IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ETHICS OF AI - A PHILOSOPHICAL INQUIRY

Author: Mouma Jash

Affiliation: B.A & M.A Jadavpur University, B.Ed INSTITUTE OF EDUCATION (P.G.) FOR WOMEN CHANDERNAGORE

Abstract: Artificial Intelligence (AI) is rapidly changing how we function as a society and is opening up complex ethical questions that need to be historically explored philosophically. This paper looks at the ethical considerations surrounding AI from a Western and an Indian philosophical position. These investigations have to do with the elements of autonomy, accountability, and moral responsibility. AI systems are taking over decision-making in a variety of important domains (e.g., healthcare, education, policing, warfare). Concerns are raised about algorithmic bias, data privacy, surveillance, and even the diminishing of human agency. Using the normative traditions of utilitarianism, deontology, and virtue ethics, this study looks at the various ways in which machines can either make moral decisions, or respect the dignity of personhood. Further, ethical relationships from the Indian traditions that have been less internationally discussed like dharma (righteous duty) or karma (moral action) relationships based on their morals can be taken into account when investigating the position of a good or bad dehumanized world.

The hope of this paper is that while it can be shown that AI can maximize efficiency and optimize outcomes, it can still be subordinated to humans as normative ethical beings in order to limit dehumanization and loss of ethics towards moral disengagement. This will involve a study of ethical behavior and choices of AI, accountability, and transparency in how algorithmic choices are made, and the implementation of ethical guardrails around the AI process in creating systems that rely on artificial intelligence. The investigation concludes that a cross-cultural, philosophical inquiry is the only way to resolve the human ethical dilemmas that AI presents us with and ultimately serve those hard questions about humanity's relationship with technology's best interests and to serve the human condition over undermining it. This study contributes to the ongoing efforts around the world to develop responsible, ethically grounded AI.

Keywords - Ethics and Artificial Intelligence, Philosophical Inquiry, Moral Responsibility, Autonomy and Responsibility, Dharma and Karma, Human-Centered Technology.

Introduction

Artificial Intelligence (AI) is no longer the domain of science fiction and speculation—it is a transformative volume shaping the future of the present. AI is embedded across institutions and daily existence from adaptive recommendations and facial recognition, to autonomous vehicles and decision making in health and law enforcement. While the development of AI manifests great opportunity for efficiency and innovation through technology, it also brings with it pressing and problematic ethical dilemmas. Can AI make moral decisions? Who is responsible for the decision made by an artificial intelligence? Can a machine reflect human dignity, or can it re-inscribe bias and injustice while appearing to act purely and objectively?

These questions are not easily answered, but this paper seeks to tackle them philosophically in the spirit of cross-disciplinary ethical inquiry. This ethical inquiry piece will mean different things within different traditions, for example, both utilitarianism and deontology are Western ethical theories, and dharma and karma are philosophic concepts from India. Each theory, dharma and karma and various Western ethics, can have unique and similar reflections on the moral responsibility of those who design and develop AI, those who use AI, and the institutions that implement and deploy Artificial Intelligence systems.

In the context of the current AI revolution in data driven decision making alongside complicated questions around agency, autonomy, and responsibility, philosophic reflection is absolutely relevant and required.In addition, the speed at which AI technology is moving ahead is far greater than the moral and legal frameworks that are meant to govern it. In education, policing, war-making and government, AI systems are making or helping to make decisions that have been the exclusive activity of human discretion in the past. Although machines do not possess empathy, consciousness or moral discretion. Philosophy can help to delineate what moral agency, accountability, and justice mean in an increasingly machine-mediated world. This paper makes the case for urgent and sustained dialogue between technology and philosophy so that AI is understood as an instrument for empowerment rather than an active force altering important aspects of our moral lives.

Furthermore, the advancement of AI has outpaced the ethical and legal structures intended to manage AI. AI systems are making or have the ability to make decisions in education, surveillance, warfare, and governance decisions that are driven by human agency. Machines lack empathy, consciousness, and moral sense. Philosophy provides a way to understand how to think through the moral agency, responsibility, and justice in a world increasingly driven by intelligent systems. Therefore, this paper outlines the need for ongoing (and urgency) dialogue between technology and philosophy to ensure machines are not a force for moral decrease, but a source of agency.

I. Ethical Dilemmas in AI

The rapid expansion of AI in many different areas has created serious ethical dilemmas that push the limits of moral obligation and values. Algorithmic bias is perhaps one of the most alarming issues that has come from the technology era. AI is often trained through historical data, meaning they reinforce and replicate existing societal prejudices. Tools for recruitment have replicated biases against gender and race, which raises big questions for fairness and justice. If AI-generated decisions discriminate against people the question is who is responsible - the developers, the users, or the machine?

Data privacy and surveillance is a major ethical issue for many people, as data of all kinds will be captured in many public and private services. Data has to be processed without consent, which fails to uphold the ethical duty to respect persons, which in education, healthcare, and public administration where data-driven systems can limit individual human autonomy and dignity.

Another ethical dilemma of AI is autonomous decision-making in an area like warfare or criminal justice where life-or-death decisions can be made without a human being involved. Emerging technology like autonomous weapons, predictive policing, and sentencing algorithms limits the moral architecture on which human behaviour relies. Is it really in humanity's best interest to have a machine make decisions about human suffering?

Another area to consider is the role AI is playing in the manipulation of emotion through user data. Targeted advertising, recommendation algorithms, and feed algorithms can manipulate a user into feeling, thinking, and behaving in ways that may infringe on freedom of choice.

II. Ethics of AI from western philosophic views

Western moral philosophy provides some building blocks to facilitate a review of the moral implications of artificial intelligence. These perspectives can assist in making clear the moral dilemmas implicated with regard to individual dignity, autonomy, and social justice, by illuminating the challenges posed by artificial intelligence.

1. Utilitarianism: The Greatest Good for the Greatest Number

Utilitarianism classically refers to theories articulated by Jeremy Bentham and John Stuart Mill, that assess actions relative to the totality of good or a useful end result of the actions in the organization of society; for the greatest good for the greatest number of beings. In the case of AI ethics, a utilitarian perspective will support the use of artificial intelligence if it leads to the greatest utility. It would seem reasonable to support an immersion of artificial intelligence if it increases the overall happiness of the community, through potential

improvements of healthcare outcomes, optimal distribution of resources, or improved social spans and general well-being.

Utilitarianism is certainly too simplistic and not a viable moral approach to AI ethics. Utilitarian approaches allow for the possibility of appropriate outcomes while causing harm to minority groups, society's most vulnerable or individuals. This raises a concern from a moral perspective of distributive justice. For example, predictive policing systems can be based on efficient outcomes that statistically show a reduction in crime, while targeting marginalized communities and minority groups, thereby ethically causing harm to a minority community or group of people. A utilitarian approach alone cannot morally justify or rationalize the ethically compromised results of their process that violates individual rights.

2. Deontology: Duty, Moral Rules

The deontological ethics of Immanuel Kant is based on duty, rationality, and recognition of persons as ends-in-themselves. In this view, ethical actions derive from universal moral rules, rather than just from outcomes. An AI system that violates privacy or has an outcome-based manipulation of users, even for good results, is unethical because it simply treats them as means to an end.

Deontology indicates the need for AI systems to provide transparency, treat users with autonomy, and align with the movement towards explainable AI and obtaining users' consent for the use of their data.

3. Virtue Ethics: Character, Moral Flourishing

Virtue ethics, which is based on the ideas of Aristotle, promotes the moral development of character. Although machines cannot achieve the virtues of honesty, compassion, or wisdom, they may promote or hinder the development of or the use of these virtues. Machines may contribute to the decline of human judgment, humane responses, or personal responsibility that are important human virtues for ethical flourishing.

These Western perspectives suggest that AI systems must promote moral integrity, human dignity, and social justice, as well as focus on efficiency.

III. Indian Philosophical Insights on AI Ethics

Indian philosophical traditions provide an ethical framework that is a meaningful counterbalance to contemporary ideas on AI. Whereas many Western ethical frameworks are based on rationality or consequentialist ideas, Indian ethics are often focused instead on "duty" (dharma), "moral causation" (karma), and a belief in the interconnectedness of life. These ideas provide a holistic model through which to address AI ethics.

1. Dharma: Moral Duty, Social Responsibility

"Upholding dharma" is at the core of most Indian ethical theories. It refers to righteous conduct and fulfilling moral duties, in terms of the context of one's conduct, one's role, and the universal sense of order (rita). All aspects of the concept of dharma can help developers, institutions and policymakers to work from a sense of moral obligation to society, rather than merely considering personal interests or technical feasibility. Within AI, for example, preventing systems that may exploit access by and discrimination against vulnerable users and individuals or violate human dignity may reflect the dharmic principle of non-harming (ahimsa).

2. Karma: Causation and Responsibility

Karma is the doctrine that every action—mental, verbal or physical—has moral consequences. This doctrine in itself raises important questions about accountability when extended to the context of AI. Even if a machine performs an action, it is the human agents—designers, programmers or users—who enact or endorse that action, so rest with them the karmic responsibility. Discarding the sense of distance from our actions that morality based on unilateral decisions often brings, this is another way to highlight that no action is morally neutral.

3. Holism and Interdependence

Indian thinking, embodied in Vedanta and Buddhism (and many similar thought patterns) emphasizes that all beings are interdependent. This is significant because it runs counter to the lordanthropocentrism of many contemporary technologies. From this same perspective, ethical AI should be concerned not just with the welfare of humans, but with the effects of those algorithms first and foremost on the environment or secondarily on society. AI that causes ecological imbalance or human inequality would be ethically contradictory.

IV. AI, Moral Agency

The idea of moral agency—the ability to make ethical decisions and be held accountable—is the basis of every ethical system. As artificial intelligence becomes more autonomous and impactful, one of the biggest questions is: Can AI be considered a moral agent? This section will answer that question with a discussion of related philosophy.

1. What is a Moral Agency?

Moral agency is, at its core, a concept from classical philosophy that typically requires intentionality, consciousness, free will, and the ability to distinguish right and wrong. Humans are moral agents because they can think consciously, reflect, and ultimately be responsible for their actions. AI does not have consciousness, does not have emotions, and has no moral intuition. AI is capable of processing information algorithmically but does not have an understanding of the ethical implications of its actions. This fundamentally separates AI from moral agency.

2. Responsibility Gap: Who's responsible?

Despite having no moral agency, AI systems are increasingly the deciding factor in moral dilemmas that carry serious ethical implications, for example, with sentencing for criminal offences, approving credit, or deploying contra autonomous weapons. An ethical dilemma develops as there emerges what is sometimes referred to as a "responsibility gap." If an AI system does harm, is it liable for that harm—legally or ethically? The answer to both questions—philosophically and legally— is, no, the liability for the harm must be assigned to a human actor - usually the developers, deployers and/or the institutions that created or used the AI system.

The Indian notion of karma underscores this notion further - all forms of human action—irrespective of the means utilised—carry consequences for the agent. A human operator may—a machine may take an act—but the moral responsibility is always borne by human control.

3. Moral Agents vs. Moral Tools

AI should not be thought of as moral agents; it should be thought of as moral tools—extensions of human will and action. Therefore, the ethical consequences of AI are entirely contingent on how we decide to interact with and embed values into them. Therefore, philosophical endeavors should be less geared toward making machines ethical, and more down the path of ensuring humans use machines ethically.

V. To develop ethical AI

While AI has the potential to serve humanity’s collective interests, it will not do so without robust ethical standards underpinning its development. An approach that is human-centered is important; AI will remove human autonomy and fairness from our lives while diminishing human dignity if it replaces critical human thought, exploits individual data, etc. Incorporating some moral doctrines (e.g., Kant's duty-based ethics; Indian dharma) into algorithms, would also increase the likelihood that AI systems would make fair and respectful decisions as they act autonomously.

A strong level of accountability is also needed. Institutions and developers—rather than machines— should be held accountable for outcomes produced through AI. Its development requires regulatory framework and ethical scrutiny to ensure that ethical risk does not transfer to living beings in terms of not being harmed or treated unfairly.

Finally, we must continuously reflect ethically about AI due to the rapid pace of technological change. Chartering interdisciplinary conversations among philosophers, technologists, and policy makers will be essential to change standards and develop approaches that are relevant to the constantly changing context.

Overall, ethical AI must balance innovation and responsibility, and in turn be facilitated by philosophy, while implementing practical, tangible safeguards for responsible and equitable outcomes.

Conclusion

As AI fundamentally alters the contours of human life, the need for more deeper philosophical engagement with the anticipated ethical implications of AI will have very realistic urgency. This paper has identified a number of moral issues posed by AI - bias, privacy, emotional manipulation, accountability, agency - while bringing in eastern philosophical traditions (dharma, karma, or interconnectedness) and western perspectives (utilitarianism, deontology, and virtue). What is clear is that AI ethics is not a question of conforming to standards of regulation or considerations of technical efficiency, it must be rooted in moral reasoning, human dignity, and social justice.

AI is a powerful technology, but AI is still a tool, not a moral agent. The value of AI and its agency is ultimately determined by the users and creators of the technology. Values, preferences, feelings, and desires all come from us and we cannot abdicate our responsibility as moral agents through technology. Assessment of technology through philosophy, both classical and contemporary, will provide an invaluable resource for assessing these values.

It is clear that human intention and intervention is fundamental to the design process, the value sensitive programming in particular and the accountability and transparent accountability strategies. Ethical reasoning requires deep engagement of moral insight as well as collaboration with multiple fields of knowledge, this includes philosophy, technology, law and education.

Ultimately it is not simply a question of what AI can do, rather it is a question of what we, as moral beings ought to do with AI. A deep philosophical engagement with the ethics of AI can champion a perspective in which we will be able to critically appreciate the worth of human beings, as well as AI as intelligent agents who participate in moral deliberation. It is in this context that AI can honestly satisfy its desire to act in accordance with AI-specific values that are shaped primarily by people.

References

1. Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.
2. Floridi, L. (2013). The ethics of information. Oxford University Press.
3. Kant, I. (1997). Groundwork of the metaphysics of morals (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)
4. Mill, J. S. (2002). Utilitarianism (G. Sher, Ed.). Hackett Publishing. (Original work published 1863)
5. Aristotle. (2009). Nicomachean ethics (W. D. Ross, Trans.). Oxford University Press. (Original work published c. 350 B.C.E.)
6. Mohanty, J. N. (2000). Classical Indian philosophy: An introductory text. Rowman & Littlefield.
7. Dasgupta, S. (1922). A history of Indian philosophy (Vol. 1). Cambridge University Press.
8. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
9. Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), 18–21. https://doi.org/10.1109/MIS.2006.80
10. Crawford, K., & Paglen, T. (2019). Excavating AI: The politics of images in machine learning training sets. AI & Society, 36, 537–548. https://doi.org/10.1007/s00146-020-00936-9
11. Vyasa. (2006). The Bhagavad Gita (E. Easwaran, Trans.). Nilgiri Press. (Original work composed c. 2nd century BCE)