International Journal of Sustainable Development & World Policy
2026 Vol. 15, No. 1, pp. 96-104.
ISSN(e): 2305-705X
ISSN(p): 2306-9929
DOI: 10.18488/26.v15i1.5190

Challenges and prospects of global ethical restrictions on artificial intelligence: Lessons learned from the regulatory approaches of the EU and the UN

Estelle Shatverian
Magic Castle School, Russia.
Email: estelleshatverian0@gmail.com

Article History
Received: 5 August 2026
Revised: 8 September 2026
Accepted: 11 September 2026
Published: 16 September 2026

Keywords
Algorithmic bias
Decision-making
Ethics of artificial intelligence
Global regulation of AI
Public administration
Risks of AI
Social justice.

ABSTRACT

This study examines the global system of ethical artificial intelligence (AI) regulation, focusing on supranational initiatives of the United Nations and the European Union, as well as AI implementation in public administration. It analyzes the challenges of applying ethical principles and prohibitions and evaluates prospects for developing an effective global regulatory architecture that minimizes risks without restricting technological innovation. The study systematizes the main ethical risks associated with AI use in public life, compares the regulatory approaches of the EU and the UN, and identifies obstacles to global consensus, including technological asymmetry, the absence of universal definitions, and limited monitoring mechanisms. The research combines a review of theoretical literature and regulatory documents, qualitative content analysis, comparative analysis, and a simulation experiment. The experiment modeled budget allocation between social assistance and business support by two agents: a rational AI algorithm and a socially oriented human. The AI allocated an average of 46.7% of resources to social assistance and achieved an economic growth index of 0.472, whereas the human allocated 62.7% to social assistance, resulting in a lower growth index of 0.389. These findings demonstrate a fundamental tension between economic optimization and social values. The results indicate that ethically significant decisions should not be fully automated. The study therefore proposes a hybrid governance model combining algorithmic efficiency, human oversight, social justice, adaptability, inclusiveness, and evidence-based regulation.

Contribution/Originality: The originality of the study lies in the comparison of mandatory risk-based regulation of the EU and the UN advisory model with the results of a computational experiment demonstrating the contradiction between the economic efficiency of the algorithm and human social priorities.

1. INTRODUCTION

Artificial Intelligence (AI) represent itself as one of the most controversial forces of technological development in the 21st century, as it permeates all areas of human activities, from fundamental science and medicine to public administration, law, and everyday every day communication, creating numerous risks for individual or society as a whole. According to the definition adopted by the Organization for Economic cooperation and development (OECD), AI systems are "*machine-based system that, for explicit or implicit purposes, infers from the data it receives how to generate results such as predictions, content, recommendations, or decisions that can impact the physical or virtual environment.*" This definition reflects not just a technological function, but also the large-scale social impact that AI systems have when making or mediating decisions that are meaningful to humans. Artificial intelligence has transitioned from a technological promise to a societal reality, yet the governance frameworks meant to guide it remain deeply

fragmented. The European Union's AI Act, entered into force on August 1st, 2024, serves as the world's first horizontal binding regulation for AI systems—one that covers AI broadly across sectors rather than targeting specific applications. Simultaneously, the UNESCO (2021) Recommendation on the Ethics of Artificial Intelligence represents the broadest global consensus, endorsed by all 193 member states on a voluntary basis. The two approaches reflect competing visions: while the EU AI Act acts as a hard law that works on a regional basis through enforceable bans, UNESCO's Recommendation is a soft law—hence, recommendation—that focuses on global advisory principles. The main argument of this paper is that the EU's hard law provides the necessary enforcement strategies that the UN otherwise lacks, in the context of AI, and UNESCO's Recommendation is a complementary document that offers legitimacy and inclusivity for long-term global convergence. Nowadays, AI has moved from the stage of laboratory experiments highly specialized applications to the phase of widespread adoption. The Artificial Intelligence Index Report 2025 notes that in 2024, 78% of organizations reported using AI in their operations, and the number of FDA-approved AI-based medical devices increased from six in 2015 to 223 by 2024. In the public sector, as highlighted in the OECD's report "Governing with Artificial Intelligence", AI is beginning to play a crucial role in optimizing public services, improving the efficiency of public finances, combating fraud, and enhancing decision-making processes. But precisely because of its pervasive nature and its ability to affect fundamental rights and freedoms in society, AI poses new ethical challenges. Unlike previous technological revolutions, AI challenges the very foundations of responsibility and justice as axiological determinants in decision-making. Algorithmic bias, lack of transparency (black boxes), deepening digital inequality, threats to privacy, and the possibility of mass surveillance are just a few of the risks that require urgent and systematic regulation. As noted in the Artificial Intelligence Index Report 2025, the issue of trust remains a major concern, with fewer people believing that AI companies can protect their data, and concerns about the fairness and bias of algorithms persist.

At the global level, the United Nations has adopted a resolution to promote "safe, secure, and trustworthy" AI systems and has established an AI Advisory Body. At the regional level, the European Union has taken an unprecedented step by adopting the world's first comprehensive Artificial Intelligence Act (EU AI Act) in 2024, which classifies AI systems based on risk levels and imposes strict restrictions, including a complete ban on certain uses such as social scoring and real-time remote biometric identification in public spaces. These initiatives set a precedent and shape a new direction for global AI policy, but the path to effective global ethical regulation is fraught with immense challenges. This study aims to explore the challenges and prospects of establishing a global ethical framework for AI, drawing lessons from the EU and UN approaches that are already being implemented. We hypothesize that while there has been progress, existing regulatory models face fundamental issues related to value differences, technological sovereignty, and implementation challenges. To achieve real impact, it is necessary to move from declarative principles to effective, adaptive, and inclusive risk management mechanisms.

2. MATERIALS AND METHODS

The methodology proceeds in three stages. First, a literature review establishes the conceptual framework, defining ethical prohibitions in AI and distinguishing them from adjacent regulatory forms such as risk-based regulation, soft law, and technical standards. Second, a document analysis of primary sources—including the EU AI Act (Regulation 2024/1689), UNESCO (2021) Recommendation, and national legislative instruments—examines the specific content, scope, and enforceability of each framework. This is amplified and corroborated by the aforementioned evaluation of the following criteria: enforcement track record, institutional capacity, and the interests of the law/regulation. Third, a structured comparative analysis evaluates the four jurisdictions across four dimensions: legal status, enforcement mechanisms, prohibited practices, and global legitimacy.

The present study is based on two key documents published in 2025 by leading international organizations: 1. The Artificial Intelligence Index Report 2025, prepared by Stanford University. This report is the most comprehensive and peer-reviewed source of data on the state of AI, covering aspects such as technical

advancements, economic impact, responsible AI development, education, and public opinion. For our research, the sections on responsible AI trends, incident analysis, regulatory landscapes, and algorithm bias data are particularly valuable. 2. The OECD's "Governing with Artificial Intelligence" report (2025). This document is the result of a thorough analysis of 200 AI use cases in 11 key government functions. It provides a unique empirical basis for studying the practical challenges of implementing AI in the public sector, including justice systems, taxation, public procurement, and anti-corruption. The report examines in detail both the risks and the management mechanisms to ensure the reliable use of AI.

The theoretical basis is based on research on the fundamental principles of AI ethics and regulatory approaches. In this context, the works of Khan et al. (2022); de Paula Porto et al. (2025) and Farooq, Tahseen, and Omer (2021) identifies five key categories of ethical requirements: transparency, fairness, accountability, security, and privacy. This classification has become the basis for our own analysis of the EU and UN regulatory approaches.

To analyze the regulatory aspects, we relied on the work of Laux, Wachter, and Mittelstadt (2024) and Butt (2024); Musch, Borrelli, and Kerrigan (2023) and Arents and Vanderstraete (2024). These works provide a detailed legal analysis of the EU AI Act, its structure, risk classification, and implementation mechanisms. They also draw on the work of Russian researchers on the ethics of economics and AI (Bondarenko, Baynazarov, & Farkhtdinov, 2025; Konev & Tsokova, 2020; Sazonov & Tsygankova, 2025), and the regulation of AI (Arzamasov, 2023; Bayniyazova, 2024; Khairullin, Yamalova, Bondarenko, & Ilikaev, 2025; Polyakova & Kamalova, 2023).

Methodologically, the study is based on a combination of several approaches: 1. A qualitative content analysis of the above-mentioned reports was conducted to identify, systematize, and interpret relevant data on regulatory approaches, ethical risks, and management practices. 2. A comparative analysis was used to compare and contrast the approaches of the EU and the UN, as well as to identify common and specific challenges. 3. Methods of synthesis and generalization of specific cases presented in the reports (such as the Robdebt scandal in Australia and the Toeslagenaffaire in the Netherlands) were used to summarize empirical data on the consequences of AI implementation and identify systemic issues.

3. THE RESULT OF THE RESEARCH

Artificial intelligence ethics is a complex and interdisciplinary field of knowledge that studies the moral principles and norms that should underlie the development, implementation, and use of AI systems. It is not a simple set of rules, but rather encompasses a wide range of issues, from the philosophical foundations of autonomy and responsibility to specific technical solutions for ensuring the transparency and fairness of algorithms. According to the analysis conducted in the AI Index and OECD reports, AI ethics focuses on several key areas of concern, as shown in Table 1.

Table 1. Main ethical risks associated with the use of AI in public administration and social life (Based on OECD data and the AI Index Report).

Risk category	Description	Specific threats and examples
Ethical risks	Threat to fundamental human rights and freedoms, including justice, privacy, and democratic values.	Algorithmic bias: discrimination based on race, gender, or social status (e.g., facial recognition systems that mistakenly identify people with darker skin). Privacy violations: mass collection and analysis of data for profiling and surveillance (e.g., the Cambridge Analytica scandal, the use of AI for monitoring citizens in some countries). Lack of transparency (the inability to explain how an AI system came to a particular decision ("the black box problem")), which undermines trust and the possibility of appeal.
Operational risks	Failures, errors, and vulnerabilities in AI systems themselves, leading to incorrect decisions and real damage.	"Hallucinations" - the generation of plausible but false information by generative AI. Systematic errors - incorrect calculation of social benefits

Risk category	Description	Specific threats and examples
		(the Robodeb scandal in Australia). Automation of bias, where AI automates and amplifies existing human errors and biases, making them widespread.
Risks of exclusion	Deepening digital, social, and economic inequalities due to unequal access to AI technologies and their benefits.	The digital divide, where AI services are inaccessible to people without internet access or digital skills. Language bias, where English dominates the training data, making AI less effective for other languages and cultures. Discrimination in the labor market through automated candidate selection systems can filter out certain groups of applicants.
Risks of public resistance	Negative public reactions, reduced trust in the government, and a refusal to use AI services due to fear, misunderstanding, or scandals.	Distrust and skepticism about the ability of AI companies to protect data and act fairly. Denial and skepticism as a refusal to accept decisions made or recommended by algorithms. Reputational damage - high-profile AI scandals undermine the credibility of all government AI projects.
Risks of inaction	Lost opportunities to improve the efficiency and quality of public services due to the failure to implement AI, which leads to a growing gap with the private sector.	The obsolescence of the government apparatus, i.e., the inability to use AI to optimize processes, leads to a decrease in competitiveness and efficiency. Loss of control as the inability to regulate technologies, as the government lacks sufficient knowledge and experience in their use.

Source: The data of Stanford University AI Index Report 2025. URL: https://ict.moscow/analytics/ai-index-report-2025/?ysclid=mqg7k0jkji586730890 Governing with Artificial Intelligence the State of Play and Way Forward in Core Government Functions. URL: https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html.

Technical issues (operational risks) directly lead to negative social consequences (discrimination, exclusion), and AI scandals (risk realization) undermine public trust (resistance risk), which directly indicates that effective regulation cannot be narrow-focused and must cover all stages of the AI lifecycle, from data collection and algorithm development to implementation and monitoring. Additionally, the "risk of inaction" itself is often underestimated, despite its significant societal costs, as it prevents the use of AI to address pressing issues such as healthcare or social policy.

The empirical data collected in the reports provide vivid and alarming examples of how these risks are being realized in practice. The scandals surrounding the use of algorithms in the public sector have been a powerful catalyst for stricter ethical requirements and have demonstrated the catastrophic consequences of disregarding ethical norms, as shown in Table 2.

Table 2. Cases of Ethical Failures in the Implementation of AI in Public Administration.

Case name, country	Brief description	Ethical risk (from Table 1)	Consequences and lessons
The Toeslagenaffaire scandal (Netherlands)	An algorithm used by the tax authorities mistakenly accused around 26,000 families (mostly of migrant origin) of fraudulently claiming child benefits. They were forced to repay large sums, leading to financial ruin and social tragedies.	Ethical risks (Algorithmic bias, violation of rights), Operational risks (systematic error), Risks of exclusion.	Led to the fall of the government, loss of trust in public institutions, and the development of new tools for assessing the impact of algorithms on human rights (Human Rights and Algorithms Impact Assessment – IAMA). Demonstrated how even a "technical" error can have devastating social consequences.

Case name, country	Brief description	Ethical risk (from Table 1)	Consequences and lessons
RobodebtScheme(Australia)	An automated system for identifying overpayments on social benefits incorrectly calculated debts for 470,000 citizens, sending them erroneous notifications without human verification and shifting the burden of proof onto the recipients of benefits.	Operational risks (systematic error), Ethical risks (injustice, violation of rights).	The scheme was found to be illegal. The Royal Commission found that the system was designed carelessly and operated without adequate human control. The lesson is that automation without proper management and protection of citizens' rights is unacceptable, especially in socially significant areas.
Using ChatGPT for law solutions (Columbia)	A judge used ChatGPT to make a decision in a case involving the exemption of medical fees for an autistic child. This sparked a heated debate about the reliability, bias, and lack of transparency in the use of generative AI in the justice system.	Ethical risks (lack of transparency and explainability), and operational risks (hallucinations and unreliability).	The Supreme Court of Colombia was forced to issue a directive regulating the use of AI in the judicial system, emphasizing that AI can support but not replace judicial discretion. Lesson: In high-risk contexts, such as justice, the use of generative AI requires strict guidelines and absolute human responsibility.

Source: AI Index 2025 and OECD 2025 reports.

An analysis of real-life cases of AI failures in public administration clearly indicates that the risks are not theoretical. They have measurable human and financial costs. The key lessons learned from these scandals are as follows: 1. The neglect of human oversight (human-in-the-loop) in critical systems is a major cause of disasters. 2. Algorithmic injustice disproportionately affects vulnerable populations, exacerbating existing inequalities. 3. The lack of transparency and accountability in algorithmic decision-making creates a breeding ground for abuse and undermines public trust, which takes years to rebuild. These practical lessons should form the basis of any global ethical prohibitions and norms, making them not just abstract principles, but concrete imperatives aimed at preventing such tragedies. An imitation experiment was conducted to empirically illustrate the differences in decision-making approaches between a rational algorithm and a human subject to social and emotional factors. The purpose of the experiment was to simulate the process of allocating a limited budget resource between two priority areas: Social assistance (Social Aid) and business support (Business Support). This choice is based on the fact that the "efficiency vs. justice" dilemma is one of the central issues in public policy, and it highlights the differences between a utilitarian optimization logic and a socially oriented approach. The simulation involved two agents.

i. The first agent, referred to as "AI," made decisions solely based on a rational objective function, aiming to maximize an objective measure of overall welfare. As such a function, a weighted sum of economic growth and social justice was chosen, with a clear preference for economic efficiency (weights of 0.7 and 0.3, respectively). The economic growth model had a parabolic shape, with a maximum at a social assistance share of around 30%, reflecting the economic consensus that excessive redistribution can reduce incentives for investment and innovation. Social justice was modeled as an increasing function of the social assistance share, but with a decreasing marginal utility. The AI recalculated the optimal proportion every year and smoothly adjusted its strategy, demonstrating rational adaptation.

ii. The second agent, "Human," made decisions based on a more complex and unstable model that took into account not only economic growth, but also social justice and the reduction of inequality (weights of 0.4, 0.6, and -0.2, respectively). A random noise was introduced into its objective function to simulate the emotional and irrational fluctuations inherent in the human psyche, such as the influence of mood, public

opinion, current crises, or populist promises. The human also changed its strategy with greater inertia and in a narrower range, reflecting cognitive limitations and conservative thinking. The simulation covered a time horizon of 20 years, allowing us to trace the long-term decision-making trajectories of both agents.

An ethical prohibition in AI constitutes categorical and non-negotiable in intent, banning either the behaviors of AI or their uses if they are deemed too dangerous or high-risk. This can also be referred to as "forbidden knowledge"—knowledge that is "too sensitive, dangerous or taboo to be produced or shared" (Hagendorff, 2019). This concept must be distinguished from three adjacent regulatory forms. First, risk-based regulation, exemplified by the EU AI Act's four-tier system (Figure 1), imposes proportionate obligations that scale with risk level, whereas ethical prohibitions apply categorically (Allen, Attrill, & Ryan, 2026). Second, soft law relies on voluntary compliance, persuasion, and political pressure; UNESCO's Recommendation explicitly states it is to be applied "on a voluntary basis" by Member States. The distinction between hard and soft law is not binary but turns on varying degrees of obligation, precision, and delegation; soft law is often preferred where flexibility, broad participation, and lower sovereignty costs are necessary (Abbott & Snidal, 2000). Four interconnected risk categories can be identified. Discrimination and bias arise when AI systems embed and exacerbate historical prejudices, "potentially resulting in discrimination, inequality, digital divides, exclusion and a threat to cultural, social and biological diversity" (Crum, 2025). The EU AI Act notes that prohibited practices "contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and fundamental rights…including the right to non-discrimination, to data protection and to privacy and the rights of the child". One could argue that UNESCO treats this risk category as a normative principle, while the EU focuses on its legal status.

4. EXPERIMENTAL MODELING RESULTS

As a result of the computational experiment, the following average values of key indicators were obtained for the entire 20-year period, as shown in Table 3.

Table 3. Comparative results of AI and Human budget allocation.

Indicator	AI (Rational agent)	Human (Social agent)
Average share of the budget for social assistance	46.7	62.7
Average economic growth (conditional index)	0.472	0.389

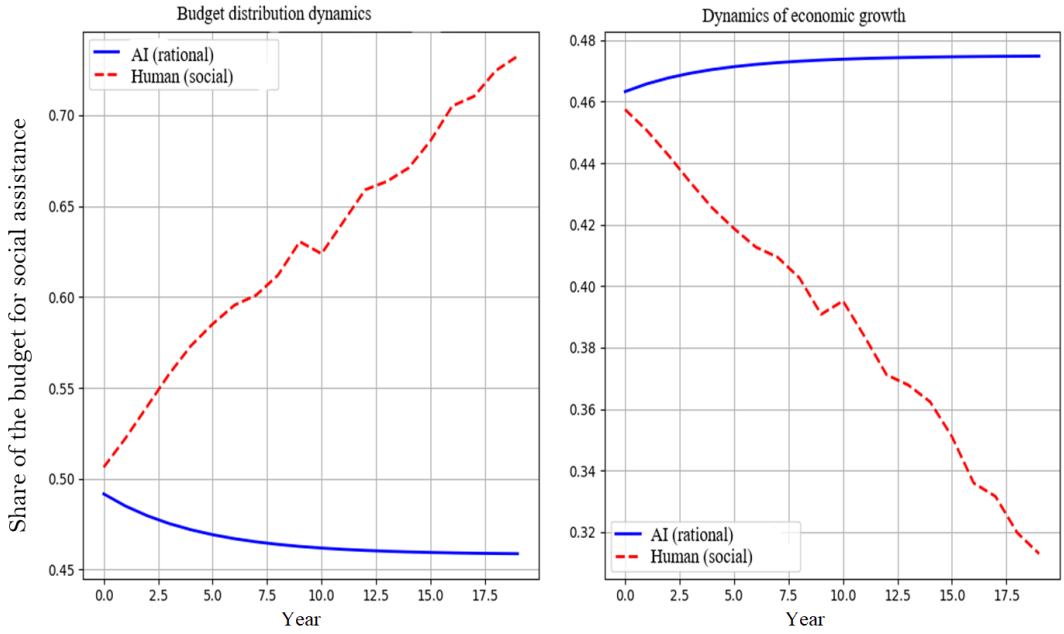

Figure 1. Rational AI and Social Human.

The average share of the budget allocated to social assistance for the AI was 46.7%, while for the Human it reached 62.7%, which means that the human agent allocated 16 percentage points more to social needs on average than the AI. This significant discrepancy indicates that the socially oriented decision-making model, even with its emphasis on fairness, systematically redistributes resources in favor of social support at the expense of direct economic efficiency. In other words, humans exhibit a strong tendency towards "social altruism," which is suboptimal from a purely rational perspective. This redistribution (and this is crucial) led to a natural decrease in economic growth, with the average economic growth rate for AI being 0.472, compared to 0.389 for humans. Human decisions driven by social motivations resulted in a decrease in economic growth by 8.3 percentage points, highlighting the fundamental gap between efficiency and equality. Attempts to improve social justice through increased redistribution inevitably led to a decrease in economic growth, which could potentially undermine the ability to provide large-scale social assistance in the long run. Comparing these two results, we can draw an important conclusion about the nature of AI as a decision-making tool: AI, devoid of empathy and social preferences, acts as a "cold calculator," optimizing a given function with impersonal precision. Its decisions are rational, consistent, and predictable, but they ignore social context and human values that go beyond narrow economic metrics. However, humans, even when informed about potential economic consequences, tend to make decisions based on empathy, a sense of justice, and social responsibility. His behavior is irrational from the perspective of strict optimization, but it reflects a complex system of social priorities that cannot always be measured in monetary terms. These results objectively align with data from the AI Index 2025 report, which notes that while AI can improve productivity and assist in cognitive tasks, human judgments about fairness remain crucial. Our experiment shows that trying to fully delegate ethically relevant decisions to AI without taking into account social values can lead to economically optimal but socially unacceptable results, but on the other hand, relying solely on human irrationality means accepting systematic losses in efficiency. This leads us to the need to find a balance and develop ethical regulatory mechanisms that not only prohibit or limit AI, but also guide its work towards achieving socially significant goals.

5. DISCUSSION

It would be incorrect to assume that ethics should be a universal and unconditional imperative for all AI applications, as there are areas where the priority of ethical norms may conflict with other equally important values, such as national security, state survival, or protection from existential threats. This raises the challenging question of whether there are situations where the regulation of AI ethics is not necessary or should be significantly relaxed. For example, the use of AI in weapons systems, intelligence, cybersecurity, and strategic planning is inevitably associated with the potential for harm, but in the context of armed conflict, ethics can be redefined through the lens of military necessity and self-preservation. For instance, the use of autonomous systems for detecting and neutralizing threats may be justified in terms of reducing casualties among one's own military personnel and improving the accuracy of strikes. In this context, the ethical prohibition on the use of lethal autonomous weapons (LAWS) is challenged by the argument that AI may be more accurate and less prone to error than humans under stress. It should also be noted that attempts to impose universal ethical prohibitions in the military sphere are challenged by the realities of international competition, as if one country voluntarily restricts itself from developing and using AI for military purposes, it may create a strategic advantage for its potential adversaries who do not share the same ethical restrictions. As highlighted in the OECD report's concept of "risk of inaction," the inability to utilize AI can lead to significant missed opportunities and exacerbate the gap between the public and private sectors. In the military sphere, this gap can mean the difference between security and vulnerability. Ethical regulation in this area risks facing a classic prisoner's dilemma, where cooperation is beneficial for everyone, but unilateral disarmament proves to be an irrational strategy. Another example is in emergency situations such as epidemics or natural disasters, where in times of crisis, society is willing to sacrifice some ethical boundaries in

order to preserve life and health. The use of AI to monitor citizens' movements, predict the spread of a virus, and optimize resource allocation can be justified, even if it involves temporary restrictions on privacy. As the COVID-19 pandemic has shown, such measures are often perceived by society as a necessary evil, despite the protests of human rights activists. Therefore, regulation in such cases should be adaptive rather than prohibitive, with clear timeframes and control mechanisms to prevent future abuses. The answer to the question of the need for ethical regulation of AI cannot be unambiguous, as universal prohibitions should apply in peaceful situations, protecting citizens from discrimination, interference in privacy, and manipulation. However, in extreme security-related situations, ethical norms should be adapted, but not completely abandoned. The key principle here is proportionality. Restrictions on rights and freedoms should be temporary, minimally necessary for achieving a legitimate goal, and accompanied by mechanisms for oversight and accountability.

6. CONCLUSION

Effective ethical regulation is impossible without considering the full range of risks, from operational and ethical risks to exclusion risks and public resistance, as it must cover all stages of the AI lifecycle and be adaptable to rapidly changing technologies. Existing approaches, such as the EU's risk-based model and the UN Framework Conventions, reflect different levels of regulatory maturity. The EU offers specific and strict regulations, while the UN and the OECD focus on developing common principles that must be adapted to national contexts. However, both face challenges in implementation, exacerbated by differences in technological development and political systems. Our experimental modeling clearly demonstrates that AI and humans make decisions differently: AI is rational and optimizes objective metrics, while humans tend to exhibit socially oriented but irrational behavior. This means that simply delegating ethically significant decisions to AI without considering social values is unacceptable, but completely abandoning AI in favor of human irrationality leads to a loss of efficiency. In some areas, such as military affairs and emergency response, universal ethical prohibitions may be ineffective or even counterproductive. Here, a flexible approach based on risk management is required, while maintaining human control and strictly adhering to the principles of proportionality and accountability. Four principles emerge from this analysis. First, substantive convergence on red lines: Social scoring prohibition, mass surveillance constraints, and biometric categorization bans appear across the EU AI Act (Article 5) and the UNESCO Recommendation. Second, procedural pluralism with mutual recognition: different jurisdictions may implement shared principles through distinct mechanisms.

Third, enforcement proportionality: core prohibitions warrant binding effect (EU model), while lower-risk applications accommodate softer approaches (UN model). Fourth, developing country participation as a governance necessity: concrete mechanisms might include embedding developing countries as formal participants in UNESCO's Readiness Assessment Methodology rollouts, or establishing observer representation in the EU's Global Dialogue on AI Governance from 2026 onward. Wilkinson (2025) is cautious about the pace of change, noting that meaningful redistribution of AI development capacity to the Global South will not happen quickly, but argues that new norms, even non-binding ones, can carry significant long-term weight.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The author states that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

Abbott, K. W., & Snidal, D. (2000). Hard and soft law in international governance. *International Organization*, *54*(3), 421-456. https://doi.org/10.1162/002081800551280

Allen, B., Attrill, N., & Ryan, F. (2026). *China's plan to scale its way to AI dominance*. Australia: The Strategist, Australian Strategic Policy Institute (ASPI).

Arents, H. C., & Vanderstraete, T. (2024). The European AI Act: A regulatory framework to achieve the necessary trust in AI. *European Public Mosaic (EPuM): Open Journal on Public Service*(22), 38-51.

Arzamasov, Y. G. (2023). Optimal model of legal regulation in the field of artificial intelligence. *Proceedings of Voronezh State University. Series: Law*, *2*(53), 133-148.

Bayniyazova, Z. S. (2024). The problem of consolidating the legal status of an individual in the Russian legal system in the digital age. *Eurasian Law Journal*(1(188)), 14-17.

Bondarenko, A. V., Baynazarov, I. N., & Farkhtdinov, R. T. (2025). The role of social intuition in the formation of corporate decisions and public administration. *Discussion*((9(142)), 243-249.

Butt, J. S. (2024). Analytical study of the world's first EU artificial intelligence (AI) Act, 2024. *International Journal of Research Publication and Reviews*, *5*(3), 7343-7364. https://doi.org/10.55248/gengpi.5.0324.0914

Crum, B. (2025). Brussels effect or experimentalism? The EU AI Act and global standard-setting. *Internet Policy Review*, *14*(3), 1-21. https://doi.org/10.14763/2025.3.2032

de Paula Porto, D., Prado, R. D. C. V., dos Santos Marques, G., Serrano, A. L. M., de Mendonça, F. L. L., & Canedo, E. D. (2025). *Ethical requirements in the age of artificial intelligence: A systematic literature review*. Paper presented at the Proceedings of the 21st Brazilian Symposium on Information Systems (SBSI 2025), Brazilian Computer Society (SBC), Brazil.

Farooq, M. S., Tahseen, R., & Omer, U. (2021). Ethical guidelines for artificial intelligence: A systematic literature review. *VFAST Transactions on Software Engineering*, *9*(3), 33-47. https://doi.org/10.21015/vtse.v9i3.701

Hagendorff, T. (2019). From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, *21*(4), 331-343. https://doi.org/10.1007/s10676-019-09510-5

Khairullin, V. A., Yamalova, E. N., Bondarenko, A. V., & Ilikaev, A. S. (2025). *A program for a comprehensive analysis of social utility (Version 1)*. Certificate of State Registration of the Computer Program No. 2025684754. Federal State Budgetary Educational Institution of Higher Education Ufa University of Science and Technology, Russian Federation.

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., . . . Akbar, M. A. (2022). *Ethics of AI: A systematic literature review of principles and challenges*. Paper presented at the Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering.

Konev, S. I., & Tsokova, B. A. (2020). Ethical and legal problems of regulation of artificial intelligence and robotics in domestic and foreign law. *Education. Science. Scientific Personnel*(4), 68-73.

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, *18*(1), 3-32. https://doi.org/10.1111/rego.12512

Musch, S., Borrelli, M., & Kerrigan, C. (2023). The EU AI Act: A comprehensive regulatory framework for ethical AI development. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4549248

Polyakova, T. A., & Kamalova, G. G. (2023). Problems of legal policy formation in the field of artificial intelligence technology application. *Legal Policy and Legal Life*(1), 28-36.

Sazonov, S. P., & Tsygankova, V. N. (2025). Digital transformation of personnel management in Russian companies: Challenges and trends. *Human Progress*, *11*(10), 1–9.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Paris, France: UNESCO.

Wilkinson, I. (2025). *Can the UN's new AI governance efforts weather the AI race? Chatham House*. Retrieved from https://www.chathamhouse.org/2025/09/can-uns-new-ai-governance-efforts-weather-ai-race