Facing constitutive and normative aspects of different philosophical currents when approaching AI Ethics

Fernando Pinto[1], Ana Cristina Garcia[2]

[1]Universidade Federal do Rio de Janeiro (UFRJ)
[2] Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

fernandobpinto@gmail.com, cristina.bicharra@gmail.com

Abstract. *Integrating artificial intelligence (AI), particularly machine learning, into legal decision-making is gaining prominence across various jurisdictions. While AI systems can offer efficiency and support in legal decisions, they also raise significant ethical concerns. This paper explores how AI systems, designed with underlying philosophical frameworks such as deontology and utilitarianism, can lead to conflicting outcomes when applied to legal judgments. The paper argues for a multidimensional ethical approach to AI in law that aligns with the legal philosophy of the respective jurisdiction while ensuring transparency, auditability, and fairness in decision-making processes.*

1. Introduction

Artificial intelligence (AI) across various sectors, including legal systems, quickly transforms decision-making processes. From assisting court clerks with administrative tasks to supporting complex judgments by judges, AI is increasingly used to enhance efficiency and consistency. However, as AI becomes more embedded in these processes, profound ethical concerns arise, particularly concerning the philosophical principles underpinning decision-making across different domains. These concerns extend beyond legal frameworks and touch upon many areas, such as healthcare, finance, and employment, where biased algorithms can exacerbate systemic inequalities, posing a significant challenge to societal fairness.

This paper critically examines how AI systems, influenced by deontological, utilitarian, and consequentialist ethical frameworks, can yield conflicting results across sectors. It also addresses the challenge of bias in AI, highlighting how historical prejudices encoded in data can skew outcomes, whether in courtrooms, hospitals, or workplaces. This paper strongly advocates for a multidimensional ethical approach to AI design and deployment in exploring these concerns. This approach integrates the philosophical underpinnings of various fields while ensuring transparency, accountability, and fairness, promoting more just and equitable decision-making processes, and mitigating existing disparities.

2. Ethical Frameworks in Decision-Making

Artificial intelligence (AI) has ignited extensive debate over the ethical implications of its use in critical decision-making processes. Examining the core ethical frameworks that anchor the ethical challenges posed by AI is essential. Three major ethical traditions—deontological, utilitarian, and consequentialist—provide lenses through which we can assess the morality of AI systems and their applications.

2.1. Deontological Ethics

Deontological ethics, associated with Immanuel Kant, focuses on the inherent morality of actions rather than their consequences. Kant's categorical imperative instructs individuals to act only according to universalized maxims. In decision-making processes, this framework supports the idea that certain rights and principles—such as human dignity or privacy—are inviolable, irrespective of the potential benefits of ignoring them.

In AI, the deontological framework emphasizes the need for systems that respect fundamental rights. For example, Binns [Binns 2018] argues that deontological principles are essential in ensuring that AI respects privacy and avoids harm, even when there are societal benefits to infringing on such rights.

2.2. Utilitarian Ethics

In contrast, utilitarianism, developed by philosophers such as Jeremy Bentham and John Stuart Mill, evaluates the morality of actions based on their outcomes. The "greatest happiness principle" suggests that the morally right action is the one that produces the greatest good for the most significant number of people.

AI's ability to aggregate and analyze vast amounts of data lends itself naturally to utilitarian approaches. However, this approach can conflict with individual rights, as it may justify decisions that harm minorities in pursuing the greater good. Studies such as Floridi et al. [Floridi et al. 2018] explore this tension, focusing on AI's ethical challenges when optimizing for efficiency and utility at the expense of individual freedoms.

2.3. Consequentialist Ethics

Consequentialism, a broader ethical framework that includes utilitarianism, also assesses the morality of actions based on their outcomes. However, unlike strict utilitarianism, consequentialism does not always prioritize happiness or utility as the highest good. Consequentialist approaches can, for example, prioritize minimizing harm or promoting justice, depending on the ethical goals set for the AI system. In law, consequentialist reasoning informs doctrines such as negligence, where the foreseeability of harm and the reasonableness of actions are crucial considerations.

Research has explored the ways consequentialist ethics can guide AI decision making, particularly in areas where predicting outcomes is complex. For instance, Mittelstadt et al. [Mittelstadt et al. 2016] explore how AI can be designed to account for the broader social consequences of automated decisions, particularly in high-stakes environments like healthcare, where both individual and collective harms must be considered.

The authors argue that we should assess AI systems for their efficiency and the broader, long-term impacts they may have on society.

3. Related Research on AI and Ethics

The ethical frameworks discussed above are central to much of the current discourse on AI and decision-making. Research on AI ethics by scholars like Jobin, Ienca, and Vayena [Jobin et al. 2019] outlines a comprehensive landscape of ethical concerns, including transparency, fairness, and accountability.

One study by Whittlestone et al. [Whittlestone et al. 2019] proposes a hybrid ethical approach that balances deontological rights with utilitarian outcomes in AI decision making.

4. A Path Forward: A Multidisciplinary Approach to Ethical AI

A fundamental starting point for the ethical deployment of AI is the establishment of clear and transparent ethical principles. What principles should guide AI systems in balancing competing priorities? These principles must address the tension between protecting individual rights (a deontological concern) and maximizing societal benefits (a utilitarian concern).

- Can we develop universal principles that apply across sectors?, or
- Should ethical guidelines be tailored to specific contexts?

4.1. Promoting Transparency and Explainability

Transparency and explainability are essential for building public trust in AI, but how much transparency is required to make these systems ethically acceptable? Should developers be obligated to fully disclose how their systems work, even at the risk of exposing proprietary information or security vulnerabilities?

These challenges become even more pressing with the rise of generative AI technologies, such as deepfakes and voice cloning, which create realistic but often fabricated content. As these technologies advance, they are increasingly being used for malicious purposes, like identity theft and misinformation [Shoaib et al. 2023].

• How do we balance the need for transparency with the potential harm these tools can cause?
• Should AI systems be required to disclose when content is artificially generated, especially when deepfakes and cloned voices can manipulate public opinion or deceive individuals [Seow et al. 2022]?

As generative AI continues to evolve, the need for transparency becomes even more urgent. Ensuring that AI systems are explainable, secure, and ethically sound will be vital in safeguarding individual rights and maintaining public trust in AI-driven systems ([Tsamados et al. 2021]. Ultimately, the challenge lies in balancing these competing priorities to create AI systems that are both transparent and responsible.

4.2. The Need for an Integrated Ethical Approach

While deontological and utilitarian ethics provide valuable perspectives on AI's moral challenges, more than a framework is required. A deontological focus on rights may need to account for broader social benefits, while a utilitarian approach may justify harm to individuals in the name of collective good. Therefore, an integrated approach that combines elements of both frameworks is essential.

However, what does this integrated approach look like in practice? How do we reconcile conflicting ethical principles, such as respecting individual privacy while promoting public safety? One possible solution is to develop context-specific ethical frameworks that allow flexibility in designing and deploying AI systems. For example, in healthcare, AI systems prioritize individual autonomy and patient consent, while utilitarian principles focused on sustainability take precedence in environmental management.

The integration of these ethical frameworks also raises questions about AI's role in decision-making.

• Should AI systems be allowed to make decisions autonomously, or should they serve as decision-support tools for human actors?
• If AI systems are to operate autonomously, how can we ensure that they act in ways that are both ethically justified and socially acceptable?

5. Conclusion

The quest for ethical AI development and application requires a nuanced, multidisciplinary approach incorporating insights from multiple ethical traditions. Deontological ethics emphasizes the importance of individual rights, while utilitarian ethics focuses on the broader societal good. However, neither approach alone is sufficient to navigate the complex moral terrain of AI. By developing clear ethical principles, promoting transparency, establishing accountability mechanisms, and fostering inclusive public debate, we can work toward an AI-driven future that is both ethical and equitable. Nevertheless, ongoing engagement with these questions is critical as AI technologies evolve, challenging our ethical frameworks and societal norms.

References

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In Conference on fairness, accountability and transparency, pages 149–159. PMLR.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al. (2018). Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. Minds and machines, 28:689–707.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of ai ethics guidelines. Nature machine intelligence, 1(9):389–399.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2):2053951716679679.

Seow, J. W., Lim, M. K., Phan, R. C., and Liu, J. K. (2022). A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities. Neurocomputing, 513:351–371.

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., and Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier ai, generative ai, and large ai models. In 2023 International Conference on Computer and Applications (ICCA), pages 1–7. IEEE.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., and Floridi, L. (2021). The ethics of algorithms: key problems and solutions. Ethics, governance, and policies in artificial intelligence, pages 97–123.

Whittlestone, J., Nyrup, R., Alexandrova, A., and Cave, S. (2019). The role and limits of principles in ai ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pages 195–200.Dyer, S., Martin, J. and Zulauf, J. (1995) "Motion Capture White Paper", http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html, December.