

5

Ethics of Artificial Intelligence: A Philosophical Inquiry into Human Values, Knowledge and Responsibility

Dr. Vinita Sharma*

Assistant Professor (Head), Department of Philosophy, S.S. Jain Subodh P.G. College, Jaipur.

*Corresponding Author: vinitajpr254@gmail.com

Abstract

Artificial Intelligence (AI) has moved from being a specialized technological field to becoming a major force shaping knowledge, education, governance, employment, healthcare, communication and everyday decision-making. The rapid expansion of generative AI creates not only technological opportunities but also fundamental philosophical questions: What is knowledge in an age of machine-generated information? Can an artificial system be regarded as intelligent in the same sense as a human being? Who is responsible when an algorithm causes harm? How should autonomy, dignity, justice, privacy and truth be protected in technologically mediated societies? This paper examines these questions through a philosophical framework combining contemporary AI ethics with selected insights from Indian and Western philosophy. The paper argues that AI should be understood as a powerful instrument rather than an independent moral agent, and that human responsibility cannot be transferred to an algorithm. It further proposes that Indian philosophical ideas concerning dharma, ahimsa, loka-saṅgraha, self-restraint and the pursuit of knowledge can enrich contemporary debates on responsible technology. The paper concludes that ethical AI requires not merely technical accuracy but value-sensitive design, transparency, accountability, human oversight and an education that develops critical and moral judgment.

Keywords: Artificial Intelligence, Ethics, Knowledge, Human Values, Responsibility, Indian Philosophy, Dharma, AI Ethics, Digital Transformation, Human Dignity.

Introduction

The twenty-first century is increasingly characterized by the interaction of human intelligence and computational systems. Artificial Intelligence can analyse large datasets, generate text and images, recognize patterns, recommend decisions and perform tasks that previously required substantial human cognitive effort. Such developments are often described in terms of efficiency, innovation and digital

transformation. Yet the philosophical significance of AI lies deeper than its technical capabilities. AI raises questions about the meaning of intelligence, the status of knowledge, the limits of human agency and the foundations of moral responsibility.

Research Problem

The central problem of this study is whether technological intelligence can be separated from the moral and epistemic responsibilities traditionally associated with human beings. When AI systems influence educational evaluation, recruitment, financial decisions, public administration or the circulation of information, the consequences are social and moral, not merely technical. The problem therefore requires a philosophical framework capable of connecting technological innovation with human dignity and social responsibility.

Objectives

- To examine the philosophical meaning of intelligence and knowledge in the context of AI.
- To analyse major ethical issues associated with AI, including bias, privacy, accountability, autonomy and misinformation.
- To explore relevant ideas from Indian philosophy for developing an ethically responsible approach to AI.
- To connect contemporary AI ethics with the principles of human dignity, social justice and responsible innovation.
- To suggest a human-centred philosophical framework for the responsible use of AI in education and society.

Research Methodology

The paper follows a qualitative, conceptual and philosophical method. It uses analytical and interpretative approaches to examine AI ethics through selected philosophical concepts. Relevant books, scholarly discussions, policy-oriented ethical frameworks and classical philosophical sources are used for conceptual clarification. The study is not an empirical survey; rather, it develops a normative argument concerning responsible AI.

Artificial Intelligence and the Question of Intelligence

The word intelligence ordinarily suggests capacities such as understanding, reasoning, learning, judgement and purposeful action. AI systems can reproduce some outputs associated with these capacities. However, functional performance does not automatically settle the philosophical question of whether a machine understands. A system may generate a meaningful-looking answer without possessing consciousness, lived experience or moral awareness. This distinction is important because human intelligence is not merely information processing; it is also embedded in embodiment, relationships, emotions, values and responsibility.

Alan Turing's famous question concerning machine intelligence shifted attention from an abstract definition of thinking to observable performance. John Searle, through the Chinese Room argument, later challenged the inference from syntactic manipulation to genuine understanding. These debates remain philosophically relevant in the age of generative AI. They suggest that technical competence and consciousness should not be treated as identical concepts.

AI and the Philosophy of Knowledge

Generative AI has intensified the distinction between information, belief, justification and knowledge. AI can produce fluent statements, but fluency is not equivalent to truth. A philosophical theory of knowledge asks not only whether a proposition is true, but also how it is justified and whether the knower has adequate grounds for accepting it. AI-generated content therefore requires verification and epistemic responsibility.

Indian epistemology offers an especially useful perspective. The traditions of Nyāya, Mīmāṃsā and Buddhist thought developed sophisticated discussions of *pramāṇa*—means of valid cognition. Perception, inference, testimony and other accepted means of knowledge were examined in relation to conditions of validity. In contemporary AI use, a comparable discipline of verification is required: What is the source? What is the evidence? Is the inference valid? Can the claim be independently checked?

Major Ethical Challenges of AI

- **Bias and Justice**

AI systems can reproduce or amplify biases present in data, institutional practices or design choices. If an algorithm is used for selection, evaluation or allocation, biased outputs can produce unequal opportunities. From a philosophical perspective, this is a question of distributive justice and procedural fairness. Responsible AI therefore requires attention to the social conditions from which data are produced, not only to algorithmic accuracy.

- **Privacy and Human Autonomy**

The extensive collection and analysis of personal data can create a tension between technological convenience and individual autonomy. Privacy is not merely secrecy; it is connected with the ability of persons to control aspects of their identity and life. Philosophically, excessive surveillance can weaken autonomy by changing how people behave when they know that they are continuously observed or evaluated.

- **Accountability and Responsibility**

One of the most difficult questions is: who is responsible for an AI-mediated decision? The developer, institution, user, data provider or decision-maker may each

have some role. The concept of responsibility should therefore remain human-centred. An algorithm cannot be used as a moral shield to avoid accountability.

- **Truth, Misinformation and Epistemic Harm**

Generative systems can produce plausible but inaccurate content. The philosophical danger is not only that individuals may receive false information, but that repeated exposure can weaken trust in reliable knowledge itself. The problem is therefore epistemic as well as social. Education in critical thinking and source evaluation becomes indispensable.

- **Human Dignity and the Risk of Dehumanisation**

When human beings are reduced to data points, scores or predictions, there is a risk of treating persons merely as objects of calculation. A philosophical ethics of dignity insists that human beings possess value that cannot be exhausted by their measurable characteristics. AI should assist human judgement, not erase the human subject from decisions that significantly affect life and opportunity.

Indian Philosophical Resources for AI Ethics

Indian philosophy should not be treated as a ready-made technological manual. Its value lies in providing ethical and epistemic orientations that can be interpreted for contemporary problems. Several concepts are particularly relevant.

- **Dharma and Responsible Action**

Dharma may be understood, in a broad philosophical sense, as a principle of right conduct, order and responsibility. In the context of AI, dharma can encourage developers, teachers, administrators and users to ask whether a technological action contributes to social well-being and whether it respects legitimate duties. Technology without responsibility can become an instrument of harm; technological power must therefore be accompanied by ethical discipline.

- **Ahimsa and Non-Harm**

Ahimsa provides a powerful ethical lens for evaluating technological consequences. In AI governance, non-harm should include physical, psychological, social, economic and epistemic harms. The principle can therefore support impact assessment before deploying systems in sensitive domains.

- **Loka-saṅgraha and Social Welfare**

The Bhagavad Gītā's idea of loka-saṅgraha, broadly associated with the maintenance and welfare of the social order, offers a useful principle for responsible innovation. AI should not be evaluated solely by private profit or technical novelty. Its wider effects on vulnerable communities, employment, education and public welfare must also be considered.

- **Self-Restraint and Ethical Use**

The ethical problem of AI is partly a problem of human desire and power. Even a technically safe system can be misused by individuals or institutions. Traditions of self-discipline and restraint in Indian philosophy remind us that not everything technically possible is ethically desirable. The ability to refrain from harmful use is itself an ethical achievement.

Western Philosophical Perspectives

Kantian ethics emphasizes the intrinsic worth of persons and the prohibition against treating persons merely as means. This provides a strong foundation for rejecting manipulative AI systems and purely instrumental treatment of human beings. Utilitarian approaches, by contrast, draw attention to consequences and overall welfare; they can help assess whether the benefits of AI outweigh foreseeable harms. Virtue ethics adds another dimension by asking what kind of human character and social practices technology cultivates.

These approaches should not be treated as mutually exclusive. A robust philosophy of AI can combine dignity, welfare, justice, responsibility and virtue. Such pluralism is especially useful because AI affects diverse social contexts and cannot be ethically evaluated by a single principle alone.

AI in Higher Education: A Philosophical Perspective

Higher education is one of the most important fields affected by generative AI. AI can support lesson planning, translation, accessibility, formative feedback, research assistance and personalized learning. At the same time, uncritical use can weaken originality, academic integrity and students' capacity for independent reasoning. The central educational question is therefore not whether AI should be banned or blindly adopted, but how it can be integrated without replacing intellectual formation.

Philosophy has a distinctive role here. Students should learn to distinguish generated answers from justified knowledge, persuasive language from valid argument, and convenience from intellectual growth. AI literacy should therefore include epistemic literacy and ethical literacy. A student who can use an AI tool but cannot evaluate its output is technologically equipped but intellectually vulnerable.

Towards a Human-Centred Ethical Framework

Based on the preceding analysis, a responsible philosophy of AI may be organized around seven principles:

- Human dignity: AI must respect the intrinsic worth of persons.
- Human oversight: Significant decisions should remain subject to meaningful human review.

- Transparency: Users should be informed about important AI-mediated processes.
- Accountability: Institutions and individuals must remain answerable for consequential decisions.
- Justice and inclusion: AI systems should be assessed for discriminatory or exclusionary effects.
- Epistemic responsibility: AI-generated claims should be verified before being treated as knowledge.
- Social and ecological responsibility: Technological benefits should be evaluated in relation to wider social and environmental consequences.

Discussion

The philosophical challenge posed by AI is ultimately a challenge concerning the direction of human civilisation. The question is not simply what machines can do, but what humans should choose to do with machines. A technology-centred approach tends to define progress through speed, scale and efficiency. A philosophical approach adds another measure: whether technological progress enhances human flourishing, justice, freedom and responsible coexistence.

This is why ethics cannot be added as an afterthought once a system has already been designed. Ethical reflection should be incorporated at the stages of problem definition, data selection, system design, testing, deployment and evaluation. Philosophy can contribute by asking normative questions that technical metrics alone cannot answer.

Findings

- AI can simulate or perform forms of intelligent behaviour without establishing that it possesses human-like consciousness or understanding.
- AI-generated information should not be equated automatically with knowledge; verification and justification remain essential.
- Bias in AI is a question of justice and social responsibility, not merely a technical defect.
- Human responsibility should not be transferred to algorithms; meaningful accountability must remain with human actors and institutions.
- Indian philosophical concepts such as dharma, ahimsa, loka-saṅgraha and self-restraint can provide complementary ethical resources for contemporary AI discourse.
- In higher education, AI should augment teaching and research while preserving independent thinking, academic integrity and human judgement.

- Ethical AI requires interdisciplinary cooperation among philosophy, computer science, law, social sciences, education and public policy.

Conclusion

Artificial Intelligence is one of the defining developments of contemporary research and society. Its philosophical importance arises from the fact that it transforms not only what humans can accomplish, but also how humans understand knowledge, agency, responsibility and themselves. The most serious risk is not that machines will simply become more capable, but that human beings may become less reflective about the purposes for which those capabilities are used.

A philosophically responsible approach therefore requires a shift from the question “Can AI do this?” to the deeper questions “Should AI do this?”, “Under what conditions?”, “Who is accountable?”, and “What kind of society will this create?” Indian and Western philosophical traditions together show that technological power must be guided by ethical reason. AI can become a force for educational improvement, knowledge exchange, inclusion and social welfare when human dignity, truth, justice and responsibility remain at its centre. The future of AI should therefore be understood not as a choice between technology and humanity, but as a project of making technology serve humane and responsible ends.

Suggestions

- Universities should develop AI ethics and AI literacy modules for students and teachers.
- Institutions should establish clear policies on disclosure and responsible use of generative AI in academic work.
- High-stakes decisions should include meaningful human review and an avenue for appeal.
- AI systems used in education and public services should undergo periodic bias, privacy and impact assessments.
- Philosophy departments should collaborate with technology, law, education and social science departments on interdisciplinary AI research.
- Research and innovation should be guided by human welfare, social justice and ecological responsibility rather than technological novelty alone.

References

1. Aristotle. (2000). *Nicomachean ethics* (R. Crisp, Trans.). Cambridge University Press.
2. Floridi, L. (2013). *The ethics of information*. Oxford University Press.
3. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).

4. Gandhi, M. K. (2009). The collected works of Mahatma Gandhi. Publications Division, Government of India.
5. International Labour Organization. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. ILO.
6. Kant, I. (1997). Groundwork of the metaphysics of morals (M. Gregor, Trans.). Cambridge University Press.
7. Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.
8. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424.
9. Sen, A. (2009). The idea of justice. Harvard University Press.
10. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.
11. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO.
12. Bhagavad Gītā. (Various editions). Bhagavad Gītā.
13. Vātsyāyana. (Various editions). Nyāya Bhāṣya.
14. Radhakrishnan, S. (1923). Indian philosophy (Vols. 1–2). George Allen & Unwin.
15. Hiriyanna, M. (1993). Outlines of Indian philosophy. Motilal Banarsidass.

