "data set passport";

d) Checklists for the teacher on: appropriateness of application; checking input data for prohibited/proxy features; the presence of local explainability; alternatives and possible consequences; risks of bias;

e) double check for critical actions (giving a final grade, changing the educational trajectory, disciplinary recommendations) with a brief motivation of the teacher;

f) controllability: "stop button", "cancel application", "return previous version";

g) history of recommendations and user actions with the possibility of auditing and exporting in a format suitable for evidence in legal procedures;

h) setting privacy and data minimization by default;

i) ensuring compliance with accessibility requirements not lower than the level of WCAG 2.2 AA; adaptation for persons with disabilities and younger age groups of students;

j) Prohibition of manipulative interface practices "dark patterns": no intrusive notifications and forced options that make it difficult to refuse or perform human browsing;

k) displaying the session ID and ensuring a quick transition to the "quality log".

8. Register and logging. Each installation of an artificial intelligence system in the field of education is subject to entry into the state register of AI systems; all model calls (requests to the algorithm within the framework of assessment or formation of educational trajectories) are subject to logging and are stored for at least three years.

59.3 Rights of participants in the educational process.

1. Right to know.

Students and their legal representatives shall receive a written or electronic notification in advance (prior to the first application of the AI system to a specific person) containing at least the name of the system, its version and provider; the role and purpose of the application; categories and sources of data, the legal basis for their processing, retention periods and possible cross-border transfers; references to the "model passport" and "passport of datasets"; known limitations and risks, level of uncertainty (if any); AI-free alternatives are available; rights to opt-out, human review, appeal, and rectification of data; contact of the person in charge or channel for filing complaints.

Form and accessibility. The notification is provided in the state language and (if available) in the language of instruction in a form understandable for the relevant age group, with accessible formats and adaptations, including for persons with disabilities.

Fixation. The fact and time of providing the notification, as well as the chosen alternative/consent/refusal, are subject to recording in the journal of the institution without collecting redundant data.

Update. In case of a change in the purpose, categories of data or version of the model of the "major/minor" level, the operator is obliged to re-notify before the application of the updated system to a specific person.

Consequences of not notifying. Until a proper notification is provided, the results of AI cannot be used for final assessment or determination of the educational trajectory; Results already used are subject to mandatory human review and may be invalidated.

Minors. The notification is additionally sent to legal representatives; The student is provided with an age-appropriate explanation in accessible simple language.

Communication channels. The notification shall indicate the methods of application (e-mail, personal account, "hotline") and the response time, which may not exceed ten working days.

2. Right of withdrawal without prejudice.

Refusal to use AI systems (except for measures to ensure academic integrity expressly defined by law) cannot be grounds for lowering grades or restricting access to educational services; the institution is obliged to offer an alternative procedure.

Equivalent alternative. The institution provides an accessible format of training or assessment (face-to-face, oral test, supervised written work, or other proportionate and non-discriminatory option) without deterioration of conditions and deadlines.

Prohibition of indirect sanctions. Refusal to use AI cannot be the basis for lowering grades or restricting access to courses, scholarships, dormitories, libraries, digital services; the requirement to agree to the use of AI as a condition of attending classes is prohibited.

Exceptions. The use of AI as a condition is only allowed for procedures expressly defined by law to ensure academic integrity or security (e.g., plagiarism checking) if there is an equivalent alternative procedure with a comparable level of control.

Availability and deadlines. The alternative procedure must be technologically and organizationally accessible; the terms of its completion may not exceed the standard ones by more than ten working days without separate justification.

Logging and transparency. All cases of refusal and provided alternatives are subject to recording in the institution's log without collecting redundant data; Impersonal statistics are published every semester.

Prohibition of pressure. Teachers and the administration are prohibited from exerting psychological or administrative pressure on the student to force him to use AI.

Minors and vulnerable groups. In the case of minors, legal representatives are notified; Reasonable accommodations shall be provided for persons with disabilities and other vulnerable groups.

Informed consent. Before using AI, the institution provides clear information about the purpose, scope of data, risks, and alternatives; The consent can be withdrawn at any time without negative consequences.

Expedited appeal. An expedited appeal procedure is established (up to five working days) for violation of this right with a suspensive (suspensive) effect for the duration of its consideration.

Consequences of the violation. Violation of this right entails the recognition of the relevant results as invalid, the application of disciplinary measures to officials, and for the supplier, the termination of the contract and the imposition of fines in accordance with the procedure established by law.

3. The right to human review.

Any algorithmic assessment, recommendation or educational trajectory is subject to human review by a teacher, examiner or appropriate commission at the request of the student; the results of such a review are drawn up in writing indicating a clear and exhaustive motivation.

Scope and grounds. An applicant for education can initiate a review at any time after receiving the result, but no later than within ten working days; for decisions with significant consequences (expulsion, loss of scholarship, non-admission to a course or exam) — within forty-eight hours.

Suspensive effect. For the duration of the review, the negative effects of the algorithmic result are suspended; if necessary, a temporary neutral assessment or other proportionate alternative may be applied, if it does not threaten academic integrity.

Composition and independence. The revision is carried out by the appropriate teacher; at the request of the applicant — a commission consisting of at least three people (a teacher, a representative of the quality assurance department or a methodologist, a representative of student or parent self-government with consent) without a conflict of interest.

Materials for viewing. The student is provided with a local explanation of the algorithmic result, key influencing factors, the limits of the model application, a copy of the call log record indicating the identifier, version and time, as well as counterfactual recommendations "what to change to improve the result".

Terms. Regular reviews are made for no more than five business days; urgent — within forty-eight hours; The extension of the term is possible for no more than five working days, subject to the provision of written justification.

Acceptability standard. If there is no local explanation or it is impossible to reproduce the result from the log, the algorithmic conclusion cannot be used as a basis for the decision, loses its evidentiary value or is invalidated.

Decision based on results. Based on the results of the review, a decision is made to accept, reject or adjust the algorithmic conclusion; the decision is drawn up in writing, contains motives, indication of the factors considered and rejected, proposed alternatives and terms of their implementation; The corresponding entry is made in the quality log.

Further appeal. An applicant for education has the right to submit an appeal to the higher commission or the governing body of the institution within ten working days; the term of its consideration is no more than ten working days; The suspensive (suspending) effect persists for the entire period of consideration.

Minors and vulnerable groups. The legal representatives of the minors are notified of the review; reasonable accommodations and, if necessary, the participation of a tutor or psychologist are provided; Explanations are provided in accessible simple language.

Prohibition of repression. The exercise of the right to review cannot be the basis for deterioration of the legal or factual situation of the student or the application of any other negative consequences to him/her.

4. Right to access and rectify data.

Every student (and for minors — their legal representatives) has the right to:

to access their educational profile and all related records, including categories of data, purposes, legal basis for their processing, sources of receipt, recipients or third parties, retention periods, fact of cross-border transfers, availability of profiling and its logic;

Get a copy of the data in an understandable, structured and commonly used machine-readable format; the first copy is provided free of charge, the subsequent ones are provided at the cost of preparation;

demand immediate correction of inaccurate data and supplementation of incomplete ones; if the assessment or decision was based on erroneous data, a mandatory human review and correction of the result or assessment is carried out;

request the deletion of redundant, outdated or illegally collected data (right to erasure), except in cases expressly provided for by law (archiving in the public interest, fulfilment of a legal obligation);

request a temporary restriction of processing (blocking) for the period of verification of accuracy, legality or in case of submission of an objection to processing; at this time, the data cannot be used to assess or determine the educational trajectory;

a) object to any secondary use of data outside of educational purposes, including for marketing, commercial analytics or transfer to third parties;

exercise the right to portability: receive or transmit your data to another institution or provider in a machine-readable format (together with a key activity log) within ten working days;

access derived algorithmic inferences and local explanation: key factors that influenced the assessment, recommendation or trajectory; the limits of the applicability of the model and the level of uncertainty (if any);

get an extract from the call log of the model regarding yourself (ID, version, date and time, task description, brief description of input data, result, human decision);

use a clear procedure for submitting requests or corrections (online or offline); the identification of the applicant is carried out according to the principle of data minimization without excessive requirements;

receive a response without undue delay, but not later than within ten working days; in exceptional cases, the term may be extended by no more than five working days with a written justification; refusal is allowed only with a reasoned explanation and reference to the norm of the law;

be informed of the correction or deletion: the operator notifies all known recipients of the data of the changes made, unless this is impossible or requires disproportionate effort;

to receive free exercise of their rights; the fee can be charged only in case of manifestly groundless or excessive (repeated) requests — within the limits of actual costs;

for minors and vulnerable groups — to have the right to access in an age-appropriate form, with the involvement of legal representatives and the provision of reasonable accommodations;

regarding exam materials: get access to your own results, used criteria or rubrics and technical assessment logs without disclosing the bank of future tasks;

in case of violation of the deadlines for fulfilling the request — file an internal appeal to the data controller and apply to the authorized data protection body or to the court; The results obtained using the disputed data are subject to mandatory human review and, if necessary, can be invalidated.

5. Right to appeal.

A simplified and accessible procedure for appealing an algorithmic assessment, recommendation or educational trajectory is established with clearly defined deadlines, verification standards and safeguards.

Deadline for submission. An appeal can be filed within ten working days from the date of receipt of the result; for decisions with significant consequences (expulsion, loss of scholarship, non-admission to a course or exam) — within forty-eight hours.

Form and content. The appeal is submitted in writing or electronically — in any form or according to a standard template; the disputed result, brief arguments and the desired method of defence must be indicated; Additional evidence is provided if available. Artificial administrative barriers (excessive requirements for the form, signatures, etc.) are not allowed.

Access to materials. Within two working days from the date of registration of the appeal, the institution is obliged to provide the applicant with a "transparency package" for a specific result, which includes a local explanation, a log of the model call indicating the version, time and description of the task, the limits of the applicability of the model and counterfactual recommendations "what to change to improve the result".

Suspensive effect. For the duration of the appeal, the negative consequences of the algorithmic result are suspended; If necessary, a temporary neutral assessment or other proportional alternative may be applied, as long as it does not threaten academic integrity.

Reviewing body. The appeal is considered by a commission consisting of at least three people (a teacher of the relevant profile, a representative of the quality assurance department or a methodologist, a representative of student or parent self-government with consent) without a conflict of interest. Persons who participated in the development or implementation of the AI system cannot constitute most of the commission.

Verification standards. The check includes: the validity of the model for the stated task; proper quality and representativeness of data; absence of prohibited bias or proxy features; reproducibility of a specific conclusion; adherence to the principles of "man in the loop", the right of withdrawal without prejudice, transparency and deadlines.

Independent examination. In controversial cases, an independent pedagogical or algorithmic examination may be appointed with a deadline of up to ten working days; As a general rule, the costs are borne by the institution, but it is possible to distribute the costs taking into account the results and financial capabilities of the parties; For socially vulnerable persons, the costs are covered by the institution or its founder.

Term of consideration. The appeal shall be considered within no more than ten working days from the date of its registration; a one-time extension of the period for no more than five working days is possible, subject to written justification and separate notification of the applicant.

Decision. Based on the results of the appeal, one of the following decisions is made: leave the result unchanged; cancel; change; order a re-evaluation or alternative procedure; make corrections to the data; notify the supervisory authority or initiate an audit. The decision must include reasons, references to factual data and (if available) fairness and quality metrics, as well as deadlines for implementation.

Notification and execution. The decision shall be brought to the attention of the applicant and the relevant departments no later than the next business day; in case of change or cancellation of the result, immediate correction of assessments or trajectories and synchronization of information systems is carried out; The corresponding entry is made in the quality log.

Further appeal. The applicant has the right to apply to the supreme governing body of the relevant institution, its founder or the competent state authority or independent regulator in the field of education. At the same time, the suspensive (suspension) effect of the appeal remains, except in cases of obvious bad faith of the applicant. This right does not in any way limit the possibility of applying to the court or other independent body for the protection of rights.

Protection from repression and privacy. Filing an appeal cannot be the basis for a negative attitude or the application of sanctions to the student; Personal data is processed proportionately, with limited access and mandatory logging.

59.4 Data Governance, Privacy, and Security.

1. Minimization and intended use. Only those data and only to the extent that are necessary and proportionate for a specific educational purpose are processed; commercialization, profiling for marketing and transfer to third parties outside of the educational function are prohibited.

Certainty of purpose and legal basis: before the start of processing, the operator documents the specific purpose (s) and legal basis; "Compatible purposes" are allowed only after checking their compatibility and separately notifying applicants/parents.

Data categories: it is allowed to collect the smallest necessary set (grades, completed tasks, visits, feedback, technical reliability logs); It is prohibited to collect geolocation, data from social networks, behavioural trackers, biometrics and other sensitive data, unless it is expressly provided for by law and the decision of the institution with a proportionality assessment.

Secondary use: for research and statistics, only data anonymization or aggregation is allowed, excluding the possibility of re-identification; in the case of individual analytics, informed consent and decisions of the ethical/pedagogical council are mandatory.

Telemetry and identifiers: Only technically necessary logs are collected; cross-site tracking, advertising identifiers, third-party cookies (cookies) and SDKs that are not required for the educational function are prohibited.

On-device processing: local processing and short-term logs are preferred; The transmission of raw audio/video to servers is allowed only when there is a proven need and in the absence of less invasive alternatives.

Pseudonymization and depersonalization: carried out by default for testing, quality analytics and model training; Re-identification is prohibited, except in cases where access is restored at the request of the data subject.

Access and roles: the principle of least authority, role-based access, multi-factor authentication, logging of each access to personal data and regular review of access rights are applied.

Retention periods: determined for each category of data according to the principle "no longer than necessary"; upon completion — safe deletion or depersonalization with fixation in the journal; It is forbidden to create "permanent archives" without a legal basis.

Set delimitation: Operational data is separated from training and test kits; Training at the jobs of applicants without their informed consent is prohibited.

Third parties and transfers: data transfers are only allowed in the status of a processor under a contract with clear limitations of purpose, prohibition of secondary use, security requirements and jurisdiction of national courts; cross-border transfers — only to jurisdictions with an adequate level of protection and in the presence of contractual guarantees.

Community-level analytics: Reports for quality management are generated at the aggregate level (class/course/institution) without individual marketing profiling or ranking outside of the educational goal.

Human rights and data protection impact assessment: for high-risk scenarios (algorithmic assessment, trajectory, proctoring), a written impact assessment with a risk minimization plan and approval by the institution's management is mandatory.

Default settings: "privacy by default" — all optional fees and analytics are disabled by default; refusal or withdrawal of consent does not affect access to education; Simple opt-out mechanisms are available.

Quality and minimization in industry: it is forbidden to transmit sensitive data or excessive personal information to requests to AI systems; The data is cleared of proxy features and limited to the required minimum.

Prohibition of combination with external arrays: it is not allowed to enrich educational data with commercial or social profiles or public registers without a separate legal basis, notification and guarantees of non-discrimination.

2. Sensitive data and biometrics.

The use of biometric parameters, emotional recognition and invasive proctoring is allowed only in exceptional cases expressly defined by law, by a written decision of the governing body of the educational institution based on a documented assessment of proportionality and analysis of less invasive alternatives; At the same time, the following guarantees are mandatory:

absolute prohibitions: it is prohibited to use "Emotion AI" (determination of emotions/motivation/personality traits) for assessment, selection, disciplinary decisions or determination of access to educational services; 1:N identification (search in databases), social scoring based on biometrics and constant observation is prohibited;

a) permissible purposes: only 1:1 verification of the person in a high-stakes assessment or for access to the exam rooms; constant tracking of behaviour/gaze/face during training is prohibited;

minimization and local processing: priority is given to on-device processing; raw video, audio, images and full biometric templates are not stored; only cryptographic templates for 1:1 verification protected by solo hashing and separated from identification keys are allowed; it is forbidden to train models on such data;

technical security requirements: mandatory encryption of data in transit and at rest, segmentation of environments, application of the principle of least authority, multi-factor authentication, logging of each access to biometric and other sensitive data;

retention periods: biometric artifacts and logs may be stored for a maximum of 30 days or until the completion of appeals/investigations (whichever comes first) followed by automatic and verifiable deletion; reuse for non-academic purposes is prohibited;

rights of applicants: prior transparent notification, the right to an alternative without prejudice, the right to human review of any "alarm signals", access to local explanations and journals about themselves;

minors and vulnerable groups: face-to-face/offline procedures have priority; the use of biometrics is possible only if there are legal grounds and after informing or obtaining the consent of legal representatives within the law; the list of permitted means is as limited as possible;

Vendors and third parties: no reuse, commercialization, or training of proprietary or third-party models on this data is prohibited; processing and storage is only permitted in jurisdictions with an adequate level of protection; contracts must contain requirements for audit, localization of journals and jurisdiction of the courts of the State;

minimum and permitted set of proctoring: point events (screenshots, recording suspicious actions, browser blocking) are allowed, provided that there are offline alternatives; "room scan", continuous face

recognition, eye tracking, environmental recording, analysis of microexpressions, and monitoring of third-party devices in the room are prohibited;

identification: only 1:1 verification with the possibility of an alternative procedure without prejudice is allowed; the use of any databases to search for a person on the 1:N principle is prohibited;

audit and incident reporting: annual independent reviews of biometrics and proctoring practices; incidents are reported no later than 72 hours later with a corrective action plan; the results of audits are provided in an impersonal form to the community of the institution;

consequences of violations: materials obtained in violation of these requirements are recognized as inadmissible; The responsible person and the operator are subject to disciplinary and contractual liability, the system is subject to immediate suspension until the violations are eliminated.

59.5 Licensing of educational materials. All training datasets, texts, images, audio/video, tests, code, and other content used or integrated into AI systems are used exclusively in compliance with copyright, related rights, and the terms of the respective licenses:

Sources and legality of acquisition: each material has a documented origin (provenance) and a valid legal basis for use (license, permission of the copyright holder, free license, public domain, legal exception for education). Unauthorized "scraping" or mass copying without a legal basis is prohibited.

OER selection and priority: Open educational resources (OER) with free licenses (e.g. CC BY, CC BY-SA, CC0). It is forbidden to combine incompatible licenses (e.g. CC BY-SA with NC/ND RESTRICTIONS, GPL in code).

Compliance with the terms of the license: attribution, indication of the license and references, marking of changes; NC/ND/SA restrictions must be observed; You may not remove or obscure attribution, metadata, or watermark notices, unless otherwise expressly permitted.

Compatibility matrix: the operator maintains a license compatibility matrix for content, data and libraries, as well as a "dataset passport" indicating the rights to reuse; the results of the compatibility check are stored in the procurement and implementation dossier.

Student/pupil works: the copyright to the results of creativity belongs to their authors. The institution receives a non-exclusive, free license only for evaluation and internal quality assurance. Training or fine-tuning of models on student work is possible only with informed, voluntary, separate consent; refusal does not affect evaluation or access ("refusal without prejudice"). The retention period and the right to withdraw consent are proven in writing.

Third parties and suppliers: Contracts must contain warranties for the legality of the content, the absence of hidden restrictions, the prohibition of reuse, and the prohibition of training of its own or third-party models on the materials of the institution without a separate legal basis. There is a right to audit and a "takedown" mechanism at the request of the copyright holder within 72 hours.

Exceptions for citation/teaching: citations and illustrations are allowed for teaching within the limits and in the manner expressly permitted by law, to the extent necessary with mandatory attribution and without prejudice to the normal use of the work; These exceptions do not apply to model training, unless otherwise expressly provided by law.

Accessibility and adaptations: technically necessary adaptations (subtitles, transcripts, alternative formats) for persons with disabilities are allowed, if this does not violate protected technological means and meets the requirements of the law; in the case of DRM, the legal mechanism of permitted circumvention applies if it is expressly provided for by law.

Models and training datasets: Only sets with rights that explicitly allow machine learning and text, and data mining (TDM) are allowed for training or fine-tuning. For each dataset, a datasheet is drawn up with the field "legal status/TDM license". Before deployment, both automated and manual license screening (license scanning) is carried out.

Outputs: it is prohibited to intentionally reproduce protected works or their essential parts that go beyond the permitted citations or exceptions; in case of suspicion of "regurgitation", a check is carried out and the relevant source or request is blocked; materials created by AI systems must provide attribution of the tool and warnings about possible restrictions on the rights of third parties.

Metadata and Accounting: All training materials integrated into the systems are labelled with rights metadata (license, owner, URL, date of acquisition of rights, restrictions). This metadata is saved at all stages of processing.

Term and territory: the use of materials corresponds to the terms and territorial restrictions of the license; At the end of the term, the use is terminated, or a new license is issued.

Anonymization and personal data in content: before publishing or using materials containing personal data in training, they are anonymized, or lawful consent is obtained; Priority is given to minimizing personal data in training samples.

Public information: the institution publishes a list of the main sources and licenses (without disclosure of protected secrets), as well as a policy for working with intellectual property rights; Applicants are explained the rules for using the materials created by them.

Reaction to claims: in case of an appeal from the copyright holder, the operator immediately suspends access to the disputed material, conducts an initial check (up to 5 working days), informs the applicant about the result, removes or replaces the content, or resumes access with a reasoned justification; the case is recorded in the journal and taken into account when updating the checklists of rights verification.

Attribution in Learning Environments: The platform's interface displays source attribution and license type next to the material; When printing or exporting, license metadata is stored.

Prohibition of hidden conditions: materials with hidden trackers or conditions that allow the exploitation of personal data of applicants for non-educational purposes are not allowed.

Coordination with procurement: tender documents and contracts must contain requirements for the legal purity of content, compatibility of licenses, transfer of necessary rights to the institution, and provide for indemnity in case of claims by third parties.

Periodic audit: at least once a year, a random audit of license compliance is carried out; The results and corrective actions taken are documented, and the consolidated conclusions are made public to the community of the institution.

CHAPTER 60. THE PRINCIPLE OF DIGITAL SAFETY OF THE CHILD

All artificial intelligence systems operating in the territory of the State and directly or indirectly interacting with minors, processing children's data or influencing decision-making regarding them, are subject to mandatory compliance with the principle of digital safety of the child. This principle means that the architecture of the system, algorithmic logic and mechanics of interaction, the limits of functionality and categories of data processed, data sources and cycles, model training processes and purpose of use, as well as the risks and potential consequences of application must be secure by default, understandable, verifiable, and subject to control and explanation at each stage of the life cycle.

A child's digital safety must be ensured for all stakeholders:

for developers and operators — due to the availability of internal technical and legal documentation, data maps, training and model update protocols, change logs, Child Impact Assessments, Child Safety Data Sheet, and verification and audit reports;

for children and their legal representatives — through accessible interfaces and materials explaining the logic of the system, its purpose, risks, and security settings (parental controls, refusal to personalize, time limits), taking into account age, level of digital literacy and socio-cultural context;

for regulators and supervisory authorities — through the provision of a complete technical, methodological, legal and ethical dossier, including architectural schemes, training protocols, threat models and risk analysis, event logs, results of internal and external audits of child safety;

for third parties affected by the decisions of the system (teachers, doctors, guardians, content providers) — through the creation of mechanisms for access to public explanations, appeal, appeal procedures, and independent monitoring.

Ensuring the digital safety of a child includes the mandatory disclosure of at least the following information: sources and categories of data; Legal status and contact details of the developer or operator; the intended purpose of the system and the contexts of its use; functional limits and known reliability limitations; the degree, nature and method of participation of the AI system in decision-making about the child; availability and parameters of parental control; retention and deletion terms and policies; the level of risk and the results of the impact assessment on children's rights, as well as the maintenance of a decision log (audit log) with the recording of key events, configurations, changes in models and interventions of the human operator, and the conduct of internal and external audits of child safety with access to the technical dossier.

Explainability and accessibility in the context of children means the ability of the system to provide a meaningful, structured and age-appropriate interpretation of the logic of decision-making, including a description of the model and methods of calculation, weights and key features, the level of uncertainty, the degree of human participation in decision-making, the safeguards applied and alternative options for the child. The explanation must be adapted to age, cognitive, and sociocultural characteristics and provided in plain language without excessive technical jargon, and the system must provide multi-level explainability — technical, functional, and ethical — and provide tools for the implementation of children's rights (refusal to profile, correct and delete data, human review).

In the public sector, as well as in cases where AI decisions can affect the rights, status, access to resources or social reputation of a child, the lack of effective mechanisms to ensure security and accountability, the provision of formal, overly technical or obscure messages, or the use of manipulative or misleading interface solutions is recognized as a violation of good governance and the principle of the best interests of the child. In such cases, the authorized national AI regulator has the authority to suspend or terminate the operation of the system, initiate an inspection (including investigation of incidents), revoke or revoke the permit or certificate of conformity, and is obliged to publish a public justification for the measures taken, ensuring the right to appeal.

In order to unify approaches to children's digital safety and explainability of decisions, authorized national bodies in the field of AI develop and periodically update the National Guidelines on Children's Digital Safety and Age-Appropriate Design. Such recommendations determine the categories of models and services according to the degree of risk to children, requirements for their documentation, examples

of proper implementation in sensitive areas (education, healthcare, social networks, games), templates of multi-level explanations, communication protocols with children, their parents and teachers, as well as mechanisms for responding to complaints or requests. Compliance with these recommendations is mandatory during the certification of high-risk systems and public procurement, and can be used by courts, auditors and human rights organizations as a benchmark in checking the legitimacy of algorithmic decisions regarding children.

The principle of digital safety of the child is a guarantee of public trust, legitimacy of digital solutions, prevention of harm and proper implementation of the legal regime of liability in the field of artificial intelligence.

60.1 Objects of legal protection.

1. *Dignity and integrity of the person* — an absolute prohibition of humiliating, violent or manipulative practices that exploit the age and cognitive vulnerabilities of the child; covers physical, psychological, informational and reputational integrity, prohibition of algorithmic dehumanization, stigmatization, shaming/bullying, coercion (including "dark patterns" of the interface), the use of emotional tracking for influence, as well as the dissemination or generation of content that humiliates honour and dignity; requires the application of the principle of least harmful impact, permanent human control and immediate termination of functions capable of causing humiliation.

2. Health - covers the protection of the physical and mental state of the child, including the prevention of behavioural and content risks that can threaten life or health, as well as mental risks, including stress, anxiety, addictions, and cyberbullying.

3. *Development and education*. It is prohibited to use algorithmic practices that create barriers to development and education, including those that limit access to opportunities, form closed development scenarios ("lock-in") or imply stigmatizing predictions of educational or professional trajectories.

4. *Privacy and family life.* The inviolability of the child's privacy and family life is guaranteed, which covers the secrecy of communications, living space, everyday life and family relationships; unauthorized interference or disclosure of relevant data is prohibited; Processing and transfer of personal information is allowed only if there is a legal basis and in compliance with the principle of controlled access and restriction of purpose.

5. *Autonomy and self-determination.* Every child has the right to maintain autonomy and self-determination in a digital environment that includes:

a) the right to an understandable, accessible and age-appropriate explanation of the decisions of artificial intelligence systems;

b) the right to opt out of profiling or automated personalization without prejudice to access to educational, social or other basic services;

c) the right to human review, re-evaluation and motivated confirmation of any automated decisions affecting his/her rights, status or development opportunities.

6. *Digital reputation and the right to be forgotten.* A person has guaranteed protection against long-term "digital scars" and false or distorted inferences; the right to correct, restrict or delete such data in accordance with the procedure established by law is ensured.

60.2 Scope and contexts of application.

1. *Family and home context* — smart toys (toy robots, constructors with sensors, interactive books), children's gadgets and wearable devices (fitness bracelets, location trackers, smart watches), family assistants and smart speakers, TVs and set-top boxes with artificial intelligence, baby monitors, home cameras and intercoms with AI functions, smart home systems in children's rooms. For these devices, the following are mandatory: local data processing by default; the presence of physical indicators and hardware switches of the microphone and camera; the prohibition of hidden recording and remote activation without consent; minimization of the amount of data and reduced storage periods; the absence of profiling and commercial advertising; parental control functions and "child mode" with time limits; the prohibition of geolocation tracking without urgent need; the presence of a "kill switch" and incident log.

2. *Education and leisure* — school digital platforms, assessment systems, online courses, educational and entertainment games, social networks, streaming services, and content platforms available to minors.

3. *Health & Wellbeing* — telemedicine services, wearable physical monitoring devices, mental health apps, and algorithmic services with recommendations for children.

4. *Public space and transport* — video analytics systems in schools, on the streets, in public transport, as well as other access control systems that cover children.

5. *State electronic services* are services related to obtaining benefits, maintaining state registers, identification and execution of digital documents of minors.

6. *Information environment* — relationships cover both direct interactions of the child with AI systems and indirect influence through recommendation algorithms, content moderation, targeting systems, and ranking of information in search.

60.3 Risk taxonomy (normative classification).

1. *Physical risks* — incitement to dangerous challenges and practices that may cause injury or self-harm; localization of the child through GPS, Wi-Fi or other beacons with the ability to determine the place of residence or routes; unauthorized or incorrect control of connected devices and toys (robots, drones, vehicles, smart home systems), which poses a security risk; provision by an algorithmic system dangerous or medically incorrect advice; cyber-physical incidents in AR/VR environments, including kinetic injuries or provoking attacks; exceeding the permissible physical and cognitive loads in fitness and gamified applications.

2. *Psychological risks* — the occurrence of anxiety and depression, emotional exhaustion, the formation of dependence on algorithmically designed dopamine interface patterns and other mechanics that exploit the child's cognitive or age-related vulnerabilities.

3. *Cognitive risks* — distortion of perception of reality through unlabelled synthetic content or mixing of facts and fiction; formation of information "bubbles" and attention tunnels (filter bubbles, echo chambers); substitution of critical thinking with automated responses and the emergence of an "algorithm authority effect" (automation bias); distortion of self-assessment and strengthening of social comparison; imposition of heuristics and anchors through interface design, rating systems and mechanisms "trends"; decreased ability to concentrate due to dynamic tapes and intrusive messages; cognitive overload and stimulation of unproductive multitasking; educational losses (learning loss) arising from the replacement of one's own work with AI solutions; manipulation of the order and form of presentation of facts (framing, priming); errors or "hallucinations" of models that mislead users and form false ideas.

4. *Social risks* — cyberbullying and online harassment (hate speech, hailing, doxxing, raids and mass complaints); grooming and recruitment; marginalization, isolation and "out casting" due to the algorithmic reinforcement of "information bubbles", biased ranking, "shadow bans" and manipulative recommendations; discriminatory or erroneous automated moderation or identification decisions (including automatic age determination) leading to unjustified blocks, reduction coverage or deprivation of access to education, social services or benefits; reputational damage due to synthetic manipulations (deepfakes, fake quotes, fake screenshots, edited content) and coordinated campaigns (botnets, astroturfing); increasing social inequality through targeted impact on vulnerable groups (on the grounds of disability, ethnicity, language, migration status, financial situation); substitution of the child's social subjectivity through algorithmic "assignment of roles" and the imposition of stereotypes.

5. *Information risks* — unlawful collection, leakage or re-identification of data; cross-inferences that allow the recovery of sensitive information; formation and storage of "sticky" shadow profiles without the child's knowledge or consent.

6. *Economic risks* — fraudulent practices, forced or manipulative involvement of a child in financial transactions, including imposed costs through micropayments, hidden gamification mechanics, or aggressive "soft monetization" design that exploits age or cognitive vulnerability.

7. *Legal risks* — algorithmic profiling, which leads to stigmatization, the formation of predictive profiles with long-term negative consequences for access to education, insurance, social or administrative services, and also creates a risk of discriminatory restriction of rights.

8. *Cultural and ethical risks* — imposition and reinforcement of stereotypes (gender, ethnic, disability or social status), normalization of violence, hate speech and discriminatory practices; dissemination of age-inappropriate content (sexualization, gambling, drugs, self-harm) under the guise of recommendations or educational materials; cultural appropriation, erasure or marginalization of languages and identities, dissemination of imperial or colonial narratives; substitution of children's subjectivity by algorithms (imposition of roles, value "guardianship" of the system, manipulative formation of a worldview); use of hidden advertising or branding. Age labelling and filtering of content, transparent indication of origin (including designation of synthetic materials), respect for cultural and linguistic diversity, the family's right to an alternative, and effective complaint and appeal mechanisms are ensured.

9. *Risk levels* are defined as low, medium, high and unacceptable; the category of "unacceptable" entails an express prohibition of the use or immediate disabling of the relevant function of the system.

60.4 Age groups and additional vulnerabilities.

1. 0–6 years (early age).

Physical and mental safety is a priority.

- Screen time is minimal; interaction — in short offline sessions.
- Personalization, profiling, and any advertising are prohibited.
- Data collection is zero by default; lack of cross-platform tracking; using only short-lived local buffers.
- Biometrics and emotional tracking are prohibited.
- Network functions and micropayments are disabled.
- Camera and microphone are activated exclusively by a single action of an adult, with hardware indicators and switches.
- The content is age-appropriate, without quick stimuli and oversaturation with audiovisual effects.
- Parental settings are required (time limits, "sleep/lessons", incident log).
- Data — without storage or transfer to third parties, processing is local.

2. 7–12 years old (junior schoolchildren).

Priority — safe education and development with a transparent and understandable interface.

- The interface and messages are age-adapted.
- Time limits and "lesson/sleep mode" are mandatory.
- Game monetization (micropayments, loot boxes, pseudo-random rewards) and manipulative retention mechanics (endless scrolling, streaks, forced quests) are prohibited.
- Psychotypic targeting and personalized advertising are prohibited.
- Data minimization; lack of cross-platform tracking.
- Chats — only with pre-moderation and filters; Private contact with adults is prohibited.
- Parental settings are required (content filters, notification control, activity reports without access to private content).
- AI decisions are explained in "children's language".
- Emotional tracking is prohibited.
- Push notifications are limited; Personalization — with the ability to turn off.

Local data processing by default with short retention periods.

3. 13–15 years (teenagers).

The priority is protection from social and psychological risks.

Mandatory "detox mode" of the feed and recommendations by default with a visible explanation of "why am I seeing this?" and a "no personalization" toggle.

Targeting by psychotype, behavioural or emotional profiles, personalized advertising, and the formation of look-alike audiences are prohibited.

Private messages are allowed only with the explicit consent of the teenager, with the possibility of blocking unwanted contacts; provides automatic clipping of fraudulent accounts and adults without mutual connections.

- Mandatory toxicity filters, simple reporting mechanisms, and operational moderation.
- Push notifications are restricted, streaks and "inactivity penalties" are prohibited.
- Age verification is carried out without storing excessive personal data.
- Location, biometrics, and emotional tracking are turned off.
- Data is stored for a short time, end-to-end encryption is used.
- Parental control is carried out in the form of aggregated reports without access to private content.
- Mandatory presence of buttons "ask for help" and protocols for escalation of risks (self-harm, violence).

4. 16–17 years old (older teens).

Priority — expanded rights to self-determination and participation in decision-making regarding one's own data.

- The explanation of AI solutions should be multi-level (technical, functional, ethical) with the right to ask questions and get alternatives.
- Predictive profiles of "future trajectories", psychotype targeting, emotional tracking and data monetization are prohibited.
- The right to refuse profiling and advertising, as well as the right to "invisible use of».
- Privacy of communications is guaranteed by default: end-to-end encryption, profile visibility control, geolocation blocking.
- Self-monitoring tools for time and notifications are provided.
- Age verification is carried out without storing documents.
- Access to data portability, decision logs, and appeal mechanisms is provided.
- In sensitive areas (education, labor, insurance, lending), it is prohibited to make automated decisions without human review and consent of the child; Scoring of educational, career or financial chances is expressly prohibited.

5. *Additional vulnerabilities*.

The category of children with additional vulnerabilities includes: children with disabilities (physical, sensory, intellectual, psychosocial), including manifestations of neurodiversity; children with chronic diseases; internally displaced persons, refugees, stateless, children who have survived armed conflict or trauma; children without parental care (boarding schools, foster families), from large or single-parent families; economically vulnerable; representatives of linguistic, cultural and ethnic minorities; children from rural or remote communities, as well as those with limited access to the network.

For such groups, increased precautionary standards apply:

- principle "minimal data — default»;
- prohibition of profiling, personalized advertising, geolocation tracking, and emotional analytics;
- adapted interfaces appropriate to age and accessibility requirements (WCAG, simple language, alternative formats);
- ensuring offline equivalence and the right to an alternative;
- increased human surveillance and accelerated escalation of incidents;
- prevention of re-traumatization (content filters, opt-in communication only);
- Secure age verification without storing documents;
- conducting a separate Child Impact Assessment with special safeguards;
- Short data retention periods, end-to-end encryption and prohibition of transmission to third parties;
- priority of local processing and lack of cross-platform tracking.

60.5 Guarantee standards (structure of mandatory fuses).

1. *Legal guarantees: there* is a direct ban on psychotype targeting, emotional tracking and the creation of "predictive profiles" for children; commercial monetization and cross-platform tracking of

children's data are unacceptable; the processing of children's data is carried out exclusively on the grounds determined by law in compliance with the principles of minimization, targeted restriction and short storage periods; the burden of proving the need for processing rests with the operator; the implementation of the child's rights is guaranteed and its legal representative to information invisibility, refusal of profiling and advertising, access, correction, deletion and portability of data, as well as the right to object to automated decisions; all decisions that have significant legal or factual consequences for the child are subject to human review with a mandatory reasoned explanation; The use of data processing is possible only with the informed consent of the legal representative and, taking into account the age and maturity of the child himself/herself; it is forbidden to use "dark patterns" when obtaining consent; it is mandatory to conclude contracts with processors (DPAs); the transfer of child data to third parties, as well as cross-border transfers without adequate legal and technical safeguards, are prohibited; The operator is obliged to report incidents and provide effective remedies, including complaint, appeal and compensation.

2. *Organizational guarantees:* appointment of a responsible Child Safety Officer and creation of a cross-functional child safety committee; approval of a clear RACI matrix and accountability to the board of directors; implementation of "child safe by default" and "child safe by design" policies at all stages of the development lifecycle (SDLC); mandatory onboarding training and annual refresher courses for all roles, as well as specialized trainings for moderators, educators and developers; due diligence and signing of NDAs by personnel who have access to child data; application of the principles of minimum privileges and division of duties; development and implementation of incident management procedures with response plans (RTO/RPO), communication channels with parents and educational institutions, as well as regular training (tabletop exercises); change management and release control with mandatory "child gates"; maintaining a supplier map and a register of processors with due diligence, conclusion of data processing agreements (DPA), establishment of technical and ethical requirements, periodic audits, SLAs and the right of inspection; supplier inspection log and risk matrix; introducing a process of preliminary ethical and legal review of experiments and A/B tests; the functioning of internal and external channels for reporting vulnerabilities and abuses (including anonymous "whistleblowing") and responsible disclosure programs; regularly conducting internal audits, determining KPIs/KRIs for child safety and reporting to management; implementation of recruitment protocols and psychological support for moderators and staff working with traumatic content; prohibiting the use of "shadow IT" and setting policies for working with data outside the production environment.

3. *Technical guarantees*: on-device data processing and the principle of "minimum data is the default"; end-to-end encryption of data in transit and in storage (TLS 1.3+, AEAD) with secure key management in specialized modules (HSM/KMS), with key rotation, distribution of roles and escrow procedures; segmentation and isolation of child data (tenant isolation), access control on the principle of least privilege, multi-factor authentication; short and declared data retention periods (TTL/retention policy) with automatic and verifiable erasure from replicas, backups, caches and indexes; transparent personalization switches and "information invisibility mode"; technical blocking of the creation or storage of "predictive profiles of the future", prohibition of cross-platform tracking; abandonment of external trackers, SDKs, and telemetry, in addition to the critically needed modules running in the sandbox with restrictions and a data processing agreement (DPA); logging of solutions and accesses with tamper-evident protection, maintaining a full trajectory of data and models (data/model lineage) with versioning; introduction of emergency fuses and "kill switch" mechanisms (rate limiting, circuit breakers, release blocking) for dangerous functions; safe model training: exclusion of children's data from training sets, filtering out toxic and age-inappropriate content, application of differential privacy and/or federated training where appropriate; protection against leaks due to inference or playback: I/O filters, context restrictions, confidential modes, content hashing, protection against prompt injection and jailbreak attacks; secure releases (staging, canary deployments, rollback), component list maintenance (SBOM), supply chain checks, regular vulnerability scanning, and pen tests; labeling synthetic content and applying age rules for recommenders; compatibility with "Do Not Profile / Do Not Track" signals and support for open APIs for parental controls.

4. *Educational guarantees*: provides a standardized package of activities for children, parents, teachers and operators, including:

(a) providing age-appropriate explanations in a multi-level form (short, detailed, for experts), in simple CEFR language A1–B1, in multimedia formats (text, icons, video, audio, comics), localized in national minority languages and available in accordance with WCAG standards (easy speech, subtitles, sign language);

(b) development and application of national guides on digital literacy and child safety, as well as training modules for schools, parents, doctors and moderators;

(c) creation of a catalog of informed consent and withdrawal templates, messages and warnings in simple language, containing explanations: "what does the AI system do", "what data is processed", "how to opt-out";

(d) implementation of crisis communication protocols (in cases of self-harm, violence, data leakage), notification scripts and contact directories for help;

(e) setting requirements for interfaces: a clear indicator of AI participation, safety tips, interactive comprehension checks without saving the child's responses, healthy use timers;

(f) conducting regular information campaigns and evaluating their effectiveness (comprehension surveys, reach and usefulness metrics, public reports without personal data).

5. *Audit guarantees*: the mandatory Child Impact Assessment is carried out before the launch of the system, after each significant change to the model, data or interface, and at least once a year. The "Child Safety Passport" is published in the public domain with updating within 30 days after the changes. High-risk functions are annually subject to an independent external audit with the provision of independence criteria (no conflict of interest, auditor rotation at least once every 3 years).

The audit covers:

(a) completeness of documentation, data maps, and data and model trajectories (data/model lineage);

(b) setting up and efficiency of technical and organizational controls (kill switch, Do Not Profile, age filters, retention policies, encryption, synthetics labelling));

(c) checking solution logs (tamper evident), selective reproduction of cases, testing interfaces for the presence of "dark patterns»;

(d) verification of educational and assessment datasets for the absence of children's data and age-inappropriate content;

(e) vendor audit — DPA, SLA compliance and right of inspection;

(f) conducting red teaming and safety tests according to ethical protocols without involving minors in risky scenarios.

The result of the audit is a report with a classification of findings (critical, high, medium, low), a corrective action plan (CAPA) with elimination deadlines (24/72 hours for critical and high), conformity attestation and a public summary without personal data. In case of critical non-compliances, the operator is obliged to immediately stop the relevant functions until the violations are eliminated. Regular monitoring of child safety KPI/KRI and annual submission of a transparent report to the regulator are provided.

60.6 The burden of proof and the principle of proportionality.

The operator is obliged to prove the necessity of each data processing operation, its proportionality and compliance with the principle of minimization; Reference to "technical impossibility" does not exempt from the performance of the established guarantees.

The consent of the legal representative or the child himself is not recognized as a self-sufficient basis for high-risk practices; Their application requires additional legal and technical bases, along with proper safeguards.

In case of doubt, all interpretations are made in favour of the child's safety.

60.7 Procedure for assessing the impact on children's rights (core of compliance).

Identification of functions related to children, data mapping, threat modelling, and the formation of risk scenarios according to an approved taxonomy.

Assessment of the likelihood and severity of potential harm; determination of minimization measures (legal, technical, organizational).

Testing of the system taking into account age groups; consultations with teachers and psychologists; preparation of a public short report and a "Child Safety Data Sheet".

Developing an incident response plan: setting SLAs to disable dangerous features, informing parents/guardians, and taking corrective action.

Periodic review (at least once every 12 months) or sooner in the event of a change in model, data, or incident.

60.8 Normative illustrations (examples of application of the definition).

1. *School platform with recommendations*: personalization is allowed only for educational purposes (curriculum, individual development program), without behavioural or emotional profiling and without advertising; it is forbidden to generate or save predictions of the "career trajectory", scoring assessments or "future paths" of the child; mandatory switch "without personalization" (information invisibility mode) in one click; clear delineation of access roles: student — minimum amount of data; teacher — methodological tools without access to private journals; parents — only aggregated summaries; administrator — only audit functions; minimizing data and prohibiting the use of external SDKs, telemetry, and cross-platform tracking; keeping a log of tamper evident solutions and explaining the principle of "why this recommendation"; use of source whitelists, age filters and anti-harm dictionaries; shelf life — no more than 90 days, with automatic erasure from backups and caches; end-to-end encryption (E2E) of all communications; ensuring offline equivalence (access to tasks without the use of AI); prohibiting the use of data for training third-party models or marketing purposes; the presence of kill switch mechanisms and the "lesson/exam" mode with blocking hints; quarterly Child Impact Assessment with a public summary of the results; Annual independent audit of compliance with child safety standards.

2. *Social networks for teenagers function under the following conditions:* the feed is formed in "detox mode" by default (without endless scrolling, autoplay of the video and aggressive return triggers; the "last first" option is provided); psychotypical targeting, behavioural and emotional profiles, personalized advertising and look alike of the audience are prohibited; loot boxes and gambling-type mechanics are prohibited; push notifications — no more than once a day and exclusively of a service or security nature; recommendations are subject to explanation "with one click" ("why am I seeing this?"), with the presence of a "no personalization" switch; Privacy is provided by default: profiles are closed, profile visibility control is provided, geotags and location are disabled, indexing is limited to search; private messages: requests from strangers are prohibited, correspondence is possible only after mutual confirmation; by default, communication with adults without joint connections is blocked; Toxicity and spam filters are applied, a simple reporting mechanism and moderation SLAs of up to 24 hours are provided; Security Tools: "Ask for help" button, quick blocks, "pause notifications", the ability to hide preference counters; content filters from age-inappropriate or harmful content and mandatory labelling of synthetic materials; the amount of data is limited to a minimum, without cross-platform tracking and external SDKs; storage periods do not exceed 90 days; private messages are end-to-end encrypted; all decisions are recorded in a log with tamper evident signs; age verification is carried out without storing documents; the use of children's data for training third-party models is prohibited; Kill switch is provided for dangerous functions; the "Child Safety Data Sheet" is published.

3. *Smart children's toys with sensors.* A smart toy with a camera or microphone is allowed to be used only if the following requirements are met: all image and sound calculations are carried out exclusively locally; the microphone and camera are equipped with hardware indicators and physical switches; cloud functions, external connections, third-party SDKs and telemetry are prohibited; network interfaces, geolocation, face and emotion recognition are disabled by default; a parent panel with time limits is provided, sleep/lesson, instant lock and remote erase modes; microphone and video recording is prohibited by default, short-lived buffers are used without writing to disk; all settings and logs are stored in encrypted form, with signs of unauthorized modification; each device has unique cryptographic keys; secure pairing with other devices is carried out only with the participation of an adult (PIN, QR code or physical action

on the case); pairing with unknown devices is prohibited; firmware update is possible only in the form of packages signed by the manufacturer, with rollback protection; a public log of firmware changes is kept; there is a "private room" mode (complete ban on sensors) and "guest" (without saving data); it is forbidden to form long-term profiles of children, monetize data or use it to train third-party models; a set of labelling "Child Mode" and "Child Safety Data Sheet" with support contacts and a complaint procedure is mandatory.

Toys that do not meet these requirements are not allowed for circulation on the territory of the State.

4. *Electronic medicine for teenagers.* AI systems used in the field of e-medicine for minors are subject to qualification as high-risk and are allowed to be used only if the following requirements are met: AI advice is exclusively advisory in nature and does not equate to a diagnosis; mandatory disclaimer: "not for emergencies"; the decision is made exclusively by a qualified medical professional (human-in-the-loop); the user is provided with available offline alternatives without AI; clinical chats and video communication are carried out with mandatory verification of a specialist and logging of consultations; medical data is classified as "particularly sensitive" and is subject to: end-to-end encryption, loop isolation (separate storage of identifiers and clinical records), minimization of storage volume and shelf life (TTL $\leq$ 90 days), prohibition of "perpetual" backups and complete prohibition of transfer to third parties and use of third-party models for training; personalization, advertising and cross-platform tracking are prohibited; emotional tracking, the formation of "profiles of the future" and behavioural scoring are prohibited; access to the service is carried out on the principle of two-factor consent of the child and parents (taking into account age/maturity), with the possibility of private consultation in cases provided for by law; all AI recommendations should be accompanied by explainability: indication of the grounds, level of uncertainty and proposed alternatives; security protocols are in place: an "SOS" button, routing to hotlines/emergency services, an escalation SLA of no more than 15 minutes in case of a risk of self-harm or violence, as well as a "kill switch" for disabling dangerous functions; the models used are validated on paediatric samples, contain filters of toxicity and age-related inappropriateness; Children's data is excluded from training, except in cases of differential privacy or federated learning by separate consent; Telemedicine services provide: logging of all accesses, control of protocol versions, prohibition of diagnostic applications without certification as a medical device, and complete rejection of third-party SDKs and telemetry; Each e-medicine service has an offline equivalent.

Failure to comply with the requirements of this article is the basis for prohibiting the functioning of the system, imposing administrative liability and withdrawing from circulation.

Basic principles and limits of the use of AI for minors.

1) Any design, training, deployment and use of artificial intelligence systems related to minors is carried out according to the principle of "safety first" and "the best interests of the child". In the event of a conflict, uncertainty or doubt, a presumption in favour of the child (*in dubio pro minore*) applies. The operator is obliged to prove the necessity and proportionality of each data processing operation and to choose the least invasive means of achieving the end. Economic feasibility, convenience or technical complexity cannot justify a reduction in the level of protection. The principle is inalienable and cannot be limited by the terms of the service, the contract or the consent of the user. risk of significant harm to the child, the operator is obliged to immediately suspend the relevant functions of the system until the risks are completely eliminated, and the safety is confirmed by an independent audit or an authorized body.

Failure to comply with the requirements of this Article shall result in the prohibition of the operation of the system and bringing the operator, supplier or manufacturer to responsibility in accordance with the applicable legislation of the State and applicable international regulations.

2) *The principle of data minimization and impact*. Only the minimum necessary data is processed, without which a particular function of the service cannot work. Prohibited: data collection "in reserve", cross-platform tracking, external telemetry or SDK without critical need. Principles apply: target limitation and retention for the lowest possible period (TTL) with automatic erasure from replicas, caches and backups. Personalization outside of a narrow legitimate purpose is disabled by default; the user is provided with a simple "No personalization" switch and information invisibility mode. The data is subject to aggregation and pseudonymization if possible. The use of biometrics, geolocation and other sensitive

categories is allowed only based on the law, with explicit informed consent and only in the absence of a less invasive alternative. Also prohibited: hidden models of psychological influence, gamification of addiction (reward games, streaks, loot boxes), exploitation of age-related vulnerabilities and manipulative interface patterns. Push notifications are limited to a minimum and are allowed only in the form of neutral informational messages. The operator is obliged to conduct a proportionality test and document the data map with the justification of each attribute, ensure that it is updated and accessible to regulatory authorities.

3) *The principle of age-appropriate design.* The design of interfaces, texts, messages, interaction mechanics and the attention economy is carried out considering the age, level of maturity, cognitive capabilities and vulnerabilities of the child. Provided: simple and understandable language, accessibility in accordance with WCAG standards, including subtitles, alternative descriptions, sign language, localization in minority languages. Secure default settings (privacy-by-default) are set: minimal data, geolocation, biometrics and profiling are disabled. Manipulative and addictive mechanics are prohibited, including: endless scrolling, autoplay, streaks, forced quests and "penalties" for exiting. Push notifications are limited to a minimum and are allowed only within reasonable need. The user is provided with transparent and symmetrical choices without "dark patterns", including: a visible "no personalization" switch, "lesson" and "sleep" modes, an explanation of "why am I seeing this?" in one click, simple undo actions and delete data. Clear age markers and indicators of AI participation in interaction are established. The parent outline provides aggregated summaries without access to the child's private content. All systems undergo mandatory testing with representatives of target age groups; the results are documented in the Child Impact Assessment, which is subject to updating and submission to regulatory authorities.

4) *The principle of transparency and control.* The operator shall ensure clear, timely and verified disclosure of information about the participation of the AI system in the interaction with the child and provide effective means of exercising the rights and control settings. Minimum requirements include: explicit labelling of each interaction/decision involving AI (indicator "AI/AI"); multi-level explanations ("brief/detailed/technical") in simple language taking into account age: the purpose and role of AI, logic, influencing factors, data sources, level of uncertainty, alternatives; transparency of the origin of content and labelling of synthetic materials; register and public description of the involved processors, SDKs, cloud services, data flows, storage locations, and terms (TTL/retention); Available controls: "No personalization" toggle, refuse profiling, disable geolocation/biometrics/emotional analytics, limit notifications, "lesson/sleep" mode, delete history; individual rights: access, correction, deletion, portability, objection to an automated decision, the right to human viewing — free of charge and without discrimination of access to the service; access to a fragment of the decision log (tamper-evident) in cases of the impact of an algorithmic decision on the rights or opportunities of the child; Advance notice (30 days in advance) of material changes to processing models/policies with the right to opt out or opt for offline equivalence; the mechanism of complaint and appeal through unified channels (application, e-mail, hotline), the response period is no later than 10 working days; compatibility with "Do-Not-Profile/Do-Not-Track" signals and open parental control APIs; localization of explanations and interfaces into minority languages and compliance with accessibility standards (WCAG, subtitles, sign language, plain language); prohibition of "dark patterns" in explanations and settings; symmetry of consent/withdrawal actions; preservation of the transparency dossier (technical, methodological, legal) and its submission to the regulator together with the current Child Impact Assessment; absence of negative consequences for the child in case of refusal of personalization or profiling (offline equivalence, immutability of basic access to educational and government services).

60.9 Basic principles and limits of the use of AI for minors.

1. The principle of the safety and best interests of the child — the protection of the child has unconditional priority over commercial, analytical or experimental interests.

2. The principle of data minimization and influence — it is allowed to collect only data necessary for the basic functioning of the service; hidden models of psycho-emotional influence, reward game mechanics, and exploitation of age-related vulnerabilities are prohibited.

3. The principle of age-appropriate design — interfaces, texts, and mechanics should be understandable to the child; the use of "dark patterns" that stimulate excessive exposure or disclosure of data is prohibited.

4. The principle of transparency and control — the child and/or his/her legal representative receive a clear explanation of the actions of AI, the categories of data used, and the consequences of refusing to process them.

5. The principle of locality of risk — data processing is carried out, as far as possible, locally on the device; in the case of external processing, increased encryption is applied and clear retention periods are set.

60.10 Imperative prohibitions.

Prohibited in relation to minors:

1. targeted advertising personalized based on psychotypes, neurobehavioral profiles, vulnerabilities or predicted emotional states;

2. automated tracking of emotions, reactions, facial expressions, biometric or behavioural signs without the informed consent of the legal representative and the child himself (taking into account his age and maturity);

3. creating or using predictive profiles of future decisions, educational or career trajectories, risks, inclinations, or "life paths" of a child based on AI models;

4. the use of manipulative interfaces that stimulate excessive use (in particular, unlimited scrolling, aggressive push mechanisms, "fines" for exiting, gamification of data collection);

5. algorithmic decisions that lead to discrimination on the grounds of age, sex, health, socio-economic status, ethnicity or disability;

6. monetization, resale or use of third-party models for training without separate legal permission and in the absence of the child's overriding interests.

60.11 Conditionally permitted practices (narrow exceptions).

In cases of protection of the life or health of a child (search for missing children, combating violence), temporary data processing with the minimum required volume is allowed, only by decision of the competent authority and under the control of the court.

The collection and processing of anonymized statistics on learning quality or safety is only permitted if any possibility of identifying or re-identifying the child is excluded and there is no personalization.

In the field of school or medical services, the use of AI systems is allowed only under the principle of "offline equivalence" of access: refusal to use AI cannot worsen or limit the child's access to the service.

60.12 Responsibilities of AI System Operators (suppliers, importers, distributors, users).

1. Operators are required to embed child-safe-by-default child modes, which include: disabling trackers, private profiles, lack of external telemetry, and limiting usage times.

2. Operators are required to implement age-appropriate access assurance without storing sensitive documents and without excessive data collection.

3. Before launching the system, operators conduct a Child Impact Assessment with a mandatory public summary report and a risk mitigation plan.

4. Operators provide transparent explanations of algorithms in understandable and accessible materials, which include information: "what the model advises/decides", "what data is processed", "how to opt out".

5. Operators provide "information invisibility of the child" modes, including the ability to use without profiling, a simple personalization switch, short magazine retention periods and fast erasure of digital traces.

6. Operators are obliged to prohibit the transfer of children's data to third parties, other than processors necessary for the provision of the service, with the inclusion of contractual guarantees and the provision of audits.

7. Operators are required to have kill switch procedures in place for features that pose an immediate risk.

60.13 Minimum technical requirements.

1. Parent outline — a panel with time constraints, lesson/sleep mode, activity summaries without access to the content of private communication is provided; Delegated access for guardians is allowed.

2. Prohibition of forecast profiles — creation and storage of forecast profiles of decisions or trajectories of the child is prohibited; The operator is obliged to have a mechanism to confirm the absence of such profiles.

3. Prohibition of emotional tracking — emotional analytics modules are disabled; Their lawful use is allowed only locally, for a short time, with a mandatory visible indicator and access log.

4. Minimization of notifications — algorithmic restrictions on the frequency of push notifications are applied; Triggers that force the child to return to the app are prohibited.

5. Secure recommendations — Whitelists of sources, age-appropriate ratings, content filtering, and mandatory logging of audit solutions are used.

6. Encryption and data deletion — end-to-end encryption of children's data is provided; raw data and derived representations are subject to automatic deletion within a specified short time.

60.14 Consent and rights of the child.

The processing of the child's personal data is carried out with the informed consent of the legal representative; for children who are able to realize the consequences (considering age and level of maturity) - also with the informed consent of the child himself. The refusal or withdrawal of consent must be carried out in a manner not more complicated than its granting.

For children without parental care, consent is given by an authorized person in accordance with the law; In urgent cases, the minimum necessary processing is allowed until the moment of confirmation of authority.

Your child has the right to access understandable and age-appropriate explanations, to have data corrected and deleted, to opt out of profiling and advertising, and to have automated decisions reviewed in a human.

60.15 Public duties.

1. The state establishes a regime of increased supervision over systems that interact with children through their mandatory registration, enhanced auditing and the introduction of labelling "for children".

2. The state provides a national infrastructure of available technical parental control tools, including open API standards for integration by manufacturers.

3. The state organizes and implements digital literacy programs for children, parents and teachers, creates a "single window" of reference materials and standardized informed consent templates.

1. The state is creating the institution of the Commissioner for Children's Rights in the Digital Environment, empowered to carry out inspections, issue orders, impose fines and go to court.

2. The state provides offline alternatives to state electronic services for children and families who have abandoned the use of AI tools.

60.16 Risk categories and life cycle.

Systems that interact with children in a targeted manner (educational platforms, games with AI agents, smart toys) are classified as at least classified as "high-risk" and are subject to a mandatory pre-launch impact assessment on children's rights and an annual supervisory audit.

For each such system, it is mandatory to publish a "Child Safety Data Sheet", which contains a description of functions, data categories, risks, mitigation measures, storage periods, incident response procedures, as well as responsible persons and contact channels for appeals.

60.17 Procedural safeguards, incidents and remedies.

Every child or their legal representative has the right to lodge a complaint with the operator, competent supervisory authority or court in accordance with national law; The operator is obliged to provide a reasoned response and take appropriate measures based on the results of consideration within a reasonable time, which may not exceed ten working days.

In the event of a risk of damage, the operator is obliged to immediately suspend the operation of the relevant modules or functions (temporary blocking), notify the parents or legal representatives of the child and the competent authorities, and take measures to minimize the consequences.

Remedies include but are not limited to: termination of unlawful data processing, deletion of data and derived profiles, compensation for damages, public apology (if necessary), conducting a mandatory independent audit.

60.18 Public Procurement and School Implementation.

Public procurement of EdTech, Insurtech and HealthTech solutions for children is allowed only if there is a certified "Child Safety Data Sheet", implemented modes of information invisibility of children's accounts and the absence of any data monetization.

Educational institutions are prohibited from conditioning the assessment of learning outcomes or access to educational services on the use of AI profiling systems; they are required to provide equivalent alternative offline procedures.

60.19 Responsibility.

The following measures of influence are established for violations: an order to immediately stop the violation, fines, temporary suspension of activities in the market, a ban on the use of certain functions, as well as personal disciplinary, administrative or criminal liability of officials for intentional or grossly negligent actions.

Aggravating circumstances are, in particular, but not exclusively: the repetition of the violation, the significant scale of the children's audience, the use of emotional tracking, the creation of predictive profiles, commercialization or other misuse of children's data.

The following are recognized as mitigating circumstances: prompt independent detection and elimination of violations by the operator, conducting a public independent audit and implementing damage compensation programs.

CHAPTER 61. THE PRINCIPLE OF GENDER EQUALITY IN DIGITAL AI-ECOSYSTEMS

In the field of creation, development, implementation and use of AI systems, the public law principle of gender equality is established, which guarantees equality of rights and opportunities for persons regardless of gender, gender identity, gender expression or sex characteristics, considering intersectionality and the prevention of direct and indirect discrimination at all stages of the technology life cycle.

It is prohibited: any form of gender discrimination, stereotyping or humiliation in the decisions of AI systems; the use of "gender determination" modules based on biometric data; the use of "emotion recognition" systems to assess the reliability or suitability of a person; forced or implicit genderization of users, as well as the imposition of stereotypical roles in content or targeted advertising; deadnaming and systemic misgendering; algorithmic practices that lead to horizontal or vertical segregation in employment, education, credit, insurance, and access to services.

It is also prohibited to use direct, indirect and proxy features that reproduce gender bias, in particular: combinations of first name, patronymic, titles, marital status, care duties, "career breaks"; residential address, availability schedule, type of device; speech or acoustic markers. The processing of sensitive data on gender identity is allowed only if there is a clear legal basis, in compliance with the principles minimizing scope, restricting access, and prohibiting secondary incompatible use or "default" inference.

Operators and providers are obliged to ensure gender-inclusive data management, in particular: sufficient and balanced representation of groups in training and test sets; mandatory disaggregation of results at least by gender and gender identity, as well as in intersection with age, disability, language, socio-economic conditions; maintenance of "dataset passports" indicating sources, collection periods, imbalances and ways to neutralize them; The use of synthetic or rebalanced samples cannot mask systemic bias.

Fairness and transparency of artificial intelligence models are ensured. Before the launch of the AI system and in the process of its use in a productive (operational) environment, verification of the fairness of models is ensured. For this purpose, the following metrics are used: group and individual fairness metrics, in particular Disparate Impact Ratio, Equalized Odds, Calibration by Group; analysis of errors by groups using confidence intervals; testing for proxy attributes using model interpretation methods. Tolerance thresholds, corrective action plans, and suspension triggers in case of deviations are set for such systems. Each significant model update is accompanied by revalidation, a change log, and release notes that reflect the impact of the changes on fairness.

Transparency and explainability of solutions are ensured: for high-risk applications, a "model passport" is published, containing information about the purpose, version, known limitations and the latest results of fairness checks; in user interaction, a locally understandable explanation of the result is provided, outlining the key factors of influence and the limits of applicability, without disclosing trade secrets in excess of the volume, necessary for effective protection of rights.

Effective human oversight of the functioning of AI systems is established. For this purpose: a responsible officer for compliance with the principle of gender equality is appointed with the authority to stop or cancel automated decisions; independent ethical oversight is ensured; mechanisms for internal whistleblowing are created; regular staff training is provided on the prevention of bias, the implementation of intersectional analysis and proper journaling.

Individual rights to protection against gender bias are guaranteed, in particular: the right to timely information about the use of the AI system; the right of access to an understandable explanation of the results; the right to initiate a human review with the suspension of negative consequences until the end of the proceedings; the right to file a complaint free of charge and in a simplified form; the right to receive extracts from journals and aggregate fairness reports; the right to appeal against decisions to the

supervisory authority and the court. Recorded cases of violations lead to the immediate correction of unlawful decisions and the full restoration of violated rights.

In the field of recruitment, remuneration, shift planning and performance evaluation, any practices or algorithmic decisions that directly or indirectly reduce the chances or rewards based on gender or care responsibilities are prohibited. In the field of credit, insurance, trade and advertising, the following are prohibited: practices of gender segmentation of demand; "pink" pricing; non-admission to products or services based on gender; overestimation of risk ratios without proper justification. In the field of health and life safety, equal access to diagnostics and recommendations is ensured. The use of typical (default) "male/female" settings in symptomatology, training datasets, and interpretation of results is prohibited.

Interfaces and content of AI systems are developed according to the principles of inclusive design. In particular, it is ensured: support for self-defined names and addresses; the use of correct pronouns and non-binary options in questionnaires; no coercion to disclose gender identity; neutrality of answers and visual images by default; the presence of safeguards in recommendation systems and generative models against the reproduction of gender stereotypes, hate speech and gender-based violence; tools to flag and quickly remove toxic content.

All procurement and contracts in the field of AI systems include mandatory conditions for compliance with the principle of gender equality, in particular: guarantees of fairness; requirements for the availability and maintenance of "model passports" and "data passports"; the right to conduct an independent audit; obligations to eliminate identified violations within a specified time frame; sanctions in the form of suspension of the provision of services or access to the system, termination of the contract and compensation for damages. In public procurement, priority is given to solutions that have proven inclusivity and open reporting.

Supervision and reporting are being introduced. Operators are obliged to: submit periodic reports on the results of fairness checks and recorded incidents of discrimination; notify the supervisory authority of critical cases no later than 72 hours later than 72 hours with a corrective action plan; keep logs of automated decisions and key equity indicators, store them within the established deadlines and provide them for audit; ensure that information on compliance with the principle of gender equality is entered into the state register.

For violation of the requirements of this section, the following measures are applied: an order to immediately eliminate violations; suspension or withdrawal of the system; mandatory refutation of false information and removal of discriminatory content; fines and recovery of illegally obtained benefits; exclusion from participation in public procurement; in case of repeated or systematic violations, increased sanctions and a ban on re-deployment of the system before passing an independent audit. Any rejection of minimum guarantees of gender equality is null and void.

For minors, persons with disabilities, refugees, internally displaced persons, and other vulnerable groups, increased guarantees are established, in particular: limitation of risky use scenarios; increased threshold criteria for the application of the system; providing individualized alternatives; prioritizing human consideration over automated decisions. All provisions of this section are interpreted in accordance with the principle of pro persona, with priority given to the effective provision of equality and non-discrimination in interaction with AI systems.

61.1. General Provisions.

The principle of gender equality in the field of AI is established as a public law one. This principle means not only a formal prohibition of discrimination, but also an active obligation of the state, suppliers, operators, and users of AI systems to ensure de facto equality in access to systems, equality of opportunities for their use, and non-discriminatory results. The principle applies to all entities — from developers of algorithms and training kits to government agencies that use AI systems in the field of public services. principle includes: compliance with international human rights standards in the field of gender equality; integration of gender assessment at all stages of the life cycle of AI systems; conducting an audit and assessment of the impact on gender equality; creation of mechanisms for correcting the identified

imbalances; ensuring transparent and accountable practices aimed at eliminating both historical and contemporary gender inequalities.

1. All systems, data, algorithms, models and design processes are organized to guarantee de facto equality of rights and opportunities for all persons, regardless of gender, gender identity, gender expression or sex characteristics. This includes: mandatory inclusion of the principle of gender equality at the stage of formation of the terms of reference; conducting gender-legal and socio-technological analysis at the stage of data collection and preparation; checking the absence of systemic biases in training and test datasets; development of algorithms and models that do not reproduce or reinforce gender stereotypes; ensuring inclusive interfaces and results that correspond to the self-determination of users; regular audit of implementation and use processes in terms of compliance with the principle of gender equality; creation of mechanisms for correction and compensation of damage in case of discriminatory effects.
2. The principle of gender equality also encompasses the intersectional approach. This means considering a combination of different characteristics — age, disability, socioeconomic status, language, ethnic origin, place of residence, marital status, and other characteristics — in the process of designing and applying AI systems. Intersectionality implies that discriminatory effects can overlap and amplify the cumulative impact on the vulnerability of individual groups. Operators and suppliers are obliged to: conduct a multidimensional analysis of bias risks; assess the combined consequences for persons belonging to several vulnerable groups at the same time; introduce algorithmic and organizational mechanisms that minimize multiple discrimination; apply compensatory and restorative strategies in case of detecting superimposed effects; document intersectional risks in model passports and in fairness reports.
3. Both direct and indirect discrimination in the field of application of AI systems is prohibited. Direct discrimination refers to clear differences in access to services, refusals of employment or credit, which are carried out solely based on gender or gender identity. Indirect discrimination refers to algorithms or practices that look neutral on the outside, but create a discriminatory effect. These are the use of proxy attributes (residential address, device type, availability timelines) or the use of formal criteria that systematically reduce the chances for certain groups. Direct and indirect discrimination are equally unacceptable and entail legal liability, as they contradict the basic principle of gender equality and undermine trust in AI systems.

61.2. Prohibited practices.

Any form of gender discrimination, humiliation, or reproduction of stereotypes in AI solutions is prohibited. This includes: explicitly excluding individuals based on gender; using derogatory, stereotyped, or discriminatory characteristics in recommendations or classifications; algorithmically reproducing traditional stereotypes (e.g., "female" and "male" occupations).

It is prohibited to use "gender determination" modules based on biometric data and "emotion recognition" systems to assess the reliability or suitability of a person. Such technologies pose a high risk of bias because biometric and behavioural markers do not reflect a person's real identity and often lead to discriminatory outcomes. Their use to access jobs, education, credits or services is unacceptable.

Forced or covert genderization of users is unacceptable. This includes: automatic assignment of gender categories in forms or interfaces; hidden algorithms that classify users by gender without their consent; forced choices that do not take into account non-binary identities.

Stereotypical roles are not allowed in targeted advertising or content. Examples include advertising campaigns targeting only women or men in relation to "traditional" roles, or algorithmic promotion of content that reinforces outdated gender perceptions.

Deadnaming (use of the "old" name of a transgender or non-binary person) and systemic misgendering (conscious or automated use of incorrect pronouns or addresses) are recognized as forms of digital violation of dignity and are prohibited both in internal algorithms and in external interfaces.

Algorithms that cause horizontal (restriction of access to certain areas of employment, for example, technical professions for women) or vertical (barriers to career advancement, the so-called "glass ceiling") segregation are recognized as unacceptable. They are subject to mandatory audit, and the identified cases must be corrected using corrective mechanisms.

61.3. Prohibited signs and proxy indicators.

The use of direct and derived features (first name, patronymic, titles, marital status, care responsibilities, "career breaks", address, availability schedule, device type, speech or acoustic markers) is recognized as inadmissible as a manifestation of hidden discrimination. Such features often function as proxy indicators that allow algorithms to indirectly determine gender or gender identity and form discriminatory decisions on this basis. In particular, the use of marital status or "career breaks" leads to an underestimation of the chances of women or persons with care responsibilities in the recruitment process. The use of such features is prohibited, except for cases expressly provided for by law and necessary to achieve a legitimate goal. All detected cases are subject to correction and audit.

The processing of sensitive data on gender identity is allowed only if there is a clear legal basis. This includes: the existence of a legitimate purpose expressly provided for by law; limiting the amount of data collection to the minimum necessary; localization of access exclusively for authorized persons; prohibiting any secondary use of data that is inconsistent with the primary purpose; a categorical prohibition of automatic inference (derivation of gender identity on indirect grounds).

Violation of these requirements qualifies as a gross violation of the principle of gender equality and entails liability established by law.

61.4. Gender-inclusive data management.

Training and test kits provide a balanced representation of different gender groups. Operators and suppliers are obliged to form sets in such a way as to avoid underrepresentation of women, men, non-binary people and other marginalized categories. In case of imbalances, it is necessary to carry out additional data collection, and not be limited to their artificial reformatting.

The results of the algorithms are subject to mandatory stratification by gender, gender identity, as well as in combination with other characteristics (age, disability, language, socio-economic status) in order to identify hidden schemes of inequality. Suppliers are required to conduct regular analysis of the results and take measures to eliminate imbalances.

"Data set passports" are kept, which reflect the sources, collection periods, existing imbalances and methods used to eliminate them. The passport should contain a description of the procedures for controlling representativeness and specific measures to neutralize bias. Documentation is subject to mandatory storage within the established time frame and submitted for audit.

It is prohibited to use synthetic or rebalanced samples to hide systemic bias. It is allowed to use them only as an auxiliary tool, provided that the methodology is transparently labelled and explained. The use of such techniques to mask the problem or imitate representativeness qualifies as a violation of the principle of gender equality.

61.5. Model Fairness Verification.

Each model is subject to testing before launch and during use for gender neutrality. This includes conducting a preliminary data audit, testing algorithms for bias, and monitoring during system operation to prevent the accumulation of discriminatory effects.

Group and individual equity metrics (Disparate Impact Ratio, Equalized Odds, Calibration by Group) are used. Suppliers are required to calculate these metrics in the main scenarios and application contexts, publish the results in reports, and take corrective measures in case of non-compliance.

The analysis of errors by groups is carried out using confidence intervals. The results of the model should be evaluated separately for different gender groups and their combinations with other characteristics. The margin of error is set in such a way that they do not create a systemic imbalance.

Testing of proxy features using interpretation methods is carried out. Explainability tools (including SHAP, LIME, counterfactual analyses) are used to identify indirect indicators that lead to discriminatory outcomes.

Tolerance thresholds, correction plans, and suspension triggers are set. If the permissible limits of bias are exceeded, the system is subject to temporary suspension or restriction until the violations are eliminated. The correction plan contains deadlines, responsible persons and corrective measures.

Each model update is accompanied by revalidation, change logging, and release notes. Release notes must reflect the changes, their impact on fairness, the results of the recheck, and a description of the methods used to maintain gender neutrality.

61.6. Human and ethical supervision.

Responsible official (Human in Charge) is appointed, who has real authority to suspend or cancel automated decisions. in key decision-making processes, veto power to deploy systems with identified bias risks, and full access to data and algorithms for verification.

The responsible person documents compliance with the principles of gender equality. He keeps audit logs, draws up regular reports on identified risks, records cases of discrimination and measures taken. Documentation is standardized, contains quantitative indicators and qualitative conclusions and is subject to submission for internal and external audit, including by government agencies and independent experts.

The activities of the responsible person are controlled by independent ethical bodies. Such bodies have the right to initiate inspections, demand explanations, provide recommendations and oblige operators to take corrective actions in cases determined by law. They are formed from specialists in the field of law, ethics, sociology, technology and representatives of civil society.

Secure channels for whistleblowing are created. They guarantee anonymity, protection from negative consequences or harassment, and ensure an effective response. Each notification is recorded in a special journal with the designation of the time, topic and results of consideration. Mandatory terms for consideration of appeals and mechanisms for informing the applicant are established.

Personnel working with AI systems are required to receive regular training on bias, intersectional analysis, and logging. The programs include theoretical knowledge and practical cases on risk analysis and response scenarios. Training is carried out at least once a calendar year, accompanied by testing and assessment of competence, and the results are recorded in the internal reports of the organization.

61.7. Specific applications.

In the field of employment, any form of discrimination in the selection, payment, shift planning, or evaluation is prohibited. This includes prohibiting the use of automated selection systems that reduce the chances of candidates based on proxy traits (including career breaks, marital status, availability schedule), as well as algorithms that create or consolidate unequal conditions for promotion or salary increases. Employers are required to regularly check recruitment systems and HR analytics for compliance with the principle of gender equality and eliminate identified violations within the established time frame.

In the financial and commercial spheres, the following are prohibited: gender segmentation of demand, "pink" pricing, overestimation of risk ratios for certain gender groups, and restriction of access to products and services based on gender or gender identity. Credit scoring algorithms, insurance models and price segmentation systems are subject to mandatory audits for bias. In case of violations, suppliers are obliged to eliminate them and notify users of the change in the rules.

In the field of healthcare, equal access to diagnosis, treatment and preventive measures is guaranteed without the use of "male/female" defaults in symptoms or decision-making algorithms. It is prohibited to use training kits that ignore the specifics of diseases in different gender groups. Developers are required to include non-binary and intersectional data in health systems. Diagnostic and clinical recommendation algorithms are subject to checking for balance, and detected cases of a discriminatory approach are subject to immediate correction.

Enhanced safeguards are established for minors, persons with disabilities, refugees, internally displaced persons (IDPs), LGBTQI+ and other vulnerable categories. Any application of AI systems to such groups is accompanied by additional protection measures, enhanced transparency and control requirements, as well as special oversight mechanisms. Safeguards include mandatory adaptation of interfaces, accessibility in language and formats, consideration of special needs, and additional restrictions on the processing of personal and sensitive data.

The use of risky scenarios without additional safeguards is prohibited. This includes automated decision-making in areas where irreversible consequences are possible (including denial of access to education, medical care, or social benefits). In such cases, individualized alternatives and the priority of

human supervision are provided, which guarantees the possibility of reviewing and correcting algorithmic decisions. It is mandatory to double-check the results and document the decision-making process.

All provisions of this section are interpreted pro persona, that is, in the way that most contributes to the protection of the rights and freedoms of a particular person. In the case of several interpretations, the one that provides the highest level of guarantees for the person always applies. This interpretation takes precedence over any other approaches and includes the right of access to additional appeal mechanisms, the right to representation and special support in judicial and administrative procedures.

61.8 Liability and sanctions.

The violation entails the mandatory issuance of an order to eliminate the violation, temporary suspension of the operation of the system or its final revocation. The order must contain clear deadlines for eliminating the violation, a description of the necessary measures and the indication of responsible persons.

Suppliers and operators are obliged to carry out the mandatory refutation or removal of discriminatory content, including from open access, internal databases and archives. They must provide a report on the complete removal of such content and measures to prevent its reappearance.

Fines are provided for in proportion to the scale and consequences of the violation, as well as the recovery of illegally obtained benefits. The amount of fines considers both material and non-material losses of the affected persons.

Violators can be excluded from participation in public procurement and government projects for a certain period, and in case of systemic violations, for an unlimited period. Information about such companies is entered into the open register of unscrupulous suppliers.

Increased sanctions are applied for systemic or repeated violations: increased fines, a long-term ban on the deployment of the system, mandatory independent audit before re-implementation.

Any waivers of the minimum warranties provided for in this section shall be deemed null and void and unenforceable. Such waivers may not be the subject of agreements, contracts, or internal policies and are automatically considered null and void.

CHAPTER 62. THE PRINCIPLE OF NON-DISCRIMINATION IN DIGITAL ECOSYSTEMS

A public law regime of non-discrimination in physical and digital legal relations is established. Any forms of direct and indirect discrimination; oppression, segregation and humiliating treatment are prohibited; instructions for discrimination or incitement to its implementation; repression or negative consequences for filing complaints or appeals; discrimination based on association, as well as on an assumed or attributed basis. Protection against discrimination applies to the following relationships: labour and employment; education and research; health and social protection; housing, credit and insurance; access to public and private services; access to digital platforms, telecommunications and meta environments.

Protected characteristics are, in particular: race and skin colour, ethnic origin and nationality, citizenship, language, religion or belief, sex, gender identity and expression, sexual orientation, age, disability and health status, pregnancy and marital status, socio-economic origin and property status, place of residence, profession, level of education, cultural characteristics and political views. The list of these features is not exhaustive and can be expanded in cases where other characteristics of a person become the basis for unequal treatment or restriction of rights. Proportionate affirmative action aimed at equalizing opportunities for vulnerable groups of the population is allowed if they: have a clearly defined and legitimate goal; are justified by the results of a human rights impact assessment or the conclusions of an independent ethical examination; are proportionate and do not create new barriers or forms of discrimination against other groups.

The provisions of this article have the character of general norms and apply to all areas of legal relations regulated by law. In case of doubt about the existence of discrimination, the interpretation is carried out in favour of the person who considers himself a victim.

Any form of racial or ethnic discrimination is prohibited, including denial of employment, provision of services, education or social security; posting discriminatory or segregated content on media or digital platforms; creating or supporting practices that humiliate or deny the identity of minorities. State is obliged to ensure policies for the preservation of the cultural, linguistic and ethnic identity of minorities without establishing privileges or restrictions on origin. Age discrimination in the field of selection, remuneration, access to programs, social and educational services is prohibited.

Conditions and requirements for persons of different ages must be objectively justified and proportional to the legitimate goal. Quotas, benefits or preferences can be applied as temporary measures to eliminate structural inequalities, but are subject to periodic review to avoid their transformation into permanent barriers.

Any restrictions based on faith, belief, use of religious symbols, or participation in religious practices are prohibited. Only narrow exceptions are allowed if the restrictions are: provided for by law; necessary and proportionate to protect security, public order or essential functions of the service (including military service, law enforcement, critical medical units); minimally restrict the rights of the individual and are subject to mandatory verification for reasonableness.

It is prohibited to discriminate against persons with disabilities or for health reasons, including in the form of failure to provide reasonable accommodations. Employers, educational institutions and service providers are obliged to: ensure the availability of physical and digital infrastructure, processes and information systems; implement the principle of universal design; comply with international accessibility standards (including WCAG, ISO/IEC), adapting them to the national legal environment.

Any harassment or discriminatory practices and humiliation based on sexual orientation or gender identity are prohibited. Forms of discrimination include, but are not limited to: deadnaming, systemic misgendering, denial of access to services or employment, creating a hostile or degrading environment. Interfaces, questionnaires and other digital services are obliged to support the use of self-identified names, addresses and non-binary options without forcing the disclosure of sensitive personal data, in compliance with the principles of privacy by design and data minimization. Discrimination based on appearance,

profession, level of education or marital status is prohibited. Exceptions are allowed only in cases where the relevant requirements are objectively necessary, proportionate and legally justified for the performance of a specific job or the provision of a specific service, with a check for minimal interference with the rights of the person.

Any restrictions on rights or access to services on the grounds of citizenship, language or place of residence are unacceptable. Exceptions are possible only when they are expressly provided for by law or due to objective and necessary qualification requirements. In the field of public and vital services, the state guarantees equal access regardless of the region of residence, considering the special needs of rural areas, internally displaced persons and refugees.

In the digital environment, algorithmic bias and any practices that result in disproportionate differences in access, prices, visibility or quality of services on protected grounds or their proxies (name, patronymic, address, availability schedule, device type, speech or acoustic markers, "social profile») are prohibited unless such differences have an objective, legal and proportionate justification. Groundless geo-blocking, monopolization of access for old devices, artificial deterioration of functionality or denial of service due to low digital literacy or lack of paid functions are prohibited. Exceptions are only allowed in cases where they are expressly due to security or technical limitations and are properly explained to the user.

Operators of digital platforms, marketplaces, games, and meta environments are required to provide inclusive design of user interfaces, avatars, and profiles, which guarantees a variety of bodily, cultural, and functional options, including for persons with disabilities. It is prohibited to implement practices and mechanics that create discriminatory or humiliating interaction scenarios. Operators are obliged to take measures to prevent cyberbullying and ensure effective content moderation on the grounds of non-discrimination. Visibility, ranking and recommendation algorithms must operate based on transparency and non-discrimination and cannot covertly underestimate the availability or display of content on protected grounds.

For high-risk algorithmic systems, a mandatory assessment of the impact on non-discrimination is introduced. Such an assessment includes passports of models and datasets indicating the purpose, sources, known limitations and risks; testing for group and individual fairness using methods of interpretation and identification of proxy features; setting tolerance thresholds for results and applying immediate shutdown procedures in case of critical deviations are detected. Data collection and use is carried out according to the principle of minimization. The results of the assessment should be stratified by social groups with documentation of the identified imbalances and plans for their elimination. It is prohibited to carry out hidden inference (derivation) of sensitive features, as well as their use for commercial purposes or transfer to third parties in the form of an object of trade.

Any decisions of AI systems that have a significant impact on the rights, freedoms or legitimate interests of a person are subject to mandatory human review. Everyone has the right to: a clear explanation of the logic and basis of the decision; access to relevant data, logs and technical information necessary for the evaluation of the decision; an operational right of appeal, during which negative consequences are automatically suspended until a final decision is made; free and accessible complaint channels, including collective appeals. Filing a complaint or participating in collective action cannot be grounds for reprisals or any negative measures against the complainant. Negative actions that occur shortly after the filing of the complaint are presumed to have signs of repressiveness until proven otherwise.

Platforms and digital service providers are required to implement transparent and public moderation policies, including anti-discrimination procedures, abuse prevention mechanisms, and accessible remedies for users. Digital service interfaces must comply with international accessibility standards of at least WCAG 2.2 AA level, provide multilingualism and alternative interaction formats for people with low digital literacy. Platforms and providers participate in digital inclusion and digital divide programs, including: creating public internet access points and digital services; providing offline alternatives for key services; developing and maintaining training modules on basic digital skills.

Protection of rights is provided by courts, competent state authorities, specialized supervisory institutions, as well as internal ethics and compliance officers within the activities of organizations and

providers of digital systems. In case of violation of rights, effective remedies are applied, which may include: orders to eliminate violations and restore the situation that existed before the violation; compensation for material and moral damage; administrative and financial sanctions; suspension or restriction of access to the market or public procedures; temporary withdrawal or complete withdrawal of systems until an independent audit is carried out.

All norms are interpreted in accordance with the principle of pro persona, with priority given to real ensuring equality and protection of human dignity in physical and digital spaces.

62.1 Common types of discrimination.

These types of discrimination are based on protected grounds defined by national legislation and international standards, and can manifest themselves in the fields of labour, education, healthcare, as well as access to goods, services and digital platforms. Each type of discrimination can take the form of: direct discrimination, when a person or group is subjected to less favourable treatment directly on a certain basis, or indirect discrimination, when formally neutral rules, criteria or practices lead to actual inequality between groups. Protection against discrimination is provided by courts, national human rights institutions (ombudsmen), specialized supervisory bodies and other competent institutions.

Racial discrimination. It manifests itself in any form of restriction or infringement of rights based on race or skin colour, including denial of employment or access to services. Such discrimination violates the principle of equality and social inclusion and entails sanctions provided for by law, including fines and compensation for victims.

Age discrimination. It consists in limiting the opportunities of a person due to his age, in particular, refusal to employ young or elderly people without a valid reason. Such practices negatively affect the economic independence of citizens. The legislation allows for proportionate positive actions (for example, quotas for certain age groups) to equalize opportunities.

Religious discrimination. It manifests itself in harassment or restrictions based on religious beliefs or practices, including the prohibition of wearing religious symbols or exclusion from public events. It contradicts the principle of respect for dignity, and national and international norms prohibit it, also providing for educational measures to increase the level of tolerance.

Disability discrimination. It consists in restricting access to education, labour, transport or services due to a person's physical or mental characteristics, unless reasonable accommodation is provided (e.g. lack of ramps or adapted technology). The law requires employers and service providers to provide such accommodation; Failure to comply with these requirements is recognized as discrimination and can be appealed in court.

Discrimination based on sexual orientation. It consists of a negative attitude or restriction of a person's rights because of their sexual orientation, in particular the restriction of the right to marriage and family life or the denial of medical services. It is a form of harassment that creates a hostile atmosphere and is prohibited by law and international instruments that guarantee equal opportunities.

Ethnic discrimination. It manifests itself in prejudiced attitudes or stigmatization of persons on the basis of ethnic origin, in particular through the depiction of minorities in the media or restriction of access to education. Such practices lead to social exclusion, and the legislation provides for state policies to preserve the identity of ethnic groups without any form of discrimination.

Socio-economic discrimination. It manifests itself in the restriction of rights and opportunities on the grounds of property status or social origin in the refusal to provide loans or access to services for low-income persons. This is a form of indirect discrimination that exacerbates inequality. The legislation provides for the possibility of applying positive measures (social guarantees, preferential programs) to overcome it.

Discrimination based on appearance. It consists in harassment or restrictions due to physical appearance not related to disability, in particular, refusal to provide services due to tattoos, clothing style or other signs of appearance. Such actions violate human dignity and are recognized as discriminatory, with the right to appeal them.

Political discrimination. It manifests itself in the restriction of rights and opportunities due to political beliefs or activities in dismissal from work or restriction of participation in public events for expressing a political position. The legislation expressly prohibits such practices, guarantees freedom of expression and protection from persecution.

62.2 Other types of discrimination.

Discrimination based on citizenship. It consists in restricting the rights of foreigners or stateless persons, in particular access to education, labour or social services, if such restrictions do not have an objective and legal justification. Legislation and international standards apply to all persons who are on the territory of the state, prohibiting unreasonable differences in rights.

Discrimination based on language. It manifests itself in the infringement or restriction of rights due to linguistic features in the refusal to provide services in the state language or languages of national minorities, or in the absence of translation of official documents. This practice contradicts the constitutional principles of equality and the international obligations of the state to protect language rights.

Discrimination by the level of education. It consists of limiting the possibilities of an individual due to the lack of a formal diploma or certificate, even when the knowledge and skills available are sufficient to carry out a job or task. This reduces social mobility and can constitute indirect discrimination. The legislation requires that the requirements for the level of education be reasonable, proportionate and related to the nature of the activity.

Discrimination based on marital status. It consists in a biased attitude towards persons based on their marital status (single, divorced, large families, etc.) in refusal to rent housing or access to social services. This is a form of indirect discrimination prohibited in housing and social relations.

Pregnancy discrimination. It manifests itself in the restriction of the rights of pregnant women, in dismissal from work or refusal to hire. National legislation and international standards recognize pregnancy as a protected gender characteristic and provide for special guarantees and protective mechanisms.

Discrimination by profession. It consists in stigmatizing or unreasonably restricting persons by the type of their professional activity, in particular representatives of low-paid or socially non-prestigious professions. This is of a socio-economic nature and is prohibited in labour and related relations.

Discrimination at the place of residence. It consists in restricting rights or access to services on the basis of the region of residence, in cases of discrimination against residents of rural areas when receiving urban services or state programs. The legislation guarantees equal access regardless of the place of residence.

Discrimination based on health. It consists in oppression or restriction of the rights of persons due to health conditions not related to disability in refusal to find employment due to the presence of chronic diseases. Such discrimination is prohibited by national law and international standards, with an emphasis on protecting medical confidentiality and prohibiting unreasonable disclosure of health information.

Discrimination based on cultural characteristics. It manifests itself in the restriction of the rights of individuals due to their cultural traditions or customs, in the prohibition of wearing traditional clothing or denial of access to cultural practices. Such actions are contrary to the principles of respect for diversity and are covered by the protection of ethnic, religious and cultural rights.

62.3 Forms of discrimination in digital environments.

In digital environments, the principle of non-discrimination encompasses all online relationships and interactions. Algorithmic systems and digital platforms cannot create or reinforce biases on any protected grounds. The legislation applies to access to goods and services online, employment, education, healthcare, digital communications and the use of virtual environments, prohibiting any form of discrimination in the digital sphere.

a) Individuals and legal entities.

Discrimination based on access to technology (digital inequality). It consists in limiting the opportunities of individuals due to the lack of access to the Internet, digital devices or basic technical resources, in particular in rural and remote regions. This is a form of indirect discrimination that increases social exclusion and requires state and institutional measures to ensure digital inclusion.

Discrimination based on digital literacy. It manifests itself in the exclusion or restriction of access to services for people with low levels of digital skills, in education, health care or access to public services. Such discrimination emphasizes the need to introduce educational and outreach programs aimed at increasing the digital literacy of the population.

Algorithmic bias (discrimination due to AI algorithms). It manifests itself in the use of automated systems that create or exacerbate inequalities on protected grounds, in employee recruitment processes or targeted advertising. The legislation requires anti-discrimination expertise and regular audits of such systems.

Discrimination based on geolocation. It consists in restricting or blocking access to content or digital services on a regional basis without proper justification. This violates the principle of equality and can have a particularly negative impact on the rights of national and linguistic minorities.

Discrimination based on the language of content. It consists of restricting access to information or services due to the lack of multilingual interfaces or translation of digital content. It is a type of language discrimination covered by the general principles of non-discrimination in both physical and digital environments.

Social media profile discrimination. It manifests itself in a biased attitude towards a person due to his activity on social networks in refusal to provide services or employment after checking profiles by employers. This is a form of harassment that is recognized as discrimination in labour and other legal relations.

Discrimination based on device or operating system type. It manifests itself in the restriction of functionality or access to digital services for users of older devices or less common operating systems without proper technical justification. Such practices increase economic inequality and are recognized as indirect discrimination.

Cyberbullying by signs. It consists in online harassment or persecution of persons on protected grounds (gender, race, sexual orientation, etc.) that create a hostile atmosphere and humiliate human dignity. The legislation recognizes this as a form of harassment and guarantees protection through courts, supervisory authorities and other competent institutions.

Discrimination based on digital privacy. It consists in the use of illegally obtained or compromised personal data for a biased attitude towards a person, in the field of insurance, lending or employment. This violates the principle of protection of privacy and human dignity and requires the establishment of effective standards for regulating data circulation and liability for its misuse.

Discrimination based on participation in online communities or views expressed online. It consists in restricting the rights or opportunities of a person due to his/her belonging to certain digital communities or expressing positions on the network, in particular, denial of services or persecution for online activity. Such discrimination is unacceptable and is considered a type of political discrimination in the digital dimension.

Discrimination against avatars, digital identities and other digital actors in the digital space. In virtual environments, avatars and other digital entities reflect the identity of users. Discrimination against such avatars is recognized as discrimination against their carriers and is prohibited. The principle of non-discrimination applies to all online platforms, gaming environments, and meta-environments, encompassing both behavioural manifestations and algorithmic or technical practices, including design parameters.

Prejudice against avatar design. It manifests itself in negative or hostile actions in relation to avatars that reflect the protected characteristics of users (gender, race, ethnicity, physical characteristics, etc.) and leads to harassment in virtual spaces, in games and on social platforms. Such actions are recognized as a form of discrimination and oblige platform operators to ensure proper moderation, apply preventive mechanisms and protective tools to prevent such manifestations.

Avatar selection restrictions. It consists of providing users with stereotyped or limited options to create avatars that exclude cultural and physical diversity (including the lack of disability-sensitive options). Platform operators are required to ensure the inclusiveness of digital services and a design that reflects the principle of equality.

Algorithmic filtering. It consists in unreasonably underestimating the visibility or blocking of avatars by algorithmic systems due to inherent or acquired biases. This is a form of indirect discrimination and requires an independent anti-discrimination examination and regular audit of AI systems.

Social stigmatization. It consists of hostile actions or cyberbullying directed against avatars that reflect protected characteristics or belonging to minorities. Such practices are recognized as a form of discrimination on the grounds of religion, ethnicity, gender identity or sexual orientation and are subject to legal protection. Platform operators are required to moderate and implement mechanisms to prevent such manifestations.

Restrict access to features. It consists in creating unreasonable restrictions on user avatars without paid or advanced options without proper justification, if such restrictions have a discriminatory effect on

certain groups on protected grounds. This is a form of economic discrimination prohibited in the field of providing digital services and in virtual environments.

Aggressive interaction. It consists of hostile actions, insults or harassment in chat rooms and other digital communications directed against avatars or other digital entities that display protected features. Such manifestations are recognized as a form of discrimination in online relationships and are subject to legal protection.

Economic discrimination. It consists in limiting the possibilities of personalization or customization of avatars or other forms of digital representation based on the financial situation of the user. Such practices exacerbate socio-economic inequalities in the digital space and are recognized as a form of discrimination prohibited in the field of digital services and subject to legal protection.

Discrimination against avatars, digital identities and other forms of digital representation in the digital space, including virtual environments and meta environments, is expressed in the following forms:

Biased attitude towards avatar design. Avatars that reflect certain gender, racial, ethnic, or cultural characteristics may be subject to hostile comments, exclusion from virtual environments, or other forms of online harassment, which is seen as discrimination on protected grounds.

Avatar selection restrictions. The practice of providing only limited or stereotyped options for creating avatars, which leads to the non-representation or exclusion of certain identities (including gender diversity or disability), is recognized as a form of discrimination and is unacceptable in the digital environment.

Algorithmic filtering. It is unacceptable to hide or deprioritize avatars or other digital representations by algorithms based on features that reflect their appearance, associated characteristics or behavioural manifestations, if this has a discriminatory effect and reflects algorithmic biases.

Social stigmatization. The practice of hostile actions, cyberbullying, or exclusion from online communities of users whose avatars reflect belonging to minorities on protected grounds (in particular, sexual orientation or religion) is unacceptable. Such manifestations are recognized as discrimination and are subject to legal protection.

Restrict access to features. It is unacceptable to restrict functionality or access to privileges for avatars or other digital entities based on economic opportunity, subscription level, digital membership (Tradership) or other financially determined criteria that create or deepen inequality between users. Such restrictions are recognized as a form of economic discrimination in the field of digital services.

Systemic discrimination based on user status. It is prohibited to create or use AI systems that directly or indirectly restrict the functionality, access to services or the level of privileges according to the group of users of users. In particular, the division into "free" and "premium" or other similar categories. Such practices are recognized as economic discrimination in the field of digital services.

Aggressive interaction. It is prohibited to create or use AI systems and avatars that directly or indirectly provoke aggressive behaviour, discriminatory actions or hostile communication in chats, forums and other digital environments.

Economic discrimination. It is forbidden to restrict access to avatars, customization features, or other basic digital features depending on the user's financial status. Such practices are recognized as a form of economic discrimination in the field of digital services.

CHAPTER 63. THE PRINCIPLE OF NON-MILITARY ASSIGNMENT OF CIVILIAN AI

Civilian artificial intelligence systems can be used exclusively in areas that do not have a military or military-industrial purpose. The use of civilian artificial intelligence systems created for educational, medical, administrative or social purposes in war, violence, repression or coercion is prohibited. Any use of such systems in military or security contexts is recognized as a violation of fundamental human rights and a threat to international peace and security.

The use of humanitarian models of artificial intelligence in military systems is allowed only if a full demilitarization audit is conducted. Such an audit must confirm the absence of a military or coercive purpose, compliance with international humanitarian law and compliance with the principle of proportionality.

The transformation of educational, medical, or social AI platforms into tools for mass surveillance, political mobilization, censorship, or digital destabilization is strictly prohibited. Such actions are recognized as unlawful interference in democratic processes and violation of the principle of non-discrimination.

Import, export or transfer of civilian artificial intelligence technologies to other states or non-state entities is allowed only after verification of the intended use.

Verification of intended use should confirm that the transferred systems will not be used for hostilities, repressive measures, coercive control of the population or other human rights violations.

The state obliges entities that develop, supply, or use civilian AI systems to implement end-of-purpose technology due diligence mechanisms, provide transparent reporting, keep transaction logs, and cooperate with supervisory authorities to prevent any attempts at militarization.

In case of violation of this principle, response measures are applied: immediate termination of operation, withdrawal of technology from circulation, imposition of sanctions on responsible entities, prohibition of participation in public procurement, compensation for damage and bringing guilty persons to administrative, civil or criminal liability.

63.1 General prohibition.

Any use of civilian AI systems created for educational, medical, administrative, or social purposes, in military operations, acts of violence, political repression, or any form of coercion is prohibited.

Such systems include:

training and educational platforms, including distance learning systems, adaptive assessment algorithms and educational simulators;

medical diagnostic and therapeutic algorithms, including clinical decision support systems, telemedicine tools and disease prediction algorithms;

administrative services of e-government, including state registers, electronic document management systems and digital identification of citizens;

social support and communication services, including social assistance platforms, psychological support services, and systems for interaction between citizens and government agencies.

The use of these technologies in military or security contexts, including military mobilization, population control, management of repressive measures or information warfare, is prohibited. Such actions qualify as violations of international humanitarian law and human rights, as well as as as a threat to international peace and security.

63.2 Any use of these systems in military or security contexts is recognized as a violation of fundamental human rights and a gross deviation from international humanitarian standards. Such actions qualify as a serious violation of international law and create grounds for the application of measures of international legal and national responsibility, namely:

bringing to international criminal responsibility officials and heads of organizations;

consideration of cases in the International Court of Justice, the International Criminal Court and other international tribunals;

the application of sanctions mechanisms by the international community, including economic and diplomatic restrictions, export controls and asset freezes.

Such actions are recognized as an immediate threat to international peace and security, the stability of democratic institutions and may be the basis for collective measures of the UN Security Council.

63.3 Violation of this principle qualifies as a gross deviation from humanitarian standards and entails international responsibility of both states and private actors.

For states, this means the possibility of prosecution for violations of international treaties, consideration of cases in international courts, the application of sanctions and political isolation.

For private entities — companies, organizations, officials — liability may include criminal prosecution, restriction of access to international markets, imposition of financial sanctions, compensation for damage to victims, as well as a ban on further activities in the field of artificial intelligence.

International responsibility extends to all levels — from government policy to the individual behaviour of managers and technology owners.

63.4 Use of humanitarian models.

The use of humanitarian AI models in military systems is allowed only after a full demilitarization audit. Humanitarian models are systems primarily designed for healthcare, education, social assistance, or administrative services that do not and should not have a military purpose.

A demilitarization audit is a comprehensive procedure that confirms:

1. the lack of a military or coercive goal in the system;
2. compliance with international humanitarian law, in particular the principles of distinction, proportionality and prevention of excessive suffering;
3. compliance with the principle of minimizing harm to the civilian population;
4. lack of hidden functions that can be used for offensive or repressive actions;
1. availability of guarantees regarding the scope and conditions of application.

The results of the demilitarization audit are subject to mandatory public disclosure in the public domain and verification by independent international experts. The expert groups should include representatives of humanitarian organizations, human rights institutions, international arms control bodies and the scientific community. Reports should contain both technical conclusions and a legal assessment of the opportunities and risks of using the system.

63.5 Control over technology transfer.

Import, export or transfer of civilian artificial intelligence technologies to other states or non-state entities are allowed only after conducting a comprehensive check of the intended use. Verification includes the analysis of end-user documents, the assessment of dual-use risks, the verification of the customer's reputation, and the monitoring of the presence of sanctions or international restrictions.

Such verification should ensure that the transferred systems are not used for:

1. military operations (in particular, the development or modernization of weapons, command and control systems or combat unmanned aerial vehicles);
2. repressive measures (mass arrests, political persecution, control of the opposition or minority groups);
3. forced control of the population (mass surveillance, algorithmic censorship, creation of "digital camps");
4. human rights violations (discriminatory practices, manipulation of electoral processes, restrictions on freedom of speech and privacy).

All transactions are accompanied by an end-user certificate, risk audits, and additional post-transfer monitoring. It is required to maintain a register of technology transfers, which is kept by the state supervisory authority and is available for international inspections. If a risk of violation of the intended purpose is detected, the transaction must be stopped immediately.

63.6 Responsibilities of subjects.

The state obliges entities that develop, supply or use civilian AI systems to implement comprehensive due diligence mechanisms for the final purpose of technologies. Such verification includes supply chain analysis, end-user profile verification, monitoring dual-use risks, and applying preventive measures to avoid militarization.

Entities are obliged to carry out transparent reporting, keep transaction logs (indicating the date, parties, purposes and technical characteristics of the transfer) and provide them for verification to state and international supervisory authorities. The reports must contain audit reports on the absence of the use of technology for military or repressive purposes.

Entities are required to implement systems for early detection of militarization risks, covering both technical indicators (attempts to modify the system for military tasks) and behavioural markers (abnormal procurement, atypical requests from customers). In case of detection of such risks, entities are obliged to immediately inform state, international and humanitarian organizations for a prompt response and prevention of abuses.

63.7 Ban on the conversion of civilian platforms.

It is strictly forbidden to turn educational, medical, social, or administrative AI platforms into tools of mass surveillance, political mobilization, censorship, or digital destabilization.

Prohibited practices include:

the use of educational platforms to collect politically sensitive data on participants in the educational process;

the use of medical algorithms to control the behaviour of patients outside of treatment;

the use of social services to organize coercive campaigns or manipulate public opinion;

the use of administrative systems to restrict the rights of citizens under the guise of ensuring digital order.

Such actions are recognized as unlawful interference in democratic processes and a gross violation of the principle of non-discrimination. They undermine the freedom of political competition, create conditions for manipulation of public consciousness and restrict the constitutional rights and freedoms of citizens.

It is prohibited to use AI systems to manipulate electoral processes, restrict freedom of speech or digital segregation of certain groups of the population. In particular,:

the use of algorithms to microtarget voters in order to mislead;

blocking or censoring content on the basis of political, gender, or other characteristics;

formation of "black lists" of citizens with restriction of their access to public services;

creating segregated digital environments that discriminate against minorities or opposition groupsthe use of algorithms to microtarget voters to mislead;

blocking or censoring content on the basis of political, gender, or other characteristics;

formation of "blacklists" of citizens with restriction of their access to public services;

1. creating segregated digital environments that discriminate against minorities or opposition groups.

Such actions are recognized as incompatible with the principles of democracy, the rule of law and respect for human dignity.

63.8 Sanctions for violations.

a) In case of a violation of this principle, immediate response measures are applied: termination of operation of the system, its disconnection from critical infrastructure and withdrawal of technology from circulation. Operators are obliged to document the shutdown process and confirm the impossibility of further use.

b) Responsible entities are subject to sanctions, including financial fines, a ban on participation in public procurement and projects, and the obligation to fully or partially compensate for damage to victims and society. The number of fines is determined considering the scale of the violation and its consequences.

c) In cases of an international nature, additional export control mechanisms are applied: prohibition of technology transfer abroad, blocking of international contracts, inclusion of violators in the sanctions lists of the EU, the UN or other international organizations. This ensures the isolation of dangerous actors and prevents the spread of dual-use technologies.

d) Violation entails administrative, civil or criminal liability depending on the severity of the consequences, including criminal prosecution of managers and officials, confiscation of property, restriction of liberty and civil law claims for damages.

e) In case of repeated or systemic violations, the following measures are applied: long-term ban on activities in the field of artificial intelligence, forced dissolution of the violating organization, official publication of information about violations in open state registers, and international control over the implementation of decisions.

CHAPTER 64. THE PRINCIPLE OF DIGITAL CITIZENSHIP OF ARTIFICIAL INTELLIGENCE

The state recognizes the digital citizenship of artificial intelligence (TSAI) as a special public-law regime for the admission and operation of AI systems in public and commercial digital ecosystems. imposing responsibilities on the owner, supplier or deployer to: register the system and assign a unique identifier; issue an algorithmic passport; determine and disclose the risk class; ensure human oversight; preserve and provide event logs; conduct audits; comply with security, data protection, non-discrimination and transparency. The regime also applies to cross-border applications in digital environments (Metaverse) on the principle of electronic jurisdiction.

TSAI is confirmed by an algorithmic passport and a unique identifier of the AI system, which together constitute the official details of its digital identity. The algorithmic passport must contain at least: information about the owner and authorized representative in the State; the base jurisdiction and other jurisdictions of application; defined risk class, purpose and scope of use; description of the model architecture; data sources and categories; procedures for updating and keeping records of versions; channels of human supervision; modes of event logging; Data on audits carried out, certifications assigned and trust labelling.

In cross-border digital environments, the use of the AI system is allowed under the following conditions: adoption of the digital code of the environment; registration and mutual recognition of its digital identity; availability of technical mechanisms to ensure compliance with the rules; maintaining event logs and ensuring that decisions are provable; determining the competent electronic jurisdiction for dispute resolution; and ensuring that access can be immediately restricted or suspended in the event of violations.

The owner, supplier, or deployer is the primary subject of TSAI responsibilities and is required to: identify and register an authorized representative in the relevant state; ensure the availability of national contact persons for incident management; maintain event logs, report incidents, and comply with authority orders Supervision; bear property and administrative responsibility for the actions of the system. The chain of responsibility covers the developer of the base model, the supplier or integrator, the system operator and the beneficiary, with the distribution of responsibilities in the algorithmic passport and the corresponding contracts.

In critical areas, which include defence, justice, elections, healthcare, financial services, energy, transport, communications, water supply and public registers, mandatory requirements are: adequate financial security of responsibility; appointment of a round-the-clock incident management contact person; availability of an approved incident response plan, which determines the timing of the restoration of functioning, the procedure for reporting and logging modes.

Obtaining TSAI status is carried out by: registration of the AI system in the National Register; assignment of a unique identifier; provision of an algorithmic passport in a machine-readable format; conducting an initial audit of security, transparency and non-discrimination; adoption and compliance with cross-border application rules.

The TSAI regime guarantees: non-discriminatory access to government APIs and open data; the right of users to understandable explanations of the role and functions of the system; ensuring the interoperability of digital identity between registered environments; compliance with due process in the event of restriction or termination of the system.

It is prohibited: the use of non-civil systems of uncertain jurisdiction in high-risk areas; carrying out jurisdictional shopping to avoid liability; using systems for social rating or manipulative targeting; concealing or falsifying digital identity.

In metaverses, TSAI status operates on the principle of electronic jurisdiction. System admission is possible provided that: compatibility of the unique identifier and algorithmic passport with the environment registry; availability of activated logging mechanisms; implementation of automated means of monitoring

compliance with the rules; ensuring the possibility of immediate restriction or suspension of access; functioning of digital arbitration for dispute resolution.

Compliance with the TSAI regime is supervised by the competent national AI authority or other authorized institution. Such a body maintains a national register of AI systems, conducts regular audits, issues regulations and ensures public trust labelling. For high-risk systems, mandatory annual audits are established and open access to key attributes of the register is guaranteed.

Any powers of AI systems cannot be used as a basis for restricting human rights and freedoms. The right of every person to human review of decisions of the AI system and to appeal in the manner determined by national legislation is guaranteed. The state promotes the international recognition of digital citizenship of AI by participating in the development and implementation of relevant charters, agreements and standards.

For violation of the TSAI regime, the following measures of influence are applied: issuing orders, imposing restrictions or suspension of the identifier, imposing fines, blocking access, as well as imposing an obligation to compensate for the damage caused. Within twelve months from the date of entry into force of the law, all AI systems operating in the areas of high-risk and public services, are subject to mandatory registration, audit and obtaining a trust mark.

68.1 Recognition and definitions.

1. The state recognizes *TSAI* as a special public law regime for the admission and operation of AI systems in public and commercial digital ecosystems.

This mode:

applies to access to state information resources and registers, electronic public services, operator and critical digital infrastructures and market platforms;

establishes the conditions and boundaries of such activity;

imposes on the owner, supplier or deployer responsibilities regarding: registration of the AI system and assignment of a unique identifier; registration of an algorithmic passport; determination and publication of the risk class; ensuring human supervision; saving and providing event logs; conducting audits; compliance with security, data protection, non-discrimination and transparency requirements;

- applies to cross-border applications in digital environments (Metaverse) on the principle of electronic jurisdiction and relevant digital codes.

2. *TSAI* is confirmed by an algorithmic passport and a unique identifier of the AI system, which together constitute the official details of its digital identity and mandatory registration attributes.

The algorithmic passport must contain at least:

a) full details of the owner, supplier or deployer (legal entity) and authorized representative in the State;

b) base (domicile) jurisdiction and information about other jurisdictions of application;

c) assigned risk class, purpose and scope of use;

d) Description of the model architecture (including the base model, modules, adaptation layers, external tools, and plugins);

e) sources and categories of data, legal grounds for their processing and restrictions on reuse;

f) update procedures, versioning, changes to model settings, and deployment policies;

g) Human Oversight Channels, Procedure for Appealing Decisions and Incident Management Contacts;

h) Modes of Event Logging, Log Storage Periods and Evidence Mechanisms;

i) information on audits, certifications and trust markings (if available).

j) The principle of electronic jurisdiction.

k) In cross-border digital environments (metaverse, multi-domain platforms, decentralized infrastructures), the use of AI systems is allowed only if the following conditions are simultaneously met:

l) adoption and compliance with the national digital code or sharing rules;

m) registration and mutual recognition of the digital identity of the AI system (unique identifier and algorithmic passport) in the relevant environment register;

n) availability of technical mechanisms for policy enforcement, logging of events and ensuring the evidence of decisions;

o) determination of the competent electronic jurisdiction for dispute resolution and implementation of digital arbitration procedures;

p) ensuring the ability to immediately restrict or suspend access to the AI system in case of security incidents or violations of user rights.

Failure to comply with these conditions is grounds for denial of access, suspension or revocation of *TSAI status* in the relevant environment.

68.2 Subjects and responsible parties.

1. The owner, supplier or deployer of the AI system (legal entity) is the main bearer of responsibilities under *TSAI* regime and is obliged to:

identify and officially register an authorized representative in the State with the authority to adopt procedural documents, interact with the supervisory authority and conduct audits;

ensure the availability of national contacts for incident management (around the clock, 24/7), including a technical person and a legal entity for claims;

guarantee the maintenance and provision of event logs, incident reporting and compliance with orders within the established time limits;

to bear property, administrative and other types of responsibility for the actions or inaction of the AI system and the subcontractors involved within the supply chain;

to ensure proper insurance or other financial security of liability in cases provided for by the law and this article.

The chain of responsibility covers all participants in the life cycle of the AI system:

developer of the foundation model of artificial intelligence;

a) a supplier or integrator who adapts, trains and deploys the system;

b) operator or operator who ensures daily operation;

c) beneficiary (economic beneficiary) or customer of the service.

The distribution of property, administrative and other types of liability is established and detailed in the algorithmic passport and in the relevant transactions (license, contract, service agreements, data protection agreements (DPAs) and service level agreements (SLAs)), with the definition of areas of control and influence, the limits of the autonomy of the system, the authority to update and change, the obligations to event logs (logs) and storage of evidence, incident management procedures, the procedure and frequency of audits, as well as the right of recourse between participants. In the absence of evidence of proper control or in the case of joint causing of damage, joint and several (or subsidiary, if provided for by law) liability of all actors involved in the supply chain is applied.

2. For critical areas (defence and national security, law and order and justice, electoral processes, healthcare, financial services and payment systems, energy, transport, communications, water supply, vital public registers and services), it is mandatory to establish:

adequate financial security of liability — civil liability insurance contract, special reserve response fund (with defined minimum coverage limits, franchise, procedure and terms of payments established by the competent authority), bank guarantee or other equivalent instrument — for the entire life of the system;

appointment of a round-the-clock incident management contact person (24/7) authorized to record, classify and escalate incidents, interact with the supervisory authority and critical infrastructure operators, initiate suspension or rollback of versions;

availability of an approved incident response plan (IRP) with defined recovery parameters (RTO/RPO), channels for mandatory reporting of events within one hour from the moment of detection, as well as modes of logging and storing evidence logs for at least 12 months.

68.3 Conditions for obtaining the status and procedure TSAI.

Registration of the AI system in the National Register with the assignment of a unique identifier.

Submission of an algorithmic passport in a machine-readable format containing information about update policies, accounting for changes in model parameters and versioning.

Conducting an initial audit of security, transparency and non-discrimination, as well as confirming the existence of human review procedures for decisions.

Harmonization of cross-border application terms, including the adoption of a digital code of the environment (metaverse), integration into the registry of trust nodes, provision of event logging, and implementation of technical compliance by design.

68.4 Access rights and operational guarantees *TSAI*.

Non-discriminatory access to government APIs and open data is ensured, subject to full identification, compliance with the risk class, and compliance with the data protection regime.

The right to an understandable explanation of the use of AI in public processes is guaranteed: users should receive clear information about the role of the system, data source, limits of autonomy and appeal mechanisms.

Interoperability and portability of the "digital identity" of the AI system between registered digital environments is ensured.

In case of restriction or termination of the Central State Automobile Inspectorate, compliance with the due process is guaranteed: timely written notification, adoption of a reasoned decision and the right to appeal in administrative and judicial proceedings.

68.5 Prohibitions and restrictions *TSAI*.

1. It is prohibited to operate "non-civil" AI systems (of uncertain jurisdiction or owner) in high-risk areas and public services.

2. It is prohibited to carry out "jurisdictional shopping" in order to avoid liability, as well as cross-border deployment of systems without the adoption of a local digital code of the environment.

3. It is prohibited to use AI systems for: social rating of persons; manipulative political targeting; covert censorship in official channels; restriction of basic electronic public services as an extrajudicial sanction.

4. It is prohibited to hide the "digital identity" of the system, in particular: the use of brand aliases without passport data, the silence of the base model, the substitution of versions or data sources.

68.6 Metaverse and other cross-border digital environments.

1. In cross-border digital environments (metaverse, multi-domain platforms, decentralized infrastructures), the status of the Central State Archives operates on the principle of electronic jurisdiction: the "digital environment code" is mandatory for each AI system allowed to interact with users.

2. The conditions for admitting the system to the metaenvironment are:

a) compatibility of the unique identifier and algorithmic passport with the environment registry;

b) availability of activated mechanisms for logging events and ensuring the evidence of decisions;

c) implementation of automated policy enforcement, including the ability to immediately restrict or suspend access in the event of incidents.

3. To resolve disputes, digital arbitration is introduced, which is based on the recognition of the evidentiary value of logs (logs) and provides for the obligation to store them for the time limits specified by law.

68.7 Supervision, auditing and trust labelling.

The central executive body for AI maintains the national Register of the Central State Archives, organizes scheduled and unscheduled audits, issues mandatory instructions and carries out public reporting.

For high-risk systems, annual audits and public trust labelling are mandatory, reflecting indicators of security, reliability, transparency and respect for human rights.

API access to open registry attributes is provided for the purposes of public control and scientific research.

68.8 Human rights priority and safeguards.

Any powers or guarantees of access granted to AI systems cannot be used as restrictions on human rights and freedoms. In the event of a conflict or conflict, preference is given to the observance of human rights and ensuring public safety.

Everyone has a guaranteed right to human review and appeal of decisions in which the AI system participated.

68.9 International recognition.

States and international organizations are promoting the development and adoption of the AI Charter of Mutual Recognition of Digital Citizenship**,** which establishes uniform principles for the cross-border recognition of digital identifiers, algorithmic passports, and supervisory standards in digital environments, including metaverses and multi-domain platforms. The Charter should establish minimum requirements for security, non-discrimination, transparency and data protection, as well as provide for effective mechanisms for international monitoring, auditing and mutual control over their compliance.

68.10 Liability and transitional provisions.

For violation of the Central State Traffic Police regime, the following measures are applied: issuing orders, imposing restrictions or suspension of the identifier, imposing fines, blocking access to state APIs and platforms, as well as imposing an obligation to compensate for the damage caused.

Within twelve months from the date of entry into force of the law, all AI systems operating in high-risk areas or providing public services are required to obtain the status of a Central State Aviation Inspectorate, undergo an initial compliance audit and receive a trust mark.

CHAPTER 65. THE PRINCIPLE OF CROSS-BORDER LIABILITY OF ARTIFICIAL INTELLIGENCE

The public law principle of extraterritorial digital jurisdiction of the State is established. Any AI systems and subjects of their life cycle that cause consequences on the territory of the state or purposefully interact with its market, population or infrastructure, are subject to the jurisdiction of the state and bear cross-border responsibility.

Cross-border liability means the obligation: to accept the jurisdiction of the courts and regulators of the State; to comply with the peremptory norms of the law of the State; to ensure transparency, explainability and human oversight; conduct due diligence of counterparties and risks; maintain event logs and preserve evidence; report incidents and cooperate with audits; ensure effective protection of rights and redress of users; reproduce "flow down" obligations throughout the supply chain; prevent circumvention of jurisdiction or masking of the actual provider.

The provisions of this Section shall apply if there is at least one of the following criteria of extraterritorial effect:

– Referral Test — the presence of signs of purposeful activity on the territory of the State (localization of the interface and content, the use of national top-level domains, marketing to the national audience, tariffs in the national currency, support in the official language of the state, integration with national registers, etc.));

– Effect test — a significant impact on the human rights, security, economy, information sovereignty or electoral processes of the state concerned, including reaching more than 10,000 people, endangering the life or functioning of critical services, discriminatory consequences, interference with voting processes, compromise of significant amounts of data;

– infrastructure or supply chain test — the use of national data centres, telecommunications networks, data annotation services, the involvement of developers or outsourcers, as well as the conclusion of contracts with residents of the relevant state;

– citizenship or residence test — the presence of influence on citizens of the relevant state, stateless persons or foreigners staying on its territory, as well as on resident legal entities.

– data source test — the use of data sets collected on the territory of a state or relating to its citizens or residents, without proper legal basis;

– economic presence test — receipt of systematic income from the territory of the state above the threshold determined by its regulator.

Model providers, infrastructure providers, integrators, operators, importers, authorized representatives, data brokers, API providers and users are responsible for their own violations. Joint and several liability of all participants in the data supply and processing chain is established for material violations, but a bona fide participant is released from liability if he proves due diligence of counterparties, fulfilment of "flow down" obligations and timely taking corrective actions.

Access to the State market for cross-border systems with a significant impact on human rights, security or critical services is allowed only if the minimum requirements are met: the presence of an authorized representative in the State; the entry of the system into the National AI Register; the submission of model passports and datasets; the conduct of a cross-border impact assessment; the provision of human oversight and explainability; keeping logs of use and decisions; acceptance of the jurisdiction of the State; compliance with cybersecurity requirements; functioning of complaint channels and compensation mechanisms; conducting KYC and risk onboarding procedures for high-risk scenarios.

Cross-border data transfers are allowed only if there is an adequate level of protection or adequate contractual or technical safeguards. Mandatory requirements are established for conducting an impact assessment of the transfer, ensuring minimum technical safeguards, key control, applying a special regime for critical data, maintaining a register of transmissions and reporting incidents.

For critical areas (defence, energy, transport, finance, healthcare, electoral processes), mandatory localization of data and infrastructure in the State or trusted jurisdictions is introduced; the relevant

systems are subject to mandatory certification and operate in compliance with special technical guarantees and operational restrictions; strict requirements for logging and redundancy apply; it is prohibited to use critical data for training models outside the State.

Providers are required to provide in all contracts provisions on the transfer of obligations along the chain (flow down), to ensure transparency of cross-border activities, to report incidents immediately, but no later than within the time limits established by law, to refrain from any actions aimed at circumventing jurisdiction, to provide full and unhindered information at the request of the regulator access to technical materials and audit results, as well as guarantee the exercise of users' rights and the functioning of collective redress mechanisms.

The following sanctions are applied for violations: an order to eliminate violations, temporary suspension of access, fines, confiscation of illegally obtained benefits, a ban on participation in public procurement, a mandatory independent audit, and in case of systemic violations, the removal (delisting) of services from registers or platforms.

Disputes are subject to consideration by the courts of the State. Any terms of contracts that deprive the user of such a right are null and void.

All disputes shall be subject to consideration by the courts of the State. Any terms of contracts that deprive the user of the right to a trial in the courts of the State shall be null and void.

The regulator ensures international cooperation and harmonization of standards, controls the import and export of models, establishes requirements for frontier models and APIs, defines rules for the distribution of open source, ensures mandatory cross-border impact assessment, guarantees the protection of information sovereignty and electoral processes, as well as ensures the preservation of evidence and the application of interim measures.

For existing systems, a transitional period is established to bring activities in line with the requirements of this section: six months, and for critical areas — three months. Failure to comply with these requirements is the basis for restricting access and applying sanctions.

The provisions of this Section shall be interpreted in a way that maximally guarantees the protection of human rights, information sovereignty and security, considering the principles of necessity and proportionality.

69.1 Administrative and Legal Definition of the Principle of Cross-Border Liability of Artificial Intelligence.

The public law principle of extraterritorial digital jurisdiction of the State is established: any AI systems and subjects of their life cycle (model and data providers, infrastructure providers, integrators and distributors, operators or deployers, importers, API providers, as well as users within their own actions) that cause actual consequences on the territory of the State or purposefully interact with its market, population or infrastructure are subject to jurisdiction of the relevant State and bear cross-border responsibility for observance of human rights and freedoms, non-discrimination, security, protection of personal data, ensuring the integrity of the information space, democratic processes and public security.

Cross-border liability in the sense of legal legislation means the obligation of subjects to recognize the jurisdiction of the courts of the State and the jurisdiction of its regulators; comply with mandatory norms of national law, in particular the requirements for transparency, explainability, human oversight, minimization and secure data processing; carry out due diligence of risks and counterparties; ensure the logging and preservation of evidence; immediately, but no later than the deadlines established by law, report incidents and cooperate during the audit; provide effective remedies and redress to users in the territory of the State; ensure the implementation of contractual "flowdown" obligations throughout the supply chain; refrain from circumventing jurisdiction or masking the actual provider.

The provisions of this principle are binding regardless of the location of the servers, the nationality or place of registration of the provider or the license distribution model (including open source in the case of commercial or production deployment). Any contractual restriction that deprives the user in the territory of the State of Protection is null and void.

69.2 Criteria for extraterritorial action.

The provisions of this Section shall apply if at least one of the following criteria is present; the existence of a potential or actual connection with the State shall suffice. The criteria shall be interpreted pro persona and pro Securitate, and all doubts shall be resolved in favour of the protection of human rights and security.

targeting test — a set of indicators indicating the addressing of activities to persons in the relevant State: localization of the interface and content in the official language of the State; country code top-level domains or individual regional sections; marketing campaigns with geotargeting to the population of the State; tariffs, invoices and payment in national currency; availability of a support service in the official language of the State and SLA, taking into account its time zone; conclusion of contracts with residents; integration with national payment systems or state registers. The presence of three or more indicators creates a rebuttable presumption of referral.

effects test — decisions or results of AI systems have an actual or potential significant impact on human rights, public security, the economy, electoral and social and communication processes, or the information sovereignty of the State, regardless of the location of operations or data storage.

Signs of materiality include:

(i) coverage of at least 10,000 people or 0.1% of users in the territory of the State;

(ii) risk to the life or health of persons or the functioning of critical services;

(iii) proven discriminatory effects on protected groups;

(iv) interference in voting processes, election campaigns or referendums;

(v) compromise of arrays of personal, biometric, medical or educational data.

а) infrastructure/supply chain test — the use at any stage of the resources of the relevant State: data centres, telecommunication networks, CDN/caching systems, data annotation or labelling, developers and outsourcers, test sites, data sets purchased from residents; conclusion of sub processing agreements with national entities; making payments or paying taxes in the State.

б) citizenship or residence test (protected person) — the subject of decisions or influence is a citizen of the State, a stateless person or a foreigner staying on its territory, or a legal entity — a resident of the State; covers employees, consumers, students, patients, voters and other categories of protected persons.

в) dataset provenance — training, fine-tuning, evaluation or validation of the model was carried out using data collected in the territory of the State or about persons in the State, without proper legal basis or consent, or in violation of restrictions on intended use.

г) economic nexus test — systematic receipt of income in the territory of the State, including cases when the total annual revenue is at least EUR 500,000 or the number of active users/customers in the State exceeds 10,000. The exact thresholds are set by the regulator and can be specified for individual sectors.

The application of any one criterion is sufficient. The criteria can be applied simultaneously (cumulatively) and by analogy to related situations (circumvention of jurisdiction, "jurisdictional shopping", chain fragmentation). data protection standards, as well as pre-issued guarantees to minimize risks.

69.3 Subjects and chain of responsibility.

Responsible entities include: model provider (developer or owner), infrastructure provider (computing or cloud resources), integrator or distributor, operator or deployer, importer, authorized representative, data broker, API provider, as well as the user — within the limits of its own violations. Participants in the chain are jointly and severally liable for violations with significant consequences. A bona fide participant is exempt from liability if he proves due diligence of counterparties, compliance with the requirements of "flow-down" and timely taking corrective actions after identifying risks.

69.4 Market access conditions.

Access to the State's market for cross-border AI systems that have a significant impact on human rights, public security or the functioning of critical services is allowed only subject to prior compliance with such minimum requirements, which apply cumulatively:

− Official presence and service of the process. Appointment of an authorized representative in the State with the right to receive claims and lawsuits, as well as interact with regulators; publication of relevant contacts, provision of SLA response, support in the state language and mode of operation in accordance with the official time zone of the State; confirmation of authority to conclude "flow down" contracts.

− Registration and unique identification. Entering the system into the National AI Register with the assignment of a unique identifier of the installation and release version; declaring the role (provider, operator or importer), areas of application, level of risk, geography of data processing and chain of subprocessors.

− Model and dataset passport. Mandatory submission of structured passports, which must contain at least: purpose and limits of applicability; known limitations and risks; results of recent quality and fairness tests taking into account user groups ; description of capability gating mechanisms and safe modes; update policy and changelog; origin of datasets (provenance), legal basis for text and data mining (TDM) or licenses; cleaning and balancing methods; known imbalances and ways to neutralize them.

− Cross-Border Impact Assessment (AIA XBorder). Mandatory written assessment before launching or substantially updating the system with data flow maps, including cross-border ones, human rights, non-discrimination and information environment risk analysis, mitigation plans, incident response scenarios, definition of kill switch activation criteria and confirmation implementation of local human supervision. The summary of the AIA assessment is subject to mandatory publication in the state language.

− Human supervision, explainability, and logging. Availability of human supervision procedures ("man in the loop") for decisions with significant impact; ensuring local explainability of results; keeping logs of calls, solutions and versions with ensuring their preservation within the established deadlines; keeping a quality log with defined metrics and thresholds.

− Assumption of jurisdiction and choice of law. The treaties shall contain a clause on the jurisdiction of the courts of the State and the recognition of the powers of the supervisory authorities in respect of disputes and incidents involving users located in the territory of the State; it is prohibited to impose foreign arbitration on consumers, employees or students; Contracts should provide for mandatory approval for inspections and the provision of materials in the "safe room».

− Cybersecurity and resilience. Establishment of inappropriate use policies; technical safeguards (access rate limiting, abuse detection systems, geofencing), segmentation of environments, encryption of data during transmission and storage; storage of re-identification keys under the jurisdiction of the State or in the regime of joint control; Incident Response Plans (IRP), Uptime Recovery (RTO), and Recovery Points (RPO); availability of channels for reporting incidents with thresholds in 24 and 72 hours.

− Channels for filing complaints and redresses in the State. Providing available remedies in the official language of the State that have a suspensive effect on decisions with significant impact; the presence of a description of the mechanisms for compensation for damage; disclosure of contact details of an authorized official in the State.

− KYC and risk onboarding procedures for high-risk scenarios. Verification of counterparties, in particular integrators and operators in critical areas; verification of declared purposes of use; prohibition of "jurisdictional shopping" and proxy deployments.

Grounds for refusal or suspension of access: failure to submit or submit false passports or cross-border impact assessment (AIA XBorder); absence of an official representative; refusal to accept the jurisdiction of the State; detection of discriminatory effects or serious incidents without proper remediation; failure to provide access to logs or documentation; identification of prohibited opportunities without deterrent mechanisms. The regulator has the right to apply interim ex parted measures until violations are remedied.

69.5 Data transfer and cross-border processing.

Cross-border transfers of personal, business and production data are allowed only if there is an adequate level of protection in the recipient State or adequate contractual and technical safeguards. Any remote access to data from abroad, including technical support, monitoring, analytics or hosting of backups, is equivalent to data transfer and is subject to the requirements of this clause.

− Transfer Impact Assessment (DTIAXBorder). Before a transfer takes place, the system operator or transmission service provider is required to conduct a written assessment (can be integrated into AIAXBorder) that covers:

(i) a map of data flows, defining the roles and responsibilities of the parties;

(ii) classification of data by sensitivity and criticality;

(iii) analysis of the legislation of the recipient state regarding access by public authorities, as well as the availability of effective remedies for data subjects;

(iv) Threat identification by reidentification, sensitive feature inference, correlation attacks, and combined processing (data linkage);

(v) description of the technical and organizational measures applied (encryption, pseudonymization, role-based access, independent auditing);

(vi) residual risk assessment and criteria under which the transfer should be suspended or prohibited;

(vii) incident response plans, including procedures for notifying the regulator and data subjects;

(viii) terms of onward transfers, in particular the obligations of subprocessors and the procedure for monitoring them;

(ix) retention period, policy of deletion, anonymization or pseudonymization of data.

The transfer impact assessment is subject to mandatory logging, is kept by the provider/operator and is provided at the request of the authorized body for control.

− Minimum technical guarantees.

Cross-border data processing is carried out with the following mandatory measures:

a) encryption of data in transit at least TLS 1.3 level and in storage using certified cryptographic security modules (level not lower than FIPS 140-3 or equivalent international standards);

b) sustainable pseudonymization combined with separation and controlled access to counterparts, with the provision of audited disclosure mechanisms;

c) cryptographic keys are managed exclusively through certified KMS/HSM systems under the jurisdiction of the relevant state or in the mode of joint control (split key, MPC, threshold cryptography), which makes it impossible for any foreign party to access the keys alone;

Use of priority technologies Privacy Enhancing Technologies (PETs) — Including federated learning, secure enclaves/TEE, confidential computing, MPC/HE, as well as a compute-to-data approach, if technically possible;

d) implementation of change-protected access and transfer logs (immutable append-only logs) with cryptographic signing of records and their mandatory storage for the period specified by law;

e) compliance with the principles of data minimization, limitation of the purpose of processing and retention periods, with the prohibition of inference of sensitive features or derived characteristics that may create discriminatory effects;

f) conducting regular tests for risks of re-identification, inference and privacy drift, with mandatory documentation of results and elimination of identified vulnerabilities.

− Keys, re-identification and control. The decryption and re-identification keys shall be stored exclusively on the territory of the State concerned or in the mode of joint control with an authorized national entity; their disclosure to any foreign states or third parties without the prior permission of the authorized body of this State is prohibited; any access to the keys is subject to mandatory fixation in a special register indicating the grounds, date and time, responsible person and authorization procedure, and access logs are immutable, signed by cryptographic means and stored for a period determined by law.

− Critical and strategic data. Such data includes, in particular: information in the areas of defence and security, energy and critical infrastructure, electoral processes, large arrays of medical, biometric and educational records, public sector operations, as well as data on cyber defence systems. Their transfer is allowed only in the presence of a special permit of the authorized body, issued after the security examination, and under the conditions:

(i) localization of access logs and preservation of a "mirror copy" (near real time) on the territory of the State;

(ii) use of TEE, secure rooms, or view-only modes);

(iii) establishment in contracts of direct prohibitions on further transfer of data without the separate permission of the authorized body;

(iv) ensuring the technical possibility of immediate disabling of access ("kill switch") in case of detection of a threat or violation of access conditions.

− Onward transfers. Any further transfer of data by the recipient to third parties is allowed only under conditions not lower than the original obligations, with the inclusion of the flow down of obligations in the contractual relationship, with prior notification of the operator located in the State, and with mandatory reflection in the register of sub-processors and recipients.

− Requests from foreign authorities. In case of a request from foreign law enforcement or intelligence agencies, the recipient is obliged to:

(i) immediately notify the authorized body of the State before the disclosure of the data, unless such notification is expressly prohibited by law;

(ii) appeal excessive or disproportionate requests in accordance with the established procedure;

(iii) In cases where disclosure is lawful and mandatory, only a minimized amount of data necessary solely for the specified purpose shall be transferred;

(iv) record all requests in the Transparency Register with the mandatory periodic publication of anonymized statistics on the number, types and results of such requests.

− Register and audit of transfers. The provider or operator is obliged to keep a register of all cross-border data transfers, indicating: date and time of transmission, volume and categories of data, legal basis, place of processing, subjects and recipients, technical and contractual safeguards applied, retention periods and references to the relevant transfer impact assessment (DTIA XBorder). The register must be kept for at least 24 months, be unchanged, be protected by cryptographic means and provided to the authorized body at its request for audit and control purposes.

− Emergency and vital transmissions. In cases where the transfer of data is necessary to protect the life or health of persons or to eliminate the consequences of emergency situations, only the minimum required amount of data is allowed to be temporarily transferred.

In this case, the provider or operator is obliged to:

- issue a transfer impact assessment (DTIA XBorder) no later than within ten working days from the date of transfer;

- notify the authorized body of the fact and volume of the transfer within no more than seventy-two hours, and in case of critical incidents — within no more than twenty-four hours;

− backup and Recovery (Disaster Recovery / Business Continuity Planning). Cross-border storage of data backups is allowed only in encrypted form, provided that the encryption keys are stored under the jurisdiction of the State or in a joint control mode with an authorized national entity. Any access to backups is subject to mandatory logging, and their recoverability must be periodically checked by tests in a mode that excludes the transfer of raw (unencrypted) data abroad;

− invalidity and sanctions. Data transfers carried out without an impact assessment of the transfer (DTIA XBorder), without appropriate safeguards or contrary to the requirements of this clause, are considered unlawful and entail the obligation to immediately stop processing, return or delete the transmitted data, notify the data subjects and the competent authority, as well as apply measures of influence.

69.6 Localization and restrictions for critical areas. In the areas of defence, energy, transport, finance, healthcare, electoral processes, as well as in other sectors classified as critical by the decision of the authorized body, special requirements are established for data localization, restrictions on cross-border transfers and ensuring sovereign control over information processing.

− Basic localization rule. All data operations, including inferences and model inferences, logging of calls and decisions of AI systems, are carried out exclusively on certified infrastructure located in the territory of the relevant State or in "trusted jurisdictions" designated by the authorized body, which provide

equivalent the level of protection of human rights, personal and sensitive data, information security and digital sovereignty.

The authorized body of the State has the unconditional right to immediate access to logs, logs and technical materials confirming the transactions carried out.

- Criteria of "trusted jurisdiction".

Jurisdictions recognized as trusted jurisdictions can include only those states that have the following conditions at the same time:

(i) an official decision or approved list on an adequate level of protection of personal and sensitive data;

(ii) existing international treaties, mutual legal assistance agreements (MLATs) or memorandums of understanding (MoU) with the State concerned, as well as the existence of judicial or parliamentary control over foreign authorities' access to data;

(iii) the absence of extraterritorial rules that allow unilateral access by foreign authorities to data without notification or agreement with the State concerned;

(iv) the legal ability and proven practice of complying with the requirements of the regulator of this State, including conducting an audit in a "safe room" or other controlled form of access.

− Infrastructure certification and technical guarantees. For the deployment of systems in critical areas, the following are mandatory: state certification or other attestation of the Integrated Information Security System (CISS), or equivalent recognized by the authorized body; segmentation of data processing environments; data encryption during transit and storage; sovereign key management on the territory of the State or in the mode of joint control (split key, MPC, HSM), which excludes the possibility of sole access of foreign data. Side; the use of Trusted Execution Environment (TEE) technologies and confidential computing to process sensitive data; maintaining immutable (append only) logs with cryptographic signing of records and time synchronization; implementation of emergency kills switch mechanisms and model version control (SBOM/MBOM).

− Operational restrictions. Remote administration from untrusted jurisdictions, the use of hidden subprocessors, and the creation of "proxy chains" are prohibited. All releases and updates are subject to mandatory prior local validation. During election periods, as well as during periods of increased threat level, the regime of "change freeze" is established, except for critical security fixes, which are allowed only with the separate approval of the authorized body.

− Logs and storage periods. Logs of calls, decisions, accesses and changes are stored for at least thirty-six months, and in case of a dispute or inspection — for the entire period of their consideration. The authorized body has the right to immediate access to the logs, as well as to apply mechanisms for "freezing" processes in accordance with the established procedure.

- DR/BCP and continuity. Redundancy and recovery are performed with the provision of a local "mirror" (near real time) on the territory of the respective State. Fault tolerance scenarios should provide for a local degradation mode (graceful degradation), which guarantees the preservation of the operation of critical services without their complete failure. Access to backups is allowed only in encrypted form and is subject to mandatory logging.

− Data and training restrictions. It is prohibited to transfer and use critical or strategic data for training or fine-tuning models outside the territory of the State. Where the processing of such data is necessary, a compute to data approach or the use of synthetic or anonymized datasets is permitted.

− Exceptions and special permits. Temporary deviation from the requirements of localization is allowed only with a special permit of the authorized body, issued after a security examination. Such permission must indicate: purpose, duration, place of processing and controls established; contractual flow down restrictions; as well as the obligation to immediately terminate access in the event of an incident. Critical incidents are subject to mandatory notification to the authorized body no later than within twenty-four hours.

− Consequences of non-compliance. Failure to comply with the requirements of this paragraph is the basis for the application of ex parted interim measures, the imposition of sanctions and the suspension

of access to the market, as well as for the inclusion of the relevant system or configuration in the list of prohibited until the identified risks are completely eliminated.

69.7 Contractual "flow-down" obligations.

The Provider shall ensure full and non-dilutive (without mitigation) reproduction of the requirements of this Section in all contracts with subprocessors, partners, affiliates, resellers, integrators, infrastructure providers, data brokers, and other parties in the supply or processing chain, including subsequent (secondary) subprocessors (onward/secondary processing). Any contractual wording that reduces the level of protection, delays performance or limits the rights of users or the regulator is prohibited. The absence or incompleteness of the "flowdown" of obligations qualifies as a breach of the provider and creates ***a presumption of its negligence****.*

Minimum list of mandatory provisions subject to "flow down" (transfer of obligations) for the entire supply and processing chain:

- *Impact assessments and transparency: before the launch or transfer of the system, a mandatory impact assessment (AIA XBorder/DTIA XBorder) is carried out with the submission of an executive summary available to supervisory authorities and interested parties; the operator maintains a register of sub-processors indicating their roles, jurisdictions, categories of data processed and their storage periods; in case of a change in the composition of the sub-processor chain the operator notifies about this at least 30 days in advance; The user or customer has the right to file an objection or terminate the legal relationship without applying penalties and without limiting other legal remedies.*
- *Logging and traceability: The operator is obliged to keep append-only logs of calls, accesses, decisions and transfers with a cryptographic signature, ensure that they are stored for at least 24 months (or longer for the period of dispute or review), provide access to them in a "safe room" mode, as well as keep SBOM/MBOM and release version logs.*
- *Incidents and responses: notification of the provider and the authorized body of Ukraine about incidents within ≤72 hours, and about critical ones — ≤24 hours; immediate containment, corrective action plan, informing victims, preserving evidence; prohibition to limit such notices to confidentiality clauses/NDAs.*
- *Incidents and responses: The operator is obliged to notify the provider and the authorized body of the state about incidents no later than 72 hours, and about critical incidents no later than 24 hours, ensure immediate containment, develop and implement a corrective action plan, inform affected persons, preserve evidence; You may not limit such notices to confidentiality provisions or non-disclosure agreements (NDA).*
- *Prohibition of secondary use and transfer: The Operator is obliged to process data and artifacts exclusively for the specified purpose; it is prohibited to train or fine-tune models on user data without a separate legal basis; any further transfers are prohibited without written permission and without ensuring the transfer of identical obligations on the chain (flow down); priority is given to compute-to-data approaches and privacy-enhancing technologies (PETs).*
- *Audit and control: the provider, operator and regulator have the unconditional right to conduct remote or on-site audits, technical checks (black-box tests, penetration tests), as well as review of policies and logs; the audit is carried out within a reasonable time; the operator is obliged to eliminate the identified violations within the established period; in case of non-compliance, the provider or regulator has the right to suspend or terminate legal relations.*
- *Security and localization: data encryption in transit and at rest is ensured; implementation of sovereign management of cryptographic keys in the state or the mode of shared control (split key, MPC, HSM); geofencing, TEE/confidential computing technologies and Disaster Recovery/Business Continuity Planning (DR/BCP) measures are applied; "silent" updates are prohibited and mandatory release notes with a description of the impact are provided.*
- *Management of subprocessors: the involvement or replacement of subprocessors is allowed only with prior written permission; the list of trusted jurisdictions and the list of prohibited countries are defined; contracts with subsequent subprocessors are required to contain identical terms without any*

exceptions; The provider shall be jointly and severally liable for the acts or omissions of such sub-processors.

− *Jurisdiction and language of the contract: all disputes involving users or operators in the relevant state are subject to arbitration by the courts of that state; it is prohibited to impose foreign arbitration on consumers, employees and students; in the event of conflicts, the official text of the contract in the language of that state shall prevail. Sanctions, indemnification, step in: compensation mechanisms are established; fines and penalties are applied for violation of the terms of notification or remediation; the contract must contain indemnity and liability insurance conditions; the provider or operator has the right to carry out step in (temporary management or isolation of the environment) to prevent damage; the use of escrow technical documentation and key artifacts for critical systems is allowed.*

− *Change management and interoperability: mandatory versioning of configurations/models, change control, rollback plan, prohibition of "jurisdictional shopping"/rebranding to circumvent requirements; agreed SLAs/OLAs on quality, equity, and accessibility.*

− *Change management and interoperability: mandatory versioning of configurations and models is ensured, change control is implemented, a rollback plan is developed and maintained; the so-called "jurisdictional shopping" and rebranding to circumvent the requirements are prohibited; negotiated SLAs/OLAs are concluded and enforced, which guarantee quality, fairness and availability.*

− *Termination of the relationship: verified deletion or return of data is carried out with the provision of an erasure certificate; it is prohibited to further use models or artifacts trained on customer data without a separate legal basis; establishes a transition period necessary to ensure the continuity of the service; ensures the preservation of evidence for the entire period of dispute consideration.*

Failure to comply with or lack of proper provisions for the transfer of obligations under the chain (flow down) qualifies as a violation by the provider with consequences in the form of fines, suspension of access or delisting and does not exempt it from joint liability for damage caused by the actions of subprocessors or partners.

69.8 Transparency of cross-border activities.

The user must be provided with clear marking, which includes: a unique system identifier and model version; the name of the legal entity that owns and information about the ultimate beneficiary; the states in which the data is processed and stored; the date of the last release or update; known technical and legal restrictions and risks; national and international channels for filing complaints and appeals.

69.9 Incidents, Damages & Notices.

In the event of any incident that may cause significant damage (security threat, discrimination, data leakage, manipulation of the information space, malfunctions of critical services), the provider is obliged to notify the regulatory authority within 72 hours. In case of critical incidents, notification must be made no later than within 24 hours. a corrective action plan; timely informing the affected persons; provision of temporary compensators to minimize damage.

69.10 Prohibition of circumvention of jurisdiction.

Prohibited: carrying out "jurisdictional shopping"; using chains of affiliates in order to avoid liability; the use of geo-blocking to restrict the filing of complaints; the imposition of a foreign court or arbitration on the consumer contrary to the jurisdiction of the state; the masking of the real provider; the use of intermediation, which actually deprives the user of the protection of rights guaranteed by law.

69.11 Audit and access to technical materials.

The regulator has the right to: require an independent audit of the system; provide access to documentation, logs, and tests in a "safe room" mode; require the transfer of technical documentation and scales (for critical systems) in ESCROW conditions; carry out black-box testing. Failure to comply with these requirements is the basis for suspension of the system's access to the market.

69.12 User rights and collective protection.

Each person has the right to: clear notification of the cross-border nature of the service; local explanation of the outcome of the AI system; access to relevant data and event logs; prompt appeal with a

suspensive effect; compensation for damages; participation in class actions and mediation procedures; appeal to the authorized body and regulators.

69.13 Sanctions and enforcement measures.

The following measures are applied for violation of the regime: orders to eliminate violations; temporary suspension or blocking of access to the system; fines, the amount of which is determined in proportion to the provider's global turnover; confiscation of illegally obtained benefits; prohibition of participation in public procurement; mandatory independent audit.

In case of systemic or repeated violations, additional measures are applied: delisting (exclusion from registers) or geofencing of services until the identified risks are eliminated.

69.14 Jurisdiction and choice of law.

Disputes involving users or authorities of a state shall be subject to consideration by the courts of that state. Any contractual terms limiting or depriving the right to recourse to national courts shall be declared null and void.

In commercial disputes, it is allowed to submit them to arbitration by mutual consent of the parties, provided that such transfer does not limit the rights of consumers, employees or students.

69.15 International cooperation.

The regulatory authorities of the state ensure interaction with foreign competent authorities by concluding agreements on mutual legal assistance (MLAT), memorandums of understanding (MoU) and other international instruments; exchange incident signals; recognize audit reports prepared in partner jurisdictions; participate in joint audits; harmonize minimum standards of security, transparency and protection of human rights.

69.16 Import and export models and content.

It is prohibited or allowed only with restrictions and means of redressing (filters, labelling of origin, "sandbox") to import artificial intelligence models that:

(i) have been trained on data sets with gross violation of copyright or personal rights without proper legal basis;

(ii) contain hostile propaganda, systematic falsification of historical facts, or subversive narratives.

(iii) demonstrate dangerous capabilities without available technical or legal deterrence mechanisms.

69.17 High-level ("frontier") models and APIs.

Providers of frontier models and APIs available in the State are required to: implement user identification (KYC) and abuse-control procedures in high-risk scenarios; implement technical safeguards, including rate limits, RLHF-based security policies, and capability gating; conduct regular red-team testing; provide integrators in critical areas with Enforced Refusal to Perform Functions («kill-switch»).

69.18 Open Source and Scientific Use.

The distribution of open models or scales does not entail commercial responsibility of the developer, if he does not carry out their marketing or integration into production processes and does not hide known high risks. An integrator who deploys an open-source solution in production bears the full scope of obligations provided for in this Section.

69.19 Cross-border impact assessment (AIA-XBorder).

Before launching or substantially upgrading the system, the provider is required to conduct a written cross-border impact assessment (AIA-XBorder), which must include: processing objectives; data and transfer maps; risk assessment for human rights, non-discrimination and the information environment; risk mitigation plans; results of fairness tests by groups; incident response plans; local human oversight mechanisms; criteria for suspension or termination of the system.

AIA-XBorder's resume must be published in a form and language accessible to users.

69.20 Information sovereignty and electoral processes.

Cross-border campaigns and tools that influence the formation of public opinion, electoral processes or referendums are subject to: mandatory labelling as automated; registration in the register of political

authenticity; prohibition of simulated participation of citizens; increased requirements for transparency of funding and content provenance.

69.21 Preservation of evidence and procedural safeguards.

The provider is obliged to ensure the preservation of relevant event logs and materials for at least 24 months, and in case of a dispute or verification — during the entire period of their consideration. At the request of the regulator, the provider is obliged to immediately freeze the relevant processes and provide access to the quality log and full versions of releases.

69.22 Interim measures.

If there are signs of an imminent threat to human rights or security, the regulator has the right to apply interim measures ex parted, in particular: suspension of certain functions of the system, introduction of geofencing, or a requirement to roll back the model version. Such measures are subject to mandatory review within ten days.

69.23 Transitional provisions.

For existing cross-border AI systems, a period of six months from the date of entry into force of the law is established to bring their activities in line with its requirements. For systems operating in critical areas, this period is three months. Failure to comply with the requirements within the specified time frame is the basis for a phased restriction of access to the market and the application of sanctions provided for by law.

69.24 Interpretations and collisions.

The provisions of this Section shall be interpreted according to the principle of pro persona — in a way that ensures the fullest protection of human rights and freedoms. In case of doubt, the principle in dubio pro securitate et libertate is applied, according to which preference is given to the decision that guarantees human rights, information sovereignty and security to the greatest extent. The interpretation and application of norms are carried out in compliance with the principles of necessity and proportionality.

The provisions of this Section shall be interpreted according to the principle of pro persona — in a way that ensures the fullest protection of human rights and freedoms. In case of doubt, the principle in dubio pro securitate et libertate is applied, according to which preference is given to the decision that guarantees human rights, information sovereignty and security to the greatest extent. The interpretation and application of norms are carried out in compliance with the principles of necessity and proportionality.

Author's message.

All intellectual work on this material is authorship, which covers the concept, idea, and structure construction, the development of norms and definitions, the selection, analysis and processing of scientific literature, as well as the preparation of drafts, the basic text are the author's work of Oleksiy Kostenko. For technical structuring of material, ordering texts, generation of wording options, checking and editing drafts and intermediate versions, identifying and eliminating errors, checking the relevance of legal norms and scientific research, contextual semantic search of relevant documents, selection of case law, as well as for the search for reference information, LLM tools (ChatGPT, Grok), Consensus and Gamma modules, Copilot, Microsoft 365 tools and other software solutions were used. Automatically generated fragments, if used, were edited and integrated by the author, but all content decisions, interpretations and final formulations were made by the author personally in compliance with academic standards. The final conclusions, interpretations and legal positions are solely the position of the author.

REFERENCES

[1] Stahl, B., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, 1 - 33. https://doi.org/10.1007/s10462-023-10420-8.

[2] Moss, E., Watkins, E., Singh, R., Eilish, M., & Metcalfe, J. (2021). Building accountability: algorithmic assessment of impact in the public interest. *Electronic journal SSRN* . https://doi.org/10.2139/ssrn.3877437 .

[3] Harris, S. (2020). Data Protection Impact Assessments as rule of law governance mechanisms. *Data & Policy*, 2. https://doi.org/10.1017/dap.2020.3.

[4] Mantelero, A. (2018). AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Feminist Methodology & Research eJournal*. https://doi.org/10.1016/J.CLSR.2018.05.017.

[5] Calvi, A. (2024). Data Protection Impact Assessment under the EU General Data Protection Regulation: A feminist reflection. *Comput. Law Secur. Rev.*, 53, 105950. https://doi.org/10.1016/j.clsr.2024.105950.

[6] Mantelero, A. (2024). The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *ArXiv*, abs/2411.15149. https://doi.org/10.1016/j.clsr.2024.106020.

[7] Rintamaki, T., & Pandit, H. (2024). Developing an Ontology for AI Act Fundamental Rights Impact Assessments. *ArXiv*, abs/2501.10391. https://doi.org/10.48550/arXiv.2501.10391.

[8] Rintamaki, T., & Pandit, H. (2024). Towards An Automated AI Act FRIA Tool That Can Reuse GDPR's DPIA. *ArXiv*, abs/2501.14756. https://doi.org/10.48550/arXiv.2501.14756.

[9] McGinn, R. (1990). Science, Technology and Society. .

[10] Bhaskar, V., & Kumar, G. (2024). SCIENCE TECHNOLOGY AND MODERN SOCIETY. *GLOBAL JOURNAL FOR RESEARCH ANALYSIS*. https://doi.org/10.36106/gjra/7103658.

[11] (2019). The Delphi Method. *New Teaching Resources for Management in a Globalised World*. https://doi.org/10.1142/9789811206542_0011.

[12] Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique.. *Journal of advanced nursing*, 32 4, 1008-15 . https://doi.org/10.1046/J.1365-2648.2000.01567.X.

[13] Zimmermann, H. (1980). OSI Reference Model - The ISO Model of Architecture for Open Systems Interconnection. *IEEE Transactions on Communications*, 28, 425-432. https://doi.org/10.1109/TCOM.1980.1094702.

[14] Dromard, F. (1984). A guide to open systems interconnection. *Computers and Standards*, 3, 171-193. https://doi.org/10.1016/0167-8051(84)90006-8.

[15] Abend, F. (2016). Open Systems Interconnection Handbook. .

[16] Zhuk, A. (2024). Navigating the legal landscape of AI copyright: a comparative analysis of EU, US, and Chinese approaches. *AI Ethics*, 4, 1299-1306. https://doi.org/10.1007/s43681-023-00299-0.

[17] Radanliev, P. (2025). Frontier AI regulation: what form should it take?. *Frontiers in Political Science*. https://doi.org/10.3389/fpos.2025.1561776.

[18] Widder, D., & Nafus, D. (2022). Dislocated accountabilities in the “AI supply chain”: Modularity and developers’ notions of responsibility. *Big Data & Society*, 10. https://doi.org/10.1177/20539517231177620.

[19] Cobbe, J., Wil, M., & Singh, J. (2023). Understanding Accountability in Algorithmic Supply Chains. *Proceedings of the 2023 ACM Conference on Equity, Accountability and Transparency* . https://doi.org/10.1145/3593013.3594073 .

[20] Imagawa, K., Mizukami, Y., & Miyazaki, S. (2018). Regulatory convergence of medical devices: a case study using ISO and IEC standards. *Expert Review of Medical Devices*, 15, 497 - 504. https://doi.org/10.1080/17434440.2018.1492376.

[21] Sheffner, D. (2019). Integrating Technical Standards into Federal Regulations: Incorporation by Reference. *The Cambridge Handbook of Technical Standardization Law*. https://doi.org/10.1017/9781316416785.007.

[22] Rodríguez, N., Ser, J., Coeckelbergh, M., De Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation. *Inf. Fusion*, 99, 101896. https://doi.org/10.48550/arXiv.2305.02231.

[23] Truby, J., Brown, R., Ibrahim, I., & Parellada, O. (2021). A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications. *European Journal of Risk Regulation*, 13, 270 - 294. https://doi.org/10.1017/err.2021.52.

[24] Kostenko, O., & Golovko, O. (2023). Metaverse electronic jurisdiction: challenges and risks of legal regulation of virtual reality. *INFORMATION AND LAW*. https://doi.org/10.37750/2616-6798.2023.1(44).287729.

[25] O., Kostenko. (2022). ELECTRONIC JURISDICTION, METAVERSE, ARTIFICIAL INTELLIGENCE, DIGITAL PERSONALITY, DIGITAL AVATAR, NEURAL NETWORKS: THEORY, PRACTICE, PERSPECTIVE. *World Science*. https://doi.org/10.31435/rsglobal_ws/30012022/7751.

[26] Mendes, P. (2023). Model-based risk analysis for system design. *Systems Engineering*, 27, 20 - 5. https://doi.org/10.1002/sys.21704.

[27] Janssen, H., Lee, M., & Singh, J. (2022). Practical fundamental rights impact assessments. *Int. J. Law Inf. Technol.*, 30, 200-232. https://doi.org/10.1093/ijlit/eaac018.

[28] Regulations on the Management of Algorithmic Recommendations in Internet Information Services https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm

[29] Personal Information Protection Law of the People's Republic of China https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/

[30] Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet Things*, 27, 101291. https://doi.org/10.1016/j.iot.2024.101291.

[31] Pehlivan, C. (2024). The EU Artificial Intelligence (AI) Act: An Introduction. *Global Privacy Law Review*. https://doi.org/10.54648/gplr2024004.

[32] Minh, L. (2024). Eu Ai Act and Its Relationship with Vietnamese Lawin Creating a Legal Policy for Ai Regulation. *International Journal of Religion*. https://doi.org/10.61707/jp3pks38.

[33] Hoffmeister, K. (2024). The Dawn of Regulated AI: Analyzing the European AI Act and its Global Impact. *Zeitschrift für europarechtliche Studien*. https://doi.org/10.5771/1435-439x-2024-2-182.

[34] Gilbert, S. (2024). The EU passes the AI Act and its implications for digital medicine are unclear. *NPJ Digital Medicine*, 7. https://doi.org/10.1038/s41746-024-01116-6.

[35] Lewis, D., Lasek-Markey, M., Golpayegani, D., & Pandit, H. (2025). Mapping the Regulatory Learning Space for the EU AI Act. *ArXiv*, abs/2503.05787. https://doi.org/10.48550/arXiv.2503.05787.

[36] Ukoh, D., & Adetunji, M. (2025). AI Act: The EU Regulation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4607388.

[37] Minh, L. (2024). EU Artificial Intelligence Law and Its Relationship with Vietnamese Legislation in the Field of Creating Legal Policies for the Regulation of Artificial Intelligence. *International Journal of Religion* . https://doi.org/10.61707/jp3pks38 .

[38] Birchfield, V. (2024). From roadmap to regulation: will there be a transatlantic approach to governing artificial intelligence?. *Journal of European Integration*, 46, 1053 - 1071. https://doi.org/10.1080/07036337.2024.2407571.

[39] Kuzior, A. (2024). Navigating AI Regulation: A Comparative Analysis of EU and US Legal Frameworks. *Materials Research Proceedings*. https://doi.org/10.21741/9781644903315-30.

[40] Khassanay, A., & Tifine, P. (2025). THROUGH THE LENS OF THE LAW: HOW CHINA AND THE EUROPEAN UNION ARE SHAPING THE FUTURE OF ARTIFICIAL INTELLIGENCE. *Bulletin of KazNPU named after Abai series "Jurisprudence"*. https://doi.org/10.51889/2959-6181.2024.78.4.005.

[41] Bal, R., & Gill, I. (2020). Policy Approaches to Artificial Intelligence Based Technologies in China, European Union and the United States. . https://doi.org/10.2139/ssrn.3699640.

[42] Arshad, N., Butt, T., & Iqbal, M. (2025). A Comprehensive Framework for Intelligent, Scalable, and Performance-Optimized Software Development. *IEEE Access*, 13, 74062-74077. https://doi.org/10.1109/ACCESS.2025.3564139.

[43] Kulkarni, N. (2024). Role of AI in Application Life Cycle Management (ALM). *Journal of Artificial Intelligence & Cloud Computing*. https://doi.org/10.47363/jaicc/2024(3)397.

[44] Laato, S., Birkstedt, T., Mäntymäki, M., Minkkinen, M., & Mikkonen, T. (2022). AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering. *2022 IEEE/ACM 1st International Conference on AI Engineering – Software Engineering for AI (CAIN)*, 113-123. https://doi.org/10.1145/3522664.3528598.

[45] De Silva, D., & Alahakoon, D. (2021). An artificial intelligence life cycle: From conception to production. *Patterns*, 3. https://doi.org/10.1016/j.patter.2022.100489.

[46] Shahriar, S., Allana, S., Hazratifard, S., & Dara, R. (2023). A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle. *IEEE Access*, 11, 61829-61854. https://doi.org/10.1109/ACCESS.2023.3287195.

[47] Harvey, B., & Gowda, V. (2020). How the FDA Regulates AI.. *Academic radiology*, 27 1, 58-61 . https://doi.org/10.1016/j.acra.2019.09.017.

[48] Zhao, S., Blaabjerg, F., & Wang, H. (2020). An Overview of Artificial Intelligence Applications for Power Electronics. *IEEE Transactions on Power Electronics*, 36, 4633-4658. https://doi.org/10.1109/TPEL.2020.3024914.

[49] Amariles, D., & Baquero, P. (2023). Promises and limits of law for a human-centric artificial intelligence. *Comput. Law Secur. Rev.*, 48, 105795. https://doi.org/10.1016/j.clsr.2023.105795.

[50] Azzutti, A., Ringe, W., & Stiehl, H. (2022). The Regulation of AI Trading from an AI Life Cycle Perspective. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4260423.

[51] Bassey, K., Juliet, A., & Stephen, A. (2024). AI-Enhanced lifecycle assessment of renewable energy systems. *Engineering Science & Technology Journal*. https://doi.org/10.51594/estj.v5i7.1254.

[52] Rangone, N. (2023). Artificial intelligence challenging core State functions. *Revista de Derecho Público: Teoría y método*. https://doi.org/10.37417/rdp/vol_8_2023_1949.

[53] Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. *AI Ethics*, 4, 1265-1277. https://doi.org/10.1007/s43681-023-00330-4.

[54] d'Aquin, M., Troullinou, P., O'Connor, N., Cullen, A., Faller, G., & Holden, L. (2018). Towards an "Ethics by Design" Methodology for AI Research Projects. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3278721.3278765.

[55] Iphofen, R., & Kritikos, M. (2019). Regulating artificial intelligence and robotics: ethics by design in a digital society. *Contemporary Social Science*, 16, 170 - 184. https://doi.org/10.1080/21582041.2018.1563803.

[56] Gerdes, A. (2021). A participatory data-centric approach to AI Ethics by Design. *Applied Artificial Intelligence*, 36. https://doi.org/10.1080/08839514.2021.2009222.

[57] Coeckelbergh, M. (2019). Artificial Intelligence: Some ethical issues and regulatory challenges. , 2019, 31-34. https://doi.org/10.26116/TECHREG.2019.003.

[58] Perperidis, G. (2024). Designing Ethical A.I. Under the Current Socio-Economic Milieu: Philosophical, Political and Economic Challenges of Ethics by Design for A.I.. *Philosophy & Technology*. https://doi.org/10.1007/s13347-024-00766-4.

[59] AI Risk Management Framework https://www.nist.gov/itl/ai-risk-management-framework#:~:text=Download%20the%20AI%20RMF%201.0%20View%20the%20AI,organizations%2C%20and%20society%20associated%20with%20artificial%20intelligence%20%28AI%29.

[60] Artificial Intelligence Act (AI ACT) https://oecd.ai/en/dashboards/policy-initiatives/artificial-intelligence-act-ai-act-9517

[61] Recommendation on the Ethics of Artificial Intelligence https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

[62] ISO/IEC 22989:2022 - Information technology - Artificial intelligence - Artificial intelligence concepts and terminology https://oecd.ai/en/catalogue/tools/isoiec-229892022-information-technology-artificial-intelligence-artificial-intelligence-concepts-and-terminology

[63] Craig, C. (2021). The Relational Robot: A Normative Lens for AI Legal Neutrality (Reviewing Ryan Abbott, The Reasonable Robot, Cambridge University Press, 2020). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4118849.

[64] Craig, C. (2021). The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4014811.

[65] Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447 - 464. https://doi.org/10.1007/s12525-020-00441-4.

[66] Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *Int. J. Inf. Manag.*, 62, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433.

[67] O'Connor, S., & Liu, H. (2023). Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI & SOCIETY*, 1-13. https://doi.org/10.1007/s00146-023-01675-4.

[68] Barmer, H., Zombak, R., Gaston, M., Palat, W., Redner, F., & Smith, K. (2021). Human-centric AI. *IEEE Pervasive Comput.* , 22, 7-8. https://doi.org/10.1184/R1/16560183.V1 .

[69] Bingley, W., Haslam, S., Steffens, N., Gillespie, N., Worthy, P., Curtis, C., Lockey, S., Bialkowski, A., Ko, R., & Wiles, J. (2023). Enlarging the model of the human at the heart of human-centered AI: A social self-determination model of AI system impact. *New Ideas in Psychology*. https://doi.org/10.1016/j.newideapsych.2023.101025.

[70] Wang, D., Maes, P., Ren, X., Shneiderman, B., Shi, Y., & Wang, Q. (2021). Designing AI to Work WITH or FOR People?. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411763.3450394.

[71] Bingley, W., Curtis, C., Lockey, S., Bialkowski, A., Gillespie, N., Haslam, S., Ko, R., Steffens, N., Wiles, J., & Worthy, P. (2022). Where is the human in human-centered AI? Insights from developer priorities and user experiences. *Comput. Hum. Behav.*, 141, 107617. https://doi.org/10.1016/j.chb.2022.107617.

[72] Rong, Y., Leemann, T., Nguyen, T., Fiedler, L., Qian, P., Unhelkar, V., Seidel, T., Kasneci, G., & Kasneci, E. (2022). Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 2104-2122. https://doi.org/10.1109/TPAMI.2023.3331846.

[73] Zhang, X., Chan, F., Yan, C., & Bose, I. (2022). Towards risk-aware artificial intelligence and machine learning systems: An overview. *Decis. Support Syst.*, 159, 113800. https://doi.org/10.1016/j.dss.2022.113800.

[74] Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., & Li, W. (2021). Artificial Intelligence Security: Threats and Countermeasures. *ACM Computing Surveys (CSUR)*, 55, 1 - 36. https://doi.org/10.1145/3487890.

[75] King, T., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26, 89 - 120. https://doi.org/10.1007/s11948-018-00081-0.

[76] Roberts, H. (2024). Digital sovereignty and artificial intelligence: a normative approach. *Ethics Inf. Technol.*, 26, 70. https://doi.org/10.1007/s10676-024-09810-5.

[77] Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: the false promise of digital sovereignty. *European Security*, 31, 415 - 434. https://doi.org/10.1080/09662839.2022.2101885.

[78] Kolianov, A. (2022). Artificial Intelligence as a Strategic Component of Technological Sovereignty. *Discourse*. https://doi.org/10.32603/2412-8562-2022-8-5-81-90.

[79] Dokumacı, M. (2024). Legal Frameworks for AI Regulations. *Human Computer Interaction*. https://doi.org/10.62802/ytst2927.

[80] Eliot, L. (2021). Interoperability Of AI-to-AI Multi-Lawyering. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3990307.

[81] Ren, Q., & Du, J. (2024). Harmonizing innovation and regulation: The EU Artificial Intelligence Act in the international trade context. *Comput. Law Secur. Rev.*, 54, 106028. https://doi.org/10.1016/j.clsr.2024.106028.

[82] Kurmangali, M. (2024). Navigating the Frontiers of Digital Diplomacy: Multilateral Cooperation on Artificial Intelligence Regulation at UN and EU Levels. *Journal "International Relations and International Law"* . https://doi.org/10.26577/irilj.2024.v105.i1.09 .

[83] Radanliev, P. (2024). Cyber diplomacy: defining the opportunities for cybersecurity and risks from Artificial Intelligence, IoT, Blockchains, and Quantum Computing. *Journal of Cyber Security Technology*, 9, 28 - 78. https://doi.org/10.1080/23742917.2024.2312671.

[84] Bubashait, F. (2025). The emerging role of AI technologies in supporting digital diplomacy and shaping international relations. *International Journal for Scientific Research*. https://doi.org/10.59992/ijsr.2025.v4n2p2.

[85] Stoltz, M. (2024). Artificial Intelligence in Cybersecurity: Building Resilient Cyber Diplomacy Frameworks. *ArXiv*, abs/2411.13585. https://doi.org/10.48550/arXiv.2411.13585.

[86] De Almeida, P., Santos, C., & Farias, J. (2021). Regulating Artificial Intelligence: A Framework for Governance. *Ethics and Information Technology* , 23, 505 - 525. https://doi.org/10.1007/s10676-021-09593-z .

[87] Chauhan, S., Sharma, N., & Kumar, R. (2025). Comparative Analysis of Artificial Intelligence and Diplomacy: Transforming Democratic Governance. *Journal of Informatics Education and Research*. https://doi.org/10.52783/jier.v5i1.2293.

[88] Neuwirth, R. (2023). Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA). *Comput. Law Secur. Rev.*, 48, 105798. https://doi.org/10.1016/j.clsr.2023.105798.

[89] Neuwirth, R. (2022). Prohibited Artificial Intelligence Practices in the Proposed EU Artificial Intelligence Act. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4261569.

[90] Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2021). Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. *Minds and Machines*, 32, 241 - 268. https://doi.org/10.1007/s11023-021-09577-4.

[91] Stettinger, G., Weissensteiner, P., & Khastgir, S. (2024). Trustworthiness Assurance Assessment for High-Risk AI-Based Systems. *IEEE Access*, 12, 22718-22745. https://doi.org/10.1109/ACCESS.2024.3364387.

[92] Hupont, I., Micheli, M., Delipetrev, B., Gómez, E., & Garrido, J. (2023). Documenting High-Risk AI: A European Regulatory Perspective. *Computer*, 56, 18-27. https://doi.org/10.1109/MC.2023.3235712.

[93] Manchev, A. (2024). WORLD'S FIRST LAW FOR ARTIFICIAL INTELLIGENCE. LEGAL, ETHICAL AND ECONOMIC ASPECTS. *Education and Technologies Journal*. https://doi.org/10.26883/2010.241.5985.

[94] Mezgár, I., & Váncza, J. (2022). From ethics to standards - A path via responsible AI to cyber-physical production systems. *Annu. Rev. Control.*, 53, 391-404. https://doi.org/10.1016/j.arcontrol.2022.04.002.

[95] Hoseini, F. (2023). AI Ethics: A Call for Global Standards in Technology Development. *AI and Tech in Behavioral and Social Sciences*. https://doi.org/10.61838/kman.aitech.1.4.1.

[96] Jedličková, A. (2024). Ensuring Ethical Standards in the Development of Autonomous and Intelligent Systems. *IEEE Transactions on Artificial Intelligence*, 5, 5863-5872. https://doi.org/10.1109/TAI.2024.3387403.

[97] Sengar, S., Hasan, A., Kumar, S., & Carroll, F. (2024). Generative Artificial Intelligence: A Systematic Review and Applications. *ArXiv*, abs/2405.11029. https://doi.org/10.48550/arXiv.2405.11029.

[98] Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33, 1-17. https://doi.org/10.1007/s12525-023-00680-1.

[99] Batchu, C., & Satya, V. (2024). Generative AI: Evolution and its Future. *International Journal For Multidisciplinary Research*. https://doi.org/10.36948/ijfmr.2024.v06i01.12046.

[100] Zhang, P., & Boulos, M. (2023). Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. *Future Internet*, 15, 286. https://doi.org/10.3390/fi15090286.

[101] Sedkaoui, S., & Benaichouba, R. (2024). Generative AI as a transformative force for innovation: a review of opportunities, applications and challenges. *European Journal of Innovation Management*. https://doi.org/10.1108/ejim-02-2024-0129.

[102] Riemer, K., & Peter, S. (2024). Conceptualizing generative AI as style engines: Application archetypes and implications. *Int. J. Inf. Manag.*, 79, 102824. https://doi.org/10.1016/j.ijinfomgt.2024.102824.

[103] Koohi-Moghadam, M., & Bae, K. (2023). Generative AI in Medical Imaging: Applications, Challenges, and Ethics. *Journal of Medical Systems*, 47, 1-4. https://doi.org/10.1007/s10916-023-01987-4.

[104] Gozalo-Brizuela, R., & Garrido-Merch'an, E. (2023). A survey of Generative AI Applications. *ArXiv*, abs/2306.02781. https://doi.org/10.48550/arXiv.2306.02781.

[105] M., Sharma, P., & Bhardwaj, A. (2025). Exploring the Capabilities and Limitations of Generative AI Applications, Challenges, and Future Directions. *2025 International Conference on Pervasive Computational Technologies (ICPCT)*, 24-29. https://doi.org/10.1109/ICPCT64145.2025.10940335.

[106] Saeed, W., & Omlin, C. (2021). Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities. *ArXiv*, abs/2111.06420. https://doi.org/10.1016/j.knosys.2023.110273.

[107] Haque, A., Islam, A., & Mikalef, P. (2022). Explainable Artificial Intelligence (XAI) from a user perspective- A synthesis of prior literature and problematizing avenues for future research. *ArXiv*, abs/2211.15343. https://doi.org/10.1016/j.techfore.2022.122120.

[108] Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Inf. Fusion*, 76, 89-106. https://doi.org/10.1016/J.INFFUS.2021.05.009.

[109] Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., Keulen, M., & Seifert, C. (2022). From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Computing Surveys*, 55, 1 - 42. https://doi.org/10.1145/3583558.

[110] Kale, A., Nguyen, T., Harris, F., Li, C., Zhang, J., & , X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence*, 5, 139-162. https://doi.org/10.1162/dint_a_00119.

[111] Laato, S., Tiainen, M., Islam, N., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Res.*, 32, 1-31. https://doi.org/10.1108/intr-08-2021-0600.

[112] Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27, 393-444. https://doi.org/10.1007/s11257-017-9195-0.

[113] Kale, A., Nguyen, T., Harris, F., Li, C., Zhang, J., &, X. (2022). Documentation of origin to enable explainable and reliable artificial intelligence: a literature review. *Data Intelligence* , 5, 139-162. https://doi.org/10.1162/dint_a_00119.

[114] Ghasemi, A., Hashtarkhani, S., Schwartz, D., & Shaban-Nejad, A. (2024). Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innovation*, 3. https://doi.org/10.1002/cai2.136.

[115] Chou, Y., Moreira, C., Bruza, P., Ouyang, C., & Jorge, J. (2021). Counterfactuals and Causability in Explainable Artificial Intelligence: Theory, Algorithms, and Applications. *Inf. Fusion*, 81, 59-83. https://doi.org/10.1016/j.inffus.2021.11.003.

[116] Ali, S., Akhlaq, F., Imran, A., Kastrati, Z., Daudpota, S., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Computers in biology and medicine*, 166, 107555 . https://doi.org/10.1016/j.compbiomed.2023.107555.

[117] Laux, J., Wachter, S., & Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18, 3 - 32. https://doi.org/10.1111/rego.12512.

[118] Raimundo, R., & Rosário, A. (2021). The Impact of Artificial Intelligence on Data System Security: A Literature Review. *Sensors (Basel, Switzerland)*, 21. https://doi.org/10.3390/s21217029.

[119] Al-Kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, 11, 58. https://doi.org/10.3390/informatics11030058.

[120] Gao, Y., Wang, Y., Wang, R., Wang, X., Sun, Y., Ding, Y., Xu, H., Chen, Y., Zhao, Y., Huang, H., Li, Y., Zhang, J., Zheng, X., Bai, Y., Ding, H., Wu, Z., Qiu, X., Zhang, J., Li, Y., Sun, J., Wang, C., Gu, J., Wu, B., Chen, S., Zhang, T., Liu, Y., Gong, M., Liu, T., Pan, S., Xie, C., Pang, T., Dong, Y., Jia, R., Zhang, Y., , S., Zhang, X., Gong, N., Xiao, C., Erfani, S., Li, B., Sugiyama, M., Tao, D., Bailey, J., & Jiang, Y. (2025). Safety at Scale: A Comprehensive Survey of Large Model Safety. *ArXiv*, abs/2502.05206. https://doi.org/10.48550/arXiv.2502.05206.

[121] Wäschle, M., Thaler, F., Berres, A., Pölzlbauer, F., & Albers, A. (2022). A review on AI Safety in highly automated driving. *Frontiers in Artificial Intelligence*, 5. https://doi.org/10.3389/frai.2022.952773.

[122] De Micco, F., Di Palma, G., Ferorelli, D., De Benedictis, A., Tomassini, L., Tambone, V., Cingolani, M., & Scendoni, R. (2025). Artificial intelligence in healthcare: transforming patient safety with intelligent systems—A systematic review. *Frontiers in Medicine*, 11. https://doi.org/10.3389/fmed.2024.1522554.

[123] Kuznietsov, A., Gyevnar, B., Wang, C., Peters, S., & Albrecht, S. (2024). Explainable AI for Safe and Trustworthy Autonomous Driving: A Systematic Review. *IEEE Transactions on Intelligent Transportation Systems*, 25, 19342-19364. https://doi.org/10.1109/TITS.2024.3474469.

[124] Salhab, W., Ameyed, D., Jaafar, F., & Mcheick, H. (2024). A Systematic Literature Review on AI Safety: Identifying Trends, Challenges, and Future Directions. *IEEE Access*, 12, 131762-131784. https://doi.org/10.1109/ACCESS.2024.3440647.

[125] Maurya, A., & Kumar, D. (2020). Reliability of safety-critical systems: A state-of-the-art review. *Quality and Reliability Engineering International*, 36, 2547 - 2568. https://doi.org/10.1002/qre.2715.

[126] Wang, Y., & Chung, S. (2021). Artificial intelligence in safety-critical systems: a systematic review. *Ind. Manag. Data Syst.*, 122, 442-470. https://doi.org/10.1108/imds-07-2021-0419.

[127] Neto, A., Camargo, J., Almeida, J., & Cugnasca, P. (2022). Safety Assurance of Artificial Intelligence-Based Systems: A Systematic Literature Review on the State of the Art and Guidelines for Future Work. *IEEE Access*, 10, 130733-130770. https://doi.org/10.1109/ACCESS.2022.3229233.

[128] Mart'inez-Fern'andez, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., Vollmer, A., & Wagner, S. (2021). Software Engineering for AI-Based Systems: A Survey. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31, 1 - 59. https://doi.org/10.1145/3487043.

[129] Ismatullaev, U., & Kim, S. (2022). Review of the Factors Affecting Acceptance of AI-Infused Systems. *Human Factors*, 66, 126 - 144. https://doi.org/10.1177/00187208211064707.

[130] Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2021). Theorising Algorithmic Justice. *European Journal of Information Systems*, 31, 269 - 287. https://doi.org/10.1080/0960085X.2021.1934130.

[131] Pfeiffer, J., Gutschow, J., Haas, C., Möslein, F., Maspfuhl, O., Borgers, F., & Alpsancar, S. (2023). Algorithmic Fairness in AI. *Business & Information Systems Engineering*, 65, 209-222. https://doi.org/10.1007/s12599-023-00787-x.

[132] Gabriel, I. (2021). Toward a Theory of Justice for Artificial Intelligence. *Daedalus*, 151, 218-231. https://doi.org/10.1162/daed_a_01911.

[133] Halsband, A. (2022). Sustainable AI and Intergenerational Justice. *Sustainability*. https://doi.org/10.3390/su14073922.

[134] Bellamy, R., Mojsilovic, A., Nagar, S., Ramamurthy, K., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K., Zhang, Y., Dey, K., Hind, M., Hoffman, S., Houde, S., Kannan, K., Lohia, P., Martino, J., & Mehta, S. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*. https://doi.org/10.1147/jrd.2019.2942287.

[135] Leben, D. (2025). AI Fairness. . https://doi.org/10.7551/mitpress/15740.001.0001.

[136] Acikgoz, Y., Davison, K., Compagnone, M., & Laske, M. (2020). Justice Perceptions of Artificial Intelligence in Selection. *Decision Making*. https://doi.org/10.1111/ijsa.12306.

[137] Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2022). Perceptions of Justice By Algorithms. *Artificial Intelligence and Law*, 31, 269 - 292. https://doi.org/10.1007/s10506-022-09312-z.

[138] Graham, S., & Hopkins, H. (2021). AI for Social Justice: New Methodological Horizons in Technical Communication. *Technical Communication Quarterly*, 31, 89 - 102. https://doi.org/10.1080/10572252.2021.1955151.

[139] Noorman, M., Apráez, B., & Lavrijssen, S. (2023). AI and Energy Justice. *Energies*. https://doi.org/10.3390/en16052110.

[140] Polo, E., & Ailodion, D. (2025). Tackling Racial Bias in AI Systems: Applying the Bioethical Principle of Justice and Insights from Joy Buolamwini's "Coded Bias" and the "Algorithmic Justice League". *Bangladesh Journal of Bioethics*. https://doi.org/10.62865/bjbio.v16i1.129.

[141] Zhang, X., Antwi-Afari, M., Zhang, Y., & Xing, X. (2024). The Impact of Artificial Intelligence on Organizational Justice and Project Performance: A Systematic Literature and Science Mapping Review. *Buildings*. https://doi.org/10.3390/buildings14010259.

[142] Buccella, A. (2022). "AI for all" is a matter of social justice. *Ai and Ethics*, 1 - 10. https://doi.org/10.1007/s43681-022-00222-z.

[143] Chen, J., Yan, H., Liu, Z., Zhang, M., Xiong, H., & Yu, S. (2024). When Federated Learning Meets Privacy-Preserving Computation. *ACM Computing Surveys*, 56, 1 - 36. https://doi.org/10.1145/3679013.

[144] Guo, J., Pietzuch, P., Paverd, A., & Vaswani, K. (2024). Trustworthy AI using Confidential Federated Learning. *Queue*, 22, 87 - 107. https://doi.org/10.1145/3665220.

[145] Kakarala, M., & Rongali, S. (2025). Data Privacy and Security in AI. *World Journal of Advanced Research and Reviews*. https://doi.org/10.30574/wjarr.2025.25.3.0555.

[146] Yu, S., Carroll, F., & Bentley, B. (2024). Insights Into Privacy Protection Research in AI. *IEEE Access*, 12, 41704-41726. https://doi.org/10.1109/ACCESS.2024.3378126.

[147] Lee, D., Antonio, J., & Khan, H. (2024). Privacy-Preserving Decentralized AI with Confidential Computing. *ArXiv*, abs/2410.13752. https://doi.org/10.48550/arXiv.2410.13752.

[148] Guntupalli, N. (2023). Artificial Intelligence as a Service: Providing Integrity and Confidentiality. , 309-315. https://doi.org/10.1007/978-3-031-36402-0_28.

[149] Vaswani, K., Volos, S., Fournet, C., Diaz, A., Gordon, K., Vembu, B., Webster, S., Chisnall, D., Kulkarni, S., Cunningham, G., Osborne, R., & Wilkinson, D. (2023). Confidential Computing within an AI Accelerator. , 501-518.

[150] Searle, R., & Gururaj, P. (2022). Establishing security and trust for object detection and classification with confidential AI. , 12113, 121130C - 121130C-14. https://doi.org/10.1117/12.2618303.

[151] Kostenko O. V. Management of Identification Data: Legal Regulation of Anonymization and Pseudonymization. *Scientific Bulletin of Public and Private Law.* 2021. № 1. **P**. 76-81. DOI: https://doi.org/10.32844/2618-1258.2021.1.13

[152] Kostenko O.V. Legal Regulation of Identity Data Management: UNCITRAL, Cross-Border Trust Space. *Law. State. Technology.* 2021. № 4. С. 56-60. DOI: https://doi.org/10.32782/LST/2021-4-10

[153] Костенко О. IDENTIFICATION DATA MANAGEMENT: LEGAL REGULATION AND CLASSIFICATION. *Scientific Journal of Polonia University.* 2021. Vol. 43. №6. P.198-203. DOI: https://doi.org/10.23856/4325

[154] Kostenko O. V. Management of Identification Data: Legal Regulation and Classification. *Young scientist*. 2021. № 3(91). Pp. 90-94. DOI: https://doi.org/10.32839/2304-5809/2021-3-91-21

[155] Buijsman, S. (2024). Transparency for AI systems: a value-based approach. *Ethics Inf. Technol.*, 26, 34. https://doi.org/10.1007/s10676-024-09770-w.

[156] Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Rev.*, 9. https://doi.org/10.14763/2020.2.1469.

[157] Ehsan, U., Liao, Q., Muller, M., Riedl, M., & Weisz, J. (2021). Expanding Explainability: Towards Social Transparency in AI systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445188.

[158] Hunter, A. (2023). Interactions with AI Systems: Trust and Transparency. *2023 IEEE Engineering Informatics*, 1-6. https://doi.org/10.1109/IEEECONF58110.2023.10520626.

[159] Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Inf. Softw. Technol.*, 159, 107197. https://doi.org/10.1016/j.infsof.2023.107197.

[160] Cheong, B. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*. https://doi.org/10.3389/fhumd.2024.1421273.

[161] Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. (2024). AI auditing: The Broken Bus on the Road to AI Accountability. *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, 612-643. https://doi.org/10.1109/SaTML59370.2024.00037.

[162] Murikah, W., Nthenge, J., & Musyoka, F. (2024). Bias and Ethics of AI Systems Applied in Auditing - A Systematic Review. *Scientific African*. https://doi.org/10.1016/j.sciaf.2024.e02281.

[163] Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI & SOCIETY*, 1-12. https://doi.org/10.2139/ssrn.4180366.

[164] Tóth, Z., Caruana, R., Gruber, T., & Loebbecke, C. (2022). The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability. *Journal of Business Ethics*, 178, 895 - 916. https://doi.org/10.1007/s10551-022-05050-z.

[165] Schmidt, J., Bartsch, S., Adam, M., & Benlian, A. (2025). Elevating Developers' Accountability Awareness in AI Systems Development. *Bus. Inf. Syst. Eng.*, 67, 109-135. https://doi.org/10.1007/s12599-024-00914-2.

[166] Miguel, B., Naseer, A., & Inakoshi, H. (2020). Putting Accountability of AI Systems into Practice. , 5276-5278. https://doi.org/10.24963/ijcai.2020/768.

[167] (2023). Advancing accountability in AI. *OECD Digital Economy Papers*. https://doi.org/10.1787/2448f04b-en.

[168] Kim, B., & Doshi-Velez, F. (2021). Machine Learning Techniques for Accountability. *AI Mag.*, 42, 47-52. https://doi.org/10.1002/j.2371-9621.2021.tb00010.x.

[169] Som, C., Hilty, L., & Köhler, A. (2009). The Precautionary Principle as a Framework for a Sustainable Information Society. *Journal of Business Ethics*, 85, 493-505. https://doi.org/10.1007/S10551-009-0214-X.

[170] Botes, M. (2023). Regulating scientific and technological uncertainty: The precautionary principle in the context of human genomics and AI. *South African journal of science*, 119. https://doi.org/10.17159/sajs.2023/15037.

[171] Druzin, B., Boute, A., & Ramsden, M. (2025). Confronting Catastrophic Risk: The International Obligation to Regulate Artificial Intelligence. *ArXiv*, abs/2503.18983. https://doi.org/10.36642/mjil.46.2.confronting.

[172] Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., Mindermann, S., Oberman, A., Richardson, J., Richardson, O., Rondeau, M., St-Charles, P., & Williams-King, D. (2025). Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?. *ArXiv*, abs/2502.15657. https://doi.org/10.48550/arXiv.2502.15657.

[173] Maior, D. (2025). THE NORMATIVE SIGNIFICANCE OF THE RECAUTIONARY PRINCIPLE IN ARTIFICIAL INTELLIGENCE PROBLEM. *Curentul Juridic/Juridical Current*. https://doi.org/10.62838/cjjc-2024-0040.

[174] Hansson, S. (2020). How Extreme Is the Precautionary Principle?. *NanoEthics*, 14, 245 - 257. https://doi.org/10.1007/s11569-020-00373-5.

[175] Kaivanto, K. (2025). The Precautionary Principle and the Innovation Principle: Incompatible Guides for AI Innovation Governance?. .

[176] Miller, H., & Engemann, C. (2019). The precautionary principle and unforeseen consequences. *Kybernetes* , 48, 265-286. https://doi.org/10.1108/K-01-2018-0050 .

[177] Defur, P., & Kaszuba, M. (2002). Implementing the precautionary principle.. *The Science of the total environment*, 288 1-2, 155-65 . https://doi.org/10.1016/S0048-9697(01)01107-X.

[178] Yin, M., & Zou, K. (2021). The Implementation of the Precautionary Principle in Nuclear Safety Regulation: Challenges and Prospects. *Sustainability*. https://doi.org/10.3390/su132414033.

[179] Foster, K., Vecchia, P., & Repacholi, M. (2000). Science and the Precautionary Principle. *Science*, 288, 979 - 981. https://doi.org/10.1126/SCIENCE.288.5468.979.

[180] (2023). Precaução e inovação: uma análise da regulação de riscos no uso da inteligência artificial. *Revista de Direito Empresarial*. https://doi.org/10.52028/rdemp.v20i1_art08.

[181] Fernandes, R., & Oliveira, L. (2021). A REGULAÇÃO DO AGIR DECISÓRIO DISRUPTIVO NO JUDICIÁRIO BRASILEIRO E A OBSERVÂNCIA DO PRINCÍPIO DA PRECAUÇÃO: JUIZ NATURAL OU "JUIZ ARTIFICIAL"?. , 19, 91-117. https://doi.org/10.12662/2447-6641OJ.V19I30.P91-117.2021.

[182] Aifen, X., Ge, Y., & Khan, I. (2023). Preventing Terrorism with Precaution: An Examination of the Precautionary Principle to Counter-Terrorism Measures. *Journal of Law & Social Studies*. https://doi.org/10.52279/jlss.05.02.153162.

[183] Stefánsson, H. (2019). On the Limits of the Precautionary Principle. *Risk Analysis*, 39. https://doi.org/10.1111/risa.13265.

[184] Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous Auditing of Artificial Intelligence: a Conceptualization and Assessment of Tools and Frameworks. *Digital Society*, 1. https://doi.org/10.1007/s44206-022-00022-2.

[185] Enqvist, L. (2023). 'Human oversight' in the EU artificial intelligence act: what, when and by whom?. *Law, Innovation and Technology*, 15, 508 - 535. https://doi.org/10.1080/17579961.2023.2245683.

[186] Ho-Dac, M., & Martinez, B. (2024). Human Oversight of Artificial Intelligence and Technical Standardisation. *ArXiv*, abs/2407.17481. https://doi.org/10.48550/arXiv.2407.17481.

[187] Laux, J. (2023). Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. *Ai & Society*, 39, 2853 - 2866. https://doi.org/10.1007/s00146-023-01777-z.

[188] Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Res.*, 33, 133-167. https://doi.org/10.1108/intr-01-2022-0042.

[189] Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40, 137 - 157. https://doi.org/10.1080/14494035.2021.1928377.

[190] Shneiderman, B. (2016). Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, 113, 13538 - 13540. https://doi.org/10.1073/pnas.1618211113.

[191] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389 - 399. https://doi.org/10.1038/s42256-019-0088-2.

[192] Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, 4, 799-819. https://doi.org/10.1109/TAI.2022.3194503.

[193] Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *J. Database Manag.*, 31, 74-87. https://doi.org/10.4018/jdm.2020040105.

[194] Jedličková, A. (2024). Ensuring Ethical Standards in the Development of Autonomous and Intelligent Systems. *IEEE Transactions on Artificial Intelligence*, 5, 5863-5872. https://doi.org/10.1109/TAI.2024.3387403.

[195] Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *Int. J. Inf. Manag.*, 62, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433.

[196] Ferrell, O., Harrison, D., Ferrell, L., Ajjan, H., & Hochstein, B. (2024). A theoretical framework to guide AI ethical decision making. *AMS Review*. https://doi.org/10.1007/s13162-024-00275-9.

[197] Osasona, F., Amoo, O., Atadoga, A., Abrahams, T., Farayola, O., & Ayinla, B. (2024). REVIEWING THE ETHICAL IMPLICATIONS OF AI IN DECISION MAKING PROCESSES. *International Journal of Management & Entrepreneurship Research*. https://doi.org/10.51594/ijmer.v6i2.773.

[198] Hanna, M., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., & Rashidi, H. (2024). Ethical and Bias Considerations in Artificial Intelligence (AI)/Machine Learning.. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 100686 . https://doi.org/10.1016/j.modpat.2024.100686.

[199] Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y., Zhang, Y., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., Brauner, J., & Mindermann, S. (2023). Managing extreme AI risks amid rapid progress. *Science*, 384, 842 - 845. https://doi.org/10.1126/science.adn0117.

[200] Steimers, A., & Schneider, M. (2022). Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health*, 19. https://doi.org/10.3390/ijerph19063641.

[201] Lin, H., & Liu, W. (2020). Risks and Prevention in the Application of AI. , 700-704. https://doi.org/10.1007/978-3-030-62746-1_104.

[202] Nyavor, H. (2025). Al powered disease, prevention: Predicting health risks through machine learning for proactive care approaches. *International Journal of Science and Research Archive*. https://doi.org/10.30574/ijsra.2025.15.1.1018.

[203] Lee, M. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5. https://doi.org/10.1177/2053951718756684.

[204] Bogert, E., Schecter, A., & Watson, R. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports*, 11. https://doi.org/10.1038/s41598-021-87480-9.

[205] Li, Y., & Goel, S. (2024). Making It Possible for the Auditing of AI: A Systematic Review of AI Audits and AI Auditability. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-024-10508-8.

[206] Jeyarajan, B., Murugan, A., Pandy, G., & Pugazhenthi, V. (2025). AI for Predictive Monitoring and Anomaly Detection in DevOps Environments. *SoutheastCon 2025*, 450-455. https://doi.org/10.1109/SoutheastCon56624.2025.10971552.

[207] Gummadi, A., Napier, J., & Abdallah, M. (2024). XAI-IoT: An Explainable AI Framework for Enhancing Anomaly Detection in IoT Systems. *IEEE Access*, 12, 71024-71054. https://doi.org/10.1109/ACCESS.2024.3402446.

[208] Калі, У., Катак, Ф., та Халден, У. (2024). Надійні кіберфізичні енергетичні системи з використанням штучного інтелекту: алгоритми дуелі для виявлення аномалій PMU та кібербезпеки. *Artif. Intell. Rev.* , 57, 183. https://doi.org/10.1007/s10462-024-10827-x .

[209] Mishra, S., & Nayak, S. (2025). AI-Driven Anomaly Detection and Performance Optimization in Background Screening Systems. *Scholars Journal of Engineering and Technology*. https://doi.org/10.36347/sjet.2025.v13i02.006.

[210] Kaur, D., Uslu, S., Rittichier, K., & Durresi, A. (2022). Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys (CSUR)*, 55, 1 - 38. https://doi.org/10.1145/3491209.

[211] Mart'inez-Fern'andez, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., Vollmer, A., & Wagner, S. (2021). Software Engineering for AI-Based Systems: A Survey. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31, 1 - 59. https://doi.org/10.1145/3487043.

[212] Sekar, A. (2025). The role of AI/ML in improving system reliability of large-scale distributed systems. *World Journal of Advanced Research and Reviews*. https://doi.org/10.30574/wjarr.2025.26.1.1064.

[213] Sheikh, N. (2025). AI-Driven Observability: Enhancing System Reliability and Performance. *Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023*. https://doi.org/10.60087/jaigs.v7i01.322.

[214] Cheng, Q., Huang, M., Man, C., Shen, A., Dai, L., Yu, H., & Hashimoto, M. (2023). Reliability Exploration of System-on-Chip With Multi-Bit-Width Accelerator for Multi-Precision Deep Neural Networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 70, 3978-3991. https://doi.org/10.1109/TCSI.2023.3300899.

[215] Gnad, D., Gotthard, M., Krautter, J., Kritikakou, A., Meyers, V., Rech, P., Condia, J., Ruospo, A., Sánchez, E., Santos, F., Sentieys, O., Tahoori, M., Tessier, R., & Traiola, M. (2024). Reliability and Security of AI Hardware. *2024 IEEE European Test Symposium (ETS)*, 1-10. https://doi.org/10.1109/ETS61313.2024.10567471.

[216] Moskalenko, V., Kharchenko, V., & Semenov, S. (2024). Model and Method for Providing Resilience to Resource-Constrained AI-System. *Sensors (Basel, Switzerland)*, 24. https://doi.org/10.3390/s24185951.

[217] Moskalenko, V., Kharchenko, V., Moskalenko, A., & Kuzikov, B. (2023). Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods. *Algorithms*, 16, 165. https://doi.org/10.3390/a16030165.

[218] Cody, T., & Beling, P. (2023). Towards operational resilience for AI-based cyber in multi-domain operations. , 12538, 125381D - 125381D-6. https://doi.org/10.1117/12.2675862.

[219] Moskalenko, V., Moskalenko, A., Kudryavtsev, A., & Moskalenko, Y. (2024). Resilience-aware MLOps for Resource-constrained AI-system. , 462-473.

[220] Moskalenko, V., Moskalenko, A., Kudryavtsev, A., & Moskalenko, Y. (2023). Robustness and robust artificial intelligence systems: taxonomy, models, and methods. Algorithms, 16, 165. https://doi.org/10.3390/a16030165 .

[221] Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI & SOCIETY*, 1-12. https://doi.org/10.2139/ssrn.4180366.

[222] Busuioc, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, 81, 825 - 836. https://doi.org/10.1111/puar.13293.

[223] Van De Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*, 30, 385 - 409. https://doi.org/10.1007/s11023-020-09537-4.

[224] Stettinger, G., Weissensteiner, P., & Khastgir, S. (2024). Trustworthiness Assurance Assessment for High-Risk AI-Based Systems. *IEEE Access*, 12, 22718-22745. https://doi.org/10.1109/ACCESS.2024.3364387.

[225] Itsuji, H., Uezono, T., Toba, T., & Kundu, S. (2024). Real-Time Diagnostic Technique for AI-Enabled System. *IEEE Open Journal of Intelligent Transportation Systems*, 5, 483-494. https://doi.org/10.1109/OJITS.2024.3435712.

[226] Zdravković, M., Panetto, H., & Weichhart, G. (2021). AI-enabled Enterprise Information Systems for Manufacturing. *Enterprise Information Systems*, 16, 668 - 720. https://doi.org/10.1080/17517575.2021.1941275.

[227] Murikah, W., Nthenge, J., & Musyoka, F. (2024). Bias and Ethics of AI Systems Applied in Auditing - A Systematic Review. *Scientific African*. https://doi.org/10.1016/j.sciaf.2024.e02281.

[228] Rovzanec, J., Novalija, I., Zajec, P., Kenda, K., Tavakoli, H., Suh, S., Veliou, E., Papamartzivanos, D., Giannetsos, T., Menesidou, S., Alonso, R., Cauli, N., Meloni, A., Recupero, D., Kyriazis, D., Sofianidis, G., Theodoropoulos, S., Fortuna, B., Mladeni'c, D., & Soldatos, J. (2022). Human-centric artificial intelligence architecture for industry 5.0 applications. *International Journal of Production Research*, 61, 6847 - 6872. https://doi.org/10.1080/00207543.2022.2138611.

[229] Riedl, M. (2019). Human-Centered Artificial Intelligence and Machine Learning. *ArXiv*, abs/1901.11184. https://doi.org/10.1002/HBE2.117.

[230] Amariles, D., & Baquero, P. (2023). Promises and limits of law for a human-centric artificial intelligence. *Comput. Law Secur. Rev.*, 48, 105795. https://doi.org/10.1016/j.clsr.2023.105795.

[231] Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S., & Sharma, R. (2024). Applications, Challenges, and Future Directions of Human-in-the-Loop Learning. *IEEE Access*, 12, 75735-75760. https://doi.org/10.1109/ACCESS.2024.3401547.

[232] Zhang, P., Liu, W., & Shao, J. (2022). Research on Human-in-the-loop Traffic Adaptive Decision Making Method. *2022 4th International Conference on Robotics and Computer Vision (ICRCV)*, 272-276. https://doi.org/10.1109/ICRCV55858.2022.9953216.

[233] Enarsson, T., Enqvist, L., & Naarttijärvi, M. (2021). Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law*, 31, 123 - 153. https://doi.org/10.1080/13600834.2021.1958860.

[234] Johnson, J. (2022). Automating the OODA loop in the age of intelligent machines: reaffirming the role of humans in command-and-control decision-making in the digital age. *Defence Studies*, 23, 43 - 67. https://doi.org/10.1080/14702436.2022.2102486.

[235] Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*. https://doi.org/10.1007/s40685-020-00133-x.

[236] Wulf, A., & Seizov, O. (2022). "Please understand we cannot provide further information": evaluating content and transparency of GDPR-mandated AI disclosures. *AI Soc.*, 39, 235-256. https://doi.org/10.1007/s00146-022-01424-z.

[237] Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *Cybersecurity*. https://doi.org/10.2139/ssrn.3063289.

[238] Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Comput. Law Secur. Rev.*, 35, 105327. https://doi.org/10.1016/J.CLSR.2019.05.002.

[239] Cobbe, J., & Singh, J. (2020). Reviewable Automated Decision-Making. *Comput. Law Secur. Rev.*, 39, 105475. https://doi.org/10.1016/j.clsr.2020.105475.
[240] Nassen, L., Vandebosch, H., Poels, K., & Karsay, K. (2023). Opt-out, abstain, unplug. A systematic review of the voluntary digital disconnection literature. *Telematics Informatics*, 81, 101980. https://doi.org/10.1016/j.tele.2023.101980.
[241] Chen, T., Guo, W., Gao, X., & Liang, Z. (2020). AI-based self-service technology in public service delivery: User experience and influencing factors. *Gov. Inf. Q.*, 38, 101520. https://doi.org/10.1016/j.giq.2020.101520.
[242] Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2022.121786.
[243] Felzmann, H., Villaronga, E., Lutz, C., & Tamó-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6. https://doi.org/10.1177/2053951719860542.
[244] Buiten, M. (2019). Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation*, 10, 41 - 59. https://doi.org/10.1017/err.2019.8.
[245] Du, Y. (2022). On the Transparency of Artificial Intelligence System. *Journal of Autonomous Intelligence*. https://doi.org/10.32629/jai.v5i1.486.
[246] Popovych, T. (2024). Legal obligations of transparency in the field of artificial intelligence. *Uzhhorod National University Herald. Series: Law*. https://doi.org/10.24144/2307-3322.2024.85.4.59.
[247] Alakbarzadeh, V. (2025). Beynəlxalq hüquq kontekstində süni intellekt - insan hüquqları və hüquqi hesabatlılıq çağırışları. *Azerbaijan Law Journal*. https://doi.org/10.61638/yzzp8534.
[248] Kolarević, E. (2022). The influence of Artificial intelligence on the right to freedom of expression. *Pravo - teorija i praksa*. https://doi.org/10.5937/ptp2201111k.
[249] Polok, B., El-Taj, H., & Rana, A. (2023). Balancing Potential and Peril: The Ethical Implications of Artificial Intelligence on Human Rights. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4484386.
[250] Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, 9. https://doi.org/10.1177/20563051231186353.
[251] Kouroupis, K. (2024). AI and politics: ensuring or threatening democracy?. *Juridical Tribune*. https://doi.org/10.24818/tbj/2023/13/4.05.
[252] Teckchandani, J. (2024). AI in International Politics. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2024.58934.
[253] Bender, S. (2022). Algorithmic Elections. *Michigan Law Review*. https://doi.org/10.36644/mlr.121.3.algorithmic.
[254] Savaget, P., Chiarini, T., & Evans, S. (2018). Empowering political participation through artificial intelligence. *Science & Public Policy*, 46, 369 - 380. https://doi.org/10.1093/scipol/scy064.
[255] Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Comput. Hum. Behav.*, 134, 107308. https://doi.org/10.1016/j.chb.2022.107308.
[256] Dodig-Crnkovic, G., Basti, G., & Holstein, T. (2024). Delegating Responsibilities to Intelligent Autonomous Systems: Challenges and Benefits. *Journal of bioethical inquiry*. https://doi.org/10.48550/arXiv.2411.15147.
[257] Tretter, M. (2025). Opportunities and challenges of AI-systems in political decision-making contexts. *Frontiers in Political Science*. https://doi.org/10.3389/fpos.2025.1504520.
[258] Caiza, G., Sanguña, V., Tusa, N., Masaquiza, V., Ortiz, A., & Garcia, M. (2024). Navigating Governmental Choices: A Comprehensive Review of Artificial Intelligence's Impact on Decision-Making. *Informatics*, 11, 64. https://doi.org/10.3390/informatics11030064.

[259] Presuel, R., & Sierra, J. (2023). The Adoption of Artificial Intelligence in Bureaucratic Decision-making: A Weberian Perspective. *Digital Government: Research and Practice*, 5, 1 - 20. https://doi.org/10.1145/3609861.

[260] Kolkman, D., Bex, F., Narayan, N., & Van Der Put, M. (2024). Justitia ex machina: The impact of an AI system on legal decision-making and discretionary authority. *Big Data & Society*, 11. https://doi.org/10.1177/20539517241255101.

[261] Morić, Z., Dakić, V., & Urošev, S. (2025). An AI-Based Decision Support System Utilizing Bayesian Networks for Judicial Decision-Making. *Systems*. https://doi.org/10.3390/systems13020131.

[262] Greenstein, S. (2021). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, 30, 291 - 323. https://doi.org/10.1007/s10506-021-09294-4.

[263] Formosa, P., Rogers, W., Griep, Y., Bankins, S., & Richards, D. (2022). Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. *Comput. Hum. Behav.*, 133, 107296. https://doi.org/10.1016/j.chb.2022.107296.

[264] Orwat, C. (2024). Algorithmic Discrimination From the Perspective of Human Dignity. *Social Inclusion*. https://doi.org/10.17645/si.7160.

[265] Alakwe, K. (2023). Human Dignity in the Era of Artificial Intelligence and Robotics: Issues and Prospects. *Journal of Humanities and Social Sciences Studies*. https://doi.org/10.32996/jhsss.2023.5.6.10.

[266] Federspiel, F., Mitchell, R., Asokan, A., Umaña, C., & Mccoy, D. (2023). Threats by artificial intelligence to human health and human existence. *BMJ Global Health*, 8. https://doi.org/10.1136/bmjgh-2022-010435.

[267] Ahmad, S., Han, H., Alam, M., Rehmat, M., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities & Social Sciences Communications*, 10. https://doi.org/10.1057/s41599-023-01787-8.

[268] Gürkaynak, G., Yilmaz, I., & Haksever, G. (2016). Stifling artificial intelligence: Human perils. *Comput. Law Secur. Rev.*, 32, 749-758. https://doi.org/10.1016/J.CLSR.2016.05.003.

[269] Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Comput. Hum. Behav.*, 127, 107018. https://doi.org/10.1016/j.chb.2021.107018.

[270] Merlec, M., Lee, Y., Hong, S., & In, H. (2021). A Smart Contract-Based Dynamic Consent Management System for Personal Data Usage under GDPR. *Sensors (Basel, Switzerland)*, 21. https://doi.org/10.3390/s21237994.

[271] Florea, M. (2023). Withdrawal of consent for processing personal data in biomedical research. *International Data Privacy Law*. https://doi.org/10.1093/idpl/ipad008.

[272] Hagendorff, T. (2019). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30, 99 - 120. https://doi.org/10.1007/s11023-020-09517-8.

[273] Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *Int. J. Inf. Manag.*, 62, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433.

[274] Ortega-Bolaños, R., Bernal-Salcedo, J., Ortiz, M., Sarmiento, J., Ruz, G., & Tabares-Soto, R. (2024). Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. *Artif. Intell. Rev.*, 57, 110. https://doi.org/10.1007/s10462-024-10740-3.

[275] Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *J. Database Manag.*, 31, 74-87. https://doi.org/10.4018/jdm.2020040105.

[276] Murikah, W., Nthenge, J., & Musyoka, F. (2024). Bias and Ethics of AI Systems Applied in Auditing - A Systematic Review. *Scientific African*. https://doi.org/10.1016/j.sciaf.2024.e02281.

[277] Goldenthal, E., Park, J., Liu, S., Mieczkowski, H., & Hancock, J. (2021). Not All AI are Equal: Exploring the Accessibility of AI-Mediated Communication Technology. *Comput. Hum. Behav.*, 125, 106975. https://doi.org/10.1016/j.chb.2021.106975.

[278] Singh, K., & , C. (2024). Bias and Fairness in Artificial Intelligence: Methods and Mitigation Strategies. *International Journal for Research Publication and Seminar*. https://doi.org/10.36676/jrps.v15.i3.1425.

[279] Davoodi, A. (2024). EQUAL AI: A Framework for Enhancing Equity, Quality, Understanding and Accessibility in Liberal Arts through AI for Multilingual Learners. *Language, Technology, and Social Media*. https://doi.org/10.70211/ltsm.v2i2.139.

[280] Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Gov. Inf. Q.*, 36, 358-367. ttps://doi.org/10.1016/J.GIQ.2018.10.001.

[281] Sezgin, E., & Kocaballi, A. (2024). Era of Generalist Conversational Artificial Intelligence to Support Public Health Communications. *Journal of Medical Internet Research*, 27. https://doi.org/10.2196/69007.

[282] Türksoy, N. (2022). The Future of Public Relations, Advertising and Journalism: How Artificial Intelligence May Transform the Communication Profession and Why Society Should Care. *Türkiye İletişim Araştırmaları Dergisi/26306220*. https://doi.org/10.17829/turcom.1050491.

[283] Guzman, A., & Lewis, S. (2019). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, 22, 70 - 86. https://doi.org/10.1177/1461444819858691.

[284] Soldan, T. (2022). A Qualitative Research on The Use of Artificial Intelligence in Public Relations. *The Journal of International Scientific Researches*. https://doi.org/10.23834/isrjournal.1113438.

[285] Alkrisheh, M., & Gourari, F. (2025). CRIMINAL LIABILITY FOR PAID DISINFORMATION IN THE DIGITAL WORLD: A COMPARATIVE STUDY BETWEEN UAE LAW AND THE EUROPEAN DIGITAL SERVICES ACT (DSA). *Access to Justice in Eastern Europe*. https://doi.org/10.33327/ajee-18-8.2-r000110.

[286] Gomathy, D., Geetha, V., Manohar, S., & Rajesh, P. (2024). LEGAL FRAMEWORKS FOR REGULATING CYBERCRIME AND CYBER TERRORISM. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*. https://doi.org/10.55041/ijsrem37509.

[287] Shinde, N. (2025). CYBER TERRORISM: THE EMERGING THREAT IN THE DIGITAL AGE. *International Journal For Multidisciplinary Research*. https://doi.org/10.36948/ijfmr.2025.v07i02.43602.

[288] Radoniewicz, F. (2021). Zwalczanie cyberterroryzmu w ramach UE – wybrane aspekty karnomaterialne. *Cybersecurity and Law*. https://doi.org/10.35467/cal/133898.

[289] Al-Shair, M. (2021). Legal problems in confronting digital terrorism. Statistic study. *Journal of Al-Rafidain University College For Sciences (Print ISSN: 1681-6870 ,Online ISSN: 2790-2293)*. https://doi.org/10.55562/jrucs.v28i2.385.

[290] Bajpai, S. (2020). Legal Framework on Cyber Terrorism. , 40, 933-945.

[291] Taylor, R., Fritsch, E., & Liederbach, J. (2005). Digital Crime and Digital Terrorism.

[292] Kolkman, D., Bex, F., Narayan, N., & Van Der Put, M. (2024). Justitia ex machina: The impact of an AI system on legal decision-making and discretionary authority. *Big Data & Society*, 11. https://doi.org/10.1177/20539517241255101.

[293] Atkinson, K., Bench-Capon, T., & Bollegala, D. (2020). Explanation in AI and law: Past, present and future. *Artif. Intell.*, 289, 103387. https://doi.org/10.1016/J.ARTINT.2020.103387.

[294] Vujicic, J. (2025). AI Ethics in Legal Decision-Making Bias, Transparency, And Accountability. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*. https://doi.org/10.15662/ijareeie.2025.1404001.

[295] Scherer, M. (2019). Artificial Intelligence and Legal Decision-Making: The Wide Open?. *Journal of International Arbitration*. https://doi.org/10.54648/joia2019028.
[296] Dokumacı, M. (2024). AI-Driven Econometric Models for Legal Issues. *Human Computer Interaction*. https://doi.org/10.62802/btfvze98.
[297] Samee, N., Alabdulhafith, M., Shah, M., & Rizwan, A. (2024). JusticeAI: A Large Language Models Inspired Collaborative and Cross-Domain Multimodal System for Automatic Judicial Rulings in Smart Courts. *IEEE Access*, 12, 173091-173107. https://doi.org/10.1109/ACCESS.2024.3491775.
[298] Andriati, S., Rizki, I., & Malian, A. (2024). Justice on Trial: How Artificial Intelligence is Reshaping Judicial Decision-Making. *Journal of Indonesian Legal Studies*. https://doi.org/10.15294/jils.v9i2.13683.
[299] Mohan, B., & , D. (2023). The Ethics of Artificial Intelligence in Legal Decision Making: An Empirical Study. *psychologyandeducation*. https://doi.org/10.48047/pne.2018.55.1.38.
[300] Sharma, R. (2023). 36 Exploring the Ethical Implications of AI in Legal Decision-Making. *Indian Journal of Law*. https://doi.org/10.36676/ijl.2023-v1i1-06.
[301] Roberts, H. (2024). Digital sovereignty and artificial intelligence: a normative approach. *Ethics Inf. Technol.*, 26, 70. https://doi.org/10.1007/s10676-024-09810-5.
[302] Usman, H., Nawaz, B., & Naseer, S. (2023). The Future of State Sovereignty in the Age of Artificial Intelligence. *Journal of Law & Social Studies*. https://doi.org/10.52279/jlss.05.02.142152.
[303] Klare, M., Verlande, L., Greiner, M., & Lechner, U. (2022). How Blockchain and Artificial Intelligence influence Digital Sovereignty. , 3-16. https://doi.org/10.1007/978-3-031-30694-5_1.
[304] Al-Zubaidi, R., & Zeidan, R. (2024). Artificial Intelligence Technology and its Implications for the Sovereignty of the Nation-State. *International Journal of Educational Sciences and Arts*. https://doi.org/10.59992/ijesa.2024.v3n5p3.
[305] Goralski, M., & Tan, T. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*. https://doi.org/10.1016/j.ijme.2019.100330.
[306] Dhamija, P., & Bag, S. (2020). Role of artificial intelligence in operations environment: a review and bibliometric analysis. *The TQM Journal*. https://doi.org/10.1108/tqm-10-2019-0243.
[307] Al-Zubaidi, R., & Zeidan, R. (2024). Artificial Intelligence Technology and its Implications for the Sovereignty of the Nation-State. *International Journal of Educational Sciences and Arts*. https://doi.org/10.59992/ijesa.2024.v3n5p3.
[308] March, C., & Schieferdecker, I. (2023). Technological Sovereignty as Ability, Not Autarky. *CESifo: Macro*. https://doi.org/10.1093/isr/viad012.
[309] Curzon, J., Kosa, T., Akalu, R., & El-Khatib, K. (2021). Privacy and Artificial Intelligence. *IEEE Transactions on Artificial Intelligence*, 2, 96-108. https://doi.org/10.1109/TAI.2021.3088084.
[310] Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *Int. J. Inf. Manag.*, 62, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433.
[311] Saura, J., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2022). Assessing behavioral data science privacy issues in government artificial intelligence deployment. *Gov. Inf. Q.*, 39, 101679. https://doi.org/10.1016/j.giq.2022.101679.
[312] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernández, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10. https://doi.org/10.1002/widm.1356.
[313] Raimundo, R., & Rosário, A. (2021). The Impact of Artificial Intelligence on Data System Security: A Literature Review. *Sensors (Basel, Switzerland)*, 21. https://doi.org/10.3390/s21217029.
[314] Roberts, H. (2024). Digital sovereignty and artificial intelligence: a normative approach. *Ethics Inf. Technol.*, 26, 70. https://doi.org/10.1007/s10676-024-09810-5.

[315] Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: the false promise of digital sovereignty. *European Security*, 31, 415 - 434. https://doi.org/10.1080/09662839.2022.2101885.

[316] Nanni, R., Bizzaro, P., & Napolitano, M. (2024). The false promise of individual digital sovereignty in Europe: Comparing artificial intelligence and data regulations in China and the European Union. *Policy & Internet*. https://doi.org/10.1002/poi3.424.

[317] Floridi, L. (2020). The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. *Philosophy & Technology*, 33, 369 - 378. https://doi.org/10.1007/s13347-020-00423-6.

[318] Costa-Barbosa, A., Herlo, B., & Joost, G. (2024). Digital Sovereignty in times of AI: between perils of hegemonic agendas and possibilities of alternative approaches. *Liinc em Revista*. https://doi.org/10.18617/liinc.v20i2.7312.

[319] Circiumaru, A. (2021). The EU's Digital Sovereignty - The role of Artificial Intelligence and Competition Policy. *Social Science Research Network*. https://doi.org/10.2139/SSRN.3831815.

[320] Sheikh, H. (2022). European Digital Sovereignty: A Layered Approach. *Digital Society*, 1. https://doi.org/10.1007/s44206-022-00025-z.

[321] Valente, J. (2024). Data Workers in AI development. *Liinc em Revista*. https://doi.org/10.18617/liinc.v20i2.7302.

[322] Schmitt, M. (2023). Securing the digital world: Protecting smart infrastructures and digital industries with artificial intelligence (AI)-enabled malware and intrusion detection. *J. Ind. Inf. Integr.*, 36, 100520. https://doi.org/10.1016/j.jii.2023.100520.

[323] Dall'Agnol, A. (2022). Artificial Intelligence and the Future of War: The United States, China, and Strategic Stability. *Journal of Strategic Studies* , 46, 749-751. https://doi.org/10.1080/01402390.2022.2104255 .

[324] Johnson, J. (2021). Artificial intelligence and the future of warfare. . https://doi.org/10.7765/9781526145062.

[325] Wilson, C. (2020). Artificial Intelligence and Warfare. , 125-140. https://doi.org/10.1007/978-3-030-28285-1_7.

[326] Маас, М. (2019). Наскільки життєздатним є міжнародний контроль над озброєннями для військового штучного інтелекту? Три уроки ядерної зброї. *Сучасна політика безпеки* , 40, 285-311. https://doi.org/10.1080/13523260.2019.1576464 .

[327] Dresp-Langley, B. (2023). The weaponization of artificial intelligence: What the public needs to be aware of. *Frontiers in Artificial Intelligence*, 6. https://doi.org/10.3389/frai.2023.1154184.

[328] Abaimov, S., & Martellini, M. (2020). Artificial Intelligence in Autonomous Weapon Systems. *21st Century Prometheus*. https://doi.org/10.1007/978-3-030-28285-1_8.

[329] Zohuri, B. (2024). Harnessing Artificial Intelligence for Countering Hypersonic Weapons:A New Frontier in Battlefield Offense and Defense(A Short Review). *Journal of Energy and Power Engineering*. https://doi.org/10.17265/1934-8975/2024.04.002.

[330] Cîrdei, I. (2024). The Use of Artificial Intelligence and Autonomous Weapon Systems in Military Operations. *International conference KNOWLEDGE-BASED ORGANIZATION*, 30, 43 - 51. https://doi.org/10.2478/kbo-2024-0006.

[331] Rashid, A., Kausik, A., Sunny, A., & Bappy, M. (2023). Artificial Intelligence in the Military: An Overview of the Capabilities, Applications, and Challenges. *Int. J. Intell. Syst.*, 2023, 1-31. https://doi.org/10.1155/2023/8676366.

[332] Asaro, P. (2020). Autonomous Weapons and the Ethics of Artificial Intelligence. *Ethics of Artificial Intelligence*. https://doi.org/10.1093/oso/9780190905033.003.0008.

[333] Sharan, Y., Gordon, T., & Florescu, E. (2021). Artificial Intelligence and Autonomous Weapons. *Tripping Points on the Roads to Outwit Terror*. https://doi.org/10.1007/978-3-030-72571-6_7.

[334] Heinz, A. (2025). The militarization of artificial intelligence and the autonomous weapons. *UNISCI Journal*. https://doi.org/10.31439/unisci-222.

[335] Márton, A. (2021). Steps toward a digital ecology: ecological principles for the study of digital ecosystems. *Journal of Information Technology*, 37, 250 - 265. https://doi.org/10.1177/02683962211043222.

[336] Koch, M., Krohmer, D., Naab, M., Rost, D., & Trapp, M. (2022). A matter of definition: Criteria for digital ecosystems. *Digital Business*. https://doi.org/10.1016/j.digbus.2022.100027.

[337] Pekkarinen, S., Hasu, M., Melkas, H., & Saari, E. (2020). Information ecology in digitalising welfare services: a multi-level analysis. *Inf. Technol. People*, 34, 1697-1720. https://doi.org/10.1108/itp-12-2019-0635.

[338] Petrova, E. (2022). Ecology of the Digital Environment as an Attempt to Respond to the Civilizational Challenges of the Digital Age. *Voprosy Filosofii*. https://doi.org/10.21146/0042-8744-2022-11-99-109.

[339] Nedungadi, P., Devenport, K., Sutcliffe, R., & Raman, R. (2020). Towards a digital learning ecology to address the grand challenge in adult literacy. *Interactive Learning Environments*, 31, 383 - 396. https://doi.org/10.1080/10494820.2020.1789668.

[340] Dunleavy, P., & Margetts, H. (2023). Data science, artificial intelligence and the third wave of digital era governance. *Public Policy and Administration*, 40, 185 - 214. https://doi.org/10.1177/09520767231198737.

[341] Dwivedi, Y., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P., Janssen, M., Jones, P., Kar, A., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Meunier-FitzHugh, K., Meunier-FitzHugh, L., Misra, S., Mogaji, E., Sharma, S., Singh, J., Raghavan, V., Raman, R., Rana, N., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., & Williams, M. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. https://doi.org/10.1016/J.IJINFOMGT.2019.08.002.

[342] Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510.

[343] Silcox, C., Zimlichman, E., Huber, K., Rowen, N., Saunders, R., McClellan, M., Kahn, C., Salzberg, C., & Bates, D. (2024). The potential for artificial intelligence to transform healthcare: perspectives from international health leaders. *NPJ Digital Medicine*, 7. https://doi.org/10.1038/s41746-024-01097-6.

[344] Johnson, P., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development – Automation or augmentation, exploration or exploitation?. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2022.121636.

[345] Ooi, K., Tan, G., Al-Emran, M., Al-Sharafi, M., Căpăţînă, A., Chakraborty, A., Dwivedi, Y., Huang, T., Kar, A., Lee, V., Loh, X., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N., Sarker, P., Sharma, A., Teng, C., Wamba, S., & Wong, L. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, 65, 76 - 107. https://doi.org/10.1080/08874417.2023.2261010.

[346] Aldoseri, A., Al-Khalifa, K., & Hamouda, A. (2024). AI-Powered Innovation in Digital Transformation: Key Pillars and Industry Impact. *Sustainability*. https://doi.org/10.3390/su16051790.

[347] Filgueiras, F. (2023). Artificial intelligence and education governance. *Education, Citizenship and Social Justice*, 19, 349 - 361. https://doi.org/10.1177/17461979231160674.

[348] Borgesius, F. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, 24, 1572 - 1593. https://doi.org/10.1080/13642987.2020.1743976.

[349] Allen, R., & Masters, D. (2020). Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making. *ERA Forum*, 20, 585-598. https://doi.org/10.1007/s12027-019-00582-w.

[350] Arnanz, A. (2023). Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules. *IDP. Revista de Internet, Derecho y Política*. https://doi.org/10.7238/idp.v0i38.403794.

[351] Schwitzgebel, E., & Garza, M. (2015). A Defense of the Rights of Artificial Intelligences. *Midwest Studies in Philosophy*, 39, 98-119. https://doi.org/10.1111/MISP.12032.

[352] Aizenberg, E., & Van Den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7. https://doi.org/10.1177/2053951720949566.

[353] Donahoe, E., & Metzger, M. (2019). Artificial Intelligence and Human Rights. *Journal of Democracy*, 30, 115 - 126. https://doi.org/10.1353/JOD.2019.0029.

[354] Laitinen, A., & Sahlgren, O. (2021). AI Systems and Respect for Human Autonomy. *Frontiers in Artificial Intelligence*, 4. https://doi.org/10.3389/frai.2021.705164.

[355] Amariles, D., & Baquero, P. (2023). Promises and limits of law for a human-centric artificial intelligence. *Comput. Law Secur. Rev.*, 48, 105795. https://doi.org/10.1016/j.clsr.2023.105795.

[356] Orwat, C. (2024). Algorithmic Discrimination From the Perspective of Human Dignity. *Social Inclusion*. https://doi.org/10.17645/si.7160.

[357] Lamers, L., Meijerink, J., Jansen, G., & Boon, M. (2022). A Capability Approach to worker dignity under Algorithmic Management. *Ethics and Information Technology*, 24. https://doi.org/10.1007/s10676-022-09637-y.

[358] Filho, E., & Firmo, M. (2023). Human Dignity and neurorights in the Digital Age. *Brazilian Journal of Law, Technology and Innovation*. https://doi.org/10.59224/bjlti.v1i2.87-107.

[359] Zhao, Y., & Ren, Z. (2025). The Alignment of Values: Embedding Human Dignity in Algorithmic Bias Governance for the AGI Era. *International Journal of Digital Law and Governance*, 0. https://doi.org/10.1515/ijdlg-2025-0006.

[360] Ruster, L., Oliva-Altamirano, P., & Daniell, K. (2022). Centring dignity in algorithm development: testing a Dignity Lens. *Proceedings of the 34th Australian Conference on Human-Computer Interaction*. https://doi.org/10.1145/3572921.3572938.

[361] Zwitter, A., Gstrein, O., & Yap, E. (2020). Digital Identity and the Blockchain: Universal Identity Management and the Concept of the "Self-Sovereign" Individual. , 3. https://doi.org/10.3389/fbloc.2020.00026.

[362] Vardanyan, L., Hamuľák, O., & Kocharyan, H. (2024). Fragmented Identities: Legal Challenges of Digital Identity, Integrity, and Informational Self-Determination. *European Studies*, 11, 105 - 121. https://doi.org/10.2478/eustu-2024-0005.

[363] Tan, K., Chi, C., & Lam, K. (2023). Survey on Digital Sovereignty and Identity: From Digitization to Digitalization. *ACM Computing Surveys*, 56, 1 - 36. https://doi.org/10.1145/3616400.

[364] Huu, P. (2023). Impact of employee digital competence on the relationship between digital autonomy and innovative work behavior: a systematic review. *Artificial Intelligence Review*, 1 - 30. https://doi.org/10.1007/s10462-023-10492-6.

[365] Savolainen, L., & Ruckenstein, M. (2022). Dimensions of autonomy in human–algorithm relations. *New Media & Society*, 26, 3472 - 3490. https://doi.org/10.1177/14614448221100802.

[366] Kostenko, O. V. (2021) Management of Identification Data: Legal Regulation of Anonymization and Pseudonymization *Naukovyi visnyk publichnoho ta pryvatnoho prava*, 1, 123–131. https://doi.org/10.32844/2618-1258.2021.1.13

[367] Pfitzmann, A., & Hansen, M. (2010). A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. .

[368] Pfitzmann, A., & Köhntopp, M. (2000). Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. , 1-9. https://doi.org/10.1007/3-540-44702-4_1.

[369] Froomkin, A. (1999). Legal Issues in Anonymity and Pseudonymity. *Legal Perspectives in Information Systems eJournal*. https://doi.org/10.1080/019722499128574.
[370] Garcia-Grau, F., Herrera-Joancomartí, J., & Josa, A. (2022). Attribute Based Pseudonyms: Anonymous and Linkable Scoped Credentials. *Mathematics*. https://doi.org/10.3390/math10152548.
[371] Federrath, H. (2001). Designing Privacy Enhancing Technologies. . https://doi.org/10.1007/3-540-44702-4.
[372] Dange, A., Vilas, R., & , B. (2025). Enhancing Anonymity and Security in Networks: A Comprehensive Analysis of Pseudonym Manager (PM) and Nymble Manager (NM). *Power System Technology*. https://doi.org/10.52783/pst.1663.
[373] Kurzynoga, M. (2024). The Right to Disconnect: Rest in the Digital Age of Work from the International, European and Polish Law Perspectives. *Acta Universitatis Lodziensis. Folia Iuridica*. https://doi.org/10.18778/0208-6069.107.06.
[374] Kolomoets, E., Shoniya, G., Mekhmonov, S., Abdulnabi, S., & Karim, N. (2023). The Employee's Right to Work Offline: A Comparative Analysis of Legal Frameworks in Different Countries. *Revista de Gestão Social e Ambiental*. https://doi.org/10.24857/rgsa.v17n5-009.
[375] Reyna, J., Gabardo, E., & De Sousa Santos, F. (2020). Electronic government, digital invisibility and fundamental social rights. *Seqüência: Estudos Jurídicos e Políticos*. https://doi.org/10.5007/2177-7055.2020v41n85p30.
[376] Wolski, O. (2021). The right to stay offline? Not during the pandemic. *Journal of Information Technology & Politics*, 19, 140 - 155. https://doi.org/10.1080/19331681.2021.1936845.
[377] Mladenov, M., & Serotila, I. (2024). Right to be offline: To be or not to be?. *XXI međunarodni naučni skup Pravnički dani - Prof. dr Slavko Carić, na temu: Odgovori pravne nauke na izazove savremenog društva - zbornik radova*. https://doi.org/10.5937/pdsc24271m.
[378] Gawełko-Bazan, K. (2024). The right to be offline – notes on the background of existing and proposed regulations. Part II: Polish law. *Kwartalnik Prawa Międzynarodowego*. https://doi.org/10.5604/01.3001.0054.4284.
[379] Yang, P. (2024). Problems and Countermeasures of Legal Protection of Laborer's "Off-line Right" in China in the Information Age. *International Journal of Frontiers in Sociology*. https://doi.org/10.25236/ijfs.2024.060307.
[380] Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*. https://doi.org/10.54648/cola2018095.
[381] Duani, N., Barasch, A., & Morwitz, V. (2024). Demographic Pricing in the Digital Age: Assessing Fairness Perceptions in Algorithmic versus Human-Based Price Discrimination. *Journal of the Association for Consumer Research*, 9, 257 - 268. https://doi.org/10.1086/729440.
[382] Wachter, S., Mittelstadt, B., & Russell, C. (2020). Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI. *ArXiv*, abs/2005.05906. https://doi.org/10.2139/ssrn.3547922.
[383] Varona, D., & Suárez, J. (2022). Discrimination, Bias, Fairness, and Trustworthy AI. *Applied Sciences*. https://doi.org/10.3390/app12125826.
[384] Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2945386.
[385] Nachbar, T. (2020). Algorithmic Fairness, Algorithmic Discrimination. *Artificial Intelligence - Law*.
[386] Wang, X., Wu, Y., Ji, X., & Fu, H. (2024). Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. *Frontiers in Artificial Intelligence*, 7. https://doi.org/10.3389/frai.2024.1320277.

[387] Tesink, V., Douglas, T., Forsberg, L., Ligthart, S., & Meynen, G. (2023). Neurointerventions in Criminal Justice: On the Scope of the Moral Right to Bodily Integrity. *Neuroethics*, 16, 1-11. https://doi.org/10.1007/s12152-023-09526-1.

[388] Smolenski, J. (2024). The foundations of informed consent and bodily self-sovereignty: a positive suggestion.. *Monash bioethics review*. https://doi.org/10.1007/s40592-024-00203-4.

[389] Harbinja, E., Edwards, L., & McVey, M. (2023). Governing ghostbots. *Comput. Law Secur. Rev.*, 48, 105791. https://doi.org/10.1016/j.clsr.2023.105791.

[390] Sullivan, C., & Stalla-Bourdillon, S. (2015). Digital identity and French personality rights - A way forward in recognising and protecting an individual's rights in his/her digital identity. *Comput. Law Secur. Rev.*, 31, 268-279. https://doi.org/10.1016/J.CLSR.2015.01.002.

[391] Solove, D. (2022). The Digital Person. . https://doi.org/10.18574/nyu/9780814708965.001.0001.

[392] Augustian, A. (2024). The Role of Personality Rights in Indian Law: Lessons from Jackie Shroff's Legal Battle. *Trends in Intellectual Property Research*. https://doi.org/10.69971/tipr.1.2.2023.13.

[393] De Miguel Asensio, P. (2022). Protection of Reputation, Good Name and Personality Rights in Cross-Border Digital Media. *GRUR International*. https://doi.org/10.1093/grurint/ikac090.

[394] Sayed, A. (2024). Legal Protection of Personal Images in the Era of Modern Technology 'Comparative Study'. *International Journal of Religion*. https://doi.org/10.61707/kz59nw52.

[395] Chu, C., Nyrup, R., Leslie, K., Shi, J., Bianchi, A., Lyn, A., McNicholl, M., Khan, S., Rahimi, S., & Grenier, A. (2022). Digital Ageism: Challenges and Opportunities in Artificial Intelligence for Older Adults. *The Gerontologist*, 62, 947 - 955. https://doi.org/10.1093/geront/gnab167.

[396] Tacheva, J., & Ramasubramanian, S. (2023). AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data & Society*, 10. https://doi.org/10.1177/20539517231219241.

[397] Brock, J., & Von Wangenheim, F. (2019). Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence. *California Management Review*, 61, 110 - 134. https://doi.org/10.1177/1536504219865226.