

3

RECONCEPTUALIZING AI ETHICS IN THE LIGHT OF DHARMA

Subhodeep Mukhopadhyay ¹, Sudarshan Therani Nadathur ²

ISBN: 978-1-1054690-0-8 | DOI: 10.25215/110546900X.03

Abstract:

Conversations around the ethical aspects of Artificial Intelligence (AI) have grown in tandem with the increased adoption of AI technologies. Current approaches to AI ethics have been criticized for being ineffectual and for serving corporate interests. Mainstream ethics, which forms the dominant basis of AI ethics, is largely rooted in Western thought. Perspectives from the Global South, including Indian ethical traditions, remain largely absent from this discourse. This study proposes a reconceptualization of AI ethics rooted in Bharatiya epistemic traditions, using *Nyaya-shastra*'s *pancha-avayava* (five-step reasoning) framework. Adopting a bottom-up analytical approach, the study begins with the foundational concept of *dharma* and progressively extends through general ethics and digital ethics to AI ethics. A *Nyaya*-based framework, with *dharma* as its foundation, envisions an ethical trajectory aligned with an Indian epistemic worldview. This in turn will help India develop indigenous and culturally-sensitive AI literacy programs.

Keywords: *AI Ethics, Nyaya, Pancha-avayava, AI Literacy, Generative Artificial Intelligence*

Introduction:

With the increased usage of Artificial Intelligence (AI) technologies in the last decade, we have also witnessed growing discussions on the ethical challenges. The challenges are multifaceted. Digital concerns like privacy violations,

¹ Doctoral Candidate, School of Education, GlobalNxt University, Kuala Lumpur, Malaysia, Email Id: subhodeepm.infinity@gmail.com

² Director, Siddhanta Knowledge Foundation, Tamil Nadu, India, Email Id: tnsudarshan@gmail.com

data leaks, unauthorized private data usage, and lack of transparency have become even more acute. On the socioeconomic front, there is a widespread belief that AI will take over jobs leading to major social upheavals (Shafik, 2026). There are major concerns about the environmental impact of training huge machine learning models (Yu et al., 2024). AI systems, in many cases, have reinforced biases (Benjamin, 2019; Mukhopadhyay & Reddy, 2023; Noble, 2018).

In response, both corporate entities and academic institutions have developed different AI ethics guidelines. The ostensible aim is to prevent miscreants from doing harm. Unfortunately, the reality is different. There is an increasing body of scholarship which say that such initiatives are ineffective and another example of corporate “ethics washing” where guidelines focus more on political correctness than addressing root issues (Mittelstadt, 2019; Munn, 2023; Wagner, 2018).

At the same time, current AI ethics formulation is essentially a re-contextualization of mainstream ethics for AI, the latter being principles and values rooted in western thought, which itself is an uneasy blend of Biblical morality and Greco-Roman rationalism (Malhotra, 2013). Socrates, Plato and other classical Greek philosophers laid the groundwork of ethics with ideas of justice and reason in the early centuries BCE (MacIntyre, 2007). Later Christian thinkers such as Augustine and Aquinas redefined ethics through the lens of divine will and salvation (Finnis, 1998). Then came the European Enlightenment movement in the eighteenth and nineteenth centuries which tried to infuse secular moral reasoning into the discipline. But by and large, ethics still remained within a Christian-influenced metaphysical framework (Taylor, 2010).

Contemporary ethical discourse remains shaped by these Western legacies. Starting in the twentieth century we find ethics as a discipline dealing with global issues like human rights, sustainability, and AI. Feminist and postmodern critiques have also pushed for more context-sensitive approaches (Gilligan, 1993). The prevailing narrative is that this version of ethics, represents a universal global standard and the non-West is expected to adopt them to align with an imagined universal norm. As noted by Roche et al. (2023), “voices from the Global South and consideration of alternative ethical approaches are largely absent from the conversation” around the legal, social, ethical and policy aspects of AI.

The Hindu concept of *dharma* is entirely missing from mainstream ethical discourse. Derived from the root *dhri* (“that which upholds”), *dharma* is a core principle in Hindu thought. It guides conduct and sustains harmony across the individual, societal, and natural realms. Any serious ethical discussion, whether societal, digital, or AI, must engage with ideas of *dharma*. Yet they remain ignored.

An important question is whether the current Western-inspired formulation is *capable* of providing solutions in the AI space. Does it have the necessary frameworks, and methodologies to tackle moral dilemmas inherent in the digital realm? Context-sensitive ethical frameworks that recognize different moral traditions are the need of the hour. One has to move beyond the one-size-fits-all model rooted in Western thought. To illustrate this, we adopt a *Nyaya*-based framework grounded in *dharma*, to demonstrate how the current AI ethics stack is inadequate and must be reimagined in terms of Indian knowledge systems (IKS).

Methodology:

Nyaya is a traditional Indian knowledge framework dealing with logic and reasoning. It provides a rigorous methodical framework to establish what knowledge is (*prama*), how knowledge can be gained (*pramana-shastra*), and the different modes of acquisition like debate (*vada*) and others (Gopinath & Sharma, 2022). Although it is often classified as one of the six Hindu philosophical systems, one must not be mistaken in thinking that *Nyaya* is a religious work.

The *Pancha-avayava* framework (five-part syllogism) is one such tool within *Nyaya* that provides a systematic way to arrive at a reasonable conclusion. The five components are *pratijna* (proposition), *hetu* (reason), *udaharana* (example), *upanaya* (application) and *nigamana* (conclusion) (Sarukkai, 2005). We apply this framework using a systematic bottom-up approach. We begin with the foundational concept of *dharma* (Table 1) and progressively move upwards from *dharma*-based ethics (Table 2), digital ethics (Table 3), to AI ethics (Table 4).

At each stage we have a new *pratijna*, which drives the *hetu* and *udaharana* of the next layer. We take one specific principle, *aparigraha*, the principle of non-hoarding or non-possessiveness, as outlined in Patanjali’s *Yoga Sutra*. Being a *shastra-pramana*, or that

which has been well-established in tradition, it is recognized as a social good. This can easily be extended to other universal dharmic principles, whether the ten *yama* and *niyama* of the *Yoga Sutra*, the ten *dharma-lakshana-s* of the *Manusmriti*, the nine from the *Yajnavalkya Smriti*, or the thirty virtues listed in the *Shrimad Bhagavatam*. Since *dharma* informs system design at the foundational level, many of the common ethical challenges encountered in AI development are inherently preempted. We subsequently present a conceptual framework outlining how a fourfold *varna*-based model for the design of digital and AI systems may proactively address and circumvent many ethical concerns (Table 5). As will be demonstrated, many of the issues prevalent in current Western formulations of AI ethics are, by design, unlikely to arise within a *dharma*-inspired framework.

Results:

Table 1 which is the injunction that *aparigraha* is a social good is presented for the sake of completeness. This foundational *yama* can guide the ethical design of digital and AI systems (Tables 2, 3 and 4): *aparigraha* → public service → digital service → AI service.

Table 1: Dharma: Aparigraha

Avayava	Explanation
<i>Pratijna</i>	<i>Aparigraha</i> is a social good.
<i>Hetu</i>	Because there is <i>dharma</i> .
<i>Udaharana</i>	When there is <i>dharma</i> there is social good, as shown by the shastras (<i>shastra-pramana</i>).
<i>Upanaya</i>	<i>Aparigraha</i> is <i>dharma</i> as seen in the <i>shastras</i> .
<i>Nigamana</i>	<i>Aparigraha</i> is a social good.

Table 2: General Ethics: Public Service for All

Avayava	Explanation
<i>Pratijna</i>	Public service for all is a social good.
<i>Hetu</i>	Because there is <i>aparigraha</i> .
<i>Udaharana</i>	When there is <i>aparigraha</i> there is social good, because it is <i>dharma</i> .
<i>Upanaya</i>	Service for all is similar to <i>aparigraha</i> which is dharmic.
<i>Nigamana</i>	Public service for all is a social good.

Public service for all, whether education, security or healthcare is a social good since it abides by the principle of *aparigraha* or non-possessiveness/non-hoarding. This stems directly from the idea that holding back services to certain groups is undesirable, and a form of discrimination. From here we can move towards a specific case: digital service for all.

Table 3: Digital Ethics: Digital Service for All

Avayava	Explanation
<i>Pratijna</i>	Digital service for all is a social good.
<i>Hetu</i>	Because it promotes public service for all.
<i>Udaharana</i>	When there is public service for all there is social good which is <i>aparigraha</i> .
<i>Upanaya</i>	Digital service for all is similar to public service for all which is <i>aparigraha</i> .
<i>Nigamana</i>	Digital service for all is a social good.

Table 4: AI Ethics: AI service for All

Avayava	Explanation
<i>Pratijna</i>	AI service for all is a social good.
<i>Hetu</i>	Because it promotes digital service for all.
<i>Udaharana</i>	When there is digital service for all, there is social good, which is similar to public service for all.
<i>Upanaya</i>	AI service for all is similar to digital service for all which is akin to public service for all.
<i>Nigamana</i>	AI service for all is a social good.

We thus arrive at the conclusion that AI service for all is a social good. Generative AI (GenAI) systems answering questions is an example of an AI service and must not be denied to certain groups. Conversely, one may say that GenAI responding to all questions is a social good. By emphasizing transparency, and non-monopolization of knowledge in the philosophical design stage itself, we can resolve the GenAI non-response at that stage. An objection may of course arise: what if someone asks questions like how to harm another person? Shouldn't there be some bare minimum policies or restrictions? We discuss this aspect in-depth in the next section.

Discussion:

In a typical Western formulation, a question which talks about harming someone else is instantly flagged as a violation. However, a *dharmic* system can potentially offer greater contextual flexibility. Unlike Abrahamic traditions that rely on fixed commandments, *dharma* is situational and guided by *guna* and *karma*.

Just as society is divided into four *varnas* (*not to be confused with the so-called caste system*), *brahmana*, *kshatriya*, *vaishya*, and *shudra*, based on qualities and roles (Bhagavad Gita 4.13, 18.41-44), AI systems too may be categorized according to the functions they best serve: intellectual, political, economic, or operational. Rather than treating AI agents as one-size-fits-all, we can align them with functional archetypes (Table 5). For instance, a *kshatriya*-type AI used in defense sector may engage in lethal planning, whereas a *shudra*-type general-purpose broad-based GenAI system like ChatGPT or Gemini AI is restricted from doing so. The moral dimension is contextually addressed. What is ethical for one may not be ethical for another.

Table 5: Conceptual Mapping of the Fourfold Model to Digital & AI Systems

Varna (Role)	Digital/AI Conceptual Mapping	Real-World Application/Usefulness
<i>Shudra</i>	Operational automation, routine task bots, background service AIs	<ul style="list-style-type: none"> • Chatbots • Ticket resolution systems
<i>Vaishya</i>	Systems optimized for trade, resource allocation, recommendation, personalization	<ul style="list-style-type: none"> • AI in e-commerce and financial analytics • Inventory and logistics AI
<i>Kshatriya</i>	AI systems for governance, cybersecurity, justice, risk management	<ul style="list-style-type: none"> • Predictive policing with dharmic oversight • AI-based disaster response systems • Governance and compliance dashboards

Varna (Role)	Digital/AI Conceptual Mapping	Real-World Application/Usefulness
<i>Brahmana</i>	Systems designed for ethical reasoning, explainability, and education	<ul style="list-style-type: none"> • AI ethics advisory agents • AI-based policy suggestion tools • Transparent AI systems explaining reasoning (e.g., Explainable AI in medical diagnosis)

Nyaya's pancha-avayava framework can thus be used to build a stacked model from *dharma* to AI *dharma*, contextualized across the four *varnas*. Each AI agent can have different use cases and applications based on its *varna*. This example is presented here to illustrate how *dharmic* categories can help address ethical challenges.

Conclusion:

In this paper, we introduce a *Nyaya*-inspired ethical framework that may better address emerging ethical issues. This is especially important as India advances toward indigenous AI systems built on culturally contextualized datasets. The integration of indigenous AI ethics will mark a crucial advancement as India rolls out massive AI literacy programs in the coming years (Mukhopadhyay, 2026). It must also be noted that *Nyaya*, as a system of reasoning, does not prescribe *how* to write code or implement systems directly. It instead serves to guide the foundational design philosophy behind digital and AI systems. With *dharma* as its core, it will ensure an ethical trajectory that aligns with an Indian epistemic model. Ethics will be intrinsic to purpose rather than as an external add-on. This in turn will have a positive impact on system design, policy formulation, AI education, and geo-strategic options.

Acknowledgement:

We thank IKSHA-JNU conference reviewers for their feedback on our paper presented at Jawaharlal Nehru University, New Delhi, July 12, 2025. This manuscript expands a specific section of the unpublished work available on Zenodo (<https://zenodo.org/records/16951169>).

References:

- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.
- Finnis, J. (1998). *Aquinas: Moral, political, and legal theory*. Oxford University Press.
- Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.
- MacIntyre, A. (2007). *After virtue: A study in moral theory* (3rd ed.). University of Notre Dame Press.
- Malhotra, R. (2013). *Being different: An Indian challenge to Western universalism*. HarperCollins Publishers India.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mukhopadhyay, S. (2026). AI literacy in India through the lens of national policy: A four-dimensional evaluation. *Journal of Artificial Intelligence Machine Learning and Neural Network*, 6(1), 1–10. <https://doi.org/10.55529/jaiml.61.1.10>
- Mukhopadhyay, S., & Reddy, D. (2023). Artificial intelligence as an enabler of Western universalism. In R. Malhotra, T. N. Sudarshan, & M. Sastry (Eds.), *The power of future machines: Essays on artificial intelligence* (pp. 241–263). BluOne Ink LLP.
- Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics*, 3(3), 869–877. <https://doi.org/10.1007/s43681-022-00209-w>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Roche, C., Wall, P. J., & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies and initiatives. *AI and Ethics*, 3(4), 1095–1115. <https://doi.org/10.1007/s43681-022-00218-9>
- Shafik, W. (2026). The future of work: AI's threat to jobs and economic stability. In W. Shafik (Ed.), *The dark side of AI: A human and societal perspective* (pp. 89–117). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-09130-7_4
- Taylor, C. (2010). *Sources of the self: The making of the modern identity* (11th printing). Cambridge University Press.
- Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In M. Hildebrandt (Ed.), *Being profiling: Cogitas ergo sum* (pp. 108–115). Amsterdam University Press.

Yu, Y., Wang, J., Liu, Y., Yu, P., Wang, D., Zheng, P., & Zhang, M. (2024). Revisit the environmental impact of artificial intelligence: The overlooked carbon emission source? *Frontiers of Environmental Science & Engineering*, 18(12), Article 158. <https://doi.org/10.1007/s11783-024-1918-y>