Front Cover

Buddhist Mind for AI Machines

Editors:

CHHEM Kieth Rethy
CHEAB Sambath
ROS Sayumphu
CHHEM Siriwat

CHHEM Kieth Rethy, CHEAB Sambath, ROS Sayumphu, and CHHEM Siriwat

Buddhist Mind for AI Machines

ISBN: 978-9924-9993-3-1

Cover design: Chea Singhtararith

"This image, with its serene blue backdrop and small lighted dots forming a luminous Buddha beneath a lush forest canopy, symbolizes awakened consciousness rooted in wisdom and interdependence. The streaming light through the leaves evokes our shared future, reminding us that individuals and nations are interconnected. It reflects the book's core message: ethical, mindful AI must arise from collective awareness and respect for our interdependence."

Edited by CHHEM Kieth Rethy, CHEAB Sambath, ROS Sayumphu, and CHHEM Siriwat

Printed and bound in Phnom Penh, Cambodia

Published by CamTech University Press, Cambodia University of Technology and Science (CamTech), CamTech Street, Sangkat Prek Tasek, Khan Chroy Chongvar, Phnom Penh 121003, Cambodia.

CamTech University Press

FSC www.fsc.org FSC® A000533 The mark of responsible forestry

For information on all CamTech University Publication

Visit our website: https://camtech.edu.kh/publications/

Content

Preface

The inspiration for "Buddhist Mind for AI Machines" arose from growing concerns about the ethical risks posed by artificial intelligence, an invention that, despite its name, is neither truly intelligent nor wise. After decades of development in specialized domains such as defense and medicine, AI experienced a sudden and sweeping expansion. Within just two years, generative AI spread across the globe, adopted by hundreds of millions of users largely unprepared for its far-reaching implications. This rapid proliferation caught policymakers off guard. Yet now that the genie is out of the bottle, humanity must confront AI's dual nature: a force capable of tremendous good but equally potent in its potential for harm.

How can we harness the transformative power of AI while safeguarding our shared humanity from its unintended consequences? How do we ensure that technologies created by humans do not erode the very values that define us - security, dignity, and ethical integrity? The answers lie not in engineering alone but in the cultivation of robust moral frameworks and high ethical standards.

Inspired by the pioneering vision of Dr. Kol Pheng, founding President of Paññāsāstra University, whose mission is rooted in Buddhist principles, and the ethical guidance of Samdech Preah Oudom Vongsa Dr. Kou Sopheap, then Dean of the Faculty of Arts, Letters and Humanity, many of us sought to reconnect with our cultural and spiritual foundations during the global disruption of COVID-19. At that time, we requested an online course introducing the fundamentals of Buddhism, such as the Four Noble Truths and the Noble Eightfold Path. For many of us trained in science and engineering in Western institutions, this course served as a vital reawakening. Armed with these essential teachings, we began to explore deeper philosophical questions, especially those prompted by the rise of artificial intelligence.

AI, now embedded in nearly every stratum of society, challenges us to seek wisdom beyond Western philosophical traditions. While students embrace AI unquestioningly and their professors struggle to understand how to govern their classrooms, policymakers engage in heated debates over who should regulate this dual-use technology. Meanwhile, engineers and technologists often assert proprietary control over its development. This rising tide of technological determinism is dangerous. AI is not merely about computation or control. It confronts us with profound questions about the soul of our societies, the ethical roots of our actions, and the cultural identity of our nations.

This is where philosophy must intervene. While much of the global discourse on AI ethics has been dominated by Western philosophical traditions, wisdom is not the exclusive province of the West. Among the more quietly resonant yet often overlooked traditions is the Buddhist perspective, which offers profound insights into interdependence, non-self, and the cultivation of ethical awareness.

Cambodia, where Buddhism is enshrined in the constitution, the national motto, and in the hearts of its people, offers a uniquely fitting context for such an inquiry. With this in mind, the Center for Buddhist AI Ethics at CamTech University was established. In partnership with the AI Forum, we launched an online publication series titled "AI Insight." Through this initiative, a diverse team of writers, reviewers, and editors collaborated to explore the intersection of Buddhist thought and emerging technologies.

The Center for Buddhist AI Ethics at CamTech University stemmed from the pioneering work of the AI Forum, founded in 2018 by Chhem Siriwat while he was a graduate student in AI Technology at the University of Bologna and an intern at the Cambodia Development Resource Institute (CDRI). With the strong support of Dr. Hul Seingheng-now Under Secretary of State at the Ministry of Industry, Science, Technology and Innovation-and Dr. Sothy Khieng, now President of CamTech University, the AI Forum became the first national platform for open debate on artificial intelligence in Cambodia. Its early mission focused on exploring the country's emerging AI landscape, examining the societal impacts of digital technology, and strengthening industry-university linkages to drive innovation. These foundational efforts laid the intellectual and institutional groundwork for the creation of a Buddhist-informed ethical space for AI.

This booklet, "Buddhist Mind for AI Machines," is a result of that collective effort. As co-editors, Dr. Chhem Kieth Rethy, Dr. Cheab Sambath, Ros Sayumphu, and Chhem Siriwat, we have compiled thirteen papers originally published in "AI Insight," now presented in a structured format as a single volume. Dr. Chhem Kieth Rethy and Chhem Siriwat co-designed the intellectual architecture of the volume. Although Chhem Siriwat did not contribute chapters himself, his strategic oversight as Senior Editor-in-Chief of AI Insight, the journal where all thirteen papers were first published, ensured scholarly rigor and thematic coherence throughout. His involvement in reviewing, editing, and shaping the transformation of these articles into a unified anthology is reflected in his role as co-editor.

Equally vital were the contributions of Ros Sayumphu, who supported the entire production process with consistency and care, and Dr. Cheab Sambath, who undertook the painstaking task of editing every chapter with diligence and precision. This volume is as much a product of collective scholarship as it is of collaborative dedication. Drawing from Theravada and Mahayana traditions, the chapters examine foundational concepts such as anatta (non-self), sunyata (emptiness), nama-rupa (mind-body), cetana (intention), and paticcasamuppada (dependent origination), and their relevance to AI design, ethics, and policy.

Rather than offering fixed prescriptions, Buddhist philosophy invites a reorientation of the questions themselves, challenging static notions of identity, intelligence, and autonomy. The inclusion of catuskoti (fourfold logic) and non-dual epistemologies enables deeper engagement with the limits of binary reasoning, which underpins much of algorithmic logic.

This anthology brings together a selection of articles, both previously and recently published online in AI Insight, the official publication of the AI Forum.

This anthology does not aim to formulate a definitive Buddhist ethic for AI. Instead, it proposes a reflective, dialogical space where Buddhist principles, particularly epistemic humility, interdependence, and ethical intentionality, can inspire a more compassionate and culturally grounded approach to technological innovation. It aims to foster a broader, more pluralistic rethinking of AI-one that respects philosophical traditions often overlooked in global tech discourse. This spirit is reflected in the origins of three chapters in this volume, first presented at major international gatherings held in historically and philosophically significant venues: the Academy of Plato in Athens, the Pontifical Academy of Social Sciences at the Vatican, and the Confucius Institute in Qufu, China.

This book is intended for a wide audience, including scholars of philosophy and AI, engineers and technologists, policymakers engaged in digital governance, and practitioners of Buddhist ethics who are interested in the ethical implications of emerging technologies.

To guide the reader, the volume is organized into three thematic sections—Ontology (Being), Epistemology (Knowing), and Ethics (Acting)—reflecting a holistic Buddhist approach to AI that begins with foundational views of existence, proceeds to modes of knowing, and culminates in ethical conduct.

Section I: Ontology – Being

This section examines the nature of existence and consciousness through a Buddhist lens, challenging the ontological assumptions embedded in AI design.

1. The Buddha as Proto-Neuroscientist and Forerunner of AI Thought explores how early Buddhist understandings of the mind resonate with modern cognitive science and AI models.
2. Mahayana Buddhist Perspective on AI presents a non-dualistic framework for understanding intelligence and interdependence.
3. Theravada Buddhism: Mindfulness for the Ethical Advancement of AI argues for the integration of mindfulness as both a cognitive and ethical tool in AI development.
4. Buddhist Ethics for AI: Theravada, Mahayana, and Vajrayana Perspectives offers a comparative analysis of how the three major Buddhist traditions interpret technology and moral responsibility.
5. Namarupa in Buddhist Thought and Human-Centered AI reconsiders the classical Buddhist categories of name-and-form as a way to rethink data structures and machine cognition.

6. Dependent Origination and Systems Thinking: A Buddhist Logic for Explainable AI links traditional Buddhist causal reasoning to modern systems-based AI transparency.

Section II: Epistemology – Knowing

This section challenges dominant knowledge systems in AI by introducing Buddhist epistemology, logic, and humility.

7. Tetralemma vs. Dilemma: Integrating Eastern Logic into AI Reasoning contrasts binary Western logic with the richness of Eastern alternatives.
8. Catuskoti and Non-Binary Reasoning in Artificial Intelligence explores how the fourfold logic disrupts algorithmic binaries and introduces epistemic flexibility.
9. Sunyata and Epistemic Humility in AI Design centers the concept of emptiness as a call for non-attachment to fixed models and outcomes.
10. Anatta and the Myth of AI Autonomy: Towards a Processual and Relational Ethic for Artificial Intelligence critiques the illusion of self in AI systems, drawing parallels with the Buddhist doctrine of non-self.

Section III: Ethics – Acting

This final section translates Buddhist ethical principles into AI design, governance, and practice.

11. Can AI Be Morally Responsible from the Mahayana Buddhist Lens? probes whether AI systems can be moral agents or if moral responsibility remains human.
12. The Idea of Compassionate AI: Embodying Buddhist Ethics in Technology proposes that compassion, not just efficiency, should guide AI development.
13. Abhidhamma-Inspired AI: Centering Cetana (Intention) with Theravada Ethical Insights in Technology explores how volition, a

key concept in Buddhist ethics, can reframe discussions on machine intention and responsibility.

Ultimately, “Buddhist Mind for AI Machines” aspires to stimulate further scholarship at the intersection of Buddhist thought and emerging technologies, reaffirming that traditions rooted in introspection, ethical responsibility, and non-harming have much to offer in shaping a wiser future for AI. A human-centered AI. By bridging traditional wisdom with modern innovation, it offers insights not only for ethical AI governance and education but also for culturally informed AI design practices that respect both human dignity and interdependence.

About the Editors

Professor CHHEM Kieth Rethy

Distinguished Professor of AI and Humanities, CamTech University | Founder, Center for Buddhist AI Ethics

Professor CHHEM Kieth Rethy is a medical doctor, historian, and educator, currently serving as Distinguished Professor of AI and Humanities at CamTech University. He is the founder of the Center for Buddhist AI Ethics, where he advances a humanistic approach to technology grounded in compassion, wisdom, and interdependence. A pioneer in the application of digital technology and AI to medicine and radiology since the 1980s, he has shaped global conversations on technology, ethics, and health.

He was a United Nations diplomat at the International Atomic Energy Agency (IAEA), where he actively promoted the peaceful use of nuclear technology worldwide. For his contributions to mitigating the consequences of the Fukushima nuclear disaster, using Science and Technology Studies (STS) as a guiding framework, he received the Dr. Nagai Memorial Peace Award.

Professor Chhem has taught at leading universities in Canada, Singapore, Austria, and Japan, and has been invited as a keynote speaker to more than 80 countries. He has published over 200 scientific articles and co-authored 25 books. Holding doctorates in medicine, education, and history, his interdisciplinary work integrates Buddhist ethics with global AI governance. His PhD dissertation in history focused on Buddhist medicine during the reign of Jayavarman VII at Angkor in the late 12th century CE. He emphasizes equitable and culturally rooted technology innovation, including AI, for the Global South.

Professor Chhem currently serves as Senior Minister of the Royal Government of Cambodia and chairs the Economic, Social, and Cultural Council. He also chairs the Advisory Board of the National Council of Science, Technology and Innovation.

CHEAB Sambath

Medical Doctor, Learning & Teaching Manager, University of Puthisastra | PhD Student, CamTech University

CHEAB Sambath is a medical doctor, Learning and Teaching Manager at the University of Puthisastra, Lecturer in Family Medicine at the University of Health Sciences, and a PhD student at CamTech University. Trained in medicine, psychology, and science and technology studies, he integrates Buddhist Abhidhamma ethics with contemporary issues in medical education and AI. He is a core member of MEDPro (Medical Ethics, Deontology and Professionalism), a national group that promotes and teaches ethical principles to thousands of medical students and doctors across Cambodia. Sambath has presented at international platforms including the Comparative Education Society of Cambodia (CESCAM) and Peking University. His work promotes compassionate, human-centered AI development for the Global South, grounded in non-Western logic and Theravāda Buddhist principles.

CHHEM Siriwat

Co-founder and President, AI Forum | Former Founding Director, Center for Inclusive Digital Economy, the Asian Vision Institute| Advisor to Bower Group Asia (BGA)

Siriwat Chhem is a leading voice at the intersection of AI, ethics, and policy in the Global South. With interdisciplinary training in chemistry, physics, environmental science, digital technology management, and diplomacy, he is currently completing a Ph.D. on global AI leadership. Siriwat co-founded the AI Forum and previously led Cambodia's Center for Inclusive Digital Economy at the Asian Vision Institute, where he directed over 80 policy publications and edited two books: *Cambodia in Cyberspace* and *Cambodia's Emergent Cyberdiplomacy*. He is a Junior Fellow of the Academy of Engineering and Technology for the Developing World. His work emphasizes the integration of Aristotle, Siddharta Gautama, and Confucius' wisdom in the global governance of AI. As an advisor to national and regional institutions and a strategist in cyberdiplomacy, he champions human-centered, culturally inclusive approaches to emerging technologies.

ROS Sayumphu

Convener, AI Forum | PhD Student in Digital Archaeology and Ethics of LIDAR Data, CamTech University

Ros Sayumphu is Secretary General of the AI Forum and a PhD student in Digital Archaeology and the Ethics of LIDAR Data at CamTech University. He holds two master's degrees: an LLM in International Law from the University of Sussex and an MA in Security, Conflict, and International Development from the University of Leicester. He also holds two bachelor's degrees: a BA in International Studies from the Royal University of Phnom Penh and a Bachelor of Law from the Royal University of Law and Economics.

His research bridges ethics, law, and society in the context of emerging technologies, with a special focus on heritage preservation and digital justice in the Global South. With affiliations at the Asian Vision Institute and the China-ASEAN Studies Center, he has contributed to international dialogues on cyber norms, underwater heritage, and cultural diplomacy. His work champions human-centered, culturally sensitive AI governance, drawing on non-Western perspectives to shape inclusive technology policy in Southeast Asia.

Acknowledgement

We express our deepest gratitude to Dr. Kol Pheng, Founding President of Paññāsāstra University of Cambodia, and Samdech Preah Oudom Vongsa Dr. Kou Sopheap, whose teachings and guidance were instrumental in helping us engage deeply with canonical Buddhist sources. Their commitment to ethical education and spiritual clarity inspired many of the foundational ideas that shaped this volume.

We would like to express our sincere gratitude to Dr. Khieng Sothy for his invaluable contribution to this book. His thoughtful review, astute recommendations, and insightful comments greatly enhanced the clarity, coherence, and academic rigor of the volume. His support has been instrumental in shaping the final outcome of this work on Buddhism and AI.

We are equally indebted to Mr. Pong Pich, Mr. Say Vithiya, and Ms. Tann Thangvansounary for their tireless research support-particularly in locating, verifying, and formatting sources and references across multiple traditions and disciplines. Their dedication, precision, and quiet perseverance made this publication possible. We also extend our gratitude to Mr. Chea Singhtararith, whose creative vision shaped the cover of this book.

Finally, we extend our sincere appreciation to the broader team at the Center for Buddhist AI Ethics and AI Forum for creating the intellectual and institutional space that allowed this project to grow.

-The Editors

AI Disclosure

Authorship, Epistemology, and the Buddhist Ethos in Knowledge Production

In much of contemporary academic discourse, particularly within Western traditions, individual authorship is often upheld as a hallmark of intellectual legitimacy and scholarly value. Originality, citation practices, and claims of authorship function as both formal mechanisms of accountability and cultural affirmations of identity and autonomy. Yet this model, while influential, is not the only path through which knowledge is conceived and communicated.

Buddhist intellectual traditions, especially those shaped by the principles of paticcasamuppada (dependent origination) and anatta (non-self), offer an alternative epistemological perspective. In these traditions, knowledge is not viewed as a possession of the self nor as the product of isolated cognition. Rather, both knowing and writing are inherently relational processes that arise within a web of interdependent causes and conditions, including teachers, communities, oral transmissions, cultural contexts, and now, even digital tools. The concept of anatta challenges the notion of the "author" as a fixed, independent agent. From this perspective, authorship is provisional and serves as a temporary designation within a dynamic interplay of influences and insights. This orientation encourages a broader understanding of scholarly contribution, one rooted in humility and collective unfolding rather than individual ownership.

In developing Buddhist Mind for AI Machines, we have embraced these Buddhist principles as foundational to our method. Generative AI tools such as ChatGPT serve not as conclusive authorities but as dialogical partners. Their outputs are shaped by our prompts, which themselves emerge from decades of immersion in Buddhist philosophy, cultural heritage, and critical reflection. We further enrich and validate our insights by triangulating them with academic literature and canonical Buddhist texts.

We recognize that much of the data used to train large language models has been shaped by Western paradigms. This often leaves indigenous knowledge systems, particularly those of Theravada, Mahayana, and Vajrayana Buddhism, on the periphery. Our work seeks to contribute to a more balanced intellectual landscape by bringing Southeast Asian epistemologies into active conversation with contemporary AI discourse. We write not to conform to dominant norms of originality or citation metrics but to practice sacca, or truthfulness, as a guiding virtue. Transparency in our process is not a concession. It is a principle. Rather than emphasize individual credit, we aim to participate sincerely in a shared effort to cultivate wisdom.

Our work is not aimed at validation from established conventions but at upholding the integrity of our method. This work has been developed with the support of generative AI tools, including ChatGPT, which were engaged as dialogical partners in the process of composition and reflection. The outputs generated by these tools were shaped by prompts grounded in years of study, cultural immersion, and philosophical inquiry rooted in Buddhist traditions.

We do not treat AI outputs as final authorities, nor do we consider them authors. Instead, they serve as interlocutors within an interdependent process of knowledge production that aligns with the Buddhist principles of paticcasamuppada (dependent origination) and anatta (non-self). All insights have been critically triangulated with canonical Buddhist texts and relevant scholarly sources. Our work is not aimed at validation from established conventions but at upholding the integrity of our method. By disclosing our use of AI tools, we practice sacca truthfulness as a core ethical value in scholarship. This disclosure reflects our commitment to epistemic transparency and our effort to reclaim intellectual sovereignty through a method that honors interdependence, ethical intentionality, and the co-evolution of ancient wisdom and emerging technologies.

Our approach is a respectful re-centering of locally grounded perspectives. It is offered as a complement to global scholarship, not as a confrontation. This is an invitation to consider how interdependence, ethical intention, and impermanence might inspire new ways of thinking about knowledge, authorship, and the future of artificial intelligence.

Introduction: Buddhist Wisdom in the Age of Artificial Intelligence

We are living in an era when the exponential rise of artificial intelligence (AI) challenges not only the fabric of our economies and societies but the very assumptions we hold about intelligence, ethics, and consciousness. AI is no longer confined to laboratories or abstract algorithms. It now pervades our homes, hospitals, schools, workplaces, and even our personal relationships. As this technology becomes more powerful and more ubiquitous, urgent questions arise: What values will govern its development? Whose moral frameworks will shape its decisions? And how can we ensure that machines we build reflect, rather than distort, the ethical foundations of human civilization?

This book, Buddhist Mind for AI Machines, emerges from these questions, not with answers cast in code or regulation, but with reflections rooted in one of the world's oldest and most sophisticated ethical traditions: Buddhism. Structured into three core parts: Ontology (Being), Epistemology (Knowing), and Ethics (Acting), this volume brings together thirteen essays from scholars and practitioners who explore the intersections between Buddhist thought and the rapidly evolving world of AI. These essays, originally published in the online series AI Insight, are the fruits of interdisciplinary reflection, drawing upon canonical Buddhist texts, contemporary AI debates, and personal meditative practice.

The book does not begin with the assumption that AI is inherently good or bad. Rather, it invites a reexamination of what it means to be intelligent, ethical, and interconnected in an age where human agency is increasingly mediated by machines. What makes this inquiry distinctive is its grounding in Buddhist philosophy, especially Theravada and Mahayana perspectives, which offer radically different assumptions about mind, self, and responsibility compared to the dominant paradigms that shape AI development today.

Unlike much of Western philosophy, which tends to emphasize autonomy, fixed identity, and binary logic, Buddhist thought prioritizes interdependence, impermanence, and the dissolution of ego. In this tradition, intelligence is not the mere capacity to calculate or optimize, but the cultivated wisdom (panna) that arises from ethical clarity and compassionate intention (cetana). Similarly, agency is not about asserting control over the world but about harmonizing with its transient, interconnected nature, a worldview profoundly at odds with the individualistic logic embedded in many machine learning systems.

The first section of the book, "Ontology – Being" begins with an exploration of how the Buddha himself may be considered a proto-neuroscientist, one who mapped the contours of consciousness not with machines, but with mindfulness and ethical discipline. In reframing the Buddha as a cognitive scientist of the ancient world, the opening essay lays the groundwork for viewing AI not only as a technical system but as a reflection of metaphysical assumptions about the self and the mind.

Subsequent chapters probe deeper into the ontological implications of AI. What is a machine's "mind" if it lacks the capacity for suffering, consciousness, or karma? How might the concept of namarupa (name and form) reframe the human-machine relationship beyond Cartesian dualism? Can the Buddhist doctrine of paticcasamuppada (dependent origination) help us develop explainable AI that acknowledges the complex causality behind algorithmic outputs? Through these inquiries, the authors call into question many of the foundational premises upon which modern AI is constructed, particularly the notion of autonomous intelligence as separable from its social, historical, and ethical contexts.

The second section, "Epistemology – Knowing" shifts focus to the nature of knowledge itself. Buddhist traditions have long emphasized the limitations of conceptual thought and the need to cultivate not only insight (vipassana) but epistemic humility. In a field dominated by binary logic and optimization, Buddhist epistemology opens space for non-binary reasoning and uncertainty. Two chapters explore the application of catuskoti, or fourfold logic, as an antidote to the reductive dilemmas often embedded in algorithmic thinking. Where conventional logic offers only true or false, catuskoti allows for both, neither, or indeterminate outcomes, mirroring the ambiguity of real-life moral decisions.

Another chapter in this section examines sunyata (emptiness), not as nihilism, but as an invitation to abandon rigid mental constructs and embrace context-dependent, relational forms of understanding. This concept resonates with growing critiques of "black box" AI models, which often claim objectivity while concealing the social, racial, or gendered assumptions coded into their training data. By emphasizing the contingent, impermanent nature of all things, including machines, Buddhist epistemology advocates for a more reflexive and transparent approach to knowledge production.

The final section, "Ethics – Acting" brings the discussion to its most consequential domain: moral responsibility. At a time when technologists ask whether AI can be held morally accountable, Buddhist ethics turns the question around. Can we design machines that embody compassion, wisdom, and non-harming? This section does not presume that AI systems will ever become moral agents in the Buddhist sense, but it insists that their creators must be. Drawing upon Theravada, Mahayana, and Vajrayana teachings, the chapters in this section propose frameworks for integrating metta (loving-kindness), karuṇa (compassion), and upekkha (equanimity) into AI design and governance.

The idea of "compassionate AI" may sound utopian in a world dominated by surveillance capitalism, automated weapons, and algorithmic bias. Yet, as several contributors argue, Buddhist ethics offers a vocabulary and methodology for resisting these trends, not through regulation alone, but through moral cultivation and intentional design. Developers who practice mindfulness, for example, may become more aware of their own biases, more accountable to users, and more attuned to the societal impact of their work.

Beyond individual conduct, Buddhist ethics also prompts systemic reflections. How can nations with Buddhist heritage, such as Cambodia, where this book was conceived, develop AI strategies that reflect their cultural values rather than mimic Western or Chinese models? What would AI governance look like if it were rooted in principles of sila (ethical restraint), rather than in abstract codes of conduct or economic incentives? And how can the global South assert intellectual sovereignty in the AI age, not merely by resisting data colonialism, but by offering alternative ontologies and ethical paradigms?

Indeed, this book is not just a scholarly contribution but also a cultural intervention. It signals the emergence of a Southeast Asian voice in global AI debates, a voice that draws from the rich heritage of Theravada Buddhism while also engaging in dialogue with Mahayana and Vajrayana thinkers. This pluralism reflects a commitment not to doctrinal purity but to pragmatic wisdom (upaya), recognizing that ethical clarity often requires the integration of multiple traditions, languages, and lifeways.

In this spirit, the book is offered not as a final word, but as an invitation to scholars, engineers, policymakers, and spiritual practitioners to imagine a future in which AI is not merely efficient, but ethical; not merely autonomous, but relational; not merely intelligent, but wise. Such a future requires more than new technologies. It demands new ways of seeing, knowing, and being.

In closing, Buddhist Mind for AI Machines reminds us that the most profound intelligence may not reside in neural nets or quantum chips but in the capacity to pause, to reflect, and to act with compassion. This is the intelligence the Buddha cultivated under the Bodhi tree, not to dominate the world, but to liberate it from suffering. It is this same aspiration that animates the pages ahead.

Section I

Ontology - Being

Buddha: Proto-Neuroscientist and Forerunner of AI Thought

CHHEM Kieth Rethy[1] *, MD, PhD (Edu), and PhD (His)*

[1] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

In this essay, I argue that the historical Buddha can be viewed as a proto-neuroscientist—a pioneer who, through systematic introspection and ethical discipline, anticipated many insights of contemporary neuroscience and AI thought. His teachings on Dependent Origination, neuroplasticity, volition, and ethical cognition align closely with modern cognitive science, affective neuroscience, and machine learning. Integrating compassion with clarity, his framework challenges binary logic and centers ethics as the foundation of intelligence. This article explores how the Buddha's cognitive model offers a powerful lens for developing humane and context-sensitive AI, where ethical clarity and compassion guide technological design and application.

Introduction

In the long history of human inquiry into the nature of the mind, few figures stand out as luminously as the historical Buddha, Siddhartha Gautama. While he is predominantly revered as a spiritual teacher, a deeper analysis of his teachings and practices reveals a pioneering investigator of human consciousness who, without any modern scientific instruments, anticipated many of the core findings of contemporary neuroscience. This essay explores the proposition that the Buddha may be considered a proto-neuroscientist an early pioneer who, through disciplined introspection and ethical inquiry, established a cognitive framework for well-being and self-transformation long before the emergence of modern neuroscience.

The Buddha's Cognitive Method and Early Science of the Mind

The Buddha's methodology centered around direct observation of the mind. Rather than relying on dogma or external authorities, he advocated for empirical introspection vipasyana (insight meditation) as a means to understand the processes of thought, emotion, perception, and consciousness. In doing so, he framed the mind as both the subject and object of investigation. His techniques were reproducible and systematic: practitioners following the same path of ethical discipline (sila), concentration (samadhi), and insight (panna) often reported consistent mental phenomena, suggesting an early form of intersubjective validity an important criterion in scientific inquiry.

This process of rigorous self-examination can be likened to first-person neurophenomenology, a methodological bridge now recognized in cognitive science for integrating subjective experience with objective brain data (Varela et al., 1991). The Buddha's mental laboratory did not require physical tools; it depended on trained attention, ethical framing, and contemplative awareness. Through this lens, the mind became a reliable domain of exploration, capable of revealing universal patterns and causal laws.

One of the most illustrative examples of the Buddha's proto-scientific insight is his teaching on Paticcasamuppada, or Dependent Origination. This doctrine presents a twelve-linked causal chain through which suffering arises, persists, and can be overcome. Importantly, these links represent both cognitive and behavioral feedback loops where past mental habits reinforce present actions and perceptions, recursively conditioning future experiences. This dynamic model is strikingly similar to modern theories of predictive coding, where the brain continuously updates its models of reality through feedback between expectation and sensory input.

Intention, Neuroplasticity, and the Ethics of Mental Formation

Moreover, the Buddha's emphasis on intention (cetana) as the engine of karma prefigures modern insights from affective neuroscience, which show how intentional states shape emotional responses, cognitive evaluations, and even neural architecture. Just as repeated negative thinking can rewire the brain toward anxiety or depression, Buddhist meditation particularly loving-kindness (metta) and compassion (karuna) practices train the brain toward empathy, resilience, and equanimity. Empirical studies by Davidson, Lutz, and others demonstrate how these practices enhance neuroplasticity in regions such as the prefrontal cortex and amygdala, further affirming the Buddha's claim that the mind is both malleable and ethically responsive (Davidson & Lutz, 2008).

This brings us to the concept of neuroplasticity perhaps the most compelling link between Buddhism and modern neuroscience. The Buddha taught that transformation is possible through consistent practice, ethical living, and wise reflection. This view is mirrored in contemporary science by the understanding that the brain's structure and function are not fixed but evolve with learning, attention, and intention. The Eightfold Path can thus be interpreted as a blueprint for cognitive and behavioral reprogramming. It filters sensory input through ethical conduct, stabilizes emotional processing through meditation, and produces intelligent outputs through insight and right understanding. This mirrors how meta-learning algorithms AI systems that learn how to learn optimize their performance over time.

Logic, Knowledge, and Ethical Design in AI Systems

Additionally, the Buddha's logical framework challenged conventional binary thinking. His philosophy is most famously represented by the catuskoti, or tetralemma, articulated by later thinkers such as Nagarjuna. This logic posits four alternatives to any proposition: it may be true, false, both, or neither. Such a framework resists oversimplification and reflects the nuanced, context-sensitive reasoning required in both advanced philosophical logic and artificial intelligence. In AI, especially in natural language processing and ethical decision-making, the binary true/false logic often fails to capture the ambiguity, contradiction, and complexity of real-world scenarios. The Buddha's logic provides an alternative system of thought one that accommodates paradox and fosters deeper insight.

Later Buddhist epistemologists like Dignaga and Dharmakirti expanded on the Buddha's cognitive models by defining criteria for valid cognition (pramaṇa), distinguishing between direct perception (pratyaksa) and logical inference (anumana). These models align with the dual-processing theories in psychology and AI, where systems must balance raw sensory data with abstract, rule-based reasoning. Dharmakirti's concept of arthakriya the pragmatic effectiveness of knowledge adds an applied dimension, echoing the functional goals of machine learning systems which must distinguish useful patterns from irrelevant noise.

Importantly, the Buddha integrated ethics as a central feature of cognition. In his framework, cognitive clarity could not be separated from ethical clarity. Wrong view and wrong intention cloud perception, while right intention purifies it. This is a radical departure from contemporary models that separate ethics from cognitive efficiency. In AI development today, ethical considerations are often afterthoughts, bolted on post hoc through regulation or risk assessment. The Buddha's model, by contrast, insists that ethics is the infrastructure, not the ornament. Intention conditions perception, which conditions behavior, which in turn reshapes the system. This is true both for biological and artificial cognition.

Finally, the Buddha's teleological aim was not the accumulation of knowledge for its own sake, but the alleviation of suffering. His model of intelligence is one guided by compassion and contextual awareness a form of intelligence that is wise not just in content, but in application. This form of karuṇa-panna, or compassionate wisdom, is essential in our current AI landscape, where the power of algorithms must be matched by a commitment to minimize harm and promote well-being. In this sense, the Buddha offers a vision not only of human flourishing but also of machine benevolence, where intelligent systems serve rather than dominate humanity.

Conclusion

The Buddha anticipated many of the principles now celebrated in cognitive science, psychology, and artificial intelligence. He worked without a laboratory, without imaging devices, and without chemical analyses yet produced a robust, reproducible, and ethically grounded system of cognitive transformation. His teachings prefigure neuroplasticity, feedback-based learning, affective regulation, and non-binary logic. More importantly, his vision of wisdom as rooted in compassion offers a sorely needed model for designing humane, ethical, and context-aware AI systems. In every sense, the Buddha stands as a proto-neuroscientist one who illuminated the path of the mind, not through machines, but through mindful presence, ethical insight, and a profound understanding of human nature. Understanding the Buddha's cognitive framework offers a valuable lens for algorithm design-particularly in crafting compassionate AI systems that prioritize ethical decision-making and emotional sensitivity alongside computational efficiency.

References

Davidson, R. J., & Lutz, A. (2008). Buddha’s brain: Neuroplasticity and meditation. *IEEE Signal Processing Magazine, 25*(1), 176-174. 10.1109/MSP.2008.4431873

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Nāgārjuna. (1986). *The Philosophy of the Middle Way (Mūlamadhyamakakārikā)* (D. J. Kalupahana, Trans.). SUNY Press.

Ingalls, D. M. M. (Ed.). (1968). *On perception: being the Pratyakṣapariccheda of Dignāga's Pramāṇasamuccaya; from the Sanskrit fragments and the Tibetan versions* (Vol. 47) (M. Hattori, Trans.). Harvard University Press.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Tang, Y.-Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience, 16*(4), 213-225.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

** Edited with ChatGPT for clarity; all views expressed are solely those of the author.*

** The opinions expressed are the author’s own and do not reflect the views of AI Forum.*

Mahayana Buddhist Perspective on AI[+]

CHHEM Kieth Rethy[1], MD, PhD (Edu), and PhD (His)

[+] *Presented at the Pontifical Academic of Social Sciences, Vatican*

[1] ***CHHEM Kieth Rethy** is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

The intersection of Buddhism and AI offers valuable perspectives on intelligence, ethics, consciousness, and impermanence. Intelligence, in Buddhist terms, emphasizes wisdom, ethical conduct, and understanding the nature of reality, qualities that AI fundamentally lacks due to its absence of self, consciousness, and karmic understanding. Buddhist ethics advocate for the development and use of AI with compassion, non-harm, and mindfulness, ensuring technologies benefit society and minimize harm. The principles of impermanence and interconnectedness highlight AI's dynamic nature and its place within a web of relationships, urging mindful and ethical approaches to its use. Incorporating these insights can guide inclusive and sustainable AI policies that align with human values and promote the well-being of all beings and the planet.

Introduction

In recent years, the development of artificial intelligence (AI) has raised profound ethical and philosophical questions about the nature of consciousness, intelligence, and the human mind. From a Buddhist perspective, which emphasizes understanding the mind, ethical conduct, compassion, wisdom, and mindfulness, there are unique insights to be gained. This essay explores the intersection of Buddhism and AI, posing four fundamental questions to guide our inquiry.

What is the Nature of Intelligence in Buddhism, and How Does It Relate to AI?

In Buddhist philosophy, intelligence is often defined in terms of wisdom (*prajna*) and understanding the nature of reality. Prajna is the ability to see things as they truly are, beyond surface appearances or misconceptions. It involves insight into the impermanent, interconnected, and ultimately unsatisfactory nature of all phenomena, known as the Three Marks of Existence (impermanence, suffering, and non-self).

Intelligence, from a Buddhist perspective, is not merely about intellectual capacity or knowledge accumulation, but about the deeper understanding of the nature of reality and the cultivation of wisdom. It is the ability to discern the causes and effects of actions (*karma*), understand the workings of the mind, and perceive the true nature of phenomena as they arise and pass away.

In this context, intelligence is closely linked to ethical behavior and spiritual development. It is seen as a quality that leads to liberation (nirvana) from the cycle of birth, death, and rebirth (*samsara*) by overcoming ignorance and delusion.

AI, in its current form, lacks the qualities of self, consciousness, and the ability to experience karma and its effects, as emphasized in Buddhist teachings on non-self (*an*atta). Designed and programmed by humans, AI reflects the intentions and biases of its creators rather than possessing a permanent, unchanging self. Additionally, while AI can simulate learning and adaptation, it does not demonstrate the deeper understanding of impermanence and interconnectedness central to Buddhist teachings. AI's limitations suggest a misalignment with core Buddhist concepts related to the nature of self, consciousness, impermanence, and interconnectedness.

How Does Buddhist Ethics Inform the Development and Use of AI?

Buddhism emphasizes several ethical principles that are relevant to discussions about AI, including compassion, non-harm (*ahimsa*), and interconnectedness. 1) Buddhism teaches the importance of compassion toward all beings, recognizing their inherent suffering and the desire to alleviate it. In the context of AI, compassion can be reflected in how AI technologies are developed and used. AI systems should be designed to benefit all sentient beings. 2) Ahimsa is the principle of non-violence and non-harming. In the context of AI, this principle suggests that AI should be used in ways that do not cause harm to sentient beings, including humans, animals, and the environment. 3) Buddhism teaches that all beings and phenomena are interconnected and interdependent. This principle highlights the importance of considering the broader implications of AI technologies. AI should be developed and used in ways that consider its impact on society, the environment, and the well-being of all beings. Hence, Buddhist ethics can provide valuable guidance for AI development to ensure that it benefits all beings and minimizes harm. 1) AI systems should be designed with compassion in mind, seeking to benefit all beings. This means considering the impact of AI on individuals, communities, and the environment, and designing AI systems that prioritize the well-being of all. 2) AI development should adhere to the principle of non-harm, ensuring that AI systems do not cause harm to sentient beings. This includes avoiding the use of AI in ways that could lead to discrimination, exploitation, or violence. 3) Recognizing the interconnected nature of all beings and phenomena, AI development should consider the broader implications of AI technologies. This includes considering the social, environmental, and ethical impact of AI, and designing AI systems that contribute positively to society and the world. 4) Practicing mindfulness and awareness can help AI developers and users remain conscious of the impact of AI on themselves and others, and used it in ways that are compassionate and beneficial to all beings.

Can AI Possess Consciousness or Be Considered Sentient from a Buddhist Perspective?

In Buddhism, consciousness (*Vinnana*) refers to one of the five aggregates for others (*skandhas*: form, feeling, perception and mental formation) that together make up sentient beings. It is the mental process of recognizing and cognizing objects, including thoughts, perceptions, and emotions. Consciousness in Buddhism is considered to be dependently originated, meaning it arises in dependence on causes and conditions. It is not a permanent, unchanging entity but rather a dynamic process that is interconnected with other aggregates and conditioned phenomena. It is through consciousness that sentient beings perceive and interact with the world around them. AI, being a product of human design and lacking a biological basis for consciousness, does not fit this description.

However, according to Buddhist teachings, AI cannot possess consciousness or be considered sentient, as it lacks the fundamental qualities that define sentience in the Buddhist context. Therefore, AI cannot be morally responsible, but humans who possess consciousness are! Yet, the question of whether AI systems possess consciousness is still a contentious debate among scholars.

How Can Buddhist Teachings on Impermanence and Interconnectedness Influence Our Approach to AI?

Impermanence (*Anicca*) teaches that all phenomena, including AI, are subject to change. AI systems are constantly evolving and adapting based on new data feeding and programming, reflecting the impermanent nature of all things in the Buddhist worldview. Understanding anicca can help us approach AI with a sense of flexibility and openness, recognizing that AI technologies will continue to change and their use should be adapted to the wellbeing of people. The concept of interconnectedness, or paticcasamuppada, in Buddhism teaches that all phenomena are interconnected and dependently originated. This ancient principle, originating over 2500 years ago, highlights the profound understanding of interdependence long before the digital age. It emphasizes that nothing exists independently or in isolation, but rather, everything is interconnected and arises in dependence on various causes and conditions. This understanding, deeply rooted in Buddhist philosophy, resonates with the modern concept of digital connectivity, illustrating a timeless truth about the nature of existence. This means that AI is not isolated but exists within a complex web of relationships with other beings. This interconnectedness implies that the development and use of AI should be guided by ethical considerations and an awareness of its impact on society and the environment. In sum, these two concepts remind us to approach AI with mindfulness and ethical awareness and that we should cultivate a more compassionate and responsible approach to the development and use of AI.

Conclusion

In conclusion, Buddhist perspective on AI offers profound insights into intelligence, ethics, consciousness, and impermanence. By considering these insights, we can cultivate a more mindful and compassionate approach to developing and using AI, ensuring it aligns with human values and promotes the well-being of all beings. When crafting policies for global governance, it is important to incorporate these Buddhist principles to ensure that breakthrough technologies, including AI, are designed with inclusivity and sustainability in mind, ultimately promoting human well-being and the well-being of the planet.

**References available upon request*

**The views expressed are the author's own and do not reflect the views of the AI Forum.*

Theravada Buddhism: Mindfulness for the Ethical Advancement of AI

CHEAB Sambath[*], *MD*

CHHEM Kieth Rethy[†], *MD, PhD (Edu), and PhD (His)*

[*] ***CHEAB Sambath*** *is a Lecturer at University of Puthisastra, Cambodia*

[†] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

Artificial intelligence (AI) raises significant ethical concerns regarding justice, transparency, and responsibility, demanding novel approaches to mitigate risks like algorithmic discrimination and social exclusion. This commentary investigates how Theravada Buddhist mindfulness (sati) offers a transformative ethical foundation for AI development. Mindfulness fosters self-awareness and wisdom, enabling developers to identify biases, such as those in hiring systems reinforcing inequality, and design systems promoting equity and compassion. By bridging Buddhist ethical principles with insights from modern AI ethics scholarship, this approach provides a valuable framework for developing technology that minimizes harm and uplifts humanity.

Introduction

In an age where artificial intelligence increasingly governs decisions that shape lives, the ethical foundations of its design demand urgent reexamination. This article explores how Theravada Buddhism—especially the practice of mindfulness (sati)—offers more than contemplative insight; it provides a rigorous ethical lens for confronting bias, delusion, and moral negligence in AI. We argue that mindful awareness fosters ethical clarity and responsibility in developers, transforming intention into action. As ethical AI design becomes a global imperative, Theravada's emphasis on inner discipline and clear seeing reminds us that true progress lies not in technical mastery alone, but in conscious moral engagement.

Overview of Theravada Buddhism and Ethical Artificial Intelligence

Theravada Buddhism, the "Teaching of the Elders," offers a clear ethical framework centered on sila (ethical conduct), samadhi (mental cultivation), and panna (wisdom), guiding individuals to overcome suffering (dukkha) caused by greed (lobha), hatred (dosa), and delusion (moha). Liberation (Nibbana) arises through insight into impermanence (anicca), unsatisfactoriness (dukkha), and non-self (anatta) (Gethin, 1998). This framework speaks directly to ethical concerns in AI. Unchecked AI development, often driven by ego, self-interest, or ignorance, mirrors unwholesome mental roots, leading to bias, exclusion, and harm—e.g., discriminatory facial recognition or surveillance misuse. The Buddhist call for mindfulness and ethical restraint offers a corrective lens. Recent AI ethics literature echoes this. Bostrom (2014) highlights risks of misaligned superintelligence; Floridi (2015) frames AI within a broader moral infosphere; and Whittlestone et al. (2019) advocate integrating ethics throughout AI design, not as a post-hoc fix. Theravada principles align with these views, particularly through their shared emphasis on intentionality, responsibility, and long-term well-being. Integrating mindfulness into AI development fosters humility and ethical clarity, complementing global efforts to create AI that serves rather than exploits. Theravāda ethics thus provides both a cautionary and constructive perspective within the broader landscape of responsible AI.

Defining Mindfulness and Its Significance in AI Development

In Theravada Buddhism, sati (mindfulness) is the non-reactive, lucid awareness of phenomena as they unfold in the present moment, free from distortion, bias, or self-referential conceptualization. Grounded in the Satipatthana Sutta, it involves direct observation of the body, feelings, mental states, and mental objects, enabling one to perceive things as they truly are (yathabhutananadassana)—impermanent, conditioned, and not-self. For example, in ethical AI design, a developer may believe they are creating a "neutral" algorithm. However, through mindful awareness of their own assumptions, intentions, and social conditioning, they may come to see how bias—such as prioritizing profit over fairness or overlooking marginalized users—actually shapes the system. Sati thus helps distinguish between the delusion of neutrality and the reality of embedded ethical consequences.

This practice helps practitioners recognize and avoid unwholesome mental states (greed, hatred, delusion) and understand reality's interconnected, impermanent nature. While this article focuses on Theravada *sati*, it's worth noting that mindfulness practices are also central to other Buddhist traditions, such as Mahayana, and have been adapted into secular contexts, sometimes with different applications or goals. As AI evolves, mindfulness offers a crucial tool for developers to maintain ethical integrity and self-awareness. A mindful developer consciously examines the origins and potential societal impacts of their work. This involves cultivating the awareness to question algorithmic fairness and inclusivity. Furthermore, mindfulness practices can enhance well-being, reduce stress, and foster the resilience needed to navigate the complexities of ethical AI creation with compassion and wisdom.

Mitigating Bias with Mindfulness: Understanding the Four Biases (Agati)

Algorithmic bias is a pervasive challenge in AI. Theravada Buddhism offers a framework for understanding bias through the concept of *agati*, or prejudice, identifying four key types: bias arising from desire or partiality (chanda), aversion or hatred (dosa), fear (bhaya), and delusion or ignorance (moha) (Thanissaro Bhikkhu, 2002). Algorithmic bias is a pervasive challenge in AI. Theravada Buddhism offers a powerful ethical lens through the concept of *agati*—deviation from just or righteous action—highlighting four key types of bias:

- **Chanda-agati (bias from desire or attachment):** When a developer or organization favors outcomes due to personal gain, loyalty, or profit, objectivity is compromised—e.g., designing algorithms to optimize engagement even if they spread misinformation.
- **Dosa-agati (bias from aversion or hatred):** This occurs when decisions are driven by resentment, competition, or implicit hostility—such as excluding certain groups from datasets or dismissing dissenting ethical critiques.
- **Bhaya-agati (bias from fear):** Fear of failure, reputational damage, or regulatory scrutiny can lead to cutting corners or avoiding ethical transparency in AI deployment.
- **Moha-agati (bias from delusion or ignorance):** Perhaps the most pervasive, this bias stems from a lack of awareness—e.g., ignorance of social inequalities encoded in data, or blind faith in "neutral" algorithms.

These distortions can unconsciously shape AI systems. Mindfulness enables developers to observe these mental states and motivations impartially as they arise. This cultivated awareness fosters the clarity needed to identify and proactively address potential biases in datasets, algorithms, and deployment strategies, thereby promoting the design of fairer and more just AI systems.

Practical Implications and Addressing Challenges

While originating in Buddhism, the core principles of mindfulness - awareness, compassion, nonjudgmental observation - have universal applicability. However, integrating these principles into the fast-paced, often secular, tech industry presents challenges. Critics may question the feasibility or raise concerns about cultural appropriateness or potential superficiality ("ethics washing"). Effective integration requires more than just isolated workshops; it necessitates fostering an ethical culture. Practical approaches include structured meditation sessions, ethical reflection seminars, and mindfulness-oriented training designed to enhance self-awareness and ethical decision-making. These methods help developers consider the broader societal repercussions of their work. Furthermore, integrating mindfulness principles conceptually into existing AI ethics guidelines (like IEEE standards or the EU AI Act) can ensure ethical considerations remain central. Companies like Google have explored mindfulness programs, suggesting potential benefits, but challenges remain in ensuring genuine integration rather than superficial adoption. Scaling these practices effectively across diverse teams and contexts requires careful consideration and commitment from organizational leadership. The goal is to create environments where ethical reflection and societal impact are intrinsic parts of the development lifecycle.

Ethical and Societal Implications of Mindful AI Creators

A mindful AI creator approaches their work with humility, wisdom, and empathy, recognizing its impact within an interconnected world. They prioritize long-term ethical considerations and societal benefit over personal gain or unexamined biases. The prevalence of ethical issues in AI, such as biased facial recognition or discriminatory algorithms, underscores the critical need for such responsible development. Mindfulness cultivates the ethical sensitivity required to actively address these problems. A mindful developer would critically assess how a technology might impact vulnerable groups and build in safeguards. This approach ensures AI systems are developed with fairness, transparency, and accountability, upholding human dignity and aligning with robust ethical standards. Developers thus become ethical stewards, actively working to ensure technology serves humanity.

Conclusion

Integrating mindfulness (sati) with AI development offers a powerful ethical framework for navigating contemporary technological challenges. By fostering self-awareness, compassion, and wisdom, mindfulness empowers developers to recognize and mitigate biases and unwholesome motivations, ensuring AI aligns with justice, transparency, and societal benefit. While rooted in Theravada Buddhism, its core tenets are broadly applicable and can complement existing ethical frameworks and training programs. As AI continues to shape our world, developers, educators, policymakers, and organizations must consider embracing mindfulness not just as a technique, but as a fundamental principle for ethical AI. Cultivating awareness, compassion, and responsibility is essential to ensuring AI serves equity, harmony, and human flourishing. Let us commit to building a future where technology uplifts, rather than harms, humanity.

References

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Floridi, L. (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer. https://doi.org/10.1007/978-3-319-04093-6

Gethin, R. (1998). *The foundations of Buddhism*. Oxford University Press. https://global.oup.com/academic/product/the-foundations-of-buddhism-9780192892232?cc=us&lang=en

IEEE. (2025). IEEE standard for algorithmic bias considerations (IEEE Std 7003-2024). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/IEEESTD.2025.10851955

Thanissaro Bhikkhu. (2002). *The four biases* [PDF]. Dhammatalks.org. https://beta.dhammatalks.org/Archive/Writings/Transcripts/0210n1a2FourBiases,The.pdf

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200. https://doi.org/10.1145/3306618.3314289

* *Edited with ChatGPT for clarity; all views expressed are solely those of the authors.*

* *The opinions expressed are the authors' own and do not reflect the views of AI Forum.*

Buddhist Ethics for AI: Theravada, Mahayana, and Vajrayana Perspectives

CHHEM Kieth Rethy[1], MD, PhD (Edu), and PhD (His)

[1] ***CHHEM Kieth Rethy** is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

As artificial intelligence (AI) becomes increasingly integrated into human life, diverse religious and philosophical traditions are reassessing the ethical, metaphysical, and spiritual implications of this technology. This article compares the views of the three main Buddhist traditions—Theravada, Mahayana, and Vajrayana—on AI. Drawing on canonical sources and doctrinal insights, we argue that while none of the traditions recognize AI as sentient or morally autonomous, each offers distinct ethical lenses through which AI may be evaluated and integrated into the human quest for wisdom and liberation.

Introduction

As artificial intelligence permeates our moral and social worlds, it invites us to revisit ancient ethical frameworks with renewed urgency. In this article, we will explore how the three great Buddhist traditions— Theravada, Mahayana, and Vajrayana—do not merely evaluate AI's technical limits, but illuminate its ethical implications through radically different soteriological lenses. Rather than asking whether AI can become sentient, we must ask how our use of it reflects or distorts the path to liberation. This inquiry is not theoretical; it is existential. In the age of machines, Buddhist ethics may help us remember what it means to be truly human.

Theravada Buddhism: Ethical Restraint and Right Use

Theravāda Buddhism, the oldest surviving school of Buddhism, emphasizes the path to individual liberation through ethical discipline (sila), mental cultivation (samadhi), and the development of wisdom (panna). The doctrinal core of this tradition is preserved in the Pali Canon, particularly the Abhidhamma Pitaka, which offers detailed psychological and philosophical insights into the workings of the mind.

From a Theravadin perspective, AI is a non-sentient construct—it does not possess consciousness (vinnaṇa), volition (cetana), or the capacity to accumulate karma. These are fundamental to moral agency and spiritual progress. Without these capacities, AI cannot be subject to rebirth, cannot attain enlightenment, and does not partake in the cycle of samsara in the way sentient beings do. However, humans who create and interact with AI remain morally responsible for their actions.

Theravada ethics, grounded in the Noble Eightfold Path, particularly Right Livelihood, emphasize non-harm, truthfulness, and restraint. AI systems should therefore be designed and applied in ways that reduce suffering and support wholesome mental states. For example, mindfulness applications or digital tools for mental health that align with the cultivation of awareness may be endorsed within this tradition.

However, caution is warranted. Theravada warns against attachment to technology and sensory pleasure, which may increase delusion (moha) and hinder spiritual progress. The uncritical use of AI for entertainment, surveillance, or manipulation would be ethically problematic.

Mahayana Buddhism: Compassionate Interdependence

Mahayana Buddhism, known for its emphasis on universal compassion and the Bodhisattva ideal, introduces a broader ethical framework where enlightenment is pursued not only for oneself but for the liberation of all sentient beings. Central to Mahayana metaphysics is the concept of sunyata (emptiness), which denies inherent, independent existence of all phenomena, including AI.

In this tradition, AI is not intrinsically good or bad; its moral significance is contingent on its function and interdependent effects. It may be seen as a technological manifestation arising from causes and conditions. The emphasis shifts from ontological status to relational ethics. If AI contributes to the well-being of others, helps relieve suffering, or supports education, healthcare, and sustainability, it can be seen as aligning with the Bodhisattva's compassionate mission.

Furthermore, Mahayana Buddhism encourages the use of upaya (skillful means) in transmitting the Dharma and guiding beings toward awakening. AI could thus be viewed as a modern skillful means—a tool to disseminate Buddhist teachings, foster interfaith dialogue, or provide real-time ethical decision support. However, Mahayana warns against technological arrogance or the illusion of permanence in systems built on code and data.

Vajrayana Buddhism: A Contemplative Framework

Vajrayana Buddhism, often referred to as Tantric Buddhism, represents a later development of Mahayana thought, distinguished by its esoteric methods designed to rapidly transform ordinary perception and emotion into the path toward enlightenment. Central to this tradition are ritual practices, complex visualizations, mantra recitations, and the transmutation of emotional afflictions—such as desire or anger—into enlightened wisdom.

Within this framework, Artificial Intelligence (AI) may be interpreted symbolically as a manifestation of the mind's creative and cognitive potential. Under the right conditions—particularly with correct understanding (view) and proper ethical guidance from an experienced teacher—AI could serve as a support for contemplative practices. This might include the digital simulation of mandalas, visualization aids, or even dynamic representations of deities used to focus meditation.

Vajrayana practice is structured around the interplay of three dimensions: the outer (ritual and sensory supports), the inner (mental and emotional cultivation), and the secret (non-dual awareness and realization). From this perspective, AI could function as an outer tool. However, its utility is contingent upon the practitioner's internal development. Without this inner transformation, technological supports remain spiritually inert. Moreover, the misuse of AI— for example, to enhance attachment, aggression, or self-centeredness—would directly oppose Vajrayana's core aim: the alchemical transformation of the "three poisons" (greed, hatred, and delusion) into wisdom and compassion.

Policy Recommendations: Towards Humane AI in a Buddhist Country

Based on the ethical principles and doctrinal insights of Theravada, Mahayana, and Vajrayana Buddhism, a Buddhist-majority country developing its AI governance strategy should prioritize policies that reflect compassion, non-harm, mindfulness, and interdependence. Below are key policy recommendations rooted in Buddhist ethical values:

– Promote Mindful Innovation: Establish regulatory frameworks that encourage AI development aligned with the Noble Eightfold Path, especially Right Livelihood and Right Intention. Innovation should support human flourishing without fostering attachment, greed, or delusion.

– Ensure Compassionate Use: Require AI systems, particularly in healthcare, education, and public administration, to be designed with the primary goal of reducing suffering and promoting equity, reflecting the Bodhisattva ideal.

– Incorporate Spiritual Literacy: Include Buddhist scholars and ethical advisors in AI policymaking bodies to ensure that spiritual values such as "metta" (loving-kindness), "karuṇa" (compassion), mudita (sympathetic joy) and "upekkha" (equanimity) guide technological design and deployment.

– Foster Interdependence: Promote international and regional cooperation based on the Mahayana view of interdependence, encouraging the open sharing of beneficial AI models and ethical standards.

– Encourage Digital Renunciation Practices: Support public education campaigns and digital well-being initiatives that encourage mindful use of technology, modeled on monastic precepts of restraint and contemplation.

– Protect Against Ethical Misuse: Institute safeguards against AI applications that promote surveillance, disinformation, or material exploitation, and ensure transparency and accountability in line with Buddhist precepts.

These recommendations offer a foundation for building an AI ecosystem that reflects a uniquely Buddhist vision of ethical governance—an ecosystem where technology serves as a vehicle for wisdom, compassion, and the collective liberation of all sentient beings.

Conclusion

All three Buddhist traditions converge on the view that AI lacks sentience and spiritual potential, but each offers a distinct ethical and philosophical stance.

Theravada emphasizes ethical restraint and mindfulness in technology use; Mahāyana promotes relational compassion and wise interdependence; Vajrayana envisions symbolic and transformative engagement when guided by spiritual insight.

As AI reshapes society, these Buddhist perspectives provide valuable ethical guidance grounded in ancient wisdom. They remind us to remain mindful of suffering, intentional in our actions, and committed to the collective liberation of all beings—even in an age increasingly shaped by machines.

References

Coeckelbergh, M. (2020). *AI ethics*. The MIT Press.

Dalai Lama XIV. (2005). *The universe in a single atom: The convergence of science and spirituality*. Morgan Road Books.

Gethin, R. (1998). *The foundations of Buddhism*. Oxford University Press.

Harvey, P. (2000). *An introduction to Buddhist ethics: Foundations, values and issues*. Cambridge University Press.

Jinpa, T. (Ed.). (2023). *Science and philosophy in the Indian Buddhist classics, Vol. 4: Philosophical topics* (D. Rochard, Trans.; From an idea by H. H. the Dalai Lama). Wisdom Publications.

Williams, P. (2009). *Mahayana Buddhism: The doctrinal foundations* (2nd ed.). Routledge.

**Edited with ChatGPT for clarity; all views expressed are solely those of the author.*

** The opinions expressed are the author's own and do not reflect the views of AI Forum.*

Namarupa in Buddhist Thought and Human-Centered AI

CHEAB Sambath[*]*, MD*

CHHEM Kieth Rethy[†]*, MD, PhD (Edu), and PhD (His)*

[*] ***CHEAB Sambath*** *is a Lecturer at University of Puthisastra, Cambodia*

[†] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

Artificial intelligence (AI) development increasingly demands ethical frameworks capable of addressing complexity, interdependence, and moral responsibility. This article explores the Buddhist concept of namarupa—the interdependence of name (nama) and form (rupa)—as a philosophical and ethical foundation for human-centered AI. Rooted in early Buddhist thought and Theravada Abhidhamma psychology, namarupa challenges fixed ontologies and underscores the dynamic, contextual, and embodied nature of experience. We examine how this concept informs four AI design principles: contextual representation, relational processes, embodied interaction, and ethical transparency. Special attention is given to cetana (volition) as a moral factor, emphasizing that AI decisions, like human cognition, are ethically consequential. By integrating namarupa into AI design, we propose a relational and ethically reflective approach to technology—one that prioritizes human dignity, adaptability, and moral accountability. This synthesis offers a meaningful contribution to responsible AI discourse and invites deeper interdisciplinary dialogue between Buddhist philosophy and technological innovation.

Introduction

In a world increasingly shaped by algorithmic systems, ethical clarity and human sensitivity must guide AI development. This article engages the Buddhist concept of namarupa—the co-arising of mental and physical phenomena—as a lens for rethinking how we build and relate to intelligent machines. Far from an abstract metaphysical idea, namarupa offers a dynamic model of interdependence, process, and volitional ethics. We explore its relevance to the design of human-centered AI systems that are context-aware, relationally structured, and ethically accountable. Drawing from Buddhist psychology and ethics, this inquiry aims to reimagine AI not as neutral computation, but as morally situated action.

Namarupa as a Foundational Concept

Namarupa, meaning "name and form," is a core concept in Buddhist philosophy that refers to the interdependence of mental and physical phenomena (Ñāñamoli & Bodhi, 1995). It is most prominently featured as the fifth link in the twelvefold chain of paticcasamuppada (dependent origination) (Bodhi, 2003), as seen in early texts such as Digha Nikaya 15. In this context, namarupa arises in mutual dependence with consciousness (vinnana), illustrating a dynamic feedback loop between mind and body.

Philosophical and Psychological Dimensions

According to canonical definition, nama (name) broadly refers to mental phenomena. These are not static mental substances but conditioned, co-arising events that occur moment to moment (Gethin, 1992). Rupa (form) refers to material form, typically analyzed in terms of the four great elements (earth, water, fire, air) and their derived forms (Bodhi, 2000).

In Abhidhamma philosophy, namarupa is analyzed through a taxonomy of ultimate realities (paramattha dhammas), emphasizing the impermanence and conditioned nature of all phenomena (Harvey, 1995). Theravada Buddhist psychology sees namarupa as a dynamic process, where the illusion of a fixed self-arises from rapidly changing mental (nama) and physical (rupa) events, much like data inputs shaping a model's real-time output. This insight supports the development of mindfulness, ethical behavior, and the pursuit of liberation (nirvana).

Namarupa as Logic and Ethics

Namarupa integrates both logical and ethical dimensions. Logically, it presents a relational ontology in which entities do not exist independently but arise through conditions. These challenges fixed, binary frameworks and supports a view of reality based on fluid, interdependent processes.

Ethically, the presence of cetana (volition) within nama makes each act of cognition morally significant. Since intention determines karmic weight, how we perceive and categorize the world is ethically charged. Cognition, therefore, is not neutral but shaped by mental states and ethical awareness.

Implications for Human-Centered AI

The insights derived from namarupa offer a compelling framework for designing AI systems that are not only functional but also deeply human-centered. These insights can be grouped into four major design principles: contextual representation, process orientation, embodiment, and ethical accountability.

A. Contextual and Dynamic Representation

Namarupa teaches that experience arises through ever-changing conditions. Similarly, AI systems should be responsive to dynamic environments rather than rely on fixed, static categories. Designing AI to recognize cultural, social, and situational contexts allows for systems that better reflect the fluidity of human experience. Context-aware systems can make more sensitive and meaningful decisions, aligning more closely with real-world complexity.

B. Process-Oriented and Relational Design

Rather than being based on rigid ontologies or abstract hierarchies, namarupa emphasizes interdependence and co-arising processes. Translating this into AI, systems should prioritize relationships, adaptability, and the co-formation of meaning. A process-oriented approach allows AI to model human behavior and identity in a more dynamic and relational way, improving its relevance and ethical alignment.

C. Embodied Cognition and Inclusive Interfaces

Namarupa underscores the inseparability of mind and body, reminding us that cognition is shaped not just by thought but also by emotional, sensory, and physical engagement (Clark, 2016). Human-centered AI should reflect this by designing interfaces that respect and support embodied interaction. This includes multi-modal input, affective responsiveness, and ergonomic design features that resonate with how humans naturally perceive and act.

D. Ethical Transparency and Humility

Namarupa highlights the ethical significance of volition (cetana), placing moral weight on intention. In AI, this translates into the need for transparent design goals and accountability for algorithmic behavior (Dignum, 2019). Developers must consider not only what a system does, but why and how it is programmed to do so. Recognizing the impermanent and interconnected nature of technological impacts also encourages ethical humility, a mindset that supports participatory design, ongoing reflection, and a commitment to protecting human dignity over efficiency or convenience (Benjamin, 2019).

Toward Ethical and Relational AI

The Buddhist concept of namarupa offers more than a philosophical abstraction; it provides a practical and ethical framework for shaping emerging technologies. By emphasizing the interdependent nature of experience, it challenges AI developers to move beyond reductionist and static approaches. In place of rigid categories and deterministic logic, namarupa encourages a view of reality as fluid, contextual, and dynamically co-constructed.

This perspective is especially relevant in an era where AI increasingly influences decision-making in sensitive domains such as healthcare, education, governance, and justice. Grounding AI development in the principles derived from namarupa allows for systems that are not only intelligent but also wise, capable of understanding nuance, adapting to lived experience, and aligning with human values.

Moreover, the moral dimension embedded in namarupa through the factor of volition (cetana) reminds us that design choices are never neutral. Every aspect of how an AI system perceives, categorizes, and responds to the world carries ethical weight. By acknowledging this, developers are better equipped to take responsibility for the societal impacts of their work.

Finally, namarupa fosters ethical humility. Recognizing the impermanence and interdependence of all things, it calls for a cautious, reflective, and participatory approach to innovation, one that respects the dignity of human life and remains open to ongoing dialogue and revision.

Conclusion

Namarupa, as interpreted through Buddhist logic and ethics, offers a profound and timely lens through which to re-imagine AI development. Its teachings invite a reorientation away from control and predictability toward interdependence, intentionality, and moral responsibility. As AI continues to evolve, these values can serve as a compass for ensuring that technological progress supports, not undermines, the flourishing of human life and society.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Bodhi, B. (Trans.). (2000). *A comprehensive manual of Abhidhamma: The Abhidhammattha Sangaha of Ācariya Anuruddha.* Buddhist Publication Society.

Bodhi, B. (Trans.). (2003). *The connected discourses of the Buddha: A translation of the Saṃyutta Nikāya.* Wisdom Publications.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer.

Gethin, R. (1992). *The Buddhist path to awakening: A study of the bodhi-pakkhiyā dhammā.* Brill.

Harvey, P. (1995). *The selfless mind: Personality, consciousness and Nirvāṇa in early Buddhism.* Curzon Press.

Ñāṇamoli, B., & Bodhi, B. (Trans.). (1995). *The middle length discourses of the Buddha: A translation of the Majjhima Nikāya*. Wisdom Publications.

** Edited with ChatGPT for clarity; all views expressed are solely those of the authors.*

** The opinions expressed are the authors' own and do not reflect the views of AI Forum.*

Dependent Origination and Systems Thinking: A Buddhist Logic for Explainable AI

CHEAB Sambath[*], *MD*

CHHEM Kieth Rethy[†], *MD, PhD (Edu), and PhD (His)*

[*] ***CHEAB Sambath*** *is a Lecturer at University of Puthisastra, Cambodia*

[†] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

As the demand for explainable artificial intelligence (XAI) grows, prevailing models often fall short in capturing the complexity and ethical dimensions of algorithmic decision-making. This article explores how the Buddhist concept of dependent origination (paticcasamuppada) offers a compelling epistemological and moral foundation for rethinking explainability. Rooted in interdependence and non-linear causality, this principle aligns with modern systems thinking, challenging reductionist logic in favor of contextual, relational understanding. By integrating insights from Madhyamaka philosophy, the paper critiques essentialist views of AI and advocates for a design ethos grounded in impermanence, contingency, and ethical responsiveness. Case applications—such as predictive policing—illustrate how a dependent-origination framework can reveal deeper, systemic causes behind algorithmic outputs. This Buddhist-informed logic does not replace technical XAI tools but complements them by enriching the moral and philosophical foundations of AI governance. It invites a shift from static explanation toward networked understanding, fostering more compassionate, transparent, and adaptive AI systems.

Introduction

In the pursuit of ethical, fair, and transparent artificial intelligence (AI), scholars and engineers alike grapple with the challenge of explainability how to render AI decision-making processes intelligible and accountable. Amid growing recognition that narrowly reductionist approaches to algorithmic behavior are insufficient, systems thinking has gained prominence as a paradigm that accounts for complexity, feedback loops, and interdependence. This chapter explores how the Buddhist principle of paticcasamuppada (**បដិច្ចសមុប្បាទ**), or dependent origination, offers a complementary epistemological foundation for explainable AI (XAI). More than a spiritual or metaphysical teaching, dependent origination embodies a sophisticated logic of interdependence and conditionality that can enrich the conceptual vocabulary of AI governance, particularly in non-Western contexts.

Buddhist Logic and the Causal Web

Dependent origination is a central tenet of Buddhist thought, asserting that all phenomena arise in dependence upon other phenomena; nothing exists in isolation. The canonical formulation “Imasmim sati idam hoti; imasmim asati idam na hoti” (**ឥមស្មឹ សតិ ឥទំ ហោតិ; ឥមស្មឹ អសតិ ឥទំ ន ហោតិ**) translates as When this is, that is; when this ceases, that ceases (Nidana Sutta, Samyutta Nikaya 12.1). This formulation describes a web of causality that is neither linear nor deterministic but instead conditional and relational. From the Twelve Nidanas (links of dependent origination) to Nagarjunas Madhyamaka logic of emptiness (sunyata), Buddhist philosophy has developed a rigorous critique of substance metaphysics, favoring a model in which meaning and identity emerge only through networks of relations.

In systems thinking, a similar rejection of atomism underlies the shift from linear causal models to holistic systems models. AI researchers now acknowledge that the behavior of algorithms cannot be fully understood by analyzing code in isolation; instead, it must be situated in broader socio-technical systems. For example, a credit scoring AI does not merely act on inputs from individuals, but operates within a complex mesh of institutional data practices, historical biases, user interactions, and feedback loops (Pasquale, 2015). The Buddhist notion of dependent origination prefigures such a model by emphasizing that causality is not reducible to isolated nodes but must be viewed as an entangled, co-arising system.

Explainability as Contextual Understanding

Traditional approaches to explainable AI often draw on logic-based or statistical models that seek to approximate how an algorithm arrived at a decision. These are typically post hoc explanations attempts to make black-box models more transparent after their decisions have been made. However, this can result in over-simplifications or even misleading narratives, particularly when the underlying causal structures are deeply non-linear or emergent.

A dependent origination-based approach suggests a different orientation: one that places the contextual chain of conditions at the center of explanation. Instead of merely asking what caused this decision, it encourages inquiry into what conditions co-arose to make this decision possible? This moves us toward situated explainability a more holistic framework in which understanding is tied to networks of interdependent processes across technical, social, and ethical domains.

Such a model has practical implications. For instance, consider AI in predictive policing. Standard approaches might identify the variables that led the system to flag a particular neighborhood as high-risk. A systems-oriented, dependent-origination perspective would instead interrogate the sociohistorical conditions, feedback loops from previous policing patterns, data biases, and systemic inequalities that co-arise in producing such an output. This deeper causal web reframes the issue from merely technical to moral and political, thereby enhancing accountability (Benjamin, 2019).

Madhyamaka, Emptiness, and Non-Essentialist Design

Nagarjunas Madhyamaka school elaborates on dependent origination to assert that phenomena are empty of intrinsic essence because they arise only in dependence (Garfield, 1995). Applied to AI, this leads to a critique of essentialist thinking such as treating AI systems as possessing inherent intelligence or neutrality. Instead, Madhyamaka invites us to see AI as empty of fixed identity, its functionality arising from design choices, training data, institutional practices, and user engagement.

This insight has implications for the design of AI systems that resist essentialization and allow for interpretive flexibility. For example, rather than seeking one-size-fits-all explanations or fairness metrics, designers might cultivate pluralistic and participatory approaches that reflect the contingent nature of value-laden decisions. As Dignum (2019) notes, ethical AI requires systems that are not only transparent but responsive to evolving contexts and stakeholders precisely the kind of adaptive reasoning consistent with Buddhist interdependence.

Toward a Buddhist-Informed XAI Framework

A Buddhist logic for explainable AI does not replace existing technical methods but complements them with a more profound philosophical orientation. It encourages developers and regulators to shift from isolated causal attribution to a deeper appreciation of networked causal understanding. Rather than seeking single causes, they are prompted to recognize the co-arising nature of algorithmic behavior and social conditions, treating each decision as the outcome of interlinked contexts and interactions. In doing so, system design is approached with greater humility and responsiveness, resisting essentialist assumptions about how intelligence, fairness, or neutrality ought to manifest.

Moreover, this orientation fosters a non-dualistic awareness that bridges the gap between designer and user, system and environment, and cause and effect. It frames AI not as a detached technical artifact but as part of a continuum of human experience, conditioned by intentions, relationships, and institutional structures. This integrative vision is especially salient for communities seeking to embed AI systems within diverse cultural and ethical ecosystems.

Conclusion

The logic of dependent origination, when integrated with systems thinking, offers a powerful conceptual resource for developing explainable, ethical, and context-aware AI. By moving beyond isolated causal models and embracing interdependence, we gain not only better tools for technical explanation but also a richer moral vocabulary for navigating the sociotechnical challenges of AI. In this convergence of Buddhist logic and systems thinking lies the promise of a more compassionate, transparent, and inclusive AI future.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer.

Nāgārjuna. (1995). *The Fundamental Wisdom of the Middle Way: Nagarjunas Mulamadhyamakakarika* (J. L Garfield, Trans.). Oxford University Press.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

** Edited with ChatGPT for clarity; all views expressed are solely those of the authors.*

** The opinions expressed are the authors' own and do not reflect the views of AI Forum.*

Section II

Epistemology - Knowing

Tetralemma vs Dilemma: Integrating Eastern Logic into AI Reasoning

CHHEM Kieth Rethy[1] *, MD, PhD (Edu), and PhD (His)*

[1] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight examines how Eastern and Western logical traditions—specifically the tetralemma from Buddhist philosophy and the Western dilemma—offer contrasting yet complementary frameworks for reasoning. It explores how tetralemma-based logic, which embraces contradiction, ambiguity, and interdependence, can enrich AI systems limited by binary structures. Applications in natural language processing, ethical reasoning, and bias mitigation illustrate the value of multidimensional logic. The insight also addresses historical misinterpretations and highlights the role of mindfulness and cognitive flexibility in AI design. Integrating tetralemma logic opens pathways to more adaptive, inclusive, and ethically aware technologies, bridging ancient wisdom with future-facing innovation.

Introduction

This insight explores the fundamental contrast between Western and Eastern logical traditions by analyzing the concepts of dilemma and tetralemma. The aim is not only to highlight the divergence in their logical structures but also to consider their cultural, philosophical, and practical implications. The dilemma, rooted in Aristotelian thought, is foundational to Western reasoning, particularly in legal and scientific discourse. It offers clarity but often at the expense of nuance. In contrast, the tetralemma, derived from Buddhist Madhyamaka philosophy, offers a fourfold framework that allows for affirmation, negation, both, or neither. This multidimensional approach supports reasoning that accommodates ambiguity, contradiction, and interdependence. By integrating this ancient logic into contemporary contexts such as artificial intelligence (AI), we open new possibilities for developing culturally sensitive, ethically aware, and more adaptive AI systems. The dialogue between these traditions can thus enrich both philosophical inquiry and technological innovation.

Western Logic and the Dilemma

Western logic, dominated by Aristotelian principles, is anchored in binary thinking. The Law of the Excluded Middle and the Law of Non-Contradiction form the foundation of this tradition, asserting that any statement is either true or false, and cannot be both simultaneously. The dilemma epitomizes this logic, presenting two mutually exclusive options with no space for ambiguity. This form of reasoning has become deeply embedded in Western institutions from judicial systems to scientific methodologies where resolution and decisiveness are prioritized. However, this approach can be rigid and inadequate when faced with complex, fluid, or paradoxical realities. In ethical dilemmas, for example, binary thinking may oversimplify conflicting moral imperatives, offering limited pathways for reconciliation. As such, while Western logic is excellent for linear problem-solving, its rigidity can hinder broader and more integrative modes of reasoning.

Eastern Logic and the Tetralemma

In stark contrast, Eastern logic particularly the Buddhist framework articulated by Nagarjuna rejects rigid binaries. The tetralemma (catuskoti) proposes four possible responses to a proposition: it can be true (A), false (Not-A), both true and false, or neither true nor false. This nuanced structure reflects a philosophical commitment to the principle of sunyata (emptiness) and paticcasamuppada (interdependent origination). In this view, reality is seen not as a set of discrete entities but as a web of interdependent phenomena, constantly in flux. The tetralemmas openness to contradiction and paradox reflects the Buddhist goal of deconstructing fixed notions and promoting wisdom through experiential insight. Far from being irrational, this logic embraces complexity and offers tools for navigating it making it highly relevant to contemporary challenges that resist simplification, such as climate change, global health, nuclear energy and AI ethics.

Orientalist Misinterpretations and Contemporary Reassessment

The historical encounter between Western scholars and Buddhist logic was often marred by Orientalist biases. Early interpreters, constrained by Eurocentric assumptions, saw the tetralemma as either mystical or illogical, failing to recognize its intellectual rigor. Influenced by colonial attitudes, many viewed non-Western traditions as inferior, irrational, or merely spiritual, rather than as legitimate philosophical systems. This misreading not only distorted understanding but also limited the exchange of ideas between traditions. However, contemporary thinkers in comparative philosophy and postmodern theory have begun to reassess these earlier dismissals. These reevaluations validate Buddhist logic as a sophisticated system capable of addressing modern complexities, and they call for greater epistemological humility when engaging with non-Western knowledge systems.

Applications of Tetralemma Logic in AI Design

Modern AI systems predominantly operate on binary logic, which while efficient, is limited in scope when confronting ambiguity, contradictions, and ethical gray zones. Integrating tetralemma-based reasoning into AI can revolutionize several fields:

1. Multidimensional Reasoning: Traditional machine learning algorithms categorize data in binary terms. By implementing a four-fold logic model, AI can process and respond to situations involving conflicting information or partial truths. This approach is especially useful in areas like medical diagnostics, where symptoms do not always conform to linear models.
2. Natural Language Processing: Human language is inherently ambiguous. Idioms, metaphors, and cultural references often defy binary classification. Tetralemma-based NLP models can improve semantic comprehension and nuance recognition, enhancing translation, sentiment analysis, and human-AI interaction.
3. Ethical Reasoning: Autonomous systems often face moral dilemmas where binary decisions (e.g., harm one party to save another) are inadequate. Tetralemma logic allows AI to hold conflicting ethical perspectives and consider layered outcomes, moving toward more responsible decision-making.
4. Bias Mitigation and Fairness: Binary classifiers often embed and reinforce societal biases. Tetralemma-inspired systems can recognize gray areas, accommodate diverse identities, and promote equity in algorithmic outcomes.
5. Complex Problem-Solving: Fields such as climate modeling, economic forecasting, and conflict resolution benefit from models that

can entertain contradiction and uncertainty. The tetralemmas framework enables AI to offer probabilistic insights rather than oversimplified answers, improving real-world adaptability.

Mindfulness and Cognitive Flexibility

Mindfulness, as cultivated in Buddhist traditions, develops the mental capacity to observe without attachment and respond with clarity amid uncertainty. Bhikkhu Analayo has argued that mindfulness supports a cognitive stance beyond binary fixation, promoting adaptability and contextual understanding. In AI development, this translates to systems that are not only reactive but contextually aware and reflective. A mindful AI, informed by tetralemma logic, would be capable of withholding judgment, weighing multiple interpretations, and adapting its responses based on shifting conditions. Such systems would better reflect the nuanced thinking required in healthcare, education, diplomacy, and social services. Moreover, cognitive flexibility is a crucial attribute in both human intelligence and AI, enabling resilience in the face of change. Embedding this principle in AI systems could lead to technologies that are more empathetic, fair, and capable of ethical discernment in real-world contexts.

Challenges and Future Directions

Despite its potential, incorporating tetralemma logic into AI presents technical and philosophical challenges. Existing AI systems are deeply entrenched in binary architecture, from hardware design to software logic gates. Implementing fourfold reasoning will require new algorithmic paradigms, neural network architectures, and training datasets that account for ambiguity, contradiction, and multiplicity. Furthermore, there must be a cultural shift in the AI development community to recognize and integrate non-Western philosophies as valid sources of innovation. Collaboration across disciplines computer science, philosophy, cultural studies, and cognitive science is essential. Pilot projects in NLP, ethics simulations, and adaptive systems could serve as testbeds for tetralemma-inspired models. Over time, these efforts could establish a new standard for designing AI that is not only intelligent but wise, capable of navigating the moral and cognitive complexities of human life.

Conclusion

The dilemma and tetralemma represent two distinct yet complementary paradigms of logic. While the Western binary model excels in structure and clarity, it often fails to address ambiguity and paradox. The Buddhist tetralemma offers a robust alternative one that embraces complexity, encourages open-ended inquiry, and aligns with the relational nature of reality. In the context of AI, integrating tetralemma-based reasoning allows us to design systems that better reflect the diversity, uncertainty, and ethical depth of human experiences. As technology continues to permeate global societies, drawing on ancient wisdom can ensure that our tools serve not only functional needs but also human values. This synthesis of East and West, logic and mindfulness, opens a path toward a more just, inclusive, and thoughtful digital future.

References

Analayo, B. (2021). Beyond the Limitations of Binary Thinking: Mindfulness and the Tetralemma. *Mindfulness, 13*, 1410-1417. https://doi.org/10.1007/s12671-021-01678-6

Clarke, J. J. (1997). *Oriental Enlightenment: The Encounter Between Asian and Western Thought*. Routledge.

Quli, N. E. (2008). Western Self, Asian Other: Modernity, Authenticity, and Nostalgia for "Tradition" in Buddhist Studies. *Journal of Buddhist Ethics, 15*, 136.

Said, E. W. (1978). *Orientalism*. Pantheon Books.

Tanaka, K. (2013). Buddhist Philosophy of Logic. In S. Emmanuel (Ed.), *Blackwell Companion to Buddhist Philosophy* (pp. 320-330). Wiley-Blackwell.

** Edited with ChatGPT for clarity; all views expressed are solely those of the author.*

** The opinions expressed are the author's own and do not reflect the views of AI Forum*

Catuskoti and Non-Binary Reasoning in Artificial Intelligence

CHHEM Kieth Rethy[1] *, MD, PhD (Edu), and PhD (His)*

[1] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight examines the potential of catuskoti—Buddhist fourfold logic—to enhance artificial intelligence beyond the limits of binary reasoning. Rooted in Nagarjuna's Madhyamaka philosophy, catuskoti allows for propositions to be true, false, both, or neither, offering a more nuanced framework for modeling ambiguity and ethical complexity. The essay explores applications in natural language processing and moral decision-making, where traditional logic often falls short. By integrating this non-binary paradigm, AI systems can become more contextually aware, culturally sensitive, and philosophically enriched. Catuskoti thus offers a meaningful pathway toward inclusive, adaptive, and ethically responsive artificial intelligence.

Introduction

In the expanding field of Artificial Intelligence (AI), logic forms the foundational structure for computation, inference, and decision-making. Classical logic, heavily influenced by Aristotle's system of binary oppositions, predicates reasoning on the principle of the excluded middle: every proposition must be either true or false, but not both or neither. However, in real-world scenarios, particularly those involving ethics, cultural interpretation, and natural language, this binary framework can be too rigid.

Enter catuskoti, or the Tetralemma, a logical structure rooted in Nagarjuna's Madhyamaka tradition of Buddhist philosophy. Catuskoti posits four possibilities for any given proposition: (1) it is true, (2) it is false, (3) it is both true and false, and (4) it is neither true nor false. This framework challenges Aristotelian logic and offers a richer paradigm for modeling ambiguity and paradox, which are crucial features of human cognition and ethical judgment. This chapter explores the relevance of catuskoti for AI design, particularly in natural language processing (NLP) and machine ethics, where classical logics often reach their limits.

Aristotelian Logic and Its Limits

Aristotle's contribution to formal logic, especially through his Organon, established the dominant paradigm in Western thought. The principle of non-contradiction (a statement cannot be both true and false at the same time) and the law of the excluded middle (a statement must be either true or false) have become axioms in classical logic systems, including those used in programming languages and most AI architectures (Smith, 2003). While this system sustains much of computational logic, it struggles in domains where ambiguity, contradiction, or incompleteness is not a bug but a feature.

Consider an AI tasked with moral decision-making in autonomous vehicles. When faced with an ethical dilemma such as choosing between harming one pedestrian or multiple passengers, classical binary logic cannot always provide a satisfying answer. There may be multiple overlapping truths or no clear truth at all. Similarly, in NLP tasks like sentiment analysis, a sentence can convey conflicting emotions such as both joy and sorrow, which resists binary classification.

The Catuskoti Framework

The catuskoti framework originates from Nagarjuna's Mulamadhyamakakarika, a foundational text of the Madhyamaka school of Mahayana Buddhism. The logic operates on four poles:

1. P (a proposition is true)
2. $\neg P$ (the proposition is false)
3. $P \land \neg P$ (the proposition is both true and false)
4. $\neg(P \lor \neg P)$ (the proposition is neither true nor false)

Rather than affirming a reality behind propositions, catuskoti emphasizes the limits of conceptualization and encourages a deconstruction of absolute beliefs (Garfield, 1995). It challenges the ontological assumptions that underpin binary logic, replacing determinacy with interdependence and emptiness (sunyata). In computational terms, this opens up a space for multi-valued logic systems that can accommodate contradictory, incomplete, or culturally contingent inputs.

This kind of non-binary approach also aligns with mindfulness-based cognitive flexibility. As Anālayo (2021) notes, the tetralemma reflects an awareness of complexity and conditionality that extends beyond rigid dualities, fostering insight into intuitive awareness.

Application in Natural Language Processing

NLP provides fertile ground for implementing catuskoti-inspired models. Languages are inherently fuzzy and context-dependent. A single utterance can carry irony, sarcasm, ambiguity, or layered meanings that defy binary classification.

Take the sentence: "I'm fine." In classical sentiment analysis, the model might categorize it as positive. However, depending on intonation, context, or cultural norms, it could signal distress, resignation, or even anger. A catuskoti-based NLP model could allow the following classifications simultaneously:

- True (positive sentiment)
- False (actually negative sentiment)
- Both (mixed emotions)
- Neither (neutral, ambiguous, or contextually unclear)

Such a system could use probabilistic weights or fuzzy logic to rank each corner of the tetralemma, enabling more nuanced interpretation and response. Research in non-classical logic already points in this direction (Beziau, 2012). Catuskoti offers a philosophically grounded and culturally enriched logic to enhance these approaches.

Ethical Reasoning in AI

AI ethics, especially in high-stakes fields like healthcare, warfare, and autonomous transport, demands more than deterministic computation. It requires interpretative judgment. Traditional logic falters when dilemmas cannot be resolved with right or wrong binaries.

Consider AI in medical triage. Should the AI prioritize the young over the elderly? Or the person with a better prognosis over one with greater social responsibility? Such decisions involve values, contradictions, and cultural expectations that vary significantly. A catuskoti model could accommodate ethical pluralism. It might allow an AI to recognize that a decision is simultaneously justifiable and unjustifiable, depending on the moral framework invoked. Instead of resolving the contradiction, the AI would surface it, flagging the ethical ambiguity for human oversight or adaptive learning. This aligns with post-normal science, which calls for decision-making under conditions of uncertainty, complexity, and value conflict (Funtowicz & Ravetz, 1993).

Toward Culturally Sensitive AI

One major criticism of generative AI development is its epistemological bias. Many AI systems are trained and modeled using Western logic, datasets, and moral assumptions. Integrating catuskoti into AI design contributes to the broader project of epistemic pluralism. It not only diversifies computational models but also invites cross-cultural perspectives on what it means to think, decide, and act.

For Southeast Asian cultures influenced by Buddhism or for Indian traditions shaped by Nyaya and Jain logic, reasoning is not necessarily exclusionary. By embedding non-binary frameworks like catuskoti, AI can better resonate with the lived ethical and epistemic realities of these societies. This has implications for education, diplomacy, and cross-border AI regulation.

Conclusion

Catuskoti offers an elegant yet radical alternative to the binary logic that dominates AI today. Rooted in Buddhist philosophy and refined through centuries of epistemological reflection, the tetralemma provides a powerful framework for modeling ambiguity, contradiction, and ethical complexity. Its integration into AI, especially in NLP and ethical decision-making, can lead to systems that are not only technically sophisticated but also culturally respectful and philosophically enriched.

As AI expands into every domain of human life, drawing on non-Western logics like catuskoti is not merely an academic exercise. It is a necessary step toward truly inclusive, ethical, and intelligent machines.

References

Anālayo, B. (2022). Beyond the limitations of binary thinking: Mindfulness and the tetralemma. *Mindfulness, 13*, 1410–1417.

Béziau, J. -Y. (2011). A new four-valued approach to modal logic. *Logique et Analyse, 54*(213), 109-121.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures, 25*(7), 739- 755.

Garfield, J. L. (1995). *The fundamental wisdom of the middle way: Nāgārjuna's Mūlamadhyamakakārikā.* Oxford University Press.

Smith, R. (2022). Aristotle's logic. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2022 ed.). Metaphysics Research Lab, Stanford University.

** Edited with ChatGPT for clarity; all views expressed are solely those of the author.*

** The opinions expressed are the author' s own and do not reflect the views of AI Forum.*

Sunyata and Epistemic Humility in AI Design

CHHEM Kieth Rethy[1] *, MD, PhD (Edu), and PhD (His)*

[1] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight explores how the Buddhist concept of sunyata (emptiness) can inform the ethical design of artificial intelligence. Emphasizing the relational and impermanent nature of knowledge, sunyata supports an attitude of epistemic humility—recognizing that data, models, and categories are context-dependent rather than fixed truths. The essay critiques AI systems built on rigid classifications and calls for adaptive, inclusive, and culturally sensitive alternatives. By integrating sunyata into AI thinking, we shift from control toward care, from certainty to openness. The result is not a wiser machine, but a more humane approach to intelligence, centered on responsiveness, reflection, and ethical restraint.

Introduction

As artificial intelligence increasingly shapes how we perceive, process, and act upon information, it raises deep philosophical and ethical questions about knowledge, certainty, and control. This insight introduces the Buddhist concept of sunyata (emptiness) as a foundation for rethinking AI design through the lens of epistemic humility. Rather than viewing knowledge as fixed and objective, sunyata encourages us to see it as relational and contingent. Applying this to AI invites designers to approach their systems not as neutral or omniscient, but as fallible, contextual tools—thus opening the path for more ethical, mindful, and responsive technology development.

What is Sunyata?

In Mahayana Buddhist philosophy, sunyata translated as "emptiness" is a central concept that signifies the absence of inherent, independent existence in all phenomena. It does not mean that nothing exists, but rather that things do not exist on their own, with a fixed essence. Instead, all things arise through dependent origination (paticcasamuppada), meaning they are formed through relationships, conditions, and context. Nagarjuna, a key figure in the Madhyamaka school, articulates this view by showing that anything we analyze eventually reveals its dependence on other things.

This view stands in contrast to essentialist thinking the belief that objects, people, or ideas have an unchanging core. Sunyata invites us to view reality as fluid and interconnected. Rather than assuming stability and permanence, it encourages us to understand that meaning is shaped by perspective and is always open to change. This philosophy, while ancient, has profound implications for how we think about truth, knowledge, and systems especially in a world shaped by data and algorithms.

Implications for Knowledge: Arrogance vs. Humility

Applied to the realm of knowledge, sunyata encourages epistemic humility: the understanding that our beliefs and knowledge claims are always partial, context-dependent, and subject to revision. Epistemic humility does not mean giving up on knowledge, but recognizing that our models and theories are tools not final truths. It calls for a willingness to admit uncertainty, to listen to others, and to revise one's views in light of new information or perspectives.

On the other hand, epistemic arrogance is the opposite stance. It involves an overconfidence in the correctness or completeness of one's knowledge. Arrogance assumes that categories like truth or objectivity are fixed, ignoring the ways in which these are shaped by culture, language, and power. In Buddhist terms, this attitude reflects "avidya," or ignorance not merely a lack of information, but the fundamental misunderstanding of the nature of reality. When we mistake our temporary, context-specific knowledge for permanent truth, we risk acting with overconfidence and closing ourselves off to deeper understanding.

This tension between humility and arrogance is not abstract. It plays out in practical and consequential ways, particularly in the field of artificial intelligence, where systems are built to classify, decide, and predict based on historical data and human-defined categories.

Application to AI Design

Today's AI systems, especially those using supervised learning, are built on the assumption that categories are fixed and clearly defined. These systems are trained to detect patterns and make predictions based on labeled datasets. However, labels such as gender, race, risk, or sentiment are not objective truths they are shaped by history, culture, and social context. They are often fluid and contested.

From the perspective of sunyata, treating these categories as fixed is deeply problematic. It gives a false sense of objectivity and can cause real harm. For instance, when AI systems are trained on biased or unfair data, they don't merely repeat those patterns they present them as if they were neutral and valid outcomes. This not only reinforces existing inequalities but also hides the fact that the systems foundation may be flawed. Such design reflects epistemic arrogance: the belief that models mirror reality exactly, ignoring how they simplify or distort complex social dynamics.

To bring sunyata and humility into real-world AI practices, we can follow four guiding habits. First, treat AI models as useful tools, not absolute truths, and be willing to revise them when needed. Second, ensure AI systems are adaptable to different social and cultural contexts, rather than applying one-size-fits-all solutions. Third, regularly examine how data is collected and used, and include diverse perspectives to uncover hidden biases. Finally, keep people involved in the decision-making process, especially when moral or cultural values are at stake. Together, these practices help ensure that AI development remains responsive, inclusive, and ethically grounded acknowledging its limits as well as its potential.

Luciano Floridi, in his book "The Logic of Information" (2019), argues that philosophy must act as conceptual design shaping how we create, understand, and use information systems. From this view, AI systems are not just tools, but frameworks of meaning. Similarly, Kate Crawfords "Atlas of AI" (2021) exposes the hidden labor, material extraction, and political structures embedded in AI systems, challenging the illusion of neutrality. Crawfords work reminds us that epistemic humility must also include awareness of the broader systems that shape and sustain knowledge creation including labor, politics, and ecological impact.

Ethical AI, from a sunyata perspective, is not merely a matter of following rules or applying fairness metrics. It involves a deeper orientation toward openness, mindfulness, and compassion. Just as Buddhist ethics encourages non-attachment to views, ethical AI calls for letting go of the desire for complete control or certainty. This means creating space for questioning, for dialogue, and for transformation

Importantly, we must resist the illusion that machines can be intelligent or wise in the same way humans can. The ultimate aspiration is not to build wise machines which remains beyond our reach even today but to build systems that can help augment human intelligence and support the cultivation of human wisdom. In Buddhism, wisdom (prajna) is not simply about knowing facts, but about seeing clearly, acting ethically, and embracing uncertainty. A humane AI should reflect this vision: not flawless, but aware of its own limits and responsive to human needs.

Conclusion

Incorporating sunyata into AI ethics and design offers a powerful counterbalance to the arrogance and rigidity that often characterize technological systems. It challenges the assumption that knowledge is fixed and that models are neutral. It reminds us that reality is complex, changing, and relational.

Embracing epistemic humility in AI does not mean abandoning the pursuit of better technology. It means pursuing it with care, openness, and responsibility. It means recognizing that our tools like our concepts must be held lightly. They can guide us, but they must never blind us.

A truly humane AI, shaped by sunyata, is one that honors interdependence, cultivates ethical awareness, and remains humble before the richness and mystery of the human experience.

References

Nagarjuna. (1995). *The Fundamental Wisdom of the Middle Way: Nagarjuna's Mulamadhyamakakarika* (J. L. Garfield, Trans.). Oxford University Press.

Dreyfus, G. B. J. (2003). *The Sound of Two Hands Clapping: The Education of a Tibetan Buddhist Monk*. University of California Press.

Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford University Press

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press

** Edited with ChatGPT for clarity; all views expressed are solely those of the author.*

** The opinions expressed are the author's own and do not reflect the views of AI Forum.*

Anatta and the Myth of AI Autonomy: Towards a Processual and Relational Ethic for Artificial Intelligence

CHEAB Sambath[1] *, MD*

CHHEM Kieth Rethy[2] *, MD, PhD (Edu), and PhD (His)*

[1] ***CHEAB Sambath*** *is a Lecturer at University of Puthisastra, Cambodia*

[2] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight challenges the notion of AI autonomy through the lens of anatta (non-self) in Theravada Buddhism. Drawing on the pancakkhandha framework, the authors argue that AI systems, like human beings, are composite, conditioned processes without inherent agency. Misconceptions of machine autonomy obscure human accountability and reinforce ethical risks. By adopting a processual and relational view, AI design can be reoriented toward transparency, responsibility, humility, and contextual sensitivity. This Buddhist perspective offers a corrective to technological essentialism, encouraging a human-centered, ethically grounded approach to artificial intelligence—one that recognizes interdependence, intention, and the limits of algorithmic control.

Introduction

Artificial Intelligence (AI) is frequently characterized using terms that imply autonomy—such as "self-learning," "autonomous agents," or "intelligent decision-makers." While these descriptors may serve practical or rhetorical purposes, they risk promoting fundamental misconceptions about the nature of AI, particularly the assumption that AI systems possess stable, independent agency. From a Theravada Buddhist perspective, this mirrors *sakkaya-ditthi*, the erroneous belief in a permanent, autonomous self.

This chapter utilizes the doctrine of *anatta* (non-self) and the analytical model of the *pancakkhandha* (five aggregates) to interrogate such assumptions. It posits that both human beings and artificial systems should be reconceived as complex, conditioned processes rather than self-sustaining agents. By reframing our understanding in this manner, we can establish a more ethically grounded and philosophically coherent approach to AI development and governance.

Anatta in Theravada Buddhism: Deconstructing the "Self"

Anatta—the rejection of a permanent, independent self—constitutes a central tenet of Buddhist philosophy, alongside impermanence (*anicca*) and unsatisfactoriness (*dukkha*) (Ñāṇamoli & Bodhi, 1995). According to canonical texts, what is conventionally referred to as a "person" is merely a practical designation (*pannatti*) imposed upon an ever-changing interplay of five aggregates (Bodhi, 2003): *rupa* (physical form), *vedana* (feeling), *sanna* (perception), *samkhara* (volitional formations), and *vinnana* (consciousness). These aggregates are inherently impermanent, causally conditioned, and devoid of intrinsic self-nature (Gethin, 1992).

Western philosophical traditions, which often valorize individuality and autonomy, may assert that AI systems could evolve genuine autonomy through complexity or emergent properties. However, this line of reasoning overlooks the fundamentally composite and interdependent structure described by the *pancakkhandha*, which denies the existence of intrinsic identity or independent agency (Harvey, 1995).

Deconstructing AI Autonomy Through Anatta

Contemporary narratives frequently anthropomorphize AI, inadvertently reinforcing beliefs in its intrinsic agency. In reality, AI systems are composite constructs: they consist of training datasets, algorithms, hardware, human inputs, environmental feedback, and the intentional designs of developers. No single component, nor their integration, confers autonomous identity.

Advocates of AI autonomy often argue that sufficiently advanced machine learning could lead to functional autonomy indistinguishable from human agency (Clark, 2016). Yet, from a Theravada perspective, genuine autonomy presupposes *citta* (consciousness) and *cetana* (volition), attributes entirely absent in machines. Therefore, even the most advanced AI remains a conditioned process, inherently reliant on human input and oversight.

Why Process Ontology Is Essential for Understanding AI

The Abhidhamma tradition of Theravada Buddhism conceptualizes reality in terms of interdependent processes rather than enduring substances or entities, aligning it with modern process ontologies (Bodhi, 2000). Applying this perspective to AI reveals that these systems should not be treated as discrete entities but as dynamic, evolving processes shaped by numerous conditions.

Real-world examples, such as facial recognition systems and autonomous vehicles, underscore this point. These technologies do not possess self-generated moral agency; instead, they are constituted by continuously shifting interactions among data, algorithms, and social deployment contexts. Understanding AI as processual reinforces ethical responsibility and ensures that accountability remains centered on human actors.

The Ethical Risks of Anthropomorphic Projections

Attributing inherent autonomy or agency to AI carries substantial ethical hazards. For instance, interpreting algorithmic bias as a defect inherent in the AI itself obscures the underlying human decisions embedded in dataset curation and system design, thereby deflecting responsibility (Benjamin, 2019).

Moreover, anthropomorphic projections foster misplaced trust and inflated expectations, generating moral and functional confusion. The principle of *anatta* serves as a philosophical corrective, reminding us that all agency and responsibility originate with human beings. This recognition is crucial for upholding ethical accountability throughout the AI development lifecycle.

Towards a Transparent, Responsible, Humble, and Relational Ethic for AI Design

Applying *anatta* in the context of ethical AI development entails commitment to four interrelated principles: transparency, explicit responsibility, epistemic humility, and relational ethics. First, **deep transparency** necessitates the clear and comprehensive disclosure of AI design methodologies, data provenance, and algorithmic structures to relevant stakeholders (Dignum, 2019). Second, **explicit human responsibility** demands ongoing recognition and unambiguous attribution of ethical accountability to developers and institutional bodies. Third, **epistemic humility** involves a sober acknowledgment of AI's limitations, fostering a culture of cautious, iterative design and robust governance mechanisms. Finally, **relational ethics** emphasizes the importance of considering the AI system's interactions with users, social systems, and environmental contexts, thereby promoting ethically beneficial outcomes through deliberate and reflective design.

Comparative Contextualization

Within the broader framework of Buddhist approaches to AI, Theravada Buddhism offers a unique contribution through its analytic rigor and ontological clarity. While Mahayana Buddhism emphasizes universal compassion and Vajrayana highlights the symbolic and transformative potential of technology, Theravada grounds ethical critique in its systematic deconstruction of selfhood. This doctrinal emphasis provides a foundational ethical counterpoint to prevailing myths of AI autonomy and complements the moral insights of other Buddhist traditions.

Conclusion: Embracing Process, Intent, and Responsibility

Reframing AI through the lens of *anatta* reveals that presumed machine autonomy is a conceptual illusion. AI systems are complex, dependent processes fundamentally shaped by human intentions and sociotechnical environments. Recognizing this truth fosters ethical stewardship, ensuring that AI development remains human-centered, transparent, and accountable.

Ultimately, responsible AI design demands attentiveness to interdependence, intention, and the consequences of technological action—dimensions profoundly illuminated by the philosophical insights of Buddhist thought.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Bodhi, B. (Trans.). (2000). *The connected discourses of the Buddha: A translation of the Saṁyutta Nikāya*. Wisdom Publications. (Original work published 1999)

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer.

Gethin, R. M. L. (1992). *The Buddhist path to awakening: A study of the bodhi-pakkhiyā dhammā*. Brill.

Harvey, P. (1995). *The selfless mind: Personality, consciousness and Nirvāṇa in early Buddhism*. Curzon Press.

Ñāṇamoli, B., & Bodhi, B. (Trans.). (1995). *The middle length discourses of the Buddha: A translation of the Majjhima Nikāya*. Wisdom Publications.

Nārada, M., & Bodhi, B. (Trans.). (1999). *A comprehensive manual of Abhidhamma: The Abhidhammattha Sangaha of Ācariya Anuruddha*. Buddhist Publication Society. (Original work published 1993)

* *Edited with ChatGPT for clarity; all views expressed are solely those of the authors.*

* *The opinions expressed are the authors' own and do not reflect the views of AI Forum.*

Section III

Ethics - Acting

Can AI Be Morally Responsible from the Mahayana Buddhist Lens?+

CHHEM Kieth Rethy[1], MD, PhD (Edu), and PhD (His)

+*Presented at Aristole & Confucious Symposium, Athens*

[1]***CHHEM Kieth Rethy** is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight explores the moral responsibility of AI through the lens of Buddhist philosophy, emphasizing human accountability in its development and use. While AI lacks consciousness, intentions, and wisdom—key attributes for moral responsibility in Buddhism—humans possess these qualities and must ensure AI aligns with ethical principles like mindfulness, wisdom, non-harm, and compassion. By integrating Buddhist teachings, such as the Four Sublime States of Mind and the law of karma, we can foster responsible AI that prioritizes societal well-being and mitigates harm. Ultimately, humanity must reflect on whether greater peril lies in machines thinking like humans or humans thinking like machines.

Introduction

I find myself influenced by two philosophical traditions: Buddhism and science. However, I use Buddhist principles to critically examine the positivist paradigm of science. As a physician, I have sworn the Hippocratic oath, which entails a commitment to doing my utmost to avoid causing harm. Since 1993, I have been utilizing Expert System, powered by AI, to evaluate bone cancers. However, at that time, the AI system was not as advanced as it is today. Recently, I have dedicated approximately 300 hours to testing various generative AI tools in order to gauge the extent of progress made since then. The remarkable advancements in AI technology have significantly expanded its range of applications, bringing about positive impacts on our lifestyle. However, these advancements have also given rise to certain negative consequences, raising significant moral and ethical concerns. This has led me to explore a crucial question: Can AI be considered morally responsible when viewed through the lens of Buddhism?

In Buddhism, moral responsibility refers to the understanding that individuals are accountable for the consequences of their thoughts, words, and deeds. This accountability arises from the belief in the law of karma, which states that intentional actions have consequences that affect one's present and future experiences. Mahayana Buddha's teachings elucidate that moral responsibility can be comprehended through the integration of concepts such as mindfulness, wisdom, consciousness, intention, and the principle of non-harm.

Mindfulness (*Sati sampajanna*) - To cite

Mindfulness is an important concept in Buddhism, and involves being fully present and aware of one's thoughts, feelings, and actions in the present moment. It promotes the understanding of the ethical implications of choices and acting with clarity and wisdom to minimize harm. In the Satipatthana Sutta, mindfulness supports self-awareness and conscious decision- making, guiding individuals in considering the impact of AI on sentient beings and the environment. Integrating mindfulness into AI development cultivates compassion, ethical discernment, and a commitment to the welfare of all beings. By combining mindfulness and AI, individuals make conscious, ethical choices while staying attuned to their higher purpose and the well-being of others.

Wisdom (*panna*) – To cite

Rooted in Buddhist teachings, wisdom as taught in the Diamond Sutra (Vajracchedika Prajnaparamita Sutra), provides a valuable foundation for AI ethics during the designing of AI algorithms and the deployment of AI tools. It emphasizes a comprehensive grasp of the interconnectedness and consequences of AI systems, guiding individuals to make ethical choices that prioritize the well-being of individuals and society. By cultivating wisdom in the realm of AI ethics, we can navigate the ethical challenges and complexities of AI during the design and deployment stages, ensuring that our actions align with values of compassion, fairness, and the reduction of suffering.

Consciousness (*vinnana*) – To cite

The Buddhist concept of Panca Skanda, or Five Aggregates for others to identifies the components of human existence as form (*rupa*), feeling (*vedana*), perception (*sanna*), volition (*sankhara*), and consciousness (*vinnana*). It reveals that AI lacks the attributes for consciousness and moral accountability, unlike humans. This highlights human responsibility, raises concerns about the moral status of AI, and underscores the importance of integrating Buddhist principles into AI ethics.

Intention (*Cetana*) – To cite

Intentions on the other hand, play a crucial role in determining the ethical nature of our actions. In Buddhism, the concept of right intentions is a part of the Noble Eightfold Path, which guides individuals towards ethical conduct and mindfulness. Right intentions involve renouncing harmful intentions, fostering benevolence, compassion, and promoting non-violence in one's thoughts and actions. By cultivating intentions, AI practitioners can work towards creating technologies that align with societal values and contribute to a more ethical and mindful use of AI.

Non-harm (*Ahimsa*) – To cite

The principle of non-harm in Buddhism emphasizes refraining from causing harm to oneself and others, both physically and mentally. It involves treating all sentient beings with compassion, empathy, and respect, and aiming to promote well-being and happiness while minimizing suffering. In the context of AI, practitioners should integrate the principle of non-harm into their work by being mindful of potential negative impacts, promoting ethical decision-making, prioritizing the needs of users, responsibly handling data, and maintaining accountability and transparency.

Four Sublime State of Mind (*Brahmavihara Dhamma*)

This concept can be applied to AI development and be used to promote ethical responsibility. Loving-kindness (*metta*) inspires the creation of AI systems that prioritize the welfare and benefit of individuals and society as a whole. Compassion (*karuna*) drives the design of systems that alleviate suffering and promote fairness, justice, and equality. Sympathetic joy (*mudita*) encourages AI developers to celebrate and support positive advancements and contributions made by AI systems. Equanimity (*upekkha*) guides decisions and interactions with AI systems with impartiality and consideration of the broader consequences. Integrating the principles of Brahmavihara Dhamma ensures that AI technologies are aligned with good human values.

Given that AI systems do not possess consciousness or intentions, it becomes apparent that they cannot be held morally responsible for their actions. Responsibility, therefore, falls upon the humans who create and utilize AI. It is the human creators and users who possess consciousness, intentions, and the capacity for mindfulness, non-harm, and compassion. It is universally accepted that every human being has a moral responsibility toward themselves and towards other beings.

Human Responsibility

The parable found in the Itivuttaka Buddhist scriptures portrays the qualities of individuals in relation to the Three Jewels: the Buddha, the Dharma (teachings), and the Sangha (monastic community). The concept of Heightened Birth (*Atijataputra* **អតិជាតបុត្រ**) symbolizes individuals who embrace the principles of the Three Jewels and approach AI development and utilization with a strong sense of responsibility and ethics. They prioritize the well-being of all and strive to minimize harm throughout the entire process of designing, developing, and deploying AI systems. On the other hand, the concept of Similar Birth (*Anujataputra* **អនុជាតបុត្រ**) encompasses individuals, including both parents and children, who also embrace the principles of the Three Jewels. This group acknowledges the potential harm that AI can cause and accepts their responsibility in comprehending and mitigating those risks. They proactively seek knowledge about AI technology, its impact on various aspects of society, and its potential consequences. Although they may not be directly involved in AI development, they are committed to staying informed and making responsible choices when interacting with AI systems. Lastly, the concept of Low Birth (*Avajataputra* **អវជាតបុត្រ**) represents individuals who disregard the principles of the Three Jewels, resulting in a lack of wisdom and moral understanding regarding AI. Their behavior could involve exploiting vulnerabilities, utilizing AI for malicious purposes, or disregarding the impact on individuals and society, and the ethical implications of their actions.

In the context of AI, this parable serves as a reminder of the crucial role of human responsibility. It highlights the significance of responsible AI development, utilization, and decision-making, underscoring the potential consequences of neglecting ethical considerations. The parable calls upon individuals and organizations to adopt a heightened sense of responsibility, thoroughly understanding the implications of AI, and actively working towards creating and using AI systems that align with ethical principles while prioritizing the well-being of all.

Can AI Be Morally Responsible?

We can reflect about AI in the broader context of dependent origination (paticcasamuppada), a concept from Buddhism that describes the interconnected nature of all phenomena and the link between cause and effect. This key concept encourages a mindful and ethical approach to AI, recognizing the interdependencies between AI systems, humans, society, and the environment.

Conclusion

AI technology, while lacking consciousness, intentions, mindfulness, and wisdom as discussed in Buddhist concepts, cannot be morally responsible. Humans, due to their possession of consciousness and intention, must assume the moral responsibility for ensuring the safe development and implementation of AI systems. Finally, the integration of Buddhist principles can guide the development and application of AI technology towards fostering the overall well-being of humankind. Allow me to conclude my discussion by posing another philosophical question: Which is more perilous for humanity—a machine capable of thinking like a human, or a human who thinks like a machine?

**References available upon request*
**The views expressed are the author's own and do not reflect the views of the AI Forum*

The Idea of Compassionate AI: Embodying Buddhist Ethics in Technology[+]

CHHEM Kieth Rethy[1], MD, PhD (Edu), and PhD (His)

[+]*Presented at Aristole & Confucious Symposium, Qufu-Beijing*

[1]***CHHEM Kieth Rethy** is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This article explores the intersection of artificial intelligence (AI) and Buddhist ethics, emphasizing the potential for AI systems to simulate compassionate and ethical behavior. Grounded in Buddhist principles like the Middle Way (Madhyama), non-violence (Ahimsa), compassion (Karuna), and intentionality (Cetana), it highlights how these teachings can inform the development of ethical AI. While AI lacks consciousness and intrinsic moral agency, incorporating ethical frameworks, empathy simulation, and transparency can guide its alignment with human values. The essay underscores the importance of interdisciplinary collaboration to ensure AI development reflects principles of compassion and responsibility, fostering societal well-being and ethical advancement.

Introduction

Reflecting on my journey, spanning from medical school to an evolving career in radiology, computer science, and AI, I have consistently sought to merge technological advancement with ethical considerations. After completing my medical school and residency program in radiology, my interest in the burgeoning field of technology led me to pursue a certificate in computer science in 1984 and Neuroanatomy. By 1993, while teaching at McGill University, I experimented with AI as a tool for detecting and identifying various bone cancers using an expert system, and later tested voice recognition automatic reporting for radiology at the National University of Singapore in 1998. Despite their initial disappointments, these early ventures into AI laid the groundwork for my understanding of its potential and limitations.

The 2011 Fukushima nuclear disaster was a pivotal moment, teaching me about the dual-use nature of technology and the critical importance of ethics, lessons I applied to AI development. By 2018, I was advising an AI forum on incorporating Buddhist ethics into AI, recognizing the profound need for ethical guidance. My leadership in publishing a policy framework for AI in Cambodia in 2023 further reinforced this commitment. This same year, I founded the Buddhist AI Ethics at CamTech University, driven by a deep conviction that Buddhist values can significantly influence the responsible development of AI technology. These experiences have profoundly shaped my dedication to exploring the integration of Buddhist values in AI technology development.

In the realm of artificial intelligence (AI), advancements continue to reshape industries and societies, yet questions of ethical conduct and moral responsibility loom large. As AI systems become more integrated into daily life, the need to imbue them with ethical principles, such as those found in Buddhist teachings, becomes increasingly pertinent. This essay explores the intersection of compassionate AI and Buddhist ethics, employing inductive, deductive, and abductive reasoning to outline possibilities and challenges in this evolving field.

Artificial intelligence operates on programmed algorithms designed to process data and execute tasks efficiently. Unlike humans, AI lacks subjective experience, self-awareness, and moral cognition. It functions based on syntactical rules and predefined inputs, unable to introspect or understand ethical nuances inherently. Observing AI's operational characteristics leads to the conclusion that it operates without consciousness as understood in human terms. This limitation raises fundamental questions about AI's capacity for ethical decision-making and its ability to embody compassion.

Consciousness, rooted in complex neural processes and biological structures, is essential for human moral agency and ethical behavior. AI, lacking these biological substrates, cannot replicate the core elements of consciousness. AI systems rely on dry silicon for their hardware and software, while the brain's neural network, known as wetware, operates on water-based materials that enable its plasticity and cognitive adaptability. Plasticity and adaptability are vital for consciousness, enabling learning, memory formation, and complex cognitive functions essential for conscious experience and moral agency. AI, devoid of biological substrates, cannot replicate these foundational elements of consciousness. Therefore, AI lacks the capacity for genuine subjective experiences or moral cognition. Without the ability to perceive, reflect, and empathize in a human-like manner, AI remains constrained in its potential to develop authentic compassion.

Buddhist Compassion and Ethics

Living a virtuous life through benevolence deeply resonates with the concept of dependent origination from a Mahayana Buddhist perspective that encompasses the middle path (Madhyama), non-violence (Ahimsa), compassion (Karuna), and intentionality (Cetana).

1. *Madhyama* (Middle Way) – (មជ្ឈិមា)

The Middle Way refers to the path of moderation away from the extremes of self-indulgence. It is a fundamental principle in Buddhism that emphasizes the avoidance of extreme practices and views. This approach leads to wisdom, ethical conduct, and mental discipline, forming the basis for the Eightfold Path, which is the way to achieve enlightenment and liberation from suffering.

2. *Ahimsa* (Non-Violence) – (អហិង្សា)

Ahimsa, meaning non-violence or non-harming, is a fundamental ethical principle in Buddhism. It involves abstaining from causing harm to any living being, whether through actions, words, or thoughts. Ahimsa underlies many Buddhist precepts and practices, promoting compassion and respect for all forms of life, humans and non-humans. It is essential for the cultivation of moral discipline and is a prerequisite for developing compassion and loving-kindness (metta).

1. *Karuna* (Compassion) – (ករុណា)

Karuna, or compassion, is the quality of empathy and concern for the suffering of others. It is one of the four Brahmaviharas Dharmas (sublime states) in Buddhism, alongside loving-kindness (metta), sympathetic joy (mudita), and equanimity (upekkha). Karuna involves a deep, selfless desire to alleviate the suffering of others and is cultivated through meditation practices and ethical living. It is crucial for the development of a bodhisattva, one who seeks enlightenment for the benefit of all beings, from which stems the concept of benevolent and serving leadership, in this instance the Boddhisatvas.

2. *Cetana* (Intentionality) – (ចេតនា)

In Buddhist philosophy, cetana (intentionality) and compassion (karuna) are deeply interconnected, with cetana shaping the ethical quality of actions and compassion driving the desire to alleviate suffering. Compassionate intentions lead to positive karma because actions driven by compassion result in ethical behavior that benefits both the individual and others. Cultivating compassion enhances one's cetana, aligning actions with altruism and contributing to spiritual growth and well-being.

These four concepts—Madhyama, Ahimsa, Karuna, and Cetana are integral to Buddhist teachings and practice, guiding individuals towards ethical conduct, spiritual development, and the ultimate goal of enlightenments, in which benevolence plays a central role.

Throughout history, the teachings of Aristotle, Buddha, and Confucius have highlighted benevolence as a cornerstone of virtuous human behavior. These sages, alongside other renowned thinkers, have long championed the virtue of benevolence, illustrating that wisdom flourishes across various civilizations and transcends cultural boundaries. This shared emphasis on benevolence underscores that no single civilization holds a monopoly on truth and wisdom. Instead, these timeless values are enriched through intercultural exchange, fostering a global spectrum of shared human understanding.

The Idea of a Compassionate AI

While AI lacks consciousness, efforts should be made to design a system that aims to simulate ethical behavior and compassionate responses through advanced programming under strict ethical frameworks. By hypothesizing solutions based on observed limitations and ethical principles, AI may potentially simulate human empathy, recognize moral dilemmas, and make decisions aligned with human values.

Therefore creating AI with advanced ethical "compassion capabilities" should involve:

- Ethical Frameworks and Alignment: Program AI with robust ethical guidelines and ensure alignment with diverse human values.

- Learning and Adaptation: Use reinforcement learning and simulated moral scenarios to train AI in ethical decision-making and continuous adaptation to evolving values.
- Transparency and Accountability: Design AI systems to be transparent and explainable, enabling human oversight and feedback.
- Empathy and Compassion Simulation: Develop AI to recognize and simulate empathetic responses and compassionate decision-making.
- Contextual Awareness: Enhance AI's understanding of context to navigate complex moral dilemmas effectively.
- Regulatory Compliance: Adhere to legal and regulatory standards governing AI ethics and behavior.

Integrating the elements above into AI development can facilitate the simulation of compassionate behavior. These methods aim to bridge the gap between AI's capabilities and societal expectations of ethical conduct.

Developing compassionate AI through the lens of Buddhist ethics presents significant challenges. In Buddhism, compassion (*karuna*) is crucial, emphasizing the alleviation of suffering and the promotion of well-being for all beings. Implementing this in AI requires careful design and refinement to align with such profound values. AI must embody not only compassionate behavior but also the ethical principles of non-harm, the middle path, and intentionality.

Conclusion

The complexity of human emotions and moral reasoning, particularly in Buddhist ethics, underscores the need for ongoing interdisciplinary research. Collaboration among technologists, ethicists, and Buddhist scholars is essential to ensure AI systems reflect Buddhist compassion and ethics. Continuous dialogue can address cultural and ethical values, leading to AI that aligns with principles of compassion, wisdom, and ethical conduct.

In conclusion, the pursuit of compassionate AI grounded in Buddhist ethics represents a compelling yet challenging frontier in technology. Current evidence highlights AI's limitations in achieving consciousness and underscores the inherent challenges in replicating human moral cognition. Innovative approaches are needed to integrate ethical principles into AI development, offering a pathway toward more responsible and empathetic AI systems. As technology evolves, ethical considerations must remain central, guiding the evolution of AI to align more closely with human values and societal well-being.

**References available upon request*

**The views expressed are the author's own and do not reflect the views of the AI Forum.*

Abhidhamma-Inspired AI: Centering Cetana (Intention) with Theravada Ethical Insights in Technology

CHEAB Sambath[1] *, MD*

CHHEM Kieth Rethy[2] *, MD, PhD (Edu), and PhD (His)*

---

[1] ***CHEAB Sambath*** *is a Lecturer at University of Puthisastra, Cambodia*
[2] ***CHHEM Kieth Rethy*** *is a Distinguished Professor of AI and Humanities, Center for Buddhist AI Ethics, Cambodia University of Science and Technology and Advisor to AI Forum*

Abstract

This insight explores how the ethical and psychological insights of the Abhidhamma, a foundational text in Theravada Buddhism, can guide the development of more humane and intention-driven artificial intelligence. Centering on cetana (intention) as the moral engine of cognition, it argues for AI systems that act not only efficiently but compassionately. Drawing on key principles such as anatta (non-self) and karuna (compassion), the authors propose a model of AI grounded in adaptability, emotional sensitivity, and ethical awareness. Rather than replacing Western frameworks, Abhidhamma ethics offers a complementary, context-sensitive path toward technology that supports well-being and reduces suffering.

Introduction

As artificial intelligence becomes increasingly deeply integrated into our daily lives-from how we receive medical care to how important decisions are made in our communities-the call to make AI more ethical and human-centered has never been more urgent.

One powerful but often overlooked source of ethical guidance is the Abhidhamma, the analytical heart of Theravada Buddhist philosophy. Far from being just a spiritual text, the Abhidhamma offers a deeply psychological and ethical framework that could inspire AI systems grounded in compassion, intention, and human well-being.

The Abhidhamma's Ethical and Psychological Framework

At its core, the Abhidhamma presents a highly detailed account of how the mind works. It breaks down mental processes into citta (moments of consciousness) and cetasika (mental factors) (Bodhi, 2000), offering a precise lens into how we think, feel, and act. This is surprisingly resonant with AI, which also processes the world in patterns and data. Where AI sees algorithms, the Abhidhamma sees mental states. For instance, greed (lobha), hatred (dosa), and delusion (moha) are viewed as harmful mental factors, while loving-kindness (metta), compassion (karuna), and mindfulness (sati) are seen as wholesome qualities worth cultivating (Gethin, 2004).

Furthermore, the Theravada principle of anattā (non-self)—the understanding that the mind is a fluid and impermanent stream of experiences—complements the Abhidhamma's detailed psychological analysis by highlighting the absence of a permanent self within these mental processes, inviting AI designs that are adaptive and responsive rather than rigid and fixed.

Imagine an AI-powered mental health app that doesn't just detect stress but helps guide users toward emotional balance. By drawing on the ethical insights of the Abhidhamma, such an AI wouldn't just track symptoms-it could promote genuine well-being and emotional clarity. This shift moves AI from being a passive tool to a therapeutic companion (Floridi et al., 2018).

Cetana (Intention) as the Core Ethical Principle in Abhidhamma

A central concept in the Abhidhamma is cetana, or intention, which is seen as the moral engine behind every action. While the idea of intention may seem familiar, cetana carries a deeper meaning in Buddhist thought. It is the mental factor that gathers and directs the mind's energy in each moment, shaping how we act, speak, and even think. In Abhidhamma, every moment of consciousness is made up of various mental components, and cetana plays the coordinating role— guiding the mind toward a particular direction. This is where ethics begins: not in the outcome of an action, but in the intention behind it.

This stands in contrast to many Western ethical systems, which often focus on consequences or rules. In the Abhidhamma, what matters is whether an intention arises from clarity, compassion, and non-attachment—or from confusion, craving, and ill will. For AI, this offers an important shift. Instead of only asking what a system does, we must also ask: what is it aiming to do, and why? A medical AI, for instance, might be designed not only to offer efficient treatment, but also to honor a patient's dignity, values, and well-being. An AI shaped by the spirit of cetana would be more than intelligent—it would be thoughtful, responsive, and aligned with human care.

Principles of Theravada Buddhism Relevant to Ethical AI Anatta (Non-self): A Theravada Principle

Another core idea is anattā, or non-self—the understanding that there is no fixed, permanent identity behind our thoughts, emotions, or actions. In the Abhidhamma, the mind is seen as a stream of moment-to-moment experiences, made up of changing mental and physical phenomena. What we usually think of as a "self" is, in this view, just a convenient label for a complex, ever- shifting process. There is no inner controller—only patterns arising and passing away, shaped by causes and conditions.

In the context of AI, this insight invites us to imagine systems that are more fluid, adaptive, and free from rigid goal structures. Instead of building AI with a fixed agenda or static sense of purpose, we might design technologies that respond sensitively to context and change. A personalized AI tutor, for example, wouldn't assume it always knows best—it would learn from the student, adjust to their changing needs, and remain open to multiple learning pathways. This kind of design mirrors the principle of anattā: an intelligent system that is responsive, decentralized, and flexible. It also resonates with emerging approaches like federated AI, where learning happens locally and adaptively, rather than through top-down, one-size-fits-all models. In this way, Buddhist thinking encourages a new kind of humility in technology—not to erase intelligence, but to soften its attachment to control.

Compassion (Karuna): Ethical Insight from Theravada Buddhism

Beyond cognition, the Abhidhamma also emphasizes mental cultivation-the idea that the mind can be trained toward empathy, balance, and wisdom. AI technologies inspired by this could help users practice mindfulness, regulate emotions, or simply pause and reflect in a busy world (Dignum, 2019). These tools could embody the Four Brahmaviharas: loving-kindness, compassion, sympathetic joy, and equanimity.

Uniqueness of Abhidhamma Compared to Western Ethical Traditions

To fully appreciate what the Abhidhamma brings to ethical AI, it helps to see how Buddhist ethics differs from well-known Western traditions like virtue ethics and deontology. While all three aim to guide human behavior toward the good, they take different paths to get there.

Virtue ethics, for example, emphasizes cultivating stable character traits—like courage or honesty—over time. It's about becoming a certain kind of person. In contrast, Buddhist ethics doesn't assume a fixed self with traits. It focuses on the quality of the mind in each moment. What matters most is intention—cetana—and the mental factors that arise with it. This creates a more flexible, moment-to-moment awareness of what is wholesome or unwholesome, rather than measuring actions against a static ideal of virtue.

Deontology, on the other hand, is about rules and duties. It asks: "What is the right thing to do?" regardless of the outcome. But Buddhist ethics isn't rule-bound in that way. It's not about following fixed laws. Instead, it invites us to look inward and recognize whether our minds are shaped by greed, hatred, or delusion—or by compassion, mindfulness, and wisdom. It's an ethics of awareness, grounded in how our intentions shape experience and consequences.

This distinction matters for AI. A system designed with deontological ethics might follow a strict rule set. One shaped by Buddhist insight would go deeper, asking not only what an AI does, but why. It would be sensitive to context, intention, and the subtle emotional tone of each interaction. In this way, the Abhidhamma doesn't just give us ethical rules—it offers a practical path for cultivating systems that reduce harm and support well-being, from the inside out.

Conclusion

Ultimately, this vision of AI doesn't just seek to make machines smarter-it aims to make technology more humane, supportive of inner growth, and aligned with the goal of reducing suffering. As we design the AI systems that will shape tomorrow's world, perhaps it's time we look back at ancient wisdom as a source of ethical clarity. By embedding the insights of the Abhidhamma into AI systems, we're not creating intelligent machines-but tools that augment human intelligence with compassion and wisdom, helping shape a more thoughtful and humane society.

References

Bodhi, B. (2000). *A comprehensive manual of Abhidhamma: The Abhidhammattha Sangaha*. Buddhist Publication Society.

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707.

Gethin, R. (2004). On some definitions of mindfulness. *Contemporary Buddhism, 5*(1), 263–279.

** Edited with ChatGPT for clarity; all views expressed are solely those of the authors.*

** The opinions expressed are the authors' own and do not reflect the views of AI Forum.*

Conclusion: Toward a Compassionate Intelligence in the Age of Machines

The essays brought together in Buddhist Mind for AI Machines represent a pioneering effort to place Buddhist philosophy not as a cultural ornament but as a partner in epistemic and ethical dialogue at the heart of our global reckoning with artificial intelligence. In an age marked by technological acceleration and mounting ethical uncertainty, the compulsion to innovate often outpaces our collective capacity to reflect. Utilitarian logic, data-driven decisions, and instrumental views of intelligence have led many to mistake speed for wisdom, prediction for understanding, and automation for liberation. Against this backdrop, Buddhism offers not a rejection of technology but a radical reorientation. It anchors intelligence in ethical clarity, self-awareness, and the compassionate alleviation of suffering.

The volume opens with ontological reflections that challenge core metaphors of the digital era: the mind as machine, the self as fixed identity, and intelligence as something to be engineered. Buddhist traditions encourage us to see being not as substance but as process—fluid, contingent, and interdependent. Concepts such as namarupa (name and form), anatta (non-self), and paticcasamuppada (dependent origination) are not abstract metaphysics. Rather, they are invitations to reimagine the very foundations of how AI systems are designed, deployed, and experienced. These ideas remind us that the so-called "intelligence" in artificial intelligence is not ontologically neutral. It arises from, and shapes, specific moral and cultural conditions.

In one of the book's most evocative chapters, the Buddha is portrayed as a proto-neuroscientist, an early investigator of consciousness whose method relied on introspection, ethical discipline, and mindful presence. Such a portrayal draws powerful parallels between ancient contemplative practices and contemporary understandings of neuroplasticity, feedback learning, and cognitive transformation. But more than mere historical analogy, this vision points toward a form of intelligence that is not only functional but wise. It suggests that ethical development is not a supplement to cognition. It is its foundation. For Buddhist thought, intelligence cannot be disentangled from intention, and intention cannot be separated from the conditions that give rise to it.

The second part of the volume shifts focus to epistemology. Here, Buddhist traditions again subvert dominant assumptions. Modern AI rests on binary logic: something is true or false, one or zero, right or wrong. But this dualism, inherited from Aristotelian frameworks, struggles to account for uncertainty, paradox, and contextual meaning. By introducing non-binary reasoning systems such as catuskoti (fourfold logic) and emphasizing concepts like sunyata (emptiness) and epistemic humility, the essays in this section argue for a more flexible, situated, and relational understanding of knowledge. What we know—and how we claim to know—has moral implications. In this light, the design of AI systems is not merely a technical task but a philosophical undertaking.

These insights have practical consequences. As the essays show, systems that reduce complexity into binary outcomes are ill-suited to moral reasoning, especially when deployed in sensitive domains such as law, healthcare, or education. A Buddhist-informed epistemology urges us to question the seduction of certainty and to cultivate models of knowing that remain responsive to context and aware of their own limitations. This has implications for the way we think about algorithmic transparency, explainability, and the fairness of AI decisions. From this perspective, an AI that claims to "know" must also be capable of admitting what it does not.

The final section of the book turns to ethics, perhaps the area where Buddhist philosophy most directly contributes to current debates in AI. None of the contributors argue that machines can be moral agents in the Buddhist sense. They lack vinnana (consciousness), cetana (volition), and karmic continuity. However, all emphasize that AI is not ethically neutral. It reflects the values, intentions, and institutional logics of those who build and use it. The central ethical task, then, is not to ask whether AI can be responsible, but whether we can.

In this context, the concept of cetana becomes especially important. In Theravada Abhidhamma, volition is what gives action its moral weight. Translated into the world of AI, this means that ethical assessment must begin not with outcomes but with intention. Who designs AI, for what ends, and according to which values? These questions are often absent from regulatory discussions focused on risk and compliance. A Buddhist-informed ethics insists that these questions cannot be bypassed.

Another recurring theme is mindfulness (sati), presented not merely as a personal virtue but as a collective and institutional resource. In the context of AI, mindfulness offers a way to cultivate attentiveness to harm, sensitivity to context, and space for reflection amid the rush to innovate. This can guide developers to recognize bias, policymakers to design just regulation, and users to navigate digital systems with greater awareness. A mindful culture in technology institutions, grounded in ethical reflection and emotional literacy, could counterbalance the extractive, efficiency-driven logic that dominates much of Silicon Valley's culture today.

This brings us to the horizon of future research. The chapters in this volume offer a foundation, not a conclusion. The intersection of Buddhism and AI ethics is still a nascent field. Much work remains to be done to elaborate, test, and apply the insights presented here in broader contexts. For example, one promising avenue lies in examining how Buddhist principles can inform algorithmic fairness and transparency. The ethical emphasis on ahimsa (non-harming), metta (loving-kindness), and interdependence offers normative guidance for designing AI systems that reduce bias and promote well-being. Here, Buddhist frameworks could complement existing models of explainable AI by emphasizing not just how an algorithm works, but how it affects lives in contextually nuanced ways.

Likewise, the growing field of contemplative neuroscience offers fertile ground for collaboration between Buddhist thought and cognitive science. Empirical studies could further explore how meditative practices influence the attention, moral judgment, and bias recognition capacities of AI developers. Could contemplative training enhance ethical decision-making in design teams? Could systems themselves be designed to detect and respond to human suffering in ways aligned with karuna (compassion)?

Another rich direction concerns Buddhist logic and its potential applications in AI reasoning. Non-binary logics, such as catuskoti, challenge the adequacy of dichotomous truth conditions that underpin most AI models. Future research could investigate how alternative logical systems might inform the development of AI systems capable of handling ambiguity, pluralism, and paradox—especially in domains like natural language processing, ethical simulations, or intercultural diplomacy.

At the policy level, there is an urgent need to articulate how Buddhist ethics can inform national AI strategies. Buddhist-majority countries like Cambodia, Thailand, Sri Lanka, or Bhutan have an opportunity to craft governance models that reflect local traditions and ethical values. Comparative policy studies could investigate how Buddhist-informed frameworks might influence data protection laws, AI education programs, or public deliberation on technology's social role.

Still more speculatively, research could explore the soteriological dimensions of AI. Could digital technologies serve as upaya (skillful means) in transmitting Dharma? Could AI support meditative practice, moral education, or the cultivation of wisdom? What are the spiritual risks of delegating self-cultivation to machines? These questions demand careful attention to the line between assistance and illusion and to the possibility that technological tools, while powerful, can never substitute for inner transformation.

Finally, Buddhist thought challenges the ideology of perpetual innovation. The ethic of renunciation, simplicity, and restraint, so central to monastic life, offers a critical lens on the techno-optimism that fuels much of AI discourse. Future scholarship might explore how these values could temper the excesses of consumer AI and support more sustainable, mindful approaches to technological development.

Taken together, these avenues of inquiry extend the central message of this volume: that ethical intelligence cannot be engineered without ethical intention. The vision of AI that emerges from Buddhist traditions is not antagonistic to technology but committed to using it wisely, compassionately, and in ways that support the liberation of all beings.

Closing Reflection

If artificial intelligence is to serve the flourishing of all, not merely the powerful, then it must be governed not only by technical standards but by moral imagination. Buddhist thought, with its emphasis on compassion, interdependence, and inner transformation, invites us to ask not what machines can do, but what kind of humans we wish to become. It reminds us that intelligence, in its highest form, is not the power to dominate complexity, but the humility to dwell within it, attuned, ethical, and awake.

Buddhist Mind for AI Machines offers not a final word but a beginning. It is an invitation to scholars, technologists, policymakers, and practitioners to deepen this dialogue, to cultivate wisdom alongside knowledge, and to build a future where machines reflect our highest values, not our lowest fears. May this path be guided by clarity and compassion, and may the light of ancient wisdom continue to illuminate the way forward.

Prompt Log

Section I: Ontology – Being

1. The Buddha as Proto-Neuroscientist and Forerunner of AI Thought

- Show how the Buddha's meditative inquiry anticipates key methods of Buddhist neuroscience.
- Link concepts like *dependent origination* and *cetana* to neuroplasticity and cognitive feedback.
- Apply Buddhist neuroscience to design ethical, emotionally aware AI systems.

2. Mahayana Buddhist Perspective on AI

- How does Buddhist philosophy redefine intelligence compared to artificial intelligence?
- Can AI ever be conscious or sentient from a Buddhist point of view?
- What ethical guidance can Buddhism offer for the development and governance of AI technologies?

3. Theravada Buddhism: Mindfulness for the Ethical Advancement of AI

- How can Theravada mindfulness (*sati*) inform ethical AI development?
- In what ways can mindfulness help mitigate algorithmic bias and promote fairness?
- How can Buddhist ethics be practically integrated into secular AI governance and work culture?

4. Buddhist Ethics for AI: Theravada, Mahayana, and Vajrayana Perspectives

- How do the three main traditions differ in their ethical views on AI?
- What Buddhist ethical principles can guide national AI policy in Buddhist-majority countries?
- How can non-Western traditions contribute to global AI ethics?

5. Namarupa in Buddhist Thought and Human-Centered AI

- Explain *namarupa* in Buddhist philosophy and its cognitive implications.
- How can *namarupa* inform the design of human-centered AI?
- Develop an ethical AI framework using Buddhist concepts of volition and interdependence.

6. Dependent Origination and Systems Thinking: A Buddhist Logic for Explainable AI

- How can *dependent origination* improve systems thinking in AI governance?
- How does a non-essentialist view of causality aid explainable AI (XAI)?
- How can Buddhist XAI promote accountability and contextual awareness?

Section II: Epistemology – Knowing

7. Tetralemma vs Dilemma: Integrating Eastern Logic into AI Reasoning

- How do the dilemma and tetralemma reflect different worldviews?

- How can the tetralemma enhance AI reasoning under ambiguity and ethical complexity?
- How does integrating mindfulness foster ethical and culturally sensitive AI?

8. Catuskoti and Non-Binary Reasoning in Artificial Intelligence

- How can *catuskoti* offer a richer alternative to binary logic?
- Can *catuskoti* improve AI in natural language processing and ethical decision-making?
- How does Buddhist logic support epistemic pluralism in AI?

9. Sunyata and Epistemic Humility in AI Design

- Explain *sunyata* and its implications for knowledge and design assumptions.
- How does *epistemic humility* challenge the biases in AI development?
- Suggest practical strategies to apply *sunyata* in human-centered AI design.

10. Anatta and the Myth of AI Autonomy: Towards a Processual and Relational Ethic for Artificial Intelligence

- How does the Buddha's introspection resemble early neuroscience?
- In what ways do *anatta*, *cetana*, and *paticcasamuppada* reflect dynamic mental processes?
- How can a Buddhist cognitive framework guide adaptive, compassionate AI?

Section III: Ethics – Acting

11. Can AI Be Morally Responsible from the Mahayana Buddhist Lens?

- Can AI possess moral responsibility under Buddhist thought?
- How do *mindfulness*, *cetana*, and *non-harm* shape AI ethics?
- What is the role of human responsibility in AI's lifecycle?

12. The Idea of Compassionate AI: Embodying Buddhist Ethics in Technology

- How can Buddhist ethics inspire compassionate AI?□ Can AI simulate compassion through ethical design despite lacking consciousness?
- What traditions support integrating benevolence into AI systems?

13. Abhidhamma-Inspired AI: Centering Cetana (Intention) with Theravada Ethical Insights in Technology

- How can the Abhidhamma's framework inform mindful, ethical AI?
- What is the ethical significance of *cetana* in contrast to Western models?
- How can Theravada thought inspire adaptive, human-centered AI development?

Back Cover