

Discussion on the Ethical Relationship between Humans and Artificial Intelligence: A Multi-stakeholder Governance Perspective

Yunru Yuan^{1*}

¹Rixin College, Tsinghua University, Beijing, China

Abstract. The rapid advancement of large language models (LLMs) and other artificial intelligence (AI) technologies has profoundly impacted our society. While existing governance frameworks address network and data security, the ethical governance of AI remains underdeveloped and cannot be solely reliant on past experiences. This article explores the human-AI relationship through a people-first ethical lens, focusing on the responsibilities of three key stakeholders: developers, regulators, and end-users. We first delineate the fundamental differences between AI and traditional automation systems. Then, we systematically review four critical ethical issues in AI governance—privacy and data security, algorithmic bias, responsibility attribution, and emotional and ethical concerns—grounding each in existing literature. Subsequently, we propose a multi-stakeholder interaction framework, incorporating technical and policy case studies to illustrate practical challenges and solutions. We conclude that building a secure, reliable, and controllable AI ecosystem requires the concerted efforts of all parties, with continuous public discourse defining the boundaries of our values.

1 Introduction

Since the release of ChatGPT, large language models and other artificial intelligence (AI) technologies have become deeply embedded in daily life. While well-established governance frameworks exist for regulating network security and data security, policies and laws specifically targeting the ethical dimensions of AI are still evolving and cannot be effectively applied based solely on past experience. This gap necessitates urgent and thorough discussion.

In this article, we intend to discuss how humans should ethically interact with artificial intelligence from a people-first perspective. This central question is divided into three interrelated aspects: Firstly, how should AI developers embed ethical considerations into their products? Secondly, how should regulators (including governments and organizations) establish boundaries and rules for AI development and deployment through laws, policies, and education? Lastly, how should ordinary users responsibly utilize AI products and collaborate with specialists on AI-related issues?

* Corresponding author: yuanr22@mails.tsinghua.edu.cn

It is important to clarify that this article does not engage with the long-term "singularity" problem. Instead, it focuses on the tangible ethical issues that have already emerged or are likely to emerge in the near future, based on the current state of technology where AI operates as a powerful but non-sentient tool.

The contributions of this article are threefold. Firstly, we propose a multi-stakeholder ethical framework for human-AI interaction, clearly delineating the roles and responsibilities of developers, regulators, and users. Secondly, we provide a structured review of five critical gaps in AI governance, supported by relevant literature and contemporary technical application cases. Thirdly, we emphasize a proactive, adaptive, and human-centric approach to AI governance that can keep pace with rapid technological iteration.

2 What Distinguishes AI from Traditional Machines and Automation Systems?

Understanding the unique capabilities of modern AI is fundamental to appreciating the novel ethical challenges it presents. Traditional machines and automation systems excel at executing predefined, repetitive tasks with high efficiency and precision, such as assembly line welding or formula-based calculations. They operate within closed-world assumptions and cannot handle unforeseen circumstances or learn from experience.

In contrast, contemporary AI, particularly machine learning (ML) and deep learning (DL) models, aims to simulate intelligent behavior. Its core capability stems from learning patterns from massive datasets. Through probabilistic models, statistics, and pattern recognition, AI systems can generalize to new situations, make judgments, and predictions. Crucially, many systems can continuously optimize their performance through new data encountered during use, enabling adaptation to novel scenarios. This allows AI to process unstructured data (images, speech, text) and tackle complex tasks requiring cognitive abilities, such as natural language understanding and image recognition.

This shift from deterministic automation to probabilistic learning is what necessitates a new governance paradigm. The "black box" nature of some complex models and their adaptive, generative capabilities introduce uncertainties that traditional mechanical systems did not present.

3 Critical Ethical Issues in AI Governance

Previous research has identified several persistent challenges. We consolidate these into four key areas, drawing connections to relevant scholarly discourse.

3.1 Privacy and Data Security

AI development and operation are inherently data-hungry. The scale and detail of data required for training LLMs, for instance, far exceed that of traditional software. However, users often unknowingly consent to extensive data collection—including location, interests, and habits—through service agreements. While this data can be used for personalization and optimization, it also poses unprecedented risks of abuse and leakage. The skill of data de-anonymisation is well-developed, so with the increasing usage of big data, risk to anonymity and privacy increases [1].

3.2 Algorithmic Bias and Fairness

The algorithms underpinning AI are not inherently objective. They can learn and amplify biases present in their training data, leading to discriminatory outcomes [2]. For example, AI models designed to support dermatological diagnoses may have less accuracy when encountering patients with darker skin, because they were trained predominantly on lighter skin tones [3].

Considering how to design equitable algorithms, it is crucial to collect diverse and equally represented data, so that bias in the data could be minimized [4]. It is also important to obtain real data,

3.3 Responsibility Attribution and Accountability

AI developers are unable to predict the outcomes and consequences of the system due to the complexity of its decision-making process [5]. When an AI system causes harm (e.g., an autonomous vehicle accident), assigning responsibility among developers, manufacturers, service providers, users, or the AI itself becomes a legal and ethical quagmire. This challenges the fundamental ethical principle of direct accountability and may create loopholes that allow corporations to evade responsibility.

3.4 Emotional and Ethical Concerns in Human-AI Interaction

As AI becomes more pervasive in roles like companionship (chatbots, social robots) and caregiving (medical robots), new emotional and ethical dilemmas arise. How humans form emotional connections with AI entities and how these interactions impact human relationships and societal norms are open questions that current ethical frameworks are ill-equipped to address.

4 Ethical Framework for Human-AI Interaction

We suppose that people are divided into three groups when interacting with AI: developers, regulators, and users. Each group has its own responsibility in improving human-AI interaction, and we also need the collaborative efforts of everyone to build a more trustworthy AI ecosystem.

4.1 Developers: Building Responsibility from the Ground Up

Developers bear the primary responsibility for integrating safety, ethics, and compliance into the technology itself. They must adhere to regulations and standards, proactively identifying and mitigating risks like data breaches and algorithmic bias.

For instance, it is suggested that AI practitioners should write a document to clarify how they have built the tool, so that it is easier to cooperate both within developers and with regulators [6].

When AI generates harmful content, developers must implement immediate mitigation measures, such as content filtering or blocking, and use such incidents for model retraining and optimization. Providing clear usage guidelines and effective feedback channels is also crucial for continuous improvement.

4.2 Regulators: Building Guardrails for Innovation

Regulators, including legislators and government agencies, must create a conducive environment for responsible innovation through agile and smart regulation.

Firstly, comprehensive laws are needed to provide clear rules for the development, deployment, and use of AI. It is important to foster cooperation among countries to promote harmonization in AI regulation [7]. Furthermore, the regulatory framework need to be more flexible and dynamic to accommodate the fast pace of technological advance [8].

Secondly, balancing innovation with public safety is crucial. Companies may consider that their profits outweighs social values, so regulators need to consider better actions to protect public advantages [9]. For instance, companies may exaggerate what their AI products could do to make more profits. In this situation, regulators need to work with specialists to identify the truth and make corresponding action [10].

Thirdly, promoting public education of AI is helpful for building trust to this new technology. This calls for regulators to help citizens discern AI-generated content. For example, regulators can support the development and validation of reliable AI content detection tools to bolster digital trust.

4.3 Users: Responsible Use and Vigilant Supervision

End-users are not passive consumers but active participants in the AI ecosystem. Users should proactively improve their AI literacy, which means that they should understand the technology's capabilities and limitations, and use it responsibly. They must refrain from using AI to generate harmful information or for illegal activities like fraud.

Users can also act as essential supervisors, by critically evaluating AI outputs and reporting errors or biases through provided channels, holding developers and regulators accountable and contributing to the iterative improvement of AI systems.

5 Conclusion

The development of artificial intelligence must be guided by the common values of human society. This requires a multi-stakeholder approach where developers embed ethics into technical design, regulators construct agile and informed governance frameworks, and users engage in responsible and critical interaction.

However, the pace of technological iteration vastly outstrips the speed of legal and ethical consensus-building. The persistent challenge of clearly defining responsibility when AI systems cause harm underscores the need for continued refinement of laws and technical safeguards. Developers must practice self-discipline in innovation, regulators must adopt sophisticated and adaptive regulatory tools, and users must continuously learn and adapt.

Ultimately, building a secure, reliable, and human-centric AI future relies on the sustained and collaborative efforts of all parties involved, guided by ongoing public discourse and a shared commitment to ethical principles.

References

1. K. Masters, Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. *Medical Teacher* **45**, 574–584 (2023).
<https://doi.org/10.1080/0142159X.2023.2186203>
2. C. Wang, S. Liu, H. Yang, J. Guo, Y. Wu, J. Liu, Ethical Considerations of Using ChatGPT in Health Care. *J Med Internet Res* **25**, (2019). doi: 10.2196/48009

3. P. Goktas, A. Grzybowski, Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI. *J. Clin. Med.* **14**, (2025).
<https://doi.org/10.3390/jcm14051605>
4. I. Nandutu, M. Atemkeng, P. Okouma, Integrating AI ethics in wildlife conservation AI systems in South Africa: a review, challenges, and future research agenda. *AI & Soc* **38**, 245–257 (2023). <https://doi.org/10.1007/s00146-021-01285-y>
5. A. Pant, R. Hoda, S. V. Spiegler, C. Tantithamthavorn, B. Turhan, Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. *ACM Trans. Softw. Eng. Methodol* **33**, Article 80 (2024). <https://doi.org/10.1145/3635715>
6. E. Ratti, M. Graves, A Capability Approach to AI Ethics. *American Philosophical Quarterly* **62**, 1–16 (2025). <https://doi.org/10.5406/21521123.62.1.01>
7. A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines. *Nat Mach Intell* **1**, 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>
8. Natalia Díaz-Rodríguez, Javier Del Ser, Mark Coeckelbergh, Marcos López de Prado, Enrique Herrera-Viedma, Francisco Herrera, Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion* **99**, (2023).
<https://doi.org/10.1016/j.inffus.2023.101896>
9. B. C. Stahl, D. Eke, The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management* **74**, (2024).
<https://doi.org/10.1016/j.ijinfomgt.2023.102700>
10. M. M. Khan, N. Shah, N. Shaikh, A. Thabet, T. alrabayah, S. Belkhair, Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. *International Journal of Medical Informatics* **195**, (2025).
<https://doi.org/10.1016/j.ijmedinf.2024.105780>