A Lens of Dharma for the Emergent AI-Ethics Discourse: A working paper[1]

Dr. Arijit Patra[2], Nitish Rai Parwani[3]

Abstract

This working paper arises from an ongoing series of discussions with scholars, practitioners, and technologists engaged in the ethics and governance of Artificial Intelligence (AI). These conversations have repeatedly highlighted that though AI is a global technology, the ethical frameworks currently shaping its development remain predominantly grounded in post-Enlightenment Western philosophical assumptions like especially individualism, rights-based liberalism, and utilitarian reasoning. While these frameworks have produced valuable insights and regulatory tools, they inadequately address the relational, systemic, and long-term ethical challenges posed by socially embedded AI systems.

Empirical mapping of global AI guidelines reveals that many such principles that are described as "universal" do infact largely reflect the normative vocabulary of Western More Economically Developed Countries (MEDCs). This asymmetry is not just geopolitical; it shapes the conceptual boundaries of AI ethics by limiting what counts as a moral problem and what solutions are thinkable.

This working paper responds to that gap by indicating perspectives from Global South, and introducing Dharma based understanding as a complementary framework for AI ethics. The Dharmic framework contributes three key shifts. An 'ontological shift' from the isolated-atomistic individual to the relational self; a 'teleological shift' from merely preventing (*non- maleficence*) harm to promoting *yogakṣema* (welfare encompassing *abhyudaya* (material prosperity) and *niḥśreyasa* (ultimate good)); and a 'methodological shift' from universalist abstraction to context-sensitive judgment guided by *deśa–kāla–pātra* (place–time–agent).

Many contributors to this debate come from philosophy, the humanities, or social theory, as well from engineers and computer scientists. A significant part of this

[1] This is a working paper and not a conclusive publication on the matter. Critique, suggestions and collaborations are invited.

[2] Arijit Patra, PhD, read for a DPhil in Engineering Science specialising in AI for healthcare at the University of Oxford as a Rhodes Scholar. He is currently a Senior Principal Scientist at UCB and a Honorary Senior Lecturer at the University of Glasgow.

[3] Nitish Rai Parwani is a D.Phil. scholar at the University of Oxford. He is the coordinator of the Oxford Dharma Centre for Research and Policy.

working paper serves to bring these communities onto shared conceptual ground, ensuring that the deliberations advanced here are framed with an adequate understanding of all the stakeholders and contributors. As a working paper, this project remains an open philosophical and practical inquiry. It aims to invite further dialogue, critique, refinement, and collaboration to collectively explore how Dharma based ethics may inform more pluralistic, context-sensitive, and inclusive approaches to the future of AI governance.

1. Introduction

Evolution of AI

Artificial Intelligence (AI) has evolved from a speculative endeavour into a technological force. It is fundamentally altering both the fabric of modern economy and of global society. In the mid-20th century, Alan Turing laid the theoretical groundwork for machine intelligence with the computational theory of mind and the Turing Test which became a benchmark for evaluating a machine's ability to exhibit human-like intelligence (Turing, 1936). The Dartmouth Conference of 1956, which is often regarded as the birthplace of AI as a formal discipline (Kline et al, 2010), marked the beginning of an era of initiatives to replicate cognition through symbolic reasoning and rule-based systems. Early programs, such as the Logic Theorist (1956) and the General Problem Solver (1957) (Gugerty et al, 2006), demonstrated that machines could perform logical deductions and problem-solving tasks within narrowly defined domains. However, these brute-force, rule-based approaches encountered limitations in handling uncertainty and real-world complexity, leading to the 'AI Winter' of the 1980s with diminished funding and a waning public scepticism on machine intelligence (Lloyd et al, 1995).

The renaissance of AI in the late 20th century was catalysed by a shift from symbolic AI to statistical learning which emphasized data-driven learning over hand-coded rules. Artificial neural networks, inspired by biological neurons, provided a new framework for adaptive, self-improving systems. A pivotal moment came in 1986 with the rediscovery of backpropagation, a gradient-based optimization technique that enabled neural networks to adjust their parameters autonomously (Rumelhart et al, 1986). This idea, combined with the exponential growth of computational power and a proliferation of digital data, set the stage for the deep learning revolution. An inflection point arrived through a convergence of massive datasets and GPU-accelerated computing in 2012 when AlexNet, a deep CNN, achieved a dramatic

improvement in performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), demonstrating the superiority of deep learning over traditional machine learning (Krizhevsky et al, 2012). This spurred a Cambrian explosion of investment in deep connectionist paradigms, leading to breakthroughs in natural language processing and reinforcement learning.

The emergence of Large Language Models (LLMs) in the late 2010s marked a fundamental shift in AI's capabilities (Wang et al, 2025). Unlike earlier NLP models reliant on task-specific architectures and supervised finetuning, LLMs used transformers introduced by Vaswani and team's seminal paper (Vaswani et al, 2017), to process and generate human-like text at an unprecedented scale. OpenAI's GPT-3, with its 175 billion parameters, showed the emergent abilities of LLMs like few-shot learning and aspects of in-context reasoning despite training only on unlabelled text (Brown et al, 2020). The scaling hypothesis that model performance improves predictably with size, data, and compute became a defining principle of modern AI (Kaplan et al, 2020), leading to an arms race in model scaling, with systems like GPT-4 (~ 1.76T parameters) pushing boundaries of linguistic fluency, reasoning, and multimodal understanding.

Parallel to the rise of LLMs, the idea of foundation models emerged as a new paradigm in machine learning. Unlike task-specific models, foundation models are pre-trained on vast, diverse datasets using self-supervision and can be adapted to a range of downstream tasks (Bommasani et al, 2021). The versatility of foundation models has led to their adoption across domains as diverse as healthcare, finance and creative industries. However, their massive computational and data needs have centralized AI power with a few well-funded organizations, raising concerns on accessibility, bias, and monopolistic control. The latest frontier in AI is generative AI, a class of models capable of creating content like text, images, and video indistinguishable from human generated outputs (Kenthapadi et al, 2023). The commercialization of generative AI by tools like MidJourney and ChatGPT, has democratized content creation but introduced new ethical and legal challenges (Zohny et al, 2023), including copyright infringement, deepfake proliferation, and an erosion of trust in digital media.

Ethics around Artificial Intelligence

Today, AI stands at the forefront of the Fourth Industrial Revolution, characterized by automation, data-driven decision-making, and human-machine collaboration. Edge AI

and federated learning have enabled decentralized, privacy-preserving deployments, addressing concerns about data sovereignty and latency in applications like autonomous vehicles, smart cities, and personalized healthcare (Jobin et al, 2019). Meanwhile, reinforcement learning from human feedback (RLHF) has improved AI alignment, allowing models to adapt to human preferences and reduce harmful or misleading outputs (Kaufmann et al, 2024). The convergence of AI with other exponential technologies, such as blockchain (for decentralized AI governance), biotechnology (AI-driven synthetic biology), and robotics (embodied AI, humanoid robots), will expand AI's societal impact, raising profound questions about the notion of agency, free will, and human-AI coexistence.

However, this phenomenal evolution and optimism around the AI are tempered by existential and ethical dilemmas. For instance, the environmental cost of AI, particularly the carbon footprint of training large-scale models, has sparked a movement toward sustainable AI, emphasizing energy-efficient architectures, model compression, and sustainable data centers (Wu et al, 2022).

The concentration of AI power among certain tech giants and nation-states may risk exacerbating global inequalities, while autonomous weapon systems and AI-driven misinformation threaten geopolitical stability and democratic integrity. The potential for artificial general intelligence (AGI), hypothetical machines with human-like cognitive abilities, has reignited debates about consciousness, rights, and control, prompting calls for proactive governance frameworks to mitigate catastrophic risks (Bucknall et al, 2022).

Therefore, in this pivotal moment, the trajectory of AI will be determined not only by technological advancements but also by how society chooses to govern, deploy, and ethically align these systems. The critical challenges ahead include developing explainable and fair AI, ensuring equitable access and benefit-sharing, mitigating environmental and societal harms, and establishing robust international governance mechanisms. The rise of LLMs and generative AI has demonstrated both the potential and the risks of unchecked AI development.

Without deliberate, inclusive, and ethically grounded interventions, AI risks perpetuating existing power asymmetries, eroding trust, and undermining human agency. Conversely, if harnessed responsibly and equitably, AI holds the potential to solve some of humanity's most pressing challenges, from climate change modelling

and precision medicine to personalised education and economic empowerment, ushering in an era of shared prosperity and flourishing.

2. The Ethical Landscape of Artificial Intelligence

Artificial intelligence has emerged as one of the most pronounced technologies of the 21st century, reshaping industries, economies, and societies. However, this technological progress has raised significant ethical concerns. Current AI ethics frameworks, such as the Asilomar Principles, the European Union's AI Act, and the IEEE's Ethically Aligned Design, are largely grounded in particular philosophical traditions. These frameworks emphasise individual rights, utilitarian outcomes, and market-driven innovation. While valuable, they often overlook the relational and collective dimensions of AI's impact. This section deals with the current ethical frameworks around AI governance, domination of certain philosophical principles therein, how global south seems to have been neglected in these deliberations, a Dharma based approach to correct the limitations of the current philosophical understandings and analysis of the approach.

2.1. Domination of certain Ethical framework.

The contemporary discourse surrounding the governance and ethics of Artificial Intelligence (AI), while ostensibly global in its ambition and reach, is empirically characterized by a profound geopolitical and philosophical asymmetry. A rigorous analysis of the current landscape reveals that the normative standards defining 'ethical AI' are overwhelmingly produced by, and reflective of, 'Western More Economically Developed Countries' (MEDCs). This hegemony is not merely a matter of geographic origin but represents a specific philosophical lineage that are rooted in post-Enlightenment rationalism, individualism, and rights-based liberalism, and those that risks marginalizing alternative ethical paradigms. To understand the necessity of philosophical interventions in AI ethics, one must first delineate the extent and nature of this domination.

Empirical Asymmetry and Geographic Centralization

The claim of Western dominance is substantiated by comprehensive empirical data. In a scoping review of the global landscape of AI ethics guidelines, Jobin, Ienca, and Vayena (2019) identified 84 distinct documents issued by private companies, governmental agencies, and research institutions. The geographic distribution of these issuers reveals a stark concentration of discursive power. The United States and the United Kingdom alone account for more than one-third of all global ethical

frameworks (23.8% and 16.7% respectively), and when combined with the European Union, these Western powers dominate and effectively monopolize the intellectual landscape (Jobin et al., 2019, p. 6). Conversely, the Global South remains largely invisible in this discourse; statistically unrepresented in the formulation of the principles intended to govern technologies that will impact them directly (Jobin et al., 2019, p. 6).

This centralization creates a feedback loop where "global" ethics become a projection of local Western values. Matilal (2002b) argues that historical intellectual relationships between the West and the non-West have often been characterized by a binary where the West views itself as the producer of rational knowledge and the East as the repository of the "mystical other." He notes the rarity, in the recent past, of a horizontal philosophical relationship between East and West" (Matilal, 2002b, p. 372). In the context of AI, this dynamic may seem to persist. The "universal" principles currently being codified are, in effect, the provincial preferences of the North Atlantic rim universalized through economic and technological hegemony.

2.2. Limitation of current ethical understandings.

The dominance of the existing philosophical thought is not limited to the location of the authors but permeates the philosophical substance of the guidelines. Jobin et al. (2019) note a global convergence around five core ethical principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (p. 7). These principles are undoubtedly extremely valuable; however, the interpretation of *several* of these, in current frameworks, prioritizes the individual over the collective.

Frameworks predicated on individualism focus on personal autonomy and rights, are sometimes at expense of community and relational ethics. For example, AI systems designed to maximize individual convenience, such as personalized advertising or social media algorithms, may undermine social cohesion by reinforcing echo chambers or polarizing public discourse. The principle of privacy, for instance, appears in nearly half of all guidelines (Jobin et al., 2019, p. 7). Privacy as a concept is well appreciated throughout the modern world. Hence, without any value judgment on the concept, we see how the currently dominant philosophical traditions understand the concept.

In Western liberal philosophy, privacy is conceptualized as a protective barrier around the autonomous individual against the state or society. Philosophically, this hegemony is the legacy of the "individualist vision" of the European Enlightenment period. As

Lucien Goldmann (1973) argues, the fundamental categories of Enlightenment thought have a structure analogous to the market economy, where "every individual appears as the autonomous source of his decisions and actions" (p. 18). In this framework, society is viewed not as an organic whole, but as a "contract" between autonomous wills (p. 20). This philosophical lineage, championed by thinkers like Locke, sought to clear away tradition as the "rubbish that lies in the way to knowledge" to build a science of man based on empirical observation of the individual mind (Berlin, 2017, p. 24, 31). This seems to align with what Creel (1972) identifies as a fundamental divergence between Western and the Eastern thought. Western ethics presupposes an autonomous individual making value choices in isolation. In contrast, traditional Dharmic thought views the individual not as an atomic unit but as a "synthesis of groups," where "an individual's life is a role in the societal scheme" (Creel, 1972, p. 158). A person is not just an individual, but also an integral part of larger society, planet, etc. By focusing almost exclusively on individual data rights, Western AI frameworks neglect the relational role of the individual within the community structure. This "individualist vision of the world", whether rationalist or empiricist, treats knowledge as autonomous and distinct from the collective historical action of mankind. Consequently, modern AI ethics, heir to this tradition, struggles to articulate obligations that transcend the atomized individual, rendering it inadequate for addressing the systemic and communal impacts of algorithmic governance.

Furthermore, Jobin et al. (2019) found that the principle of solidarity, which emphasizes social cohesion and collective responsibility, is severely underrepresented, appearing in only 6 of the 84 documents analysed by them (p. 7). Similarly, sustainability is marginalized, referenced in only 14 documents (p. 7). This neglect suggests that the dominant framework views ethics primarily through the lens of the atomic individual interacting with a system, rather than through a holistic lens of community welfare and ecological interdependence.

The Negativity Bias and the Limitations of "Doing No Harm"

The current ethical corpus displays a distinct "negativity bias." The references to *non-maleficence* (the duty to do no harm) in the global AI regulations outnumber references to *beneficence* (the duty to do good) by a factor of 1.5 (Jobin et al., 2019: 15). Referring to Taddeo and Floridi (2018), they note that "This negative characterization of ethical values is further emphasized by the fact that existing

guidelines focus primarily on how to preserve privacy, dignity, autonomy and individual freedom in spite of advances in AI, while largely neglecting whether these principles could be promoted through responsible innovation in AI". This bias reflects a rights-based legalistic framework where the primary moral imperative is "negative liberty", i.e. freedom from interference and harm. The ethical imagination is thus constrained to risk mitigation rather than actively structuring technology to cultivate human flourishing.

This approach contrasts sharply with the Dharmic tradition, which Matilal (2002a) describes as distinct from rigid, rule-based Kantian morality. In examining the epics, Matilal highlights how Dharma allows for flexibility and context-sensitivity that Western formalism often lacks. Dharmic ethics, or Dharma, is not merely about avoiding harm (non-maleficence) but involves a "lived allegiance" to an order that sustains the cosmos and society (Creel, 1972, p. 166). The current Western AI frameworks, by focusing on 'not doing harm', miss the teleological dimension of technology, what it is for, and how it serves the goals of our lives.

Excludes inputs from the Global South

A scholar working on Global South, particularly on Eastern traditions, may also argue that the domination of Western frameworks is sustained by a lingering intellectual prejudice that views Western philosophy as the sole repository of "rationality" and logic, while dismissing Eastern thought as "mystical" or "irrational." Prof. B.K. Matilal, who served as the Spalding Professor of Eastern Religions and Ethics at the University of Oxford, vigorously deconstructs this binary, noting that the "mystique of the East is essentially a Western construction" (2002b: 265), and argues against the dogma that "The East is irrational and emotional, the West rational and logical" (p. 374). Similar critiques have been registered from other Global South intellectual traditions as well. In African philosophy, thinkers working within the Ubuntu framework challenge Western moral individualism by grounding ethical reasoning in relational personhood, encapsulated in the maxim *umuntu ngumuntu ngabantu* ("a person is a person through other persons"). Menkiti (1984) argues that moral agency in African thought is constituted through community, while Ramose (2002) contends that Western ethics misconstrues personhood by abstracting individuals from their social and moral embeddedness.

Concerns like these arise from within Confucian ethics as well. Moral reasoning is structured around *ren* (humaneness), *li* (social function), and role-based obligations

rather than rights-bearing autonomy. The ethicist of this tradition emphasise that Confucian ethics is fundamentally relational and situational, prioritizing harmony and responsibility within social roles over abstract moral universalism. Contemporary scholars have therefore noted that the Confucian philosophy “characterizes a moral agent from the perspectives of relational, situational, and role specific interaction and moral cultivation” (Seok, 2024: 9)

In the context of AI, this prejudice manifests as the exclusion of non-Western ethical resources under the guise that they are "religious" or "cultural" rather than "philosophical" or "ethical." As Creel (1972) notes, that Hindus, for instance, have Dharma, while west has a category of ethics (p. 166), implying a different categorical organization of moral life by the Eastern civilisations, that integrates natural law, social duty, and cosmic order. By relying almost exclusively on Western philosophical assumptions, the current AI ethics landscape suffers from an epistemic blind spot. It attempts to govern a technology with global potential using a framework that that stems from a particular world view and prioritizes the individual over the collective, the negative right over the positive duty, and the short-term mitigation of harm over the long-term cultivation of virtue.

2.3. Need for a Dharma based Perspective

The limitations of the prevailing ethical paradigm in Artificial Intelligence (AI) signals to discuss alternate or complementary normative frameworks that can address the systemic complexities of algorithmic governance. While the principles of transparency, justice, non-maleficence, responsibility, and privacy are important and have also been crystallized into a global standard of AI ethics (Jobin, et al., 2019, p. 7), their interpretation is tethered to a post-Enlightenment ontology of atomistic individualism and legalistic compliance. This dominant paradigm is effective in defining negative liberties; however, often lacks the vocabulary to address the systemic, relational, and teleological challenges posed by algorithmic governance. To bridge the gap between individual rights and collective flourishing, and to move beyond a "negativity bias" focused on harm reduction, Dharma based frameworks may provide a solution. A Dharmic perspective offers a corrective to the atomism and solutionism of current discourse by enforcing three fundamental shifts: an ontological shift from the individual to the relational self, a teleological shift from non-maleficence to welfare (*yogakṣema*), and a methodological shift from rigid universalism to context-sensitive wisdom.

First, The most distinct divergence between the dominant Western paradigm and Dharmic thought lies in the ontology of the moral subject. Western ethics generally presupposes an autonomous individual whose rights serve as protective barriers against the collective. In contrast, the Dharmic tradition views the individual not as an atomic unit but as a "synthesis of groups," where one's existence is *also* defined by roles within a larger societal, ecological, and cosmic scheme. From this perspective, the ethical concerns of AI cannot be restricted to data privacy or individual autonomy alone; they must encompass the technology's impact on social cohesion and ecological interdependence. By treating the self as intrinsically relational, a Dharmic framework argues that AI systems be evaluated on their ability to sustain the social fabric (*lokasaṃgraha*) rather than *merely* maximizing individual utility functions.

Second, Current AI governance is heavily skewed toward legalistic compliance and risk management. Essentially it is the duty to do no harm. The Dharmic tradition proposes the objective of *Yogakṣema*, that is defined as the acquisition of well-being and its preservation. Kautilya asserts that the legitimacy of governance rests on the active promotion of the subjects' happiness (Kautilya, 1992, p. 107). Applied to AI, this principle demands a teleological reorientation. Algorithmic systems must not be passive tools checked only for safety; they must be active instruments for *Abhyudaya* (material prosperity) and *Niḥśreyasa* (spiritual welfare/ ultimate good).

Third, the integration of Dharma may challenge the "technical solutionism" prevalent in AI ethics. It is a belief that ethical dilemmas can be solved through binary code or static checklists. Bimal Krishna Matilal argues that dharma is not a rigid Kantian imperative but a dynamic negotiation between conflicting duties (Matilal, 2002b, p. 33). Ethical life involves genuine moral dilemmas where values like truth and non-violence may conflict, requiring context-sensitive practical wisdom (based on situation, time and subject: *deśa–kāla–pātra*) rather than algorithmic rigidity (Matilal, 2002a, p. 5). A Dharmic perspective accepts that moral truth may not always be captured by determinate solutions and hence argues for an adaptive AI governance model that is flexible, context-aware, and capable of navigating the subtle nature of moral responsibility.

2.4. Potential limitations of a Dharma based Framework

While a Dharma based approach to AI ethics offers a sound alternative to conventional frameworks, it confronts significant challenges that must be addressed for its effective implementation. Chief among these is the *ex-facie* inherent

subjectivity of Dharma. Its context sensitivity and plurality of meanings may introduce ambiguity in policy and practice, complicating efforts to standardize Dharmic principles within AI governance and leaving room for misinterpretation or selective application in global regulatory contexts. Another formidable obstacle is the operational complexity of integrating Dharmic ethics into technical AI systems, particularly when attempting to quantify abstract ethical values such as harmony, duty, or collective well-being, within the mathematical and algorithmic frameworks that govern machine learning and automated decision-making. Bridging this gap between philosophical ideals and computational realities requires innovative methodological approaches, such as ethically aligned reinforcement learning or value-sensitive design, which remain nascent and underdeveloped in current research. Furthermore, the philosophical concepts like utilitarian, rights-based, and individualistic ethical frameworks are deeply entrenched in international AI policies, corporate practices, and academic discourse. This institutional inertia may marginalise an alternate perspective, relegating them to peripheral considerations rather than core principles in global AI ethics, thereby limiting their influence in shaping future technological trajectories.

2.5 Counterarguments and Mitigation Strategies to potential limitations

Despite these challenges, Dharma based frameworks for AI ethics can be successfully operationalized through deliberate collaboration, adaptive governance, and strategic integration with existing ethical paradigms. One key strategy is consensus-based interpretation, where the subjective nature of Dharma is addressed not as a limitation but as an opportunity for deliberative engagement. This may be done through bringing together ethicists, policymakers, technologists, and community representatives to co-develop ethical standards through dialogues, cross-cultural workshops, and participatory policymaking. This working paper is a result of such deliberations itself. These processes can potentially yield contextually nuanced yet broadly applicable guidelines that balance Dharma principles with calculative technological needs, as has also observed at initiatives like the IEEE's Ethics Certification Program for Autonomous Systems, which increasingly incorporates non-Western perspectives. Moreover, the growing global demand for inclusive AI ethics, which is fuelled by critiques of Western dominance in tech governance and calls for decolonizing AI, creates a favourable environment for Dharma based principles to gain traction, particularly as Global South nations, indigenous communities, and marginalized

groups advocate for pluralistic alternatives that reflect their values, histories, and societal priorities; in this evolving landscape, a Dharma based framework stands poised to bridge divides, offering a more representative, and spiritually grounded path forward for AI development.

2.6. Dharma based Ethics as a Complementary Framework

The integration of Dharma based principles into the global discourse on Artificial Intelligence (AI) ethics offers a vital complementary framework that addresses the specific ontological, teleological, and methodological limitations of the prevailing Western paradigm.

Dharma based frameworks would introduce a distinct form of rationality, that is not merely calculative but sensitive to the complexities of moral life. It may enrich and enhance global AI governance by merging utilitarian efficiency with relational ethics, individual rights with duty based on the concept of unity of existence, and short-term innovation with long-term harmony.

While the current western oriented computational ethics often seeks algorithmic solutions to resolve moral conflicts into binary outcomes, Dharmic ethics acknowledges the persistence of moral dilemmas. In his analysis of the Indian epics, Matilal (2002a) demonstrates that *Dharma* is not a static code of conduct but a dynamic guide that may navigate situations where duties conflict, such as the tension between truth-telling and saving lives (pp. 26–27). The ethical AI governance must also move beyond rigid compliance checklists to accommodate the context sensitive nature of moral truth, designing systems capable of handling ambiguity and context-dependence rather than forcing determinate solutions upon complex social realities.

Furthermore, integrating this perspective also helps to bridge the gap between the technical capabilities of AI and the holistic needs of human existence. Current global guidelines, heavily focused on technical robustness and economic efficiency (Jobin et al., 2019), often neglect what Rabindranath Tagore termed the "Surplus in Man", *vis.* that aspect of human existence which exceeds mere biological survival or material utility. Tagore presses that ‘human personality cannot develop if there is a division between the human individual and nature or the world in general’ (Pop, 2014: 256). A Dharmic approach insists that technology must serve not only the "scientific conception" of the universe, but also the "humanistic conception," where the world is viewed as a unity between man, ecology and other aspects of Universe.

3. The open quest

This is a working paper developed from the ongoing discussions with scholars, and technocrats. To this stage, there has been a delineation of the geopolitical and philosophical asymmetry in the current AI ethics environment, and a theoretical noting of the necessity of a Dharma based intervention. The translation of these insights into actionable governance mechanisms remains an open quest. Significant intellectual labour also remains to fully articulate the 'Core Principles of Dharmic AI Ethics' and to undertake the complex pragmatic task of 'Operationalizing Dharmic AI Ethics' within technical systems. In the spirit of the Indian philosophical tradition of *vāda* (dialogue), which views truth as a collective discovery rather than a dogmatic imposition, we invite scholars, technologists, and ethicists to offer their critiques, collaborations, and inputs.

References

1. Berlin, I. (1969). *Four essays on liberty*. Clarendon Press.
2. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., … Liang, P. (2021). *On the opportunities and risks of foundation models* (arXiv:2108.07258). arXiv. https://arxiv.org/abs/2108.07258
3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.
4. Creel, A. B. (1972). Dharma as an ethical category relating to freedom and responsibility. *Philosophy East and West, 22*(2), 155–168. https://doi.org/10.2307/1397799
5. Goldmann, L. (1973). *The philosophy of the Enlightenment: The Christian burgess and the Enlightenment* (H. Maas, Trans.). Routledge & Kegan Paul. (Original work published 1968)
6. Gugerty, L. (2006). Newell and Simon's Logic Theorist: Historical background and impact on cognitive modeling. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 9, pp. 880–884). SAGE Publications.
7. Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial intelligence: The global landscape of ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
8. Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). *Scaling laws for neural language models* (arXiv:2001.08361). arXiv. https://arxiv.org/abs/2001.08361
9. Kaufmann, T., Weng, P., Bengs, V., & Hüllermeier, E. (2024). *A survey of reinforcement learning from human feedback* (arXiv:2312.00863). arXiv. https://arxiv.org/abs/2312.00863
10. Kautilya. (1992). *The Arthashastra* (L. N. Rangarajan, Trans.). Penguin Classics.
11. Kenthapadi, K., Lakkaraju, H., & Rajani, N. (2023, August). Generative AI meets responsible AI: Practical challenges and opportunities. In *Proceedings*

of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 5805–5806). ACM. https://doi.org/10.1145/3580305.3599550

12. Kline, R. (2010). Cybernetics, automata studies, and the Dartmouth Conference on artificial intelligence. *IEEE Annals of the History of Computing, 33*(4), 5–16. https://doi.org/10.1109/MAHC.2010.45
13. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105.
14. Lloyd, J. W. (1995). *Surviving the AI winter*. University of Western Australia.
15. Matilal, B. K. (2002a). *The collected essays of Bimal Krishna Matilal: Ethics and epics* (J. Ganeri, Ed.). Oxford University Press.
16. Matilal, B. K. (2002b). *The collected essays of Bimal Krishna Matilal: Mind, language and world* (J. Ganeri, Ed.). Oxford University Press.
17. Menkiti, I. A. (1984). Person and community in African traditional thought. In R. A. Wright (Ed.), *African philosophy: An introduction* (pp. 171–181). University Press of America.
18. Pop, M. (Ed.). (2014). *Values of the human person: Contemporary challenges* (Vol. 47). Council for Research in Values and Philosophy.
19. Ramose, M. B. (2002). *African philosophy through Ubuntu*. Mond Books.
20. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0
21. Seok, B. (2024). Digital Confucius and virtuous AI: Confucianism in the age of AI and robotics. *Journal of Confucian Philosophy and Culture, 42*, 7–27. https://jcpc.skku.edu/xml/41776/41776.pdf
22. Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752. https://doi.org/10.1126/science.aat5991
23. Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*(2), 230–265. https://doi.org/10.1112/plms/s2-42.1.230
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

25. Wang, Z., Chu, Z., Doan, T. V., Ni, S., Yang, M., & Zhang, W. (2024). History, development, and principles of large language models: An introductory survey. *AI and Ethics, 5*(3), 1955–1971.
26. Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., & Gschwind, M. (2021). Sustainable AI: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems, 4*, 795–813.
27. Zohny, H., McMillan, J., & King, M. (2023). Ethics of generative AI. *Journal of Medical Ethics, 49*(2), 79–80. https://doi.org/10.1136/jme-2022-108956.