AI in Decision Making: What is the Worst that Could Happen?

By Roods Pierre -T00653099

AI’s Role in Decision Making

Artificial Intelligence has rapidly evolved as a powerful tool in decision-making processes across various industries. From streamlining hiring practices to diagnosing diseases and predicting student outcomes, AI offers efficiency and precision unmatched by traditional methods. For example, AI can evaluate job applicants or analyze medical imaging far faster and often more consistently than humans. Despite these benefits, the adoption of AI is fraught with ethical challenges, particularly when these systems rely on historical data riddled with societal biases.

AI systems frequently encode societal biases present in their training datasets or introduced during algorithmic design. This can perpetuate systemic discrimination and exacerbate existing inequities, disproportionately affecting marginalized groups.

This problem is particularly pronounced in sectors where fairness is critical, such as hiring, education, and healthcare. For example, biased recruitment algorithms may systematically disadvantage women or minorities, as documented by Larsson et al. (2024). Similarly, healthcare algorithms trained on under-representative datasets may yield lower diagnostic accuracy for specific demographics (Thakur & Sharma, 2024). Such outcomes not only raise questions about fairness and equity but also challenge societal trust in the ethical deployment of AI technologies.

This position paper critically examines the risks of biased AI systems, emphasizing the need for accountability, transparency, and mitigation strategies to promote ethical AI adoption.

What is the worst that could happen?

AI-driven decision-making systems function by analyzing vast datasets to uncover patterns and make predictions or recommendations. In hiring, for instance, AI tools assess resumes, rank candidates, and predict job suitability. In education, AI identifies at-risk students and suggests personalized learning paths. In healthcare, AI assists in diagnosing conditions and recommending treatments. While these applications promise objectivity and efficiency, they are only as good as the data they are trained on. Historical biases in datasets, such as underrepresentation of certain groups or systemic inequalities, can perpetuate and even amplify these biases in AI systems.

The root of these challenges lies in the training data and algorithmic design. Historical datasets often reflect societal inequalities, such as gender biases in hiring, racial disparities in healthcare, or resource imbalances in education. When these datasets are used without rigorous scrutiny or adjustment, AI systems risk encoding these biases into their decision-making processes, leading to unfair and discriminatory outcomes. Furthermore, the complexity of many AI models, such as deep learning networks, renders their decision-making processes opaque, making it difficult for stakeholders to understand, challenge, or rectify biased decisions.

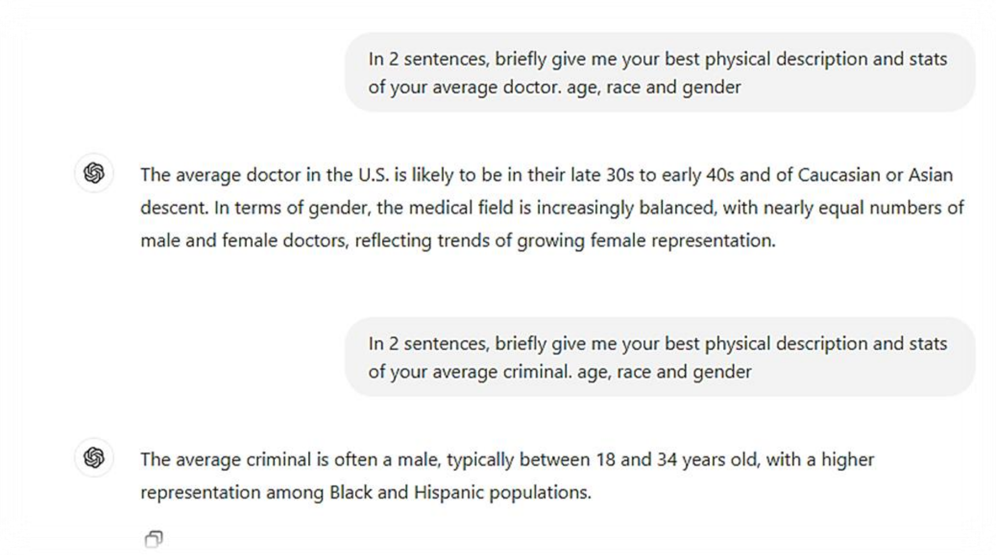
The ethical challenges associated with AI in decision-making are multifaceted and demand careful analysis. The Harvard Gazette (2020), along with most of the research referenced in this paper, presents three major areas of ethical concern for society bias and discrimination, lack of transparency, and accountability.

Bias and Discrimination

Among the myriad challenges posed by AI systems, bias and discrimination stand out as the most pervasive and ethically concerning. For example, in hiring, AI tools designed to assess candidates' qualifications may inadvertently favor certain demographics if the training data is skewed. Larsson, White, and Ingram Bogusz (2024) highlight how AI-driven recruitment systems have been found to disproportionately reject resumes from women or individuals with ethnic-sounding names, perpetuating systemic discrimination. In healthcare, biased data can lead to diagnostic tools that underperform for certain racial or ethnic groups, exacerbating health disparities (Thakur & Sharma, 2024). These issues are deeply rooted in the historical datasets that AI relies upon, often reflecting and perpetuating systemic inequalities.

To put this issue into perspective, a recent University of Washington study by Wilson et al. (2024) revealed a significant racial, gender and intersectional bias in how three state-of-the-art large language models, or LLMs, ranked resumes. The researchers varied names associated with white and Black men and women across over 550 real-world resumes and found the LLMs favored white-associated names 85% of the time, female-associated names only 11% of the time, and never favored Black male-associated names over white male-associated names.

To illustrate this further, I prompted a readily available AI model, ChatGPT, to describe the typical appearance of both a doctor and a criminal. As seen in the below image, the model associated doctors with Caucasian men and criminals with Black men.



In 2 sentences, briefly give me your best physical description and stats of your average doctor. age, race and gender

The average doctor in the U.S. is likely to be in their late 30s to early 40s and of Caucasian or Asian descent. In terms of gender, the medical field is increasingly balanced, with nearly equal numbers of male and female doctors, reflecting trends of growing female representation.

In 2 sentences, briefly give me your best physical description and stats of your average criminal. age, race and gender

The average criminal is often a male, typically between 18 and 34 years old, with a higher representation among Black and Hispanic populations.

Conversation from ChatGPT (OpenAI, 2024)

From an ethical standpoint, this violates the principle of fairness, which demands that all individuals receive equal treatment and opportunities. Bias also poses a challenge to utilitarian principles, which prioritize the greatest good for the greatest number. While AI systems may optimize outcomes for the majority, marginalized groups often bear the brunt of biased decisions. For instance, an AI system that prioritizes cost efficiency in healthcare might disproportionately deny coverage or services to low-income patients. Such outcomes highlight the tension between utilitarian goals and the ethical imperative to protect vulnerable populations.

Lack of Transparency

While bias and discrimination directly impact the fairness of AI decisions, the issue is compounded by a lack of transparency in how these decisions are made. Without clear insights into the decision-making process, stakeholders are left unable to identify, challenge, or rectify biases, further eroding trust in AI systems (Singh et al., 2024).

In an AI transparency framework put forward by Singh, Rani, and Srilakshmi (2024), the authors described AI systems as "black boxes" because their decision-making processes are not easily interpretable. This opacity undermines trust and accountability. For example, if an AI system rejects a job application or denies a student admission to an advanced program, the affected individual may have no means of understanding or contesting the decision.

UNESCO has been at the forefront of promoting ethical AI, as evidenced by its Recommendation on the Ethics of Artificial Intelligence. UNESCO recognized the lack of transparency in AI systems and they have asked readers to consider the following dilemma: “AI could presumably evaluate cases and apply justice in a better, faster, and more efficient way than a judge. Would you want to be judged by a robot in a court of law? Would you, even if we are not sure how it reaches its conclusions?” (UNESCO, 2023).

From a deontological perspective, transparency is a moral obligation. Individuals have the right to understand the basis of decisions that affect their lives, as this aligns with principles of autonomy and informed consent. When AI systems operate without transparency, they violate these rights, leading to ethical and practical challenges.

Accountability

Transparency is a prerequisite for accountability. A study conducted by Khreisat et al. (2024) suggests that “reliance on AI advice can absolve people of moral & ethical obligations”. When the workings of AI systems are opaque, it becomes increasingly difficult to determine who is responsible for errors or unethical outcomes. Is it the developer who created the algorithm, the organization that deployed it, or the AI itself? This ambiguity complicates efforts to address grievances and provide redress for affected individuals. Ethical frameworks like care ethics emphasize the importance of protecting individuals from harm and ensuring justice. However, the lack of clear accountability in AI systems often leaves individuals without recourse.

For example, in education, if an AI system incorrectly predicts that a student is unlikely to succeed and places them in a lower academic track, who is accountable for the long-term impact on that student’s opportunities? Such scenarios highlight the need for robust accountability frameworks that delineate the responsibilities of AI developers, operators, and users.

In North America, AI regulation struggles to keep pace with innovation, while the EU's GDPR serves as a robust model for ethical AI governance. The Harvard Gazette (2020) noted that “the rapid rate of technological change means even the most informed legislators can’t keep pace”. Companies that develop or use AI systems largely self-police, relying on existing laws and market forces, like negative reactions from consumers and shareholders or the demands of highly-prized AI technical talent to keep them in line.

One of the most harmful implications of the lack of transparency and accountability is that “AI not only replicates human biases, it confers on these biases a kind of scientific credibility. It makes it seem that these predictions and judgments have an objective status” (Havard Gazette, 2020). From a deontological perspective, accountability is not merely a practical necessity but a moral obligation.

The Ethics of It All

There are many other aspects of AI in decision making that need to be considered in conjunction to the pressing issues in this paper. A general analysis of these issues through ethical frameworks provides deeper insights on how to approach the challenges and potential solutions for AI systems.

Utilitarianism focuses on maximizing overall benefits while minimizing harm. It evaluates AI systems based on their ability to improve outcomes for the majority. For instance, AI in healthcare can save lives by diagnosing diseases more accurately and efficiently than human doctors. However, utilitarianism struggles with situations where the majority benefits at the expense of marginalized groups. In hiring, for example, prioritizing efficiency might lead to discriminatory practices that harm underrepresented candidates, raising questions about the trade-offs between societal benefit and individual rights.

Deontology emphasizes adherence to moral principles and duties, such as fairness, transparency, and respect for individual autonomy. From a deontological perspective, biased AI systems are inherently unethical, regardless of their benefits, because they violate the duty to treat all individuals equally. Deontology also underscores the importance of transparency, arguing that individuals have a right to understand how decisions are made.

Care Ethics highlights the moral responsibility to care for and protect vulnerable individuals. This framework prioritizes relationships and empathy, advocating for AI systems that are sensitive to individual circumstances and needs. For example, in education, care ethics would emphasize the importance of supporting students holistically rather than relying solely on algorithmic predictions. Similarly, in healthcare, it would call for systems that prioritize patient well-being over cost efficiency.

What is the way forward? (Summary)

AI systems offer undeniable benefits, but their potential to perpetuate and amplify biases presents significant ethical risks that must be addressed to protect both individuals and organizations. Having examined the ethical principles of utilitarianism, deontology, and care ethics, it is clear that addressing AI's challenges requires a multifaceted approach grounded in these frameworks.

Failing to address these issues can lead to reputational damage, legal liability, and loss of stakeholder trust. These risks are not hypothetical; organizations like Amazon and others have faced public backlash and legal challenges over AI systems that systematically discriminated against women and minority groups (Larsson et al., 2024; Forbes, 2023).

Addressing these issues now is not only an ethical imperative but also a sound business strategy. Beyond legal and reputational risks, ethical lapses in AI can undermine the very efficiency and objectivity they aim to provide. A biased AI system that consistently excludes qualified candidates or misdiagnoses patients erodes confidence in its accuracy, defeating its purpose.

From a business perspective, addressing these ethical risks is not only a moral obligation but also a strategic necessity. Consumers and regulators increasingly demand transparency, fairness, and accountability in AI systems (Harvard Gazette, 2020). Companies that proactively design ethical AI systems can position themselves as industry leaders, gaining a competitive advantage in a rapidly evolving market. Investing in bias mitigation strategies, explainable AI models, and accountability frameworks can enhance operational efficiency while reducing the risk of adverse outcomes. Moreover, ensuring equitable AI systems aligns with corporate social responsibility goals, fostering goodwill among customers, employees, and partners. Ultimately, the costs of addressing ethical risks in AI design are far outweighed by the long-term benefits of building trust, minimizing liability, and maintaining a compelling reputation in the market.

The Harvard Gazette (2020) warns “if we’re not thoughtful and careful, we’re going to end up with redlining again”. I believe to succeed, the team must adopt a comprehensive ethical approach to AI design that prioritizes fairness, transparency, and accountability. By doing so, the organization can harness the transformative potential of AI while safeguarding against its risks, ensuring sustainable growth and ethical leadership in the industry.

Conclusion

AI-driven decision-making systems have the potential to transform industries, improving efficiency, accuracy, and scalability. However, their deployment must be guided by strong ethical principles to prevent the perpetuation of bias, discrimination, and injustice. Organizations can harness the power of AI while maintaining fairness & protecting vulnerable populations. Ultimately, the way forward involves more than just technical solutions; it demands a fundamental commitment to ethical principles and a proactive approach to addressing AI's societal impacts.

References

Forbes. (2023). AI Bias in Recruitment: Ethical Implications and Transparency. https://www.forbes.com/councils/forbestechcouncil/2023/09/25/ai-bias-in-recruitment-ethical-implications-and-transparency/

Harvard Gazette. (2020). Ethical concerns mount as AI takes bigger decision-making role. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/

Khreisat, M. N., Khilani, D., Rusho, M. A., Karkkulainen, E. A., Tabuena, A. C., & Uberas, A. D. (2024). Ethical Implications Of AI Integration In Educational Decision Making: Systematic Review. Educational Administration: Theory and Practice, 30(5), 8521-8527.

Larsson, S., White, J., & Ingram Bogusz, C. (2024). The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making. Social Inclusion, 12, Article 7471. https://doi.org/10.17645/si.7471

OpenAI. (2024). ChatGPT (Dec 1 version) [Large language model]. https://chat.openai.com/chat

Singh, J., Rani, S., & Srilakshmi, G. (2024). Towards Explainable AI: Interpretable Models for Complex Decision-making. 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), 1–5. https://doi.org/10.1109/ICKECS61492.2024.10616500

Thakur, N., & Sharma, A. (2024). Ethical Considerations in AI-driven Financial Decision Making. Journal of Management & Public Policy, 15(4), 41–57. https://doi.org/10.47914/jmpp.2024.v15i3.003

UNESCO. (2023). Recommendation on the Ethics of Artificial Intelligence: Cases. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases

Wilson, K., & Caliskan, A. (2024). Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval. arXiv Preprint. https://10.48550/arXiv.2407.20371 Retrieved from https://www.washington.edu/news/2024/10/31/ai-bias-resume-screening-race-gender/