Artificial Intelligence (AI) and Buddhist Thought: An Analysis of Moral Limits

Kaumada Karunathilaka[1]

Introduction

The rapid evolution of Artificial Intelligence (AI) shapes much of contemporary societal infrastructure, altering communication patterns, organizational operations, and decision-making processes. As automated systems become embedded within critical sectors such as healthcare, legal adjudication, and national defense, the need to define ethical parameters becomes increasingly pressing. Existing frameworks in technology ethics predominantly rely on Western philosophical models, primarily utilitarianism and Kantian deontology. While these traditions offer valuable frameworks for regulatory policy and risk assessment, they tend to focus on external outcomes and rule compliance, often giving less emphasis to internal psychological states and intentionality. Within this context, Buddhist philosophy provides an alternative, mind-centered perspective for evaluating the ethical implications of technological automation.

The primary inquiry of this research addresses the ethical and conceptual issues surrounding the deployment of automated systems that simulate human reasoning without subjective awareness. Contemporary computer science prioritizes algorithmic performance and data processing efficiency, often operating independently of traditional moral philosophy. Analyzed within Buddhist psychology (*Abhidhamma*), this structural gap presents two key analytical concerns: the conceptual error of attributing moral agency to artificial systems, and the practical risks posed to human well-being. Although current AI architectures generate highly sophisticated outputs, Buddhist doctrine maintains that morally significant action (*kamma*) depends on conscious intent (*cetanā*) and subjective awareness (*viññāṇa*). Because AI mechanisms operate as physical processes (*rūpa*), they lack the mental factors required for moral intention. Consequently, moral responsibility remains anchored to the human actors who design, deploy, and utilize these technologies.

Additionally, the unregulated application of predictive and generative systems introduces specific ethical concerns. Automated tools deployed for synthetic content generation or systematic manipulation conflict directly with traditional ethical guidelines, such as Right Speech (*sammā-vācā*) and Right

1 Faculty of Arts, University of Peradeniya, Peradeniya, Sri Lanka.
nk.karunathilaka2003@gmail.com

Livelihood (*sammā-ājīva*). This study explores how early Buddhist psychological and ethical concepts can inform objective limits on AI development, examining ways to align technology with the reduction of human suffering (*dukkha*).

Objectives

The primary objectives of this study are:

- To analyze the conceptual differences between human consciousness (*viññāṇa*) and AI data processing through the framework of Buddhist psychology (*Abhidhamma*).
- To examine the ethical limits of autonomous machine actions using the core Buddhist principles of intention (*cetanā*) and moral cause-and-effect (*kamma*).
- To propose a practical governance framework based on the Five Precepts (*pañcasīla*) and Right Livelihood (*sammā-ājīva*) to guide ethical AI development and policy-making.

Research Methodology

This study uses a qualitative, descriptive, and analytical research methodology, as it focuses on conceptual and philosophical analysis rather than empirical data. Since the research interprets traditional Buddhist doctrines and applies them to modern technological issues, a text-based analytical approach is most appropriate.

The research design involves two main sources of data:

- **Primary Sources:** The core concepts of consciousness, intention, and ethics are drawn from early Pali Buddhist scriptures. Texts from the Sutta Piṭaka (such as the *Aṅguttara Nikāya* for discussions on *kamma* and *pañcasīla*) and key references from the *Abhidhamma Piṭaka* are selected based on their explicit focus on mental factors (*cetasika*) and ethical action.
- **Secondary Sources:** Contemporary academic literature is used to examine current AI capabilities, limitations, and ethical issues. Literature selection focuses on peer-reviewed journal articles, conference papers on technology ethics, and recent books exploring the intersection of computer science, philosophy, and mind studies.

The collected information is analyzed using comparative and textual interpretation methods. Current technological concepts, such as machine learning and autonomous decision-making, are critically examined against the psychological and ethical principles of Buddhist philosophy. Through this conceptual comparison, the study identifies key areas where AI applications encounter moral limits and proposes practical ethical boundaries.

Results and Findings

Based on the analysis of Buddhist texts and contemporary AI ethics literature, the study identifies three primary findings regarding the moral limits of artificial intelligence:

The Distinction Between Machine Data Processing and Intention (*Cetanā*)

Within the framework of Buddhist psychology (*Abhidhamma*), an action carries moral weight (*kamma*) only when guided by intention (*cetanā*). The findings indicate that while AI models can simulate complex decision-making through algorithms, they consist strictly of physical and computational processes (*rūpa*) lacking a conscious mind (*citta*) and mental factors (*cetasika*). Within this philosophical framework, AI systems cannot form wholesome (*kusala*) or unwholesome (*akusala*) intentions. Consequently, AI cannot be classified as an autonomous moral agent. Instead, moral responsibility remains with the human developers, operators, and institutions that design and deploy these systems.

Application of the Five Precepts (*Pañcasīla*) to AI Governance

The study finds that the principles of the *Pañcasīla* provide a clear normative framework for identifying ethical risks in AI deployment:

- **The First Precept (*Pāṇātipātā Veramaṇī*):** The deployment of AI in autonomous weapons systems, where algorithms make life-and-death decisions, conflicts with the core Buddhist value of non-harming (*ahiṃsā*). In practice, this principle supports international policy calls for human oversight in military applications.

- **The Fourth Precept (*Musāvādā Veramaṇī*):** The use of AI to generate deepfakes and automated misinformation represents a modern form of false speech (*musāvāda*). Applying this precept suggests the need for practical governance measures, such as mandatory AI content labeling and transparency standards.

Implications for Right Livelihood (*Sammā-ājīva*) and Mindfulness (Sati)

The findings highlight the broader social and psychological impacts of unregulated automation:

- **Right Livelihood (*Sammā-ājīva*):** Deploying AI solely for cost reduction without considering rapid labor displacement can cause social and economic instability. Aligning AI adoption with *sammā-ājīva* requires corporate policies that prioritize human well-being and responsible transition strategies.

- **Mindfulness (*Sati*):** Over-reliance on predictive algorithms for everyday choices may reduce personal reflection and self-awareness. Sustained dependence on automated recommendations risks weakening active mindfulness (*sati*), emphasizing the need for conscious, intentional engagement with technology.

Overall, the findings suggest that moral limits must be established externally through human ethics and policy, ensuring that AI remains a tool that supports human well-being.

Discussion

The findings of this study highlight a conceptual distinction between "intelligence" in modern technology and "consciousness" (*viññāṇa*) in Buddhist psychology (*Abhidhamma*).

Contemporary AI development often operates on the assumption that complex decision-making can be fully replicated through advanced computing. However, analyzed through a Buddhist framework, computational systems operate primarily within physical processes (*rūpa*). While machine learning models process data in ways that appear similar to perception (*saññā*) or mental formations (*saṅkhāra*), Buddhist doctrine views consciousness (*viññāṇa*) as an essential condition for subjective experience and intentionality (*cetanā*). Within this specific philosophical framework, an AI system functions as a tool reflecting human inputs rather than an independent conscious entity.

This distinction is central to understanding moral responsibility. In contemporary AI ethics, some scholars explore "machine ethics" the idea of programming moral rules directly into algorithms. However, from a Buddhist perspective, because morality requires intentionality (*cetanā*), a computational system cannot perform morally wholesome (*kusala*) or unwholesome (akusala) acts on its own. For instance, when an autonomous system causes harm, the moral responsibility (*kamma*) remains tied to the intentions such as greed (*lobha*), ill-

will (*dosa*) or delusion (*moha*) of the humans who designed, authorized, and deployed the technology.

Therefore, this analysis shifts the ethical focus from the machine back to human decision-makers. Rather than attempting to create morally autonomous machines, the immediate priority is to establish human-enforced boundaries on technology. Applying the *Pañcasīla* as design principles helps prevent technology from being used to inflict harm or manipulate truth. Additionally, integrating *sammā-ājīva* ensures that AI deployment considers worker well-being alongside economic goals. Ultimately, within the framework of Buddhist thought, AI is best understood as an instrument that requires human mindfulness (sati) and ethical responsibility to ensure it serves the common good.

Conclusion

This study demonstrates that Artificial Intelligence processes data through mechanisms distinct from human moral agency. Analyzed within the framework of Buddhist psychology (*Abhidhamma*), computational capacity is not equivalent to subjective consciousness (*viññāṇa*). Because algorithmic architectures function without mental states (*citta*) or intentionality (*cetanā*), they lack intrinsic moral responsibility and wisdom (*paññā*). Consequently, efforts to embed autonomous systems with inherent moral agency remain conceptually limited within this doctrinal tradition.

Accordingly, the establishment and enforcement of ethical boundaries remain entirely human responsibilities. Challenges arising from AI integration reflect human intent, often shaped by unwholesome psychological factors such as greed (*lobha*) and delusion (*moha*). By incorporating principles from the Five Precepts (*pañcasīla*) and Right Livelihood (*sammā-ājīva*) into technology policy, regulatory governance can establish objective limits on automation. This ensures that AI functions as a human-directed instrument designed to mitigate suffering (dukkha) and protect societal well-being without compromising human judgment or mindfulness (sati).

Keywords - Artificial Intelligence Ethics, Buddhist Psychology, Consciousness, Technology Ethics, Cetanā, Moral Limits, Kamma

References

Bodhi, B. (ed.) (2000). *A Comprehensive Manual of Abhidhamma: The Abhidhammattha Sangaha of Ācariya Anuruddha.* Kandy: Buddhist Publication Society.

Bodhi, Bhikkhu. (trans.) (2012). *The Numerical Discourses of the Buddha: A Translation of the Aṅguttara Nikāya.* Boston: Wisdom Publications.

Cooray, N. and Silva, P. (2022). '*Buddhist Perspectives on Artificial Intelligence and Autonomous Systems',* Journal of Buddhist Ethics, 29, pp. 45–78.

Gunasiri, Ven. M. (2022). *Abhidhamma and the machine: A Buddhist psychological critique of artificial intelligence.* Journal of Buddhist Studies and Contemporary Society, 14(2), 45–59.

Hershock, P. D. (2021). *Artificial intelligence and the technology of consciousness:* A Buddhist perspective. Routledge.

Hongladarom, S. (2020). *The ethics of AI and robotics:* A Buddhist viewpoint. Springer.

Karunadasa, Y. (2018). *The Buddhist analysis of matter.* Buddhist Publication Society.