РОЗДІЛ XI. МІЖНАРОДНЕ ПРАВО

УДК 341.1/.8:004.8:35.077
DOI https://doi.org/10.24144/2307-3322.2025.90.5.8

LEGAL MECHANISMS OF AUDIT AND ACCOUNTABILITY IN ARTIFICIAL INTELLIGENCE SYSTEMS IN THE CONTEXT OF INTERNATIONAL LAW

Afgan V.A.,
Baku State University, Faculty of Law, UNESCO Department of Human Rights and Information Law, PhD student, e-mail: vadiya.zadeh@gmail.com

Afgan V.A. Legal mechanisms of audit and accountability in artificial intelligence systems in the context of international law.

This article is devoted to the oversight of AI systems' activities and the provision of accountability obligations within the framework of international legal norms. The study provides a comparative analysis of existing regulatory mechanisms across various legal systems, examining the legal and normative challenges they present during implementation. The article discusses the oversight and accountability mechanisms of AI systems based on key international legal sources, including the European Union Artificial Intelligence Act, the UNESCO Recommendation on Ethics in Artificial Intelligence, the Organisation for Economic Co-operation and Development (OECD) Principles on Artificial Intelligence, and Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (draft).

The research examines the methodological foundations of technological and process-oriented types of IT audits, including algorithmic, performance, security, and ethical audits. It also clarifies the responsibilities of various participants in the accountability chain (developers, providers, users, and regulators). The article highlights the practical implications of technological challenges such as the "black box" nature of IT systems, their dynamic and adaptable nature, and information bias, as well as regulatory gaps such as inter-jurisdictional coordination problems and uncertainty about legal liability. In this context, a comparative analysis is conducted of different regulatory practices, such as the European Union's risk-based model, the United States' sector-specific approach, and China's state control model. The analysis demonstrates that these different approaches hinder the establishment of uniform global standards and create legal uncertainty for companies operating across borders. The role of international standards and certification mechanisms, as well as the importance of ensuring human control and transparency in addressing these challenges, is emphasized.

The research findings demonstrate that effective auditing and accountability of IT systems within the context of international law requires integrating technology-focused and process-focused approaches and strengthening international legal cooperation. The article offers specific recommendations to address existing regulatory gaps, including the adoption of a national legal framework, the implementation of mandatory human rights impact assessments, strengthening transparency and accountability obligations, ensuring human oversight, and the establishment of independent oversight bodies. These steps will ensure the fair and safe development and implementation of AI technologies that respect human rights.

Key words: Artificial intelligence, oversight, accountability, international law, legal regulation, legal mechanisms, legal nature, international initiatives, regional initiatives, auditing standards, certification mechanisms, control model.

Афган В.А. Правові механізми аудиту та підзвітності в системах штучного інтелекту в контексті міжнародного права.

Ця стаття присвячена нагляду за діяльністю систем штучного інтелекту та забезпеченню зобов'язань щодо підзвітності в рамках міжнародно-правових норм. Дослідження надає порівняльний аналіз існуючих регуляторних механізмів у різних правових системах, розглядаючи правові та нормативні виклики, які вони створюють під час впровадження. У статті розглядаються механізми нагляду та підзвітності систем штучного інтелекту на основі ключових міжнародно-правових джерел, включаючи Закон Європейського Союзу про штучний інтелект, Рекомендацію ЮНЕСКО щодо етики у штучному інтелекті, Принципи Організації економічного співробітництва та розвитку (ОЕСР) щодо штучного інтелекту та Рамкову конвенцію Ради Європи про штучний інтелект, права людини, демократію та верховенство права (проект).

У дослідженні також розглядаються методологічні основи технологічних та процесно-орієнтованих видів IT-аудитів, включаючи алгоритмічні, продуктивні, безпекові та етичні аудити. Також уточнюється відповідальність різних учасників ланцюжка підзвітності (розробників, постачальників, користувачів та регуляторів). У статті висвітлено практичні наслідки технологічних викликів, таких як «чорна скринька» IT-систем, їхня динамічна та адаптивна природа, інформаційна упередженість, а також регуляторні прогалини, такі як проблеми міжюрисдикційної координації та невизначеність щодо юридичної відповідальності. У цьому контексті проводиться порівняльний аналіз різних регуляторних практик, таких як ризикоорієнтована модель Європейського Союзу, галузевий підхід Сполучених Штатів та модель державного контролю Китаю. Аналіз демонструє, що ці різні підходи перешкоджають встановленню єдиних глобальних стандартів та створюють правову невизначеність для компаній, що працюють за кордоном. Підкреслюється роль міжнародних стандартів та механізмів сертифікації, а також важливість забезпечення людського контролю та прозорості у вирішенні цих викликів.

Результати дослідження показують, що ефективний аудит та підзвітність IT-систем у контексті міжнародного права вимагає інтеграції технологічно-орієнтованих та процесоорієнтованих підходів та зміцнення міжнародно-правового співробітництва. У статті пропонуються конкретні рекомендації щодо усунення існуючих регуляторних прогалин, включаючи прийняття національної правової бази, впровадження обов'язкових оцінок впливу на права людини, посилення зобов'язань щодо прозорості та підзвітності, забезпечення людського нагляду та створення незалежних органів нагляду. Ці кроки забезпечать справедливий та безпечний розвиток і впровадження технологій штучного інтелекту, що поважають права людини.

Ключові слова: штучний інтелект, нагляд, підзвітність, міжнародне право, правове регулювання, правові механізми, правова природа, міжнародні ініціативи, регіональні ініціативи, стандарти аудиту, механізми сертифікації, модель контролю.

Introduction.

The rapid development and penetration of artificial intelligence (AI) technologies into various areas of society have created new and complex challenges for international law. The increasing role of AI systems in decision-making in sensitive areas such as healthcare, finance, transportation, and even justice has raised serious concerns about their potential impact on human rights, democratic values, and the rule of law [1]. Fundamental issues such as the transparency, impartiality, and explainability of the operation of these systems, as well as the problem of determining liability for harmful consequences resulting from their activities, have necessitated the revision and adaptation of international legal norms [2].

In this context, ensuring information systems auditing and accountability obligations is of particular importance. The transnational nature of information systems—that is, their global operation beyond national jurisdictions—requires the development of a unified and harmonized international legal approach [3]. Otherwise, gaps and contradictions in regulations can lead to legal uncertainty, evasion of responsibility, and violations of fundamental rights.

The main objective of this article is to analyze the current state of information systems auditing and accountability mechanisms within the context of international law, examine the content and legal nature of international legal instruments developed in this field, compare legal approaches applied in different jurisdictions, and identify practical and theoretical challenges in this area. The main objectives of the study are as follows:

First, to analyze the legal definition of information systems and auditing mechanisms in international legal sources.

Second, to comparatively examine the legal regulatory experience of information systems auditing and accountability in different jurisdictions.

Third, to identify the legal and practical challenges encountered in information systems auditing processes.

Finally, to develop recommendations for improving effective information technology auditing mechanisms within the context of international law. From a methodological perspective, this study analyzes the content and legal nature of international legal sources, particularly recommendations from international organizations (UNESCO, OECD, Council of Europe), legislative bodies of the European Union, and national standards of the United States, using comparative law, formal legal, and systematic analysis methods. Since the issue of auditing and accountability of information technology systems is a relatively new area in international law literature, research on this topic has primarily focused on technological and ethical aspects [4]. This article aims to fill this gap and provide a systematic analysis of the legal mechanisms of auditing information technologies.

The state of scientific development of the issue. In modern times, the expansion of the application areas of information technology technologies and the deepening impact of these systems on society have necessitated the inclusion of their activities within the framework of legal regulations [5]. In particular, the introduction of automated solutions in the decision-making processes of artificial intelligence systems, the reduction of human intervention, and the generation of algorithmic decisions have created serious challenges to the application of traditional legal control mechanisms in this area [6]. These challenges are particularly evident in areas such as the lack of transparency in AI system decision-making, the technological complexity that hinders the effective functioning of accountability mechanisms, and the lack of coordination across jurisdictions. For example, during the «child benefits» scandal in the Netherlands, a risk-based algorithm used by tax authorities falsely accused tens of thousands of low-income families of social fraud, resulting in significant financial and social losses for hundreds of citizens. This incident is a striking example of how the state's use of AI systems without proper oversight leads to widespread human rights violations. The scandal's consequences were so severe that the wave of unjust algorithmic decisions sparked a crisis within the country's political leadership, culminating in the government's resignation in 2021. This fact clearly illustrates the real dangers of implementing AI systems without accountability.

The international community has taken a number of important steps to address these issues. The «Recommendation on the Ethics of Artificial Intelligence,» adopted by UNESCO in 2021, was supported by 193 countries and was recognized as the first global standard in the field of artificial intelligence [7]. This document sets out the fundamental principles for ensuring the compatibility of artificial intelligence systems with human rights, dignity, and fundamental freedoms. The Organization for Economic Co-operation and Development's Principles on Artificial Intelligence were adopted by 47 countries, establishing an international framework to promote the safe and innovative use of artificial intelligence systems [8]. The Framework Convention on Artificial Intelligence, adopted by the Council of Europe in 2024, is of historical significance as the first legally binding international agreement in this field [9]. However, despite these international and regional initiatives, serious challenges remain in the practical implementation of audit and accountability mechanisms for artificial intelligence systems. These challenges primarily arise in areas such as technological complexity and transparency, lack of coordination across jurisdictions, the lack of unification of audit standards, the ambiguity of legal accountability mechanisms, and the weakness of international cooperation mechanisms. In this article, based on a comprehensive analysis of these problems and a comparative study of existing international legal frameworks, specific recommendations for improving the audit and accountability mechanisms of IT systems will be presented.

Presentation of research material.

I. The Legal Framework of Artificial Intelligence Systems in the Context of International Law

Artificial intelligence systems are intelligent and intelligent systems. To establish international communication, control, and circulation mechanisms for artificial intelligence systems, it is necessary to first clearly define the legal definition of these systems and examine their legal qualifications. Different international organizations and regional associations have different definitions of artificial intelligence systems, which creates certain challenges in the field of legal regulation and makes it difficult to develop a unified approach [10].

Article 3 of the EU Artificial Intelligence Act (AI Act) defines an artificial intelligence system as "a machine-based system designed to operate at different levels of autonomy and capable of adapting after deployment" [11]. The key elements emphasized in this definition are that it is a machine-based system, exhibits different levels of autonomy, is adaptable after deployment, produces output from three input applications, explicit or implicit, and consequences physical and virtual ammunition, predictions, and recommendations. A particular point that draws attention in this definition is the degree of autonomy of artificial intelligence systems, which is crucial for their evaluation in the context of legal liability. The Organisation for Economic Co-operation and Development's Principles on Artificial Intelligence define an AI system as "a machine-based system that learns to produce outputs such as predictions, content, recommendations, or decisions that can affect physical or virtual environments from inputs received for explicit and implicit purposes" [12]. The well-adapted and adaptable nature of this systemic system is particularly notable, demonstrating its dynamic energies and the challenges arising from its approaches.

UNESCO's Recommendation emphasizes AI as a socio-technical system with significant ethical, societal, and human rights implications, rather than defining it solely as a technical simulation of human intelligence[13]. This definition places particular emphasis on ethical and social aspects and emphasizes the impact of AI systems on human society. UNESCO's approach considers AI systems not only as technological tools but also as complex systems with social and ethical implications.

The Council of Europe Framework Convention on Artificial Intelligence (draft) defines AI systems as "technological systems that must be fully compatible with human rights, democracy, and the rule of law" [14]. Therefore, the convention takes a human rights perspective into account in the legal regulation of AI systems and emphasizes the importance of ensuring that their operation is fully compatible with fundamental legal principles.

The differences in these definitions create certain challenges in the implementation of oversight and accountability mechanisms for AI systems. In particular, determining the level of autonomy of AI systems, assessing their adaptability, and determining the degree of human intervention are key challenges in oversight processes. To overcome these challenges, a unified international approach needs to be formulated and the legal nature of AI systems clearly defined.

The legal content and evolution of the concepts of auditing and accountability. In the context of international law, the concept of auditing derives from the concept of "independent expertise" traditionally used in finance and management and has been adapted to various fields over time [15]. In the context of information systems, auditing can be defined as "the process of independently and systematically assessing the conformity of the design, development, implementation, and use of information systems to predetermined standards, regulations, or norms" [16]. The principles of independence, systematicity, and objectivity emphasized in this definition form the legal basis of information systems auditing.

Among the fundamental components of information systems auditing is the principle of independence, which requires the person or organization performing the audit process to be independent of the parties that develop or use the AI system. This principle is essential to ensure the objectivity and reliability of audit results and is a fundamental requirement of international auditing standards. The principle of systematicity requires the audit process to be based on a structured and consistent methodology, ensuring that audit results are repeatable and comparable. The principle of transparency requires that the audit process and its results be accessible to relevant stakeholders and is essential for establishing public trust. The principle of competence requires those conducting the audit process to possess appropriate technical and legal knowledge. Accountability, on the other hand, relates to the obligations and responsibilities of various actors throughout the lifecycle of information technology systems and is increasingly complex in the context of international law [17]. Accountability in the context of international law also includes the element of legal liability, meaning that relevant actors are legally responsible for damage caused by information technology systems. This liability can encompass civil, criminal, and administrative domains and can take different forms in different jurisdictions. Transparency obligations require the provision of information about the operating principles, decision-making mechanisms, and potential risks of information technology systems. Control mechanisms include the establishment of mechanisms for continuous monitoring and control of the operation of information technology systems. Corrective measures, on the other hand, include measures to eliminate problems caused by information technology systems and to provide remediation to affected parties.

The evolution of AI regulation in international legal sources. In the international legal system, SI regulation is implemented at various levels, and the hierarchy of international legal sources is

particularly important in this field [18]. These levels include recommendations from global international organizations, legislative acts of regional associations, and bilateral international agreements. Each level has its own unique legal nature and degree of compulsion, demonstrating the complex and multifaceted nature of SI regulation.

At the global level, regulations are implemented within the framework of the United Nations, and SI issues are discussed in various committees and institutions. UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) is considered the most comprehensive global document in this field and has been adopted by 193 countries. The Recommendation emphasizes the compatibility of SI systems with human rights, dignity, and fundamental freedoms and establishes fundamental guidelines to ensure that the activities of these systems comply with ethical principles. The core principles of the UNESCO Recommendation are the protection of human rights and dignity, transparency and accountability, non-discrimination and justice, respect for human autonomy, prevention of harm, responsibility and accountability, privacy, and data protection. These principles form the basis for the development of oversight and accountability mechanisms for information technology systems. Regional Regulations: The EU Artificial Intelligence Act (AI Act), the most detailed regulatory framework at the regional level, defines a risk-based classification of information technology systems and related regulatory requirements. The law divides information technology systems into categories such as prohibited information technology applications, high-risk information technology systems, limited-risk information technology systems, and minimum-risk information technology systems. Specific auditing and reporting requirements are established for high-risk information technology systems, including the establishment of a risk management system, data quality and governance, technical documentation, record keeping, transparency and provision of information to users, human control and intervention, accuracy, robustness, and cybersecurity. The Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law is of particular importance as the first legally binding international treaty in the field of artificial intelligence. It aims to ensure that the lifecycle of artificial intelligence systems is fully compatible with human rights, democracy, and the rule of law. This convention adopts a human rights approach to the regulation of artificial intelligence and establishes binding obligations to ensure that the operation of artificial intelligence systems complies with fundamental legal principles. ECtHR case law demonstrates that States are obliged to establish effective mechanisms to protect human rights in the area of mass digital surveillance and algorithmic decision-making. For example, in Klass and Others v. Germany (1978), the Court acknowledged that covert electronic surveillance measures may be necessary in certain circumstances, while emphasizing the importance of individuals having an effective remedy under Article 13 of the ECHR. According to this decision, the State must monitor such covert surveillance programs through an independent body and provide individuals with the opportunity to challenge and appeal any violations of their rights. Another important precedent, Roman Zakharov v. Russia (Grand Chamber, 2015), examined the Russian security services' large-scale telephone tapping system (SORM). The ECtHR found that the country's existing legislation did not provide sufficient safeguards against eavesdropping, amounting to a violation of Article 8 (right to respect for private life) of the ECHR. The applicant also relied on the lack of an effective remedy in domestic law (Article 13), and the Court paid particular attention to this issue. Therefore, the conclusion drawn from cases such as Klass and Zakharov is that States must establish legal controls and effective remedies against surveillance or other autonomous decision-making processes conducted through AI systems. This is important both for protecting privacy (Article 8) and ensuring that individuals have an effective remedy in the event of breaches (Article 13).

National-level regulation is characterized by different countries adopting different approaches to regulating AI in their national legislation. NIST AI Risk Management Framework (AI RMF 1.0, 2023), developed by the United States National Institute of Standards and Technology, proposes a voluntary risk assessment mechanism for IT systems. The People's Republic of China has adopted a national strategy for IT development and established a regulatory framework for IT security. The United Kingdom, on the other hand, is developing sector-specific regulatory mechanisms by taking a principles-based approach to IT governance.

II. Audit and Accountability Mechanisms

Types of AI Audits and Methodologies. Various types and methodologies of AI system audits are implemented in international practice, and these are broadly divided into two main categories:

technology-focused audits and process-focused audits [19]. This distinction reflects the complex nature of AI systems and the different approaches required for their audits.

Technology-focused audits focus on the technical specifications, performance, and outputs of AI systems, and algorithmic auditing is particularly important in this area. Algorithmic auditing involves analyzing the algorithmic structure, decision-making mechanisms, and potential biases of AI systems and is used specifically to identify discrimination and unfairness issues in machine learning models. This type of audit process includes analyzing the dataset and checking for biases, examining the model's architecture and parameters, monitoring decision-making processes, comparing results for different demographic groups, and calculating and evaluating fairness criteria. Performance auditing involves assessing the extent to which AI systems achieve their stated objectives, accuracy indicators, and reliability levels. This type of auditing is particularly important for AI systems deployed in critical infrastructure and high-risk areas. This process includes measuring accuracy metrics, verifying that the system performs consistently under different conditions through robustness testing, ensuring that the system meets real-time requirements through latency testing, maintaining system performance under increased load through scalability testing, and assessing system behavior under extreme loads or unusual conditions through stress testing.

Security auditing involves assessing the resilience of AI systems against cybersecurity threats, information security measures, and system integrity, with particular attention paid to threats such as hostile attacks and model poisoning. Key areas of security auditing include resistance to adversarial attacks, i.e., the system's resilience to manipulated input data, model poisoning threats, i.e., compromising the system through poisoning of training data, confidentiality leakage, i.e., the risk of the system disclosing sensitive information about training data, backdoor attacks, i.e., the insertion of hidden triggers into the system, and model extraction attacks, i.e., theft of the system's intellectual property.

Transparency auditing involves assessing the extent to which AI systems' decision-making processes meet the requirements for AI to be explainable, understandable, and explainable to users. In this area, model interpretability (the comprehensibility of system decisions), feature importance (which input parameters have the greatest impact on decisions), decision paths (step-by-step follow-up of the decision-making process), counterfactual explanations (which conditions must be changed to achieve a different outcome), and user-friendly explanations (simple explanations for non-technical users) are particularly important. Process-focused audits encompass the management processes, quality management systems, and corporate governance mechanisms of organizations that develop and use AI systems. Management audits, on the other hand, encompass the evaluation of the policies, procedures, and management structures adopted by organizations for their information systems, examining the effectiveness and appropriateness of AI systems governance frameworks. In this area, the existence and effectiveness of the AI systems governance structure, the activities and powers of ethics committees, approval and control procedures for AI projects, personnel training and competency management, supplier and third-party risk management, incident management, and response procedures require special attention. A risk management audit involves analyzing the processes for identifying, assessing, and managing risks related to AI and assessing compliance with international standards in this area, such as those of the National Institute of Standards and Technology. Components of a risk management audit include the adequacy of the risk assessment methodology, the completeness and timeliness of the risk register, the effectiveness of risk mitigation strategies, the existence of acceptable levels of residual risks, risk monitoring and reporting mechanisms, crisis management, and business continuity plans.

An ethics audit involves assessing the compliance of AI systems with ethical principles, human rights, and social impacts. This type of audit is based on the principles of UNESCO and the Organization for Economic Co-operation and Development. Key areas of ethics audit include respect for human dignity and the protection of fundamental rights, ensuring social justice and equality, protecting privacy and personal data, the existence of transparency and accountability mechanisms, assessing and managing social impacts, stakeholder engagement, and public consultation processes.

A compliance audit involves verifying the compliance of AI systems with relevant legal requirements, standards, and regulatory norms. Compliance with regional legislation, such as the European Union Information Systems Act, is the primary focus of this type of audit. The scope of compliance auditing encompasses compliance with legal requirements, adherence to international standards, fulfillment of documentation requirements, certification and accreditation processes, maintenance of audit trails and records, and fulfillment of regulatory reporting obligations.

International auditing standards and certification mechanisms. The role of international standards in information systems auditing is becoming increasingly important, and various international organizations are developing standards and certification mechanisms to ensure the quality and reliability of information systems audits. The International Organization for Standardization and the International Electrotechnical Commission have developed a number of standards for information systems, the most important of which are the information systems risk management framework, information systems reliability requirements, and quality models for information systems [20].

The application of these standards to AI systems audits encourages a risk-based approach, establishes more stringent audit requirements for high-risk systems, and defines audit requirements covering all stages of the information system lifecycle, from design to decommissioning. The standards require the participation of various stakeholders in the audit process and support the continuous improvement of audit processes and learning mechanisms.

The Institute of Electrical and Electronics Engineers has developed a number of standards in the area of AI systems ethics and auditing, covering privacy engineering and risk management, mitigating AI system bias, and transparency of information systems. These standards require that AI systems' privacy requirements be considered from the design phase onward, provide methods for identifying and mitigating algorithmic bias, and establish requirements for ensuring the transparency of information systems decision-making processes.

In international practice, various mechanisms are used for digital systems certification, including third-party certification, self-assessment, and continuous monitoring mechanisms. Third-party certification is the process of independent certification bodies verifying that AI systems comply with specific standards, and the European Union **Artificial Intelligence Act (AI Act)** mandates this type of certification for high-risk AI systems. The advantages of this type of certification include objectivity and independence, international recognition, technical expertise, and continuous monitoring and updating, while its disadvantages include high cost, time requirements, technical complexity, and the rapid change of standards.

Self-assessment is the process by which organizations developing AI systems evaluate their systems through internal audit procedures, and this mechanism is particularly applicable to low- and medium-risk systems. Components of this mechanism include the establishment of internal audit teams, the use of checklists and assessment tools, the implementation of regular internal audit processes, management review and approval procedures, external validation, and peer review mechanisms. Continuous monitoring is an ongoing control and evaluation mechanism throughout the AI systems lifecycle, and this approach is particularly important for adaptive and learning systems. Elements of continuous monitoring include real-time performance monitoring, automated anomaly detection, regular model validation, deviation detection and management, incident response and remediation, stakeholder feedback, and complaint management.

The structure and legal nature of accountability mechanisms. For accountability mechanisms in information systems to function effectively, clear allocation of responsibilities and appropriate coordination must be ensured among all relevant actors at different levels and throughout the information system lifecycle.

Developers' responsibilities include adhering to ethical principles during the information system design and development phase, risk assessment, and implementation of security measures. These responsibilities include designing by responsible information system principles, bias detection and mitigation, security by design, privacy by design, comprehensive testing and validation, documentation and transparency, stakeholder engagement, continuous learning, and improvement.

Providers' responsibilities include launching information systems, implementing certification processes, and providing users with necessary information. These responsibilities include product safety and quality assurance, regulatory compliance, user education and training, technical support and maintenance, incident response and remediation, supply chain management, intellectual property protection, market surveillance, and post-market monitoring.

Users' responsibilities include ensuring the appropriate use of AI systems, ensuring human control, and managing potential risks. These obligations include appropriate use and context awareness, human control and intervention, risk assessment and management, user training and competence, data management and protection, performance monitoring and evaluation, incident reporting and management, stakeholder communication, and transparency.

Regulators' responsibilities include monitoring AI systems' compliance with legal requirements, ensuring the quality of audit processes, and implementing penalty mechanisms. These functions include regulatory framework development, standard setting and enforcement, market surveillance and control, investigation and enforcement measures, guidance and interpretation, international cooperation, research and development support, and public participation and engagement.

Technical accountability encompasses accountability mechanisms related to the technical aspects of AI systems, and algorithmic accountability requires AI algorithms to comply with the requirements for explainability and transparency of their decision-making processes. Explainable AI and interpretable machine learning methods are applied in this area. The components of algorithmic accountability include: algorithm transparency, i.e., clarity and understanding of algorithmic processes; decision auditability, i.e., traceability and verification of decisions; bias accountability, i.e., identification and resolution of bias; performance accountability, i.e., transparency of performance indicators; and change management, i.e., management of algorithmic changes.

Data accountability requires the quality, representativeness, and bias-free nature of training and testing data for AI systems, and data governance and data provenance mechanisms are key components of this area. Data accountability elements include data quality assurance, data provenance and generation, data bias detection and mitigation, data privacy and protection, data storage and deletion, data sharing and access control, data documentation and metadata, and data verification and validation.

Model accountability requires accountability for the performance, reliability, and robustness of AI models, and model validation and verification procedures are key components of this area. Model accountability elements include model performance monitoring, model validation and testing, model versioning and change control, model documentation and metadata, model interpretability and explainability, model robustness and reliability, model lifecycle management, and model retirement and replacement.

Legal liability encompasses accountability mechanisms for the legal aspects of AI systems, and legal liability requires compensation mechanisms for damages caused by AI systems. Legal concepts such as product liability and negligence are applied in this area. Elements of legal liability include strict liability (liability arising from product defects), liability in negligence (liability arising from negligence), contractual liability (breach of contractual obligations), vicarious liability (liability arising from the acts of another), damages and compensation, insurance and risk transfer, and alternative dispute resolution.

Criminal liability encompasses liability mechanisms for crimes committed as a result of the use of AI systems, and the application of legal concepts such as intent and causation in the context of AI raises problematic issues. Challenges of criminal liability include determining criminal intent, determining criminal behavior, proving causation, corporate criminal liability, international criminal law, cybercrime, and digital evidence.

Administrative liability encompasses administrative penalties for violations of regulatory requirements for AI systems, and the EU Artificial Intelligence Act provides detailed penalty mechanisms in this area. Types of administrative liability include fines, operating restrictions, license suspension and revocation, compliance orders, public disclosure and naming, corrective measures, monitoring and auditing, international cooperation, and sanctions.

III. Regulatory Practices in Different Jurisdictions

The EU's Comprehensive Risk-Based Regulatory Model. The EU has adopted the world's first comprehensive AI legislation, demonstrating the most comprehensive and systematic approach to AI regulation. The **EU AI Act**., which entered into force in 2024, sets new standards for risk-based regulation of AI systems and is considered a fundamental model for AI regulation in the context of international law [21].

A key feature of the **EU AI Act**. is the application of different regulatory requirements based on the risk level of AI systems, and this approach encompasses four main categories. Prohibited AI applications are defined as those that pose a serious threat to human dignity and fundamental rights and are completely prohibited. This category includes the use of subliminal techniques to influence people's subconscious minds, exploiting vulnerabilities such as age, disability, or socioeconomic status, real-time biometric identification systems and social credit systems in public spaces, and general behavioral assessment mechanisms. The legal basis for these prohibitions is the principle of human dignity, the right to personal autonomy and free will, the right to privacy and confidentiality, the principle of non-

discrimination, democratic values and the rule of law, and the universal nature of human rights. These principles are reflected in the European Convention on Human Rights and the Charter of Fundamental Rights of the European Union and form the basis for ensuring that the operation of AI systems complies with these fundamental legal principles. High-risk AI systems are defined as AI systems that could have a significant impact on human security and fundamental rights, and this category is further defined in the Third Annex to the Law. This category encompasses critical infrastructure, education and training, employment and labor management, essential private and public services, law enforcement, immigration and border control, justice, and democratic processes.

Critical infrastructure areas include electricity, gas, heating, and water supply systems management, transportation infrastructure and automated transportation systems, telecommunications networks and internet infrastructure, financial infrastructure, and payment systems. The use of AI systems in these areas directly impacts society's core activities, and their failure or malfunction can have large-scale social and economic consequences.

Education and vocational training include admission decisions to educational institutions, academic assessment and examination systems, vocational training program selection, and educational loan and financial support decisions. The use of AI systems in these areas directly impacts individuals' educational and career opportunities, and the risks of discrimination and injustice are particularly high.

Employment and personnel management include decision support in recruitment processes, employee performance evaluations, workplace control and monitoring systems, and dismissal or job change decisions. The use of AI systems in these areas has significant impacts on employee rights and labor relations, and the risks of bias and discrimination are particularly serious. For high-risk systems, the EU AI Code establishes detailed requirements covering the risk management system, data management and quality, technical documentation, recording and logging, transparency and user information, human control and intervention, accuracy, robustness, and cybersecurity.

A risk management system requires the establishment of risk management processes throughout the system's lifecycle, the development and implementation of a risk assessment methodology, the identification and implementation of risk mitigation measures, the maintenance of residual risks at acceptable levels, and the organization of risk monitoring and reassessment processes.

Data management and quality require compliance with quality standards for training, validation, and test data, the representativeness and completeness of datasets, the identification and reduction of data biases, the assurance of data source reliability, the quality of data labeling processes, and data security and privacy measures.

Technical documentation requires detailed documentation of the system's technical specifications, a description of the algorithmic design and architecture, documentation of the training process and methodology, the recording of test and validation results, the identification of performance metrics and limitations, and the documentation of risk assessment results. The EU Artificial Intelligence Act prescribes detailed conformity assessment procedures for high-risk AI systems, including internal control procedures and third-party assessments. Internal control procedures are the fundamental procedure applied to most high-risk systems and include the preparation of technical documentation, the establishment of a risk management system, the implementation of a quality management system, the preparation of a declaration of conformity, the affixing of the European Conformity Mark, and the establishment of a post-market surveillance system.

Third-party assessment is a mandatory procedure for certain high-risk systems and applies to biometric identification and categorization systems, systems used in critical infrastructure areas, certain systems used in law enforcement, and immigration and border control systems. The EU Artificial Intelligence Act provides for serious penalties in the field of AI regulation, including €35 million or seven percent of a company's annual turnover for the use of prohibited AI applications, €15 million or three percent of a company's annual turnover for violations of high-risk system requirements, €7.5 million or one and a half percent of turnover for violations of transparency obligations, and €7.5 million or one and a half percent of turnover for providing false information.

The United States' Sector-Specific and Voluntary Approach. The United States has adopted a different approach to information technology regulation, opting for sector-specific regulations and a system of voluntary standards over centralized legislation [22]. The core philosophy of this approach is to foster innovation and prevent technological advancement while also addressing security and ethical issues.

The Information Technology Risk Management Framework, developed by the United States National Institute of Standards and Technology, provides a comprehensive framework for voluntary risk management in information technology systems. This framework encompasses four main functions. The governance function includes establishing information technology governance structures and defining risk management policies. The framework encompasses the establishment of information technology governance and control structures, the development of risk management policies and procedures, stakeholder engagement and communication strategies, the definition of information technology ethics and responsible information technology principles, the establishment of legal and regulatory compliance mechanisms, human resources and competency management, and supplier and third-party risk management. The mapping function encompasses defining the information system context, objectives, and risks, and includes defining the information system context and use cases, identifying stakeholders and affected parties, mapping the information system lifecycle and dependencies, modeling the risk environment and threat, determining legal and regulatory requirements, assessing organizational capacity and maturity, impact assessment, and consequences analysis.

The measurement function encompasses measuring and evaluating information system performance and risks, and includes defining performance metrics and key performance indicators, applying risk measurement and quantification methods, implementing testing and verification procedures, establishing monitoring and control systems, conducting audit and evaluation activities, conducting benchmarking and comparative analysis, and assessing data quality and integrity.

The governance function includes implementing actions to manage and mitigate identified risks, risk mitigation strategies and controls, incident response and crisis management, business continuity and disaster recovery, change management and version control, training and awareness programs, continuous improvement and learning, and stakeholder communication and reporting. In the United States, AI regulations are primarily driven by existing sector-specific regulatory authorities, while those for the financial sector are handled by the Federal Reserve, the Office of the Comptroller of the Currency, and other financial regulators. Key elements of AI regulation in the financial sector include model risk management and approval, fair lending and anti-discrimination requirements, consumer protection and transparency, operational risk and resilience, cybersecurity and data protection, systemic risk and financial stability, and cross-border and international coordination.

Healthcare regulations are driven by the Food and Drug Administration (FDA), which has developed a specific regulatory framework for the use of AI in medical devices. This framework covers the classification of software as medical devices, pre-market approval and authorization, clinical validation and real-world evidence, post-market surveillance and adverse event reporting, quality management and risk management, cybersecurity and data integrity, international harmonization, and mutual recognition.

The transportation sector is regulated by the National Highway Traffic Safety Administration (**NHTSA**.), which sets safety standards for autonomous vehicles. These standards cover autonomous driving system safety standards, testing and validation requirements, cybersecurity and functional safety, human-machine interface and driver monitoring, data recording and sharing, incident reporting and investigation, and international coordination and harmonization.

The United States federal government is undertaking a number of significant initiatives in the field of artificial intelligence, including the "Artificial Intelligence Bill of Rights," announced by the White House in 2022, and the Executive Order signed in 2023. The AI Bill of Rights outlines how AI systems should be designed to protect citizens' rights and includes principles for safe and efficient systems, algorithmic anti-discrimination protections, data privacy, notice and disclosure, human alternatives, review, and takedown.

The Executive Order directs federal agencies to take steps to ensure the safe and secure development of AI systems and includes AI safety and security standards, innovation and competition, employee support and protection, prevention of bias and discrimination, privacy and civil rights protections, and international cooperation and leadership.

Centralized state control model in the People's Republic of China. The People's Republic of China adopts a centralized, state-based approach to AI regulation, reflecting its priorities of national security, social stability, and state control [23]. China's AI regulatory strategy combines the goals of achieving technological leadership and strengthening social control.

In 2017, China adopted the "Next Generation Artificial Intelligence Development Plan" to set its goal of becoming a global leader in AI by 2030. Key components of this strategy include breakthrough

and innovation in AI technology, AI industrial development and application, AI human resources education and training, AI governance and ethics, AI international cooperation and competition, and AI military and national security applications. China has adopted specific regulatory regulations for various AI application areas, and the regulation of algorithmic recommendations is handled by the "Internet Information Service Algorithmic Recommendation Management Provisions," which will come into force in 2022. These rules regulate recommendation algorithms used on internet platforms and cover requirements such as algorithm transparency and explainability, user consent and choice, content moderation and prevention of harmful content, data protection and privacy, antitrust and fair competition, national security, and social stability.

Deepfake regulation is implemented through rules adopted in 2023 that regulate deepfake and other artificial content creation technologies. Key elements of the regulation include content labeling and watermarking, user verification and authentication, prevention of harmful content, intellectual property protection, platform responsibility and liability, and cooperation with law enforcement.

Regulation of generative AI services is implemented through interim rules adopted in 2023 that regulate the provision of generative AI services, such as ChatGPT. Key requirements of the regulations include content security and prevention of harmful content, information security and privacy protection, algorithm transparency and accountability, user rights and protection, platform responsibility and oversight, and international cooperation and compliance. China has established centralized coordination mechanisms for AI regulation, with the Cyberspace Administration of China serving as the main regulatory authority. The Ministry of Industry and Information Technology is responsible for industrial development and standards. The Ministry of Science and Technology is responsible for research and innovation. The National Development and Reform Commission is responsible for national planning and policy matters. The Ministry of Public Security is responsible for security and law enforcement.

The UK's principles-based and flexible regulatory approach. The UK adopts a principles-based and flexible approach to AI regulation that combines the objectives of fostering innovation and providing international leadership [24]. The UK's approach strengthens existing regulatory structures and enhances its ability to adapt to new technologies.

The "A pro-innovation approach to AI regulation," published in 2023, sets out the country's AI strategy, and key principles include an innovation-friendly approach, proportionate and risk-based regulation, sector-specific expertise and flexibility, international leadership and collaboration, evidence-based policymaking and stakeholder engagement, and co-regulation. The UK is encouraging existing regulators to be responsible for AI regulation in their respective areas. The key regulators are the Financial Conduct Authority for financial services, the Medicines and Healthcare Products Regulatory Agency for healthcare, the Department for Science, Innovation and Technology (DSIT) for telecommunications and media, the Information Commissioner's Office for data protection, and the Competition and Markets Authority for competition and markets. Established in 2023, the AI Safety Institute conducts research and assessment on the security risks of AI systems. The Institute's core functions include AI security research and assessment, risk assessment and testing, international collaboration and standards, policy advice and guidance, public engagement and education, and industry collaboration and partnerships. The UK promotes global cooperation by organizing international summits on cybersecurity, and key initiatives include the AI Safety Summit and the Bletchley Declaration, leadership in the Global Partnership on Artificial Intelligence (GPAI), the Organization for Economic Co-operation and Development's cybersecurity principles and practices, United Nations cybersecurity governance discussions, bilateral cooperation agreements, and academic and research partnerships.

IV. Consulting Law Approaches and Implementation Challenges

Major legal and technological challenges encountered in IS auditing. The practical application of IS auditing and accountability mechanisms presents a number of significant challenges, both technological and legal, requiring the development of new approaches to address these challenges within the context of international law [25].

Technological complexity and the black box problem relate to the complex nature of modern IS systems, particularly deep learning models, making it difficult to understand decision-making processes. This «black box» problem is a fundamental obstacle in auditing processes, as auditors cannot fully understand the internal workings of the system. Despite the development of explainable IS technologies,

achieving full transparency of complex IS systems remains technically impossible, and this creates serious challenges, especially in high-risk areas.

The dynamic and adaptable nature of IS systems stems from their ability to learn and adapt, causing their behavior to change over time. This characteristic complicates the application of traditional auditing approaches because the system's behavior during an audit may differ in the future. This problem is exacerbated by the fact that in continuous learning systems, the system changes its parameters based on new data, and these changes occur in real time.

Data quality and bias arise because the performance of AI systems depends heavily on the quality of the training data. Data bias, incomplete datasets, or underrepresented groups can introduce systematic bias into AI systems. Assessing data quality is particularly challenging in auditing processes because the volume of training data can be enormous, data can be collected from different sources, the data labeling process can be subjective, and data bias can be subtle and multifaceted.

The challenge of interjurisdictional coordination stems from the fact that the cross-border nature of AI systems necessitates coordination between the legal systems of different countries. However, different jurisdictions apply different approaches to AI regulation, leading to challenges in auditing processes. For example, there are significant differences between the requirements of the European Union's Artificial Intelligence Act and the voluntary approach of the United States, and the approaches to AI regulation in China and Russia are even more divergent. This creates the problem of complying with multiple and sometimes conflicting requirements for companies operating internationally.

Uncertainty and the Problems of Legal Liability Mechanisms. Determining legal liability for damages caused by artificial intelligence systems is one of the most complex issues in modern legal systems, and international legal norms need to be adapted in this area [26].

Determining causality is related to the requirement in traditional legal systems that liability must have a direct causal relationship between the damage and the action. However, establishing this relationship is quite difficult in artificial intelligence systems because artificial intelligence systems consist of multiple components, the decision-making process is multi-stage, the system's behavior is unpredictable, and various actors contribute.

The concepts of intent and negligence are related to the requirement of an element of intent for liability in criminal law. However, applying the concept of intent to artificial intelligence systems is problematic because AI systems make decisions on their own, the connection between human intent and machine behavior is unclear, and there may be a difference between programming intent and outcome. The concept of negligence in civil law also causes problems in the context of artificial intelligence, as the «reasonable care» standard is difficult to apply to AI systems. This division of responsibility stems from the involvement of multiple actors, each with different responsibilities throughout the lifecycle of AI systems. Developer responsibility includes algorithmic design, training data selection, and testing procedures. Provider responsibility includes product launch, user information, and support services. User responsibility includes proper system use, ensuring human oversight, and risk management. Data providers' responsibilities include data quality, impartiality, and confidentiality. The precise allocation and legal definition of these responsibilities remains unresolved.

Deficiencies in international cooperation mechanisms. International cooperation is necessary for effective IT auditing and accountability, but existing mechanisms are insufficient, and new international legal regulations need to be developed in this area [27].

Information exchange issues are related to the importance of international information exchange in IT auditing, but a number of obstacles exist. These obstacles include national security restrictions, the protection of trade secrets, privacy laws, and differences in technical standards.

The lack of mutual recognition mechanisms stems from the lack of mechanisms for mutual recognition of IT audit results conducted in different countries. This translates to duplicative audit procedures and additional costs for international companies.

International arbitration and dispute resolution stem from the lack of specialized mechanisms for resolving international disputes related to IT systems. Traditional international arbitration mechanisms fail to take into account the technical complexity of IT.

Striking a balance between technological innovation and regulation. Striking a balance between regulation and innovation in the field of artificial intelligence is a crucial issue, and a breach of this balance can both slow down technological development and create security risks [28].

The problem of regulatory lag stems from the gap between the rapid development of technology and the slowness of regulatory processes, creating a «regulatory lag» problem. AI technologies are evolving so rapidly that legislation cannot keep pace.

Innovation barriers stem from the fact that overly strict regulations can slow down AI innovation. Oversight and compliance costs can be a significant burden, especially for small and medium-sized businesses.

Regulatory testbed mechanisms are associated with some countries establishing «regulatory testbed» mechanisms to allow testing of new AI technologies under limited conditions. However, the effectiveness and scalability of these mechanisms remain questionable.

V. Legal Consequences of Normative Gaps

Lack of a specific legal framework at the international level: There is currently no binding international agreement or convention in the field of artificial intelligence. The work of the Committee on Artificial Intelligence (CAI) also demonstrates that existing international legal norms are inadequate to address the challenges posed by artificial intelligence and that legal gaps exist in this area. Consequently, states' precise international obligations regarding the use of artificial intelligence systems are unclear. This gap directly impacts individual rights; individuals whose rights are violated by artificial intelligence often have difficulty applying to international protection mechanisms (because such specific mechanisms have not been established). Furthermore, states may fail to fully fulfill their positive obligations in the field of artificial intelligence (e.g., preventive measures to protect human rights) and may exploit the lack of specific international norms to evade responsibility.

Gaps in National Legislation: Many countries lack specific provisions or laws regarding the oversight and accountability of artificial intelligence systems in their domestic legislation. This normative gap makes it difficult to protect individuals' rights. For example, when an algorithm's wrongful decision violates an individual's rights, the individual cannot clearly determine who they should sue in court—the developer, operator, or the government agency using the AI system. The lack of appropriate legal mechanisms prevents victims from seeking effective legal remedies. Furthermore, due to gaps in national legislation, the responsibility and accountability of state authorities in the implementation of AI is inadequately regulated. This increases the risk of legal violations resulting from AI decisions going unpunished.

Gaps in Transparency and Accountability Requirements: Existing legislation often fails to address the "black box" nature of AI systems. Because information about algorithms' decision-making criteria and processes is not disclosed, individuals struggle to understand the reasons for automated decisions made about them, challenge them, and correct them. For example, a person receiving a negative response from AI cannot understand why that decision was made, limiting their right to a fair defense [30,37]. This gap in state accountability means that regulators and courts lack sufficient information to monitor the operation of AI systems. This makes it difficult to detect potential violations in a timely manner and lacks accountability. Therefore, international organizations emphasize the importance of legal requirements for AI systems to be auditable, traceable, and explainable. The absence of such requirements both undermines rights protections and erodes public trust. Lack of independent oversight mechanisms: In many countries, specialized regulatory bodies to oversee the implementation of AI systems are either nonexistent or lack sufficient authority. Consequently, there is a lack of independent and professional oversight of algorithms used in the public or private sectors. This gap impacts individual rights; a person complaining about AI decisions cannot find a specific institution to appeal to. For states, the lack of such an oversight body makes it difficult to prevent the misuse of AI. Research by the European Union Agency for Fundamental Rights (FRA) has shown that AI users are sometimes unaware of which national authority controls the algorithms they implement. This creates a de facto lack of oversight. Therefore, the lack of independent oversight structures both undermines individuals' right to protection and weakens the accountability of states over AI systems, limiting their ability to prevent potential abuses.

Conclusion.

An analysis of the legal mechanisms for oversight and accountability in AI systems within the context of international law reveals that this field is still in its nascent stage and presents a number of serious challenges. The main conclusions of the study can be summarized as follows.

In the area of establishing an international legal framework, international documents such as the UNESCO Recommendation on Ethics in Artificial Intelligence, the OECD Principles on Artificial Intelligence, and the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law have formed the basis of the legal framework in the field of AI. However, most of these documents are advisory in nature, have limited legal binding, and additional efforts are needed to establish a unified international approach.

Differences in Regional Regulatory Approaches: There are significant differences between the European Union's risk-based and detailed regulatory approach and the United States' voluntary and sector-specific approach [34]. These differences complicate international cooperation and harmonization processes and create the challenge of complying with multiple regulatory requirements for companies operating internationally.

Technical Challenges of Oversight Mechanisms: The complex and dynamic nature of AI systems complicates the application of traditional oversight approaches. In particular, the "black box" problem, the variable behavior of adaptive systems, and data bias pose significant obstacles to auditing processes, and new technological solutions must be developed to address these issues [29, 36].

Uncertainty of Legal Liability Mechanisms: Significant uncertainties exist in determining legal liability for harm caused by AI systems. Determining causality, applying the concept of intent, and assigning responsibility remain problematic issues, and specialized legal mechanisms need to be established in this area [32].

Weakness of International Cooperation Mechanisms: The international cooperation mechanisms necessary for the effective conduct of AI audits are not sufficiently developed. Serious deficiencies exist in the areas of information exchange, mutual recognition, and dispute resolution, and international legal instruments in this area need to be strengthened.

Based on the research findings, the following recommendations are offered:

1. Adopting a national legal framework: States should adopt specific legal regulations governing the implementation of AI systems. This framework should reflect existing international human rights norms and principles (e.g., non-discrimination, transparency, and accountability) in the context of AI and define AI-specific requirements [35]. A Feasibility Study conducted by Committee on Artificial Intelligence (CAI) also highlighted the inadequacy of existing rules and the need to develop new international legal frameworks. In this context, states can fill legal gaps by enacting separate AI laws at the national level (or amending existing laws).

2. Mandatory human rights impact assessments and AI audits: High-risk AI systems should be subject to mandatory human rights impact assessments before they become operational and throughout their lifecycle. Legislation should prescribe human rights impact assessment (HRIA) procedures for government institutions and companies, and the results of these assessments should be acted upon. Furthermore, independent audits of AI algorithms should be conducted; that is, third-party experts should verify the impartiality, accuracy, and legal compliance of the systems. For example, Amnesty International calls on states to implement mandatory human rights impact assessments to prevent human rights violations by identifying potential risks before the use of AI systems. Such audits and oversight processes will serve to identify and address violations early.

3. Transparency and explainability obligations: The disclosure and explanation of AI system decisions should be ensured by legal requirements. Legislation should impose such obligations on both public authorities and private companies to ensure that the basis for AI decisions is as clear as possible. The Artificial Intelligence Law drafted by the European Union is a key example in this area: According to Article 13 of the law, providers of high-risk AI systems must explain the purpose, functions, and potential risks of the system in detail in technical documentation and provide users with accurate and understandable instructions. Such transparency requirements will both ensure citizens understand AI decisions and facilitate regulatory authorities' assessment of their compliance. Consequently, strengthening the legal obligation for transparency will lead to increased accountability in the field of AI [31,33].

4. Ensuring human control: The principle that "humans should prevail over algorithms" should be ensured by law. In high-risk areas (e.g., law enforcement, justice, healthcare, social protection), AI decisions should not be implemented without human control. To this end, AI systems should be designed to effectively monitor the activities of competent individuals and, if necessary, intervene and correct them. Article 14 of the European Union Artificial Intelligence Act also requires this: High-risk

AI systems should be designed to enable humans to promptly detect and prevent wrongful or rights-infringing actions. To this end, in some cases, a requirement is also imposed for the decision taken by an AI system to be verified and approved by two different individuals (especially in recognition and classification systems) before it can take effect. Such legal regulations protect the human factor, enable the correction of errors in automated systems, and provide additional safeguards for the protection of rights.

5. Establishing independent supervisory authorities: States should establish dedicated regulatory institutions to ensure the oversight and public accountability of AI systems. These institutions should monitor the implementation of AI systems under ethical and legal standards, investigate complaints, and impose sanctions when necessary. Alternatively, or in addition, the mandate of existing regulatory institutions, such as data protection commissions, equality bodies, and ombudsman institutions, should be expanded to cover AI-related matters. According to FRA recommendations, states should increase the human resources and technical knowledge of existing regulatory bodies and specialize in AI, ensuring they can effectively regulate complex algorithmic systems. Strengthening independent audit and oversight mechanisms will not only increase accountability but also boost citizen trust in AI. It should be noted that the 2021 UNESCO "Recommendation on the Ethics of Artificial Intelligence" also recommends that states establish oversight, impact assessment, and due diligence mechanisms for AI systems and protect whistleblowers in this area. This means that both technical and institutional measures should be taken to ensure an effective accountability environment.

6. Clarification of Legal Responsibility and Safeguards: Legislation should clarify the distribution of liability for potential harms that AI systems may cause. If an individual's rights are violated due to an error in the AI algorithm used, the identity of the party responsible for the violation (state authority, software provider, client company, etc.) should be predetermined by law. This will ensure that the harmed individual receives effective legal protection. Furthermore, remediation mechanisms—that is, procedures that enable individuals to obtain compensation for harm caused by AI (such as dedicated compensation funds or insurance systems)—must be established. As highlighted in the FRA's "Getting the future right – Artificial intelligence and fundamental rights" report, states should establish accountability systems to monitor the negative impact of AI on fundamental rights, respond to adverse events, punish those responsible for violations, and provide effective legal assistance to victims. In practice, this means that appeals and legal action avenues must be clear and accessible to anyone who complains about an AI decision (for example, alleging that an automated decision violates a human right). Without such legal safeguards, combating AI violations can be ineffective.

REFERECES:

1. Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
2. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.
3. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.
4. Winfield, A.F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A, 376(2133), 20180085.
5. Brynjolfsson, E., & McAfee, A. (2017). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.
6. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.
7. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO.
8. OECD. (2024). OECD AI Principles. OECD Publishing.
9. Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe.
10. Smuha, N.A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International, 20(4), 97–106.

11. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
12. OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.
13. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence – Section I: Scope and Definitions. UNESCO.
14. Council of Europe. (2024). Framework Convention on Artificial Intelligence – Article 2: Definitions. Council of Europe Treaty Series.
15. Gupta, K. (2004). Contemporary Auditing (7th ed.). Tata McGraw-Hill Education.
16. Mökander, J., Axente, M., Casolari, F., Floridi, L. (2023). Auditing of AI: Legal, Ethical and Technical Approaches. Digital Society, 2, 49.
17. Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. European Law Journal, 13(4), 447–468.
18. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. Science and Engineering Ethics, 24(2), 505–528.
19. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44.
20. Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., ... & Hamid, F. (2022). Towards algorithm auditing: A survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. Computer Science Review, 43, 100435.
21. Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112.
22. NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology.
23. Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. AI & Society, 36(1), 59–77.
24. UK Government. (2023). A Pro-Innovation Approach to AI Regulation. HM Government White Paper.
25. Castelvecchi, D. (2016). Can we open the black box of AI? Nature News, 538(7623), 20.
26. Lior, Y. (2020). Can artificial intelligence be held accountable? A comparative analysis of the liability of AI systems. Computer Law & Security Review, 39, 105474.
27. Smuha, N.A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International, 20(4), 97–106.
28. Moses, L.B. (2013). How to think about law, regulation and technology: Problems with 'technology' as a regulatory target. Law, Innovation and Technology, 5(1), 1–20.
29. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.
30. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.
31. Bandy, J. (2021). Problematic machine behavior: A systematic literature review of algorithm auditing. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 440–455.
32. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68.
33. Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62.
34. Bradford, A. (2020). The Brussels Effect: How the European Union Rules the World. Oxford University Press.
35. Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer.

36. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.
37. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.