Proceedings of ICILLP 2025 Symposium: Property Law and Blockchain Applications in International Law and Legal Policy
DOI: 10.54254/2753-7048/2025.LD27419

Artificial Intelligence and Social Ethics: Opportunities, Challenges, and Boundaries — Ethical Reflections in the Age of Technological Waves

Haoyu Wen

Lianjiang Experimental School, Zhanjiang, China
wenw4078@gmail.com

Abstract. Currently, artificial intelligence technology is integrating into various sectors of society at an unprecedented pace, profoundly transforming modes of production, daily life, and decision-making. However, the complex ethical challenges it poses are becoming increasingly prominent, necessitating systematic research. This paper focuses on the core social ethical dilemmas arising from the widespread application of AI technology and corresponding strategies to address them. Using a multidisciplinary literature analysis approach, it examines existing ethical frameworks, which primarily include the erosion of personal privacy and autonomy due to large-scale algorithmic surveillance, threats to fairness posed by algorithmic discrimination and bias, drastic changes in employment structures and issues of social equity triggered by AI replacing human labor, as well as problems such as ambiguous accountability and lack of transparency resulting from automated decision-making. In the final section, the paper proposes a comprehensive global governance framework that integrates technological governance, legal regulation, multi-stakeholder participation, and public ethics education. This framework aims to guide AI technology toward a more responsible, equitable, and human-centered development, ensuring the harmonious advancement of technological progress and social well-being.

Keywords: Artificial Intelligence, Social Ethics, Ethical Governance, Human-Machine Relationship

1. Introduction

With breakthroughs in deep learning, big data, and algorithms, artificial intelligence has deeply integrated into core societal domains, such as healthcare, finance, justice, and social interaction, reshaping human production and lifestyles. However, the application of algorithms can also evolve into a double-edged sword. The rapid advancement of technology is accompanied by severe ethical challenges: algorithmic bias exacerbates social inequality, the "trolley problem" in autonomous driving presents moral dilemmas, the misuse of facial recognition infringes on privacy rights, and AI-driven displacement of human labor leads to employment crises. These phenomena highlight the "profound conflict between technological rationality and humanistic values," making "AI and social ethics" a core issue determining the trajectory of human civilization.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.LD27419

This paper mainly focuses on the "loss of ethical norms in AI technology applications," employing case studies to reveal real-world contradictions, constructing an evaluative framework based on ethical theories, and exploring governance pathways by analyzing policy documents and industry standards. The core questions in AI technology application are: How can we balance technological innovation with ethical constraints? How can universal AI ethical principles be established? How can we ensure that technological development promotes social equity rather than deepens divisions? The theoretical significance of this study lies in bridging the research gap between philosophy of technology and practical ethics, while its practical value lies in providing policymakers with risk warnings and governance references, guiding AI toward a "human-centric" and trustworthy direction, and contributing to the construction of a sustainable future where technology and society coexist harmoniously.

2. Fundamental ethical challenges in the interaction between artificial intelligence and society

2.1. Algorithmic bias and social equity

Algorithmic bias is essentially a reflection of social bias in the age of artificial intelligence. In other words, algorithms are not absolutely neutral; their objectivity is only manifested in their operational processes [1]. Bias primarily stems from training data deviations and flaws in algorithm design. For example, visually striking rural scenes are often highlighted on short-video platforms. These scenes tend to employ techniques such as color contrast, composition, and dramatic storytelling to portray rural life as fundamentally different from urban environments. While this spectacularized representation can quickly capture viewers' attention, it may also lead them to perceive rural areas as static and homogeneous spaces rather than vibrant, evolving societies [2]. Thus, algorithmic bias perpetuates social discrimination and hinders a comprehensive public understanding of complex realities.

2.2. Data privacy and information security

The rapid development of AI relies heavily on the training and application of massive datasets, yet this data-intensive nature heightens the risk of privacy breaches. Deep learning algorithms, through in-depth mining and correlation analysis of sensitive personal information, may lead to identity theft and discriminatory decision-making, raising both ethical and legal concerns. Although current privacy protection technologies, such as differential privacy, enable “data usability without visibility,” they still involve a trade-off between security and model accuracy. Existing legal frameworks, like the data minimization principle, may intensify conflicts of interest between individuals and credit agencies during personal credit information collection. They could also further imbalance supply and demand in the personal credit product market [3]. Therefore, future efforts must focus on building privacy governance paradigms that integrate technological encryption and rights protection to balance innovation with ethical safeguards.

2.3. Accountability and lack of transparency

When AI causes harm in complex decision-making scenarios, accountability becomes significantly blurred. For instance, self-driving cars raise acute attribution dilemmas in accident scenarios. Traditional legal frameworks struggle to clearly identify liable parties—as autonomy levels increase from L0 to L5, the driver’s control diminishes accordingly, and their responsibility shifts. Key issues include the collection of algorithm update data, route selection, obstacle avoidance, and the extent of

DOI: 10.54254/2753-7048/2025.LD27419

post-sale obligations of manufacturers and sellers [4]. This lack of transparency severely impedes accident investigation and accountability determination, obstructs judicial redress, and fuels public distrust.

3. Reshaping of ethical relations in specific social domains by AI

3.1. Work and employment ethics

AI's deep integration into the labor sector is triggering structural changes in work ethics. Its impact on the job market is characterized by a substitution effect that significantly outweighs its creation effect, owing to both cost and skill advantages. In the short term, intelligent automation will gradually replace certain jobs, inevitably accompanied by unemployment risks [5]. When algorithmic management reduces labor to efficiency metrics, professional loyalty is simplified into performance parameters. This shift reveals a human value dilemma—as discussed in "The Age of AI and Human Value" [6], AI-generated deepfakes and information filtering can undermine societal consensus on truth and threaten democratic values. More critically, AI-driven productivity gains exacerbate inequality, and the concentration of algorithmic power allows technology owners to extract "intelligence rents," while ordinary workers face welfare squeezes. This technological alienation calls for a reinvention of labor justice theory in the digital age: we must acknowledge AI's value in liberating humans from repetitive labor while establishing ethical safeguards to prevent the devaluation of human dignity in work.

3.2. Human–machine relations and moral agency

AI's anthropomorphic design and emotional interaction technologies are reshaping traditional human–machine relationships and sparking philosophical debates about moral agency. Through natural language generation, affective computing, and social behavior simulation, AI systems can mimic human emotional expressions and even elicit empathetic responses from users. For example, Replika, one of North America's most popular chatbots, was created to help its founder cope with the loss of a friend and alleviate loneliness. Research shows that successful and satisfying interactions with AI agents can trigger user passion [7]. However, this technological trend blurs the line between tool and agent, leading people to unconsciously assign quasi-agent status to AI. Relational ethics proposes that prolonged anthropomorphic engagement can confer "functional agency" on AI, positioning it as a quasi-moral participant within particular social contexts. Some scholars argue that if humans willingly relinquish their agential status and become "slaves to machines," and if designers continue to personify social robots, it is not impossible for robots to become ethical agents [8]. This tension highlights a core dilemma in technology ethics: how to enhance human–AI collaboration while avoiding the cognitive alienation caused by emotional simulation. Future regulatory frameworks must define the limits of anthropomorphism in design to ensure humans maintain a clear understanding of technology's essence during emotional interactions.

4. Examination of existing ethical principles and governance frameworks

4.1. Mainstream international AI ethical principles

Current international AI ethical principles (from the EU, OECD, UNESCO, etc.) have formed a basic framework centered on fairness, transparency, explainability, and accountability, while also

DOI: 10.54254/2753-7048/2025.LD27419

considering privacy protection, safety, human well-being, and controllability. Although these principles constitute the normative foundation for AI ethics governance, they face multiple practical tensions from the perspective of international soft law: AI ethical principles are often too abstract; regulatory oversight is either excessive or insufficient; and "regulatory competition" hinders global development. The root of these value conflicts lies in the lack of clear prioritization standards within existing principles, resulting in inconsistent ethical trade-offs across different application scenarios. Global AI governance must develop concrete, human-centric ethical principles and implement categorized and differentiated regulation [9].

4.2. Comparison of governance models across countries/regions

Global AI ethics governance currently exhibits diverse models. The EU adheres to a "risk-based" regulatory approach, centered on the AI Act, which classifies AI systems into four risk levels—unacceptable, high, medium, and low—based primarily on "threats to citizens' rights and discrimination" [10], emphasizing human rights protection and preemptive prevention. The US adopts a sectoral regulatory strategy, relying on non-mandatory standards (e.g., NIST AI RMF) and ex-post accountability mechanisms to balance innovation incentives and risk control. China emphasizes "simultaneous development and security," implementing regulations such as the Interim Measures for the Management of Generative AI Services to promote inclusive, prudent, and tiered oversight [11]. In comparison, the EU's model is stringent but may stifle innovation, the US approach is flexible but lacks uniformity, and China focuses on application governance but requires deeper ethical integration. The core challenge of global governance lies in reconciling regulatory differences, preventing transnational risks, and avoiding fragmented standards that hinder technological cooperation and development.

4.3. Limitations of existing governance

Current AI ethics governance still faces significant limitations. The lack of global consensus leads to fragmentation, with national differences in governance approaches stemming from ideologies, development stages, and national interests, resulting in a "polycentric, low-coordination" global AI governance landscape [12]. Diversified technical standards are difficult to harmonize, as they are rooted in different social values and ethical concepts. Cultural and structural differences impede international agreement, exacerbating competition and geopolitical divisions, thereby endangering the harmony, stability, and effectiveness of the global governance system [13]. Finally, a considerable gap exists between abstract principles (e.g., transparency, fairness) and concrete rules, with a lack of operable implementation standards and evaluation tools undermining practical effectiveness. These limitations collectively challenge the actual efficacy and sustainability of governance systems.

5. Building a path for responsible AI ethics governance

5.1. Constructing a multi-level governance system

Establishing a responsible AI ethics governance system requires a multi-level collaborative approach. Governments should lead in improving legal and regulatory frameworks, establishing specialized oversight agencies, setting mandatory technical standards, and promoting high-quality public data openness and governance. Industry organizations must develop self-regulatory agreements, promote best practices, and foster third-party ethics certification and audit mechanisms

DOI: 10.54254/2753-7048/2025.LD27419

to create internal industry constraints. Enterprises should integrate ethical requirements into the design, development, deployment, and evaluation of AI, establish internal ethics review boards, and provide systematic ethics training for development and operational teams. At the societal level, public digital literacy should be enhanced, supervision and participation channels broadened, and the critical yet constructive role of media and non-governmental organizations encouraged to form a broad social oversight atmosphere. Only through organic coordination across all levels can an agile, effective, and resilient governance ecosystem be achieved.

5.2. Key technological and institutional support

Building a responsible AI ethics governance system depends on the synergistic support of key technologies and institutions. Core technological breakthroughs lie in developing explainable AI, striving to build a dual-track governance system of "explainability + accountability," pursuing traceability and verifiability in AI decision-making, and gradually establishing risk remediation mechanisms to address the "black box" dilemma [14]. Simultaneously, privacy-enhancing computation technologies must be widely promoted. On one hand, privacy computing technologies enable secure and trustworthy circulation of data elements, meeting data security risk prevention needs; on the other hand, challenges and risks brought by technological updates must be addressed to ensure security and compliance throughout the data processing lifecycle [15]. the key is to establish transparent accountability mechanisms that clearly delineate the duties of developers, deployers, and users, while ensuring robust redress and complaint channels deliver prompt compensation and legal protection to those whose rights are violated. The deep integration of these elements forms an indispensable foundational pillar for the implementation of ethical principles.

5.3. Promoting global dialogue and cooperation

Establishing a responsible AI ethics governance path urgently requires global dialogue and cooperation beyond national borders. The primary task is to promote internationally accepted basic ethical norms and minimum standards for AI, providing a common value benchmark for transnational R&D and application. On this basis, policy coordination and risk information sharing among regulatory agencies must be strengthened to avoid regulatory arbitrage and market fragmentation, thereby building a synergistic governance ecosystem. Faced with global challenges such as ethical dilemmas, security issues, and social impacts triggered by AI, no single country can address them alone. Only through multilateral mechanisms to build consensus and integrate resources can technological risks be effectively managed, guiding AI toward fair, safe, and sustainable development and ensuring the common interests of the global digital future.

6. Conclusion

The rapid advancement of artificial intelligence is reshaping our social structures, economic models, and interpersonal relationships with unprecedented depth and breadth. The analysis presented in this paper clearly demonstrates that the wave of technological progress brings with it complex and profound socio-ethical challenges. From algorithmic biases that may exacerbate social inequities, severe threats to data privacy, and the disruptive impact of automation on labor markets to "black box" decision-making that complicates accountability and superintelligent systems raising existential concerns—these issues are not distant science fiction but pressing real-world dilemmas. Technology itself may be neutral, but its societal consequences are shaped by human choices.

DOI: 10.54254/2753-7048/2025.LD27419

Addressing the ethical crises prompted by AI cannot rely solely on technical self-improvement; it necessitates the construction of a multidimensional, collaboratively governed ethical framework. This requires: (1) Strengthening Ethical Research and Value Embedding: Integrating core ethical values such as fairness, transparency, accountability, and privacy protection deeply into the entire lifecycle of AI system design, development, and deployment. (2) Improving Legal Regulations and Standard Systems: Governments must accelerate the establishment of forward-looking, adaptable, and enforceable laws and regulations while promoting internationally recognized AI ethical standards and certification mechanisms. (3) Promoting Interdisciplinary Dialogue and Public Participation: Engineers, ethicists, social scientists, policymakers, legal experts, and the broader public must engage in sustained and in-depth dialogue to collectively define the boundaries of "AI for good" and ensure that technological development reflects shared social values. (4) Clarifying Accountability Mechanisms: Clearly defining the responsibilities of all stakeholders in the AI system (developers, deployers, and users) to provide a legal basis for redress in cases of harm. Ultimately, the central issue of AI ethics lies in ensuring that technological progress serves human well-being, upholds human dignity and agency, and promotes social equity and justice. In confronting the ethical challenges of AI, what is needed is not only cautious vigilance but also proactive action and wise guidance. Only through a collective societal effort to construct a responsible pathway for AI development can we truly harness this transformative force. In doing so, AI can become a powerful catalyst for building a better, more just, and sustainable future—rather than a source of social division or a threat to human values. Shaping the future of AI is shaping our common future, and it demands that we embrace this responsibility with a profound ethical consciousness.

References

[1] Ping Yue, Yue Miao. (2021). Social Governance: Issues and Regulation of Algorithmic Bias in the Age of Artificial Intelligence. Journal of Shanghai University (Social Sciences Edition), 38(06).
[2] Sanmin Che. (2025). Algorithmic Bias and Rural Imagination: The Shaping of Rural Image Perception Among Audiences on Short-Video Platforms. News Outpost, (07): 76–78.
[3] Xiangjuan Zhai. (2024). The Applicability Dilemma and Alternative Solutions of the Minimization Principle in Personal Credit Information Collection. Journal of Nanjing University (Philosophy, Humanities and Social Sciences), 61(03): 73–82+158.
[4] Peihong Wu. (2024). A Study on the Attribution of Criminal Responsibility in Autonomous Driving. Guizhou Minzu University. DOI: 10.27807/d.cnki.cgzmz.2024.000070.
[5] Wenyuan Sun, Qi Li. (2022). The Employment Effects of Artificial Intelligence and Risk Responses. Unity, (06): 36–39.
[6] Henry Kissinger, Eric Schmidt, Craig Mundie. (2025). The Age of Artificial Intelligence and Human Values. China Information Times, (05): 256.
[7] Hernandez-Ortega B, Ferreira I. (2021). How Smart Experiences Build Service Loyalty: The Importance of Consumer Love for Smart Voice Assistants. Psychology & Marketing, vol. 38, pp. 1134.
[8] Jianhua Li. (2023). Ethical Reflections on Human-Machine Relationships in the Intelligent Era. Theory Monthly, (09): 5–15. DOI: 10.14180/j.cnki.1004-0544.2023.09.001.
[9] Guang Ma, Liwen Wang. (2024). The Current State and Recommendations for Global AI Governance from the Perspective of International Soft Law. Customs and Economic Trade Research, 45(03): 1–19.
[10] Yifan Dong, Changni Gu. (2024). Policy Evolution, Strategic Considerations, and Impact Analysis of the EU Artificial Intelligence Act. China Information Security, (08): 88–92.
[11] Qiming Fan. (2023). Key Points Analysis and Development Suggestions for the Interim Measures for the Management of Generative Artificial Intelligence Services. Enterprise Management, (09): 19–21.
[12] Yongjiang Xie. (2025). The Current State of Global AI Governance and the Challenges and Responses for China. China Cyberspace, (04): 59–63.

DOI: 10.54254/2753-7048/2025.LD27419

[13] Yonghui Han, Gangjuan Zhou, Cuifen Xu. (2024). The Status Quo, Dilemmas, and China's Path in Global Artificial Intelligence Governance. Special Zone Practice and Theory, (06): 94–102. DOI: 10.19861/j.cnki.tqsjyll.20250106.005.
[14] Qi Wang. (2025). Transparency and Interpretability of Artificial Intelligence Technology: Solving the "Black Box" Problem. New Security, (03): 61–64.
[15] Xia Xie. (2024). Risk Issues and Institutional Improvement in the Application of Privacy Computing in China. Xinyang Normal University. DOI: 10.27435/d.cnki.gxsfc.2024.000204.