WEP Warwick Evans Publishing

International Journal of Global Economics and Management

ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijgem.v4n2.08
Journal homepage: https://wepub.org/index.php/IJGEM/index

Exploring the Governance of Artificial Intelligence Ethics: Current Issues and Challenges

Jiumei Zhang [1, 2], Weiqiang Li [2], Liping Wang [2]

[1] Brunel University London Business School, Brunel University London, Uxbridge, London, UK
[2] Zhanjiang University of Science and Technology, Zhanjiang, Guangdong, China

ABSTRACT

This paper delves into the current state, challenges, and issues in the governance of artificial intelligence (AI) ethics, proposing an ethical risk assessment model and governance strategies. The article begins by summarizing the ethical issues arising from the rapid development of AI technology and their impact on society, including privacy invasion, algorithmic bias, and unclear accountability. It then analyzes the limitations of current AI ethics governance, including deficiencies in governance systems and tools, ambiguity in ethical principles, and a lack of public awareness and international cooperation. Building on this, the paper introduces a comprehensive ethical risk assessment model that emphasizes the consideration of multidimensional factors and the importance of dynamic updates. Furthermore, it constructs a comprehensive governance framework that includes mechanisms for prevention, response, and remediation to effectively manage ethical risks. Finally, the article suggests enhancing the clarity and implementation of ethical principles, increasing public awareness, and fostering international cooperation, while looking forward to future research directions, including the application of interdisciplinary research methods to promote harmonious development between technology and ethics.

KEYWORDS

Artificial Intelligence; Risk Governance; Ethical Principles

1. INTRODUCTION

1.1. AResearch Background and Importance

The rapid development and widespread application of artificial intelligence (AI) technology have penetrated various layers of society, from everyday social media interactions to critical medical diagnoses, and even military applications at the national security level. However, the accompanying ethical issues have become increasingly prominent, sparking global attention and debate. For instance, the Cambridge Analytica scandal, which involved the unauthorized acquisition of Facebook user data, not only violated user privacy rights but also potentially undermined the democratic electoral process. The Boeing 737 MAX incidents, caused by flaws in the automated control system, resulted in hundreds of deaths, highlighting safety issues in automation technologies. Furthermore, the fatal accident involving an Uber autonomous vehicle raised direct questions about the ethical responsibility of autonomous driving technologies. These cases illustrate that ethical issues in AI are no longer merely theoretical discussions but pressing practical problems that need resolution. They relate not only to the reliability and safety of the technology itself but also to social justice, individual rights, and human well-being. Thus, in-depth research into the governance of AI ethics and the exploration of effective ethical risk assessment and management strategies are crucial for promoting the healthy

development of technology, maintaining social order, and having significant real-world importance and historical impact.

1.2. Current State of AI Ethics

The field of AI ethics currently presents a complex and evolving situation. The conveniences brought by technological advancements coexist with ethical challenges, and societal reactions are increasingly intense. For instance, while facial recognition technology has improved security monitoring efficiency, controversies surrounding Amazon's Rekognition have exposed issues of racial bias, leading to widespread concerns about its accuracy and potential misuse. Scandals involving content moderation on platforms like Facebook highlight the flaws in algorithms managing inappropriate content, as well as the regulatory responsibilities concerning hate speech and the spread of misinformation. The "one-click undressing" feature of the DeepNude application not only violated individual privacy rights but also exposed serious moral and legal hazards associated with AI technology. Additionally, while autonomous driving technologies aim to enhance traffic efficiency, fatal incidents such as the Uber crash have prompted deep reflection on their safety and ethical responsibilities. These events indicate that AI ethics have become a focal point of societal concern, urgently necessitating the establishment of more robust governance mechanisms to ensure that technological development aligns with ethical standards, safeguarding both the overall interests of society and the fundamental rights of individuals.

2. CURRENT STATUS OF GOVERNANCE IN AI ETHICS

2.1. Ethical Issues in Artificial Intelligence

2.1.1. Privacy and Data Protection

Wang Youmei et al. (2023) point out that with the widespread application of artificial intelligence in education, privacy issues have become increasingly apparent, involving the protection of student personal information. Li Mengwei et al. (2023) in their research on service robots, identify risks of privacy invasion, emphasizing concerns over the leakage of personal privacy. Fan Chunliang and Wu Yifei (2023) discuss inherent ethical issues including privacy and human rights, stressing the importance of transparency and explainability.

2.1.2. Algorithmic Bias and Fairness

Wang Youmei et al. (2023) summarize potential biases and discrimination issues in the application of AI in education. Li Mengwei et al. (2023) identify risks of bias or discrimination, involving unfair treatment caused by algorithmic biases. Chen Lei et al. (2021) note ethical challenges brought by AI technology, including issues of algorithmic bias that affect social equity. Fan Chunliang and Wu Yifei (2023) also mention algorithmic bias and discrimination as part of inherent ethical issues.

2.1.3. Accountability and Ethical Principles

Li Mengwei et al. (2023) in their study of service robots, identify risks associated with unclear or improper attribution of responsibility. Chen Lei et al. (2021) discuss issues of responsibility ethics, emphasizing the development principles to strengthen ethical accountability. Wu Hong and Du Yanyong (2021) discuss the formulation and implementation of AI ethical principles, highlighting their significant role in AI governance. Meng Lingyu and Wang Yingchun (2023) propose innovations in the AI ethics review paradigm, emphasizing the application of virtue ethics.

2.1.4. Autonomy and Governance Models

Wang Youmei et al. (2023) emphasize the need for agile governance models in managing ethical risks of educational AI. Bai Junyi discusses the real challenges in governing ethics of educational AI,

suggesting the implementation of a "soft-hard" governance model. Zhu Mingting and Xu Chong (2023) discuss the application of international soft law in AI ethics governance, pointing out the challenges of unstable cooperation and implementation difficulties. Yu Xue (2024) proposes using experimental methodologies for governance of AI ethics. Geng Xu and Huang Miaoxin explore the principles of AI ethics governance in the UK public sector, emphasizing an institutionalized, systematic ethical governance framework. Wang Chen analyzes the issues in medical AI ethics governance, suggesting a dynamic, robust, and sustainable governance mechanism.

3. LIMITATIONS OF CURRENT GOVERNANCE MODELS

3.1. Innovation and Perfection of Governance Systems and Tools

Wang Youmei et al. (2023) indicate that traditional governance methods are no longer suitable for addressing ethical issues in educational AI, citing lack of standardized guidance, systemic insufficiencies, missing humanitarian concerns, lack of innovative governance tools, an unrobust accountability mechanism, and lack of public awareness. Li Mengwei et al. (2023), through designing AI ethical risk assessment methods and analyzing risk source priorities, confirm the applicability of their method but also reflect on the existing governance model's inadequacies in assessing and responding to ethical risks. Zhu Mingting and Xu Chong (2023) discuss the limitations of international soft law and suggest constructing a mixed governance model of international soft and hard law with indirect enforcement mechanisms to enhance the enforceability of soft law. Bai Junyi believes that the governance model for educational AI ethics needs to transition from solely soft law to a combination of soft and hard laws, facing numerous challenges.

3.2. Clarity and Implementation of Ethical Principles

Bai Junyi considers the clarity of ethical principles a key issue in the transition of governance models for educational AI ethics. Wu Hong and Du Yanyong (2021) discuss the application of ethical principles in AI governance, noting challenges in their implementation and the need for measures to ensure their enforcement. Geng Xu and Huang Miaoxin, through analyzing practices in the UK, suggest that the construction and application of ethical governance principles need to be adaptively co-constructed, proposing an "evidence-based action logic of principle guidance-application-embedding" to harmonize ethical principles with governance practices.

3.3. Public Awareness and International Cooperation

Wang Youmei et al. (2023) mention the lack of public awareness as a problem in the governance system. Zhu Mingting and Xu Chong (2023) discuss the limitations of international soft law in AI ethics governance, including unstable cooperation among entities and occasional ineffective implementation, suggesting the construction of a mixed model of international soft and hard law governance. Chen Lei et al. (2021) point out that although there is general agreement among countries, regions, and international organizations on the recommendations and legislation of AI ethics, AI ethics development is still in its early stages, requiring further research and discussion, underscoring the importance of public awareness and international cooperation. Meng Lingyu and Wang Yingchun (2023) believe that traditional ethical review paradigms face challenges when applied to emerging technologies like AI and need innovation in theoretical, functional, and methodological dimensions, involving the cultivation of public awareness and the strengthening of international cooperation.

4. ETHICAL CHALLENGES AND ISSUES IN ARTIFICIAL INTELLIGENCE

4.1. Technical Challenges: Algorithm Transparency, Data Bias, and Explainability

In 2016, in the State v. Loomis case in Wisconsin, USA, the defendant Eric Loomis challenged the use of the COMPAS algorithm—a proprietary tool used to assess his risk of recidivism—which influenced his sentencing. Loomis argued that his inability to access the algorithm's workings violated his legal rights and due process. Although the Wisconsin Supreme Court dismissed his claims stating that the data used by the COMPAS algorithm was public and Loomis had the opportunity to challenge the report's findings, this decision sparked widespread concern and introspection within the legal community and the public regarding the transparency of algorithms and the ethical use and regulation in the criminal justice system.

In 2017, researchers from Stanford University published a study involving a deep neural network capable of recognizing individuals' sexual orientation from facial photos with high accuracy. The AI's accuracy in identifying male homosexuals was 81%, significantly outperforming human judges. This study ignited intense ethical debates and privacy concerns, particularly about potential misuse of such technology, for instance, in countries where homosexuality is illegal to identify and persecute LGBTQ+ individuals. Despite the researchers asserting the neutrality of technology and their aim to enhance understanding of the biological underpinnings of sexual orientation, the potential for abuse raised profound concerns among the public and academia, highlighting the ethical boundaries and social responsibilities that must be considered in AI research and application.

In March 2023, images falsely depicting former U.S. President Donald Trump's arrest, created by Bellingcat founder Eliot Higgins using the AI drawing tool Midjourney, gained significant attention. These images, which included scenes of Trump being arrested, in a prison uniform, on trial, in jail, and escaping, were initially believed by many due to their realistic portrayal despite some unnatural details like extra limbs and malformed hands. The rapid spread and impact of these AI-generated images on social media illustrated both the advancements in AI-generated image technology and the challenges the public faces in distinguishing between real and fake information. This incident spurred discussions on the role of AI in spreading misinformation and the responsibilities of social media platforms and regulatory bodies in managing AI-generated content.

4.2. Social Challenges: Public Trust, Privacy Protection, Employment Impact

Large Language Models (LLMs) have attracted attention for their powerful text generation capabilities. However, errors in basic logic and arithmetic exposed limitations in cognition and reasoning, such as when some models incorrectly answered which number between 13.11 and 13.8 is greater. These errors, referred to as "hallucinations" in the industry, not only raised public awareness but also impacted trust in the reliability of large models. This incident underscores the need for deeper research and improvement in the logical reasoning capabilities of models to ensure their accuracy and reliability in practical applications.

DeepNude, an application using AI technology through a deep learning framework and Generative Adversarial Networks (GANs) to create a "one-click undress" effect, processed clothed photos of women into nude images, leading to widespread ethical and privacy controversies. Despite being quickly taken down, the technology resurfaced under the promotion of criminals, posing severe ethical challenges and emphasizing the importance of protecting individual privacy.

The "2023 Future of Jobs Report" by the World Economic Forum indicated that due to AI, digitalization, green energy transitions, and supply chain shifts, nearly one-fourth of global jobs would change within five years. It estimated that AI could displace up to 26 million roles in record-keeping and administration, such as cashiers, ticket agents, data entry, and accounting, sparking widespread concern about the impact on low-skilled labor. However, the report also noted that with

big data analysis, management technology, and cybersecurity as major drivers of job growth, the net impact of most technologies over the next five years would be positive, suggesting that while AI might lead to job losses in the short term, it could also create new employment opportunities and drive structural changes in the labor market.

5. ETHICAL RISK ASSESSMENT AND GOVERNANCE STRATEGIES

5.1. Risk Assessment Models and Tools

Constructing effective risk assessment models and tools is crucial for ensuring that the development and application of AI technologies adhere to ethical standards. The models should incorporate multidimensional factors such as privacy protection, algorithmic bias, accountability, and autonomy. For example, the AI ethical risk assessment method designed by Li Mengwei et al. (2023) analyzed risk source priorities and provided a systematic tool for assessing and addressing ethical risks. Additionally, these models should be capable of dynamic updates to accommodate new challenges posed by technological advancements. The development of these tools should focus on user-friendliness and operability to ensure that non-professional users can also conduct effective ethical risk assessments.

5.2. Governance Framework: Prevention, Response, and Remediation Mechanisms

Constructing a comprehensive governance framework that includes prevention, response, and remediation is essential for effectively managing AI ethical risks. Prevention mechanisms focus on identifying and assessing risks beforehand through clear ethical principles and standards to guide the ethical considerations during the development and application of technology. Response mechanisms focus on rapid reactions when risks materialize, ensuring timely actions to mitigate negative impacts. Remediation mechanisms focus on post-incident remedies and improvements, through accountability and corrective measures, to restore damaged stakeholder relationships and learn from these instances to optimize future governance strategies.

5.3. Ethical Review and Regulatory Practices

Ethical review and regulatory practices are crucial for ensuring AI technologies meet societal ethical expectations. Ethical reviews should be integral throughout the development, deployment, and application of AI technologies to ensure decisions and operations adhere to ethical principles. Regulatory practices should combine legal regulations and industry standards to supervise and guide the application of AI technologies. For example, the construction of AI ethics governance principles in the UK public sector highlights the importance of an institutionalized and systematic ethical governance framework. Regulatory bodies should also be flexible and adaptable to address new situations and challenges posed by the rapid development of technology.

6. CONCLUSION, RECOMMENDATIONS, AND OUTLOOK

6.1. Conclusions and Recommendations: Key Findings and Contributions

This paper has extensively explored the current state, challenges, and issues in AI ethics governance, revealing key issues such as privacy protection, algorithmic bias, accountability, and the implementation of ethical principles. The study's main contribution lies in proposing a comprehensive ethical risk assessment model that considers multidimensional factors including privacy, bias, accountability, and autonomy, emphasizing the importance of dynamic updates. Additionally, this paper suggests a comprehensive governance framework covering prevention, response, and remediation mechanisms to effectively manage ethical risks. Future research and practice should

focus on clarifying and implementing ethical principles, enhancing public awareness education, and promoting international cooperation to build a more robust and adaptive AI ethics governance system.

6.2. Research Outlook: Future Directions and Possibilities

Looking ahead, research in the field of AI ethics governance will continue to deepen. As technology progresses, new ethical issues such as algorithm transparency, data bias, and explainability will emerge, requiring researchers to continuously update and refine ethical risk assessment tools. The construction of governance frameworks needs further refinement, especially in the specific implementation of prevention, response, and remediation mechanisms. Additionally, enhancing public awareness and strengthening international cooperation will be key to achieving effective governance. Future research should also focus on the application of AI technologies in specific fields such as healthcare, justice, and education, as well as their impact on the job market and social structure. Finally, interdisciplinary research methods will provide new perspectives and solutions for AI ethics governance, promoting harmonious development between technology and ethics.

REFERENCES

[1] Wang Chen. Issues, Dilemmas, and Solutions in the Ethical Governance of Medical Artificial Intelligence [J/OL]. Studies in Science of Science, 1-18 [2024-08-26].

[2] Yu Xue. An Inquiry into the Experimental Methodology of Artificial Intelligence Ethics Governance [J/OL]. Studies in Science of Science, 1-14 [2024-08-26].

[3] Meng Lingyu, Wang Yingchun. Exploring a New Paradigm for Artificial Intelligence Ethics Review [J]. Science and Society, 2023, 13(04): 97-113+165.

[4] Bai Junyi. Ethical Governance of Educational Artificial Intelligence: Real Challenges and Implementation Pathways [J]. Chongqing Higher Education Research, 2024, 12(02): 37-47.

[5] Fan Chunliang, Wu Yifei. A Three-Dimensional Analytical Framework for the Ethical Governance of Artificial Intelligence: Core Elements, Innovation Processes, and Main Responsibilities [J]. Studies in Dialectics of Nature, 2023, 39(10): 18-24.

[6] Chen Xiongshun. Ethical Risks of Artificial Intelligence and Their Governance—A Path Based on the Algorithm Audit System [J]. Studies in Dialectics of Nature, 2023, 39(10): 138-141.

[7] Li Mengwei, Xu Feng, Yan Qi, et al. Design and Practice of Artificial Intelligence Ethical Risk Assessment Methods in the Service Robot Field [J]. China Science and Technology Forum, 2023(10): 74-84.

[8] Zhu Mingting, Xu Chongli. The International Soft Law Governance of Artificial Intelligence Ethics: Current Status, Challenges, and Strategies [J]. Bulletin of the Chinese Academy of Sciences, 2023, 38(07): 1037-1049.

[9] Wang Youmei, Wang Dan, Liang Weiyi, et al. Agile Governance: A New Model for the Ethical Governance of Educational Artificial Intelligence [J]. Research in Distance Education, 2023, 44(07): 21-28.

[10] Geng Xu, Huang Miaoxin. Good Governance of Evidence: The Principle Construction and Evidence-Based Logic of Ethical Governance of Artificial Intelligence in the UK Public Sector [J]. China Science and Technology Forum, 2023(06): 170-179.

[11] Du Yanyong. An Analysis of Several Basic Issues in the Study of Artificial Intelligence Ethical Risk Prevention [J]. Yunnan Social Sciences, 2022(03): 12-19.

[12] Song Yan, Chen Lin, Li Qin, et al. Perception of Artificial Intelligence Ethical Risks, Trust, and Public Participation [J]. Studies in Science of Science, 2022, 40(07): 1153-1162+1171.

[13] Jia Kai, Xue Lan. A Global Comparison of Artificial Intelligence Ethical Issues and Safety Risk Governance: Chinese Practice [J]. Public Administration Review, 2021, 3(01): 122-134.

[14] Wu Hong, Du Yanyong. Artificial Intelligence Ethics Governance: From Principles to Actions [J]. Studies in Dialectics of Nature, 2021, 37(04): 49-54.

[15] Chen Lei, Wang Baicun, Huang Si Han, et al. Artificial Intelligence Ethical Guidelines and Governance Systems: Current Status and Strategic Recommendations [J]. Science and Technology Management Research, 2021, 41(06): 193-200.