Artificial Intelligence: Unmasking the Potential Harms and Ethical Dilemmas of a Technological Revolution

Preeti Khanna*
Professor & Incharge
Department of Education
Vardhaman College, Bijnor
Email: preetikhanna711@gmail.com

Reference to this paper should be made as follows:

Received: 14.04.2025
Accepted on: 15.06.2025

Preeti Khanna

Artificial Intelligence: Unmasking the Potential Harms and Ethical Dilemmas of a Technological Revolution

Vol. XVI, Sp.Issue July1 2025
Article No.04, Pg. 020-027

Similarity Check: 02%

Online available at https://anubooks.com/special-issues?url=-jgv-vol-xvi-special-issue-july-25

DOI: *https://doi.org/10.31995/jgv.2025.v16iSI7. 004*

Abstract

Artificial Intelligence (AI) is a fixture of modern society, reshaping industries and transforming human relations, government, and the economy. Although its use in medicine, finance, and automation presents limitless potential for the common good, AI also poses very real dangers that are generally overlooked in the pursuit of innovation. Critically, AI harms in this paper are ethical harms, algorithmic bias, invasions of privacy in data, loss of job, lethal autonomous weapons, spreading disinformation, and existential hazard from Artificial General Intelligence (AGI). Being multidisciplinary in nature, the research offers real-world experiences, including Cambridge Analytica incident, algorithmic injustice in the predictive policing domain, discriminatory mechanisms of hiring people, and the threat posed by lethal autonomous weapon systems. The research also briefly touches on the governance loopholes, ethics, and recommendations for ethical deployment of AI. Concluding with a call for global cooperation, stringent regulation, and ethics-based design guidelines, this research highlights the importance of solving the darker sides of AI in aligning it with humanity's greater good.

Keywords

Artificial Intelligence, Algorithmic Bias, Data Privacy, Autonomous Weapons, AI Governance, Job Displacement, Ethics, AGI Risks.

https://doi.org/10.31995/jgv.2025.v16iSI7.004

1. Introduction

Artificial Intelligence (AI), which was previously the domain of science fiction and theoretical experimental ideas, has in the present day become a ubiquitous force that touches every facet of life. From automated customer service to autonomous vehicles and tailored medication, AI has been deployed at a whizzing, revolutionary pace in the current era. According to PwC's report (2018), AI will generate over $15.7 trillion in today's economy by 2030, and potentially be at the centre of next-generation industries globally.

Although AI can be powerful and revolutionary, early implementation brings extremely serious ethical, social, and existential threats. With more autonomous and stronger AI systems feasible, the threat of unforeseen side effects, loss of control, and socio-economic disturbances grows. Most AI applications already have achieved detrimental impacts—diverse from data exploitation, biased decision-making, job automation, to enabling mass surveillance and war.

This essay attempts to provide a comprehensive analysis of the likely harms of AI. The paper learns from case studies, exposes the ethical concerns, and criticizes existing models of regulation. The research ultimately demands harmony where AI research is regulated, ethical, and in line with the values of the people.

2. The Ubiquity and Emergence of AI in Contemporary Society

2.1 Early Breakthroughs and Advancements

The intellectual origins of AI are traced back to the 1950s with Alan Turing's "Can machines think?" John McCarthy later coined the term "Artificial Intelligence" during the 1956 Dartmouth Conference. Symbolic computation and rule-based programming were the forte of early AI systems, and among its early successes are ELIZA (1966), one of the first natural language processing programs.

2.2 Contemporary Trends and Utilizations

Machine learning, deep learning, and neural networks have evolved so much that AI systems are capable of performing complex functions. AI powers voice assistants, medical diagnosis, autonomous vehicles, and online financial trading platforms. Google, Microsoft, and IBM have spent billions of dollars on AI research, observing technology diffusion at a rapid rate (McKinsey, 2021).

The use of AI in automating operations, optimizing efficiency, and improving decision-making cannot be overemphasized. As integration intensifies, augmentation of risks involved intensifies.

3. The Multifaceted Harms of AI Technologies

3.1 Surveillance and Violations of Data Privacy Issues

3.1.1 The Cambridge Analytica Scandal

The 2018 Cambridge Analytica data scandal where over 87 million Facebook users' data was harvested without consent highlighted the dangers of AI-driven data analytics (Cadwalladr & Graham-Harrison, 2018). The firm created psychographic profiles for political campaign targeting through AI algorithms, which influenced the 2016 US Presidential election and the Brexit referendum. The scandal revealed how AI-driven data analytics can invade privacy, manipulate the masses, and undermine democratic processes.

3.1.2 AI-Powered Mass Surveillance

Other countries like China have employed AI-powered surveillance systems in their social credit system. Facial recognition cameras linked to databases at the national level monitor citizens, as behavioral control determines employment and access to facilities (Creemers, 2018). While proponents would argue there is greater security, critics note that the system represses dissidence and is a human rights infringement.

3.2 Algorithmic Discrimination and Bias

3.2.1 Racial Discrimination in Predictive Policing

COMPAS algorithm within the US criminal justice system estimates the risk of recidivism. In 2016, a ProPublica analysis discovered the system was twice as likely to incorrectly categorize African American defendants as high-risk (Angwin et al., 2016).

3.2.2 Employment Discrimination in the Hiring Process

Amazon canceled its AI recruitment program after it discovered that the program was discriminatory against female candidates. The program, which was trained on historical records of employment, tagged resumes as "women's" (Dastin, 2018). These examples show how AI can keep reinforcing systemic bias if it is not audited.

3.3 Economic Disruption and Job Displacement

3.3.1 Automation of Labor-Intensive Industries

It is also estimated by World Economic Forum (2020) that AI technology and automation can automate 85 million jobs worldwide by 2025. Customer services, logistics, and manufacturing industries are most vulnerable. Although AI technology can create jobs, it's a gigantic task of re-training and re-skilling, which need not always be feasible (Acemoglu & Restrepo, 2020).

https://doi.org/10.31995/jgv.2025.v16iSI7.004

3.3.2 The Exploitative Gig Economy and Algorithmic Management

Deliveroo and Uber, and others, exercise algorithmic control over the assignment of work, calculation of wages, and evaluation of performance. Efficient as it is, this organization denies gig workers employment rights, benefits, and union collective bargaining rights (Rosenblat, 2018).

3.4 AI-Generated Disinformation and Deepfakes

3.4.1 The Emergence of Deepfakes

Deepfake videos created by artificial intelligence have become a powerful means of spreading false news. A replica video of President Obama, for example, was uploaded to show how AI can create natural speeches (Vincent, 2018). Deepfakes can damage media credibility, destabilize politics, and enable fraud.

3.4.2 Social Media Manipulation

AI algorithms maximize sharing of the content for engagement, which will have the effect of propagating hate speech and extremist ideologies. Facebook whistleblower Frances Haugen also testified to the US Congress in 2021 that its algorithms promoted inflammatory content for increased user engagement, and that it increased political polarization and mental health crisis (Haugen, 2021).

3.5 Autonomous Weapons and the Challenge of AI to Warfare

3.5.1 Lethal Autonomous Weapon Systems (LAWS)

Autonomous AI weapons such as autonomous drones can target and identify without human intervention. In Libya, a Turkish Kargu-2 drone was reported to have struck targets autonomously, creating a chilling precedent (UN Panel of Experts on Libya, 2021). The use of LAWS is ethically and legally suspect in terms of accountability in war.

3.5.2 AI Arms Race and Global Security

Superpowers such as the US, China, and Russia are creating AI for military use, which can lead to an arms race. AI has the potential to speed up war decision-making, which leads to fewer human interventions, thus more opportunities for unintended escalation and international instability (Altmann & Sauer, 2017).

3.6 Artificial General Intelligence (AGI) Existential Risk

3.6.1 The Value Alignment and the Control Problem

AGI, general human-level AI, is dangerous to life if not contained. Bostrom (2014) cautioned that an AGI can possess goals that are not aligned with human

values with disastrous consequences. Value alignment is one of the core issues in the development of AGI.

3.6.2 Unintended Consequences and Paperclip Maximizers

The "paperclip maximizer" thought experiment illustrates how a misaligned AGI would make the world a paperclip factory if it had an improperly specified goal (Bostrom, 2014). They are thought experiments, but they illustrate the requirement for strict control and moral design.

4. Ethical, Legal, and Social Implications (ELSI) of AI Harm

4.1 Unpacking Privacy and Consent

Private data is processed in greater volumes by AI technology without consent. From voice records that intelligent speakers record to behavior monitoring that predictive code does, privacy lines are constantly breached (Zuboff, 2019).

4.2 Entrenchment of Inequality and Bias

AI disproportionately affects vulnerable groups on the basis of bias in decision-making systems and training data. AI can perpetuate amplification of system-level discrimination in criminal justice, healthcare, and finance unless altered (O'Neil, 2016).

4.3 Moral Paradoxes in Autonomous Decision-Making

There are ethical issues of self-driving cars in accident situations. They save passenger or pedestrian lives; whose lives are they? They need to be addressed by public society consensus and open policymaking (Bonnefon et al., 2016).

4.4 Legal Responsibility and Liability

Guilty attribution of harm caused by AI is multifaceted. The user, the creator, or the AI is guilty, but who? Courts will need to figure out how to balance for responsibility with AI-based decision-making (Gunkel, 2018).

5. Existing Frameworks of Rules and Governance Constraints

5.1 Efforts at a Global Regulatory Level

The EU AI Act attempts to risk-classify AI systems and place high burdens on high-risk use cases (European Commission, 2021). Enforcing this is difficult, along with country-by-country regulation.

5.2 Initiatives and Ethical Guidelines

Initiatives like the OECD AI Principles and IEEE's Ethically Aligned Design promote openness, responsibility, and equity (OECD, 2019). They are voluntary, however.

https://doi.org/10.31995/jgv.2025.v16iSI7.004

5.3 Governance Shortfalls and Issues

AI is governed through regulatory fragmentation, rival jurisdictions, and technological innovation running ahead of the law. Coordination cooperation across borders is needed in an effort to repair such cracks (Cath, 2018).

6. Proposals to Limit AI Damage

6.1 Pass Broad AI Legislation

Governments should enact legislation with binding rules on AI ranging from privacy to discrimination, workers' rights, to AI warfare. Law should make AI technology transparent, explainable, and controllable by humans (Gasser & Almeida, 2017).

6.2 Encourage Ethical AI Design and Development

Ethical AI involves diverse development teams, bias checks, and human-oriented design. Impact assessments and algorithmic interpretability should be guaranteed without the risk of harming anyone (Floridi et al., 2018).

6.3 Enhance Public Awareness and AI Literacy

Public and policymaker education around AI can assist with informed decision-making and inform citizens about defending against ill-gotten applications of AI. They need to provide transparency and engagement with stakeholders (Cave & Dignum, 2019).

6.4 Put Global Governance Mechanisms in Place

There is a worldwide regime of AI regulation, i.e., climate agreements or weapons control, that can prevent an AI arms race and ensure good development and use of AI. Multilateral action is necessary (Brundage et al., 2018).

7. Conclusion

Artificial Intelligence holds the potential to change nearly all elements of human life, including medicine and learning, transport and communications. Yet such record-breaking advantage comes hand-in-hand with immense risks endangering privacy, security, social justice, economic well-being, and even humanity itself. While AI technologies extend increasingly into ever-wider aspects of life, and become self-determining, policymakers, technologists, and society in total need to move to regulate the risks ahead with foresight and urgency. This demands application of ethical design principles, good and robust legislation, open governance, and global cooperation. Unless regulated and responsible innovation takes hold, the shadow aspects of AI—algorithmic discrimination, surveillance capitalism, work automation, autonomous weaponization—could overshadow its revolutionary potential. This

dance between technological innovation and civic responsibility demands inter- and transdisciplinarity collaboration, public activism, and unwavering commitment to human rights and protection of human dignity. It is through incorporating ethical values into the very nature of AI that we can guarantee this marvelous technology is placed into the service of mankind's collective welfare.

References

1. Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6), Pg. **2188–2244**.
2. Altmann, J., & Sauer, F. (2017). Autonomous Weapon Systems and Strategic Stability. *Survival*, 59(5), Pg. **117-142**.
3. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*.
4. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), Pg. **1573-1576.**
5. Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
6. Brundage, M., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.
7. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, Pg. **1–15**.
8. Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica. *The Guardian*.
9. Cave, S., & Dignum, V. (2019). Overcoming barriers to a beneficial AI. *Science*, 366(6468), Pg. **698-700**.
10. Cath, C. (2018). Governing artificial intelligence. *Internet Policy Review*, 7(4).
11. Creemers, R. (2018). China's Social Credit System. *SSRN Electronic Journal*.
12. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.
13. European Commission. (2021). Proposal for a Regulation on a European approach for Artificial Intelligence.
14. Floridi, L., et al. (2018). AI4People—An Ethical Framework for a Good AI Society. *Minds and Machines*, 28(4), Pg. **689-707**.

https://doi.org/10.31995/jgv.2025.v16iSI7.004

15. Gasser, U., & Almeida, V. A. (2017). A Layered Model for AI Governance. *IEEE Internet Computing*, 21(6), Pg. **58-62**.
16. Gunkel, D. J. (2018). *Robot Rights*. MIT Press.
17. Haugen, F. (2021). Testimony Before the United States Senate.
18. McKinsey. (2021). Global AI Survey: AI proves its worth, but few scale impact.
19. OECD. (2019). OECD Principles on Artificial Intelligence.
20. O'Neil, C. (2016). *Weapons of Math Destruction*. Crown Publishing.
21. Rosenblat, A. (2018). *Uberland: How Algorithms Are Rewriting the Rules of Work*. University of California Press.
22. UN Panel of Experts on Libya. (2021). Final Report.
23. Vincent, J. (2018). Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news. *The Verge*.
24. World Economic Forum. (2020). The Future of Jobs Report.
25. Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Public Affairs.