



Artificial Intelligence and Ethics: Moral Challenges from a Philosophical Perspective

Gengmeng Liu*

Chengdu Experimental Foreign Languages School, Chengdu, 611134, China

*Corresponding author. Email: czdeukcf@163.com

Abstract. AI technology has not only brought significant efficiency and convenience improvements in key fields such as healthcare, transportation, and finance, but also raised a series of important ethical issues. This article aims to explore these major ethical challenges from a philosophical perspective, in order to provide profound thinking and solutions for this rapidly developing field. With the continuous advancement of AI technology, people are facing increasingly complex ethical tests: from personal privacy protection to algorithmic bias, from the moral responsibility of intelligent systems to their long-term impact on human work and social structure. This article will specifically focus on how AI technology can overturn traditional ethical concepts and explore how to understand and address these challenges from a philosophical perspective. By analyzing theories such as deontology, utilitarianism, and moral ethics, we will explore how to effectively balance technological progress with the protection of social values in the development, design, and application of AI technology, ensuring that individual rights and public interests are appropriately balanced and respected.

Keywords: artificial intelligence; ethical challenges; obligation theory; utilitarianism.

1 Introduction

With the rapid development of artificial intelligence (AI) technology, people are gradually entering a technology driven new era. AI has shown great potential not only in key fields such as medical diagnosis, autonomous driving, and financial analysis, but also gradually penetrated into various aspects of daily life. From voice assistants to smart homes, from financial investment to urban management, the application of AI is significantly changing the way people live and work^[1]. However, with the popularization and application of these technologies, a series of profound ethical issues have also emerged. How can personal privacy be fully protected in the data-driven AI era? Is there bias in algorithmic decision-making based on gender, race, and other factors? Who is responsible for any errors in intelligent systems? These issues not only test our understanding and control over technological development, but also touch upon the core morality and values of human society. In this context, this article aims to explore

the main ethical challenges brought by AI technology from a philosophical perspective, and explore how to strike a balance between technological progress and social value protection. Classic ethical theories such as ethics, utilitarianism, and moral ethics will become our analytical tools, helping us understand and solve ethical dilemmas in the development of AI technology. Through in-depth exploration of these theories, we hope to provide profound thinking and specific solutions for the design, development, and application of future AI technologies, ensuring that the promotion and use of technology not only conforms to efficiency and economic interests, but also values the protection of individual rights and the overall well-being of society. Thus, in this rapidly developing AI era, the complexity and importance of ethical issues are constantly highlighted. The promotion and application of AI technology must find a balance between technological innovation and ethical standards. We need to deeply consider how to ensure that the development of AI technology can not only create economic and social benefits, but also follow ethical principles and legal regulations, ensuring effective protection of individual rights and maintenance of social justice. This is not just a matter of technological development, but also a challenge to human values and social progress. Through philosophical exploration and ethical guidance, we hope to provide more profound and systematic solutions to ethical issues in the AI era, laying a solid moral foundation for future technological development.

2 Basic Concepts of Artificial Intelligence and Ethics

The definition and application areas of artificial intelligence: Artificial intelligence (AI) refers to the ability of computer systems to simulate human intelligence through data and algorithms^[2]. In modern society, the application of AI has penetrated into multiple key fields such as medical diagnosis, autonomous driving, and financial analysis. Narrow AI focuses on performing specific tasks, such as speech recognition systems and recommendation algorithms, and its advantages lie in high efficiency, accuracy, and specialization. These systems can significantly enhance productivity, reduce human error, and streamline processes across industries. For example, in healthcare, AI-powered tools assist doctors in diagnosing diseases with higher precision, while in finance, algorithms analyze market trends and make rapid trading decisions. In contrast, General AI, or AGI (Artificial General Intelligence), aims to replicate the broad range of human cognitive abilities, including autonomous decision-making and problem-solving across diverse tasks. Although AGI is still in its infancy and largely theoretical, its potential to revolutionize industries by performing complex, multi-domain functions autonomously is of great significance. With advancements in machine learning, neural networks, and quantum computing, the boundaries of AI capabilities are continuously expanding, which could eventually lead to breakthroughs that dramatically reshape not only the technological landscape but also societal structures, labor markets, and even human interactions. As AI integrates further into areas such as education, public safety, and environmental management, its applications will continue to grow, potentially transforming the very foundation of social life and economic systems.

The basic concepts and importance of ethics: Ethics is the science that studies moral values and behavioral norms, and its importance lies in guiding individuals and society to make correct moral choices in ethical dilemmas^[3]. With the rapid development of technology, especially in the field of artificial intelligence, ethical issues are becoming increasingly prominent. The widespread application of AI technology and its profound impact on social and economic structures make ethical guidance particularly necessary. Ethical dilemmas in AI involve critical concerns, such as how to protect personal data in an age where algorithms collect and analyze vast amounts of information, and how to address biases embedded in machine learning systems that may lead to discrimination. Beyond privacy and bias, ethical concerns extend to the accountability of decisions made by AI systems, especially in high-stakes scenarios like autonomous vehicles or medical diagnoses, where errors could result in severe consequences. The need for a robust ethical framework is essential to ensure that AI advancements do not undermine societal values, human rights, and fairness. The basic theoretical framework of ethics includes deontology, utilitarianism, and virtue ethics, each offering distinct perspectives on what constitutes correct behavior and decision-making. Deontology emphasizes duties and principles, suggesting that AI systems must adhere to universal moral laws; utilitarianism focuses on the consequences of actions, promoting the greatest good for the greatest number.

3 The Main Ethical Issues of Artificial Intelligence

The first ethical issue concerns privacy and data security. The right to privacy is closely linked to individual autonomy, which entails the ability to make personal decisions without external interference^[4]. Privacy safeguards individuals' thoughts, choices, and actions from unwanted scrutiny or control, ensuring they retain decision-making authority^[5]. Kant's concept of humanity as an end in itself posits that each person is inherently valuable and should not be used merely as a means to someone else's ends. However, many software companies upload user data to the cloud for AI training and even sell this information to other firms, turning users into commodities for corporate profit. In March 2018, a major data breach involving Facebook and Cambridge Analytica highlighted these concerns. Cambridge Analytica collected data via the "this is your digital life" app, developed by Cambridge University's Aleksandr Kogan, under the guise of a psychological test. By accessing users' personal data and that of their Facebook friends, Cambridge Analytica created highly personalized political ads aimed at influencing the 2016 US presidential election and other campaigns. Using AI learning algorithms, Cambridge Analytica analyzed user data to model psychological traits, creating targeted ads designed to alter political views and voting behavior. This constitutes a severe breach of Kantian moral deontology, where users lose autonomy over their behavior, manipulated by AI as political tools. Facebook disregarded users' intrinsic worth, allowing them to be exploited for commercial and political gain.

The second ethical dilemma involves responsibility. The complexity of AI responsibility spans multiple stakeholders: developers, users, corporations, regulators, and others^[6]. Determining liability becomes intricate when multiple parties are involved. In

Aristotelian ethics, moral responsibility hinges on voluntary, informed actions and their consequences^[7]. For instance, in the case of a malfunction in a self-driving car caused by AI, who bears responsibility? Manufacturers are accountable for design flaws in the software, while users are responsible for correct AI utilization. Regulators also share responsibility for overseeing vehicle safety. However, if the AI makes an erroneous decision due to an unforeseeable glitch, no party may be fully accountable. In Aristotelian terms, without voluntary and informed choices, moral responsibility becomes blurred.

The third issue concerns algorithmic bias, particularly gender and racial biases in hiring algorithms. For instance, certain Chinese recruitment apps filter out women without maternity records and ethnic minorities due to cultural holidays and policies^[8]. This practice violates Rawls's theory of justice, which advocates for equal liberties and opportunities for all. AI-driven biases restrict job opportunities, infringe on ethnic minorities' cultural rights, and undermine gender equality.

4 Solution from a Philosophical Perspective

Firstly, the design and implementation of AI must adhere to Kantian deontological principles of universality and necessity. According to Kant, true moral actions are universal—they apply to all people and situations without exception^[9]. Kantian ethics also stress the unconditional and imperative nature of moral duty; there should be no exemptions from moral laws—they must be upheld in every circumstance. Therefore, AI systems should be developed to ensure that their decisions and behaviors are consistent and applicable universally. Decision-making algorithms should be designed with transparent and interpretable rules and standards. Testing AI behavior across various scenarios ensures that ethical principles guide its decisions uniformly. Regardless of context, AI systems must prioritize user privacy as an absolute moral obligation, not contingent on specific situations^[10]. AI should autonomously follow pre-defined moral laws rather than responding to external influences. Embedding explicit ethical rules in AI algorithms enables autonomous adherence to these principles, rather than adjusting behavior based on external incentives or penalties.

Rawls proposed a well-known thought experiment to define principles for a just society: the veil of ignorance^[11]. This hypothetical scenario strips decision-makers of all personal knowledge—wealth, health, talents, social status—making them unaware of their own circumstances. Under this veil, decision-makers are purely self-interested but constrained from using any information that could benefit their own position. AI should integrate an algorithm akin to Rawls's Veil of Ignorance, ensuring fairness by treating everyone impartially, irrespective of race, gender, age, or other factors. By eliminating biases, AI promotes fairness and reduces social prejudice.

Thirdly, applying utilitarian principles to AI aims to maximize overall happiness. Utilitarian ethics seeks to maximize happiness or welfare as its core principle^[12]. It evaluates actions based on their consequences—whether they increase collective happiness or reduce suffering. For instance, in autonomous driving technology, utilitarian considerations prioritize societal benefits such as reducing accidents, improving traffic

efficiency, and lowering labor costs. Despite occasional errors, AI-driven vehicles generally outperform human drivers due to continuous operation capabilities, quicker reaction times, and superior route planning, significantly benefiting society. Autonomous driving not only enhances transportation efficiency but also minimizes accident risks, contributing to overall societal well-being according to utilitarianism's principles of maximizing collective happiness.

5 Conclusion

In conclusion, the rapid development of artificial intelligence (AI) presents society with unprecedented opportunities and profound ethical dilemmas. As AI systems become more integrated into daily life and decision-making processes, it is crucial to navigate these challenges thoughtfully, ensuring that technological progress aligns with human values and social justice. The ethical issues surrounding privacy violations, accountability, and algorithmic bias are not just technical challenges; they touch upon the fundamental moral principles that govern human society. In particular, Kant's deontological ethics, Rawls' theory of justice, and the utilitarian focus on maximizing collective well-being offer valuable philosophical frameworks for addressing these dilemmas. These theories provide practical tools to guide the ethical design, implementation, and regulation of AI technologies, ensuring that they serve the greater good while respecting individual rights.

However, addressing these ethical concerns is not a one-time task but requires continuous reflection and adaptation. As AI continues to evolve, new challenges will emerge, and the solutions we propose today may need refinement tomorrow. To ensure that AI development does not outpace our moral understanding, it is essential to foster interdisciplinary collaboration among technologists, ethicists, legal scholars, and policymakers. This collaboration will help create comprehensive ethical guidelines and regulatory frameworks that not only promote responsible AI development but also protect individuals and society from potential harm.

Moreover, global cooperation is paramount in establishing international ethical standards for AI. As AI technologies transcend national borders, so too must the ethical guidelines that govern them. Countries must work together to create universal principles that ensure fairness, transparency, and accountability in AI systems. By building a global consensus on AI ethics, we can prevent technological advancements from exacerbating social inequalities or infringing on fundamental human rights.

In addition to ethical considerations, the role of public awareness and education is crucial. As AI becomes more prevalent in society, individuals must be empowered to understand the ethical implications of these technologies and advocate for responsible AI practices. Public engagement can serve as a powerful force in holding developers and policymakers accountable for the ethical impact of AI, ensuring that technology serves humanity rather than exploits it. In essence, the integration of ethical principles into AI development is not merely an academic exercise but a moral imperative. By embedding ethics into the very fabric of AI systems, we can ensure that technological advancements contribute to human flourishing and social well-being. The journey

towards ethical AI is a shared responsibility—one that requires vigilance, cooperation, and a commitment to safeguarding the dignity and rights of all individuals. Only through sustained effort can we achieve a future where AI and society progress hand in hand, ensuring that innovation not only drives economic growth but also promotes justice, equality, and the common good for all.

References

1. Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, 62, 102433.
2. Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management (JDM)*, 31(2), 74-87.
3. Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9).
4. Hanna, R. (2016). Wittgenstein and Kantianism. A companion to Wittgenstein, 682-698.
5. Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, 129, 961-974.
6. Stahl, B. C. (2021). *Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies* (p. 124). Springer Nature.
7. Müller, J. (2015). Agency and Responsibility in Aristotle's Eudemian Ethics. *Phronesis*, 60(2), 206-251.
8. Scanlon Jr, T. M. (1972). Rawls' Theory of Justice. *U. Pa. L. Rev.*, 121, 1020.
9. Bowen, S. A. (2004). Expansion of ethics as the tenth generic principle of public relations excellence: A Kantian theory and model for managing ethical issues. *Journal of public relations research*, 16(1), 65-92.
10. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds and Machines*, 31(2), 239-256.
11. Huang, K., Greene, J. D., & Bazerman, M. (2019). Veil-of-ignorance reasoning favors the greater good. *Proceedings of the national academy of sciences*, 116(48), 23989-23995.
12. Nantel, J., & Weeks, W. A. (1996). Marketing ethics: is there more to it than the utilitarian approach?. *European journal of marketing*, 30(5), 9-19.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

