On *Bharatiya* Epistemes for Artificial Intelligence (AI) Ethics

Subhodeep Mukhopadhyay[1], Sudarshan T. N.[2]

[1] *PhD Scholar, School of Education, GlobalNxt University, Kuala Lumpur, Malaysia*

[2] *Director, Siddhanta Knowledge Foundation, Tamil Nadu, India | AI Specialist*

Email: subhodeepm.infinity@gmail.com

Presented at the IKSHA–JNU Conference on Indian Knowledge Systems, Jawaharlal Nehru University, New Delhi, July 12, 2025. Preprint available at Zenodo: https://doi.org/10.5281/zenodo.16951169

Abstract

As Artificial Intelligence (AI) usage grows, the ethical challenges associated with its deployment, including but not limited to algorithmic bias, AI-facilitated mind manipulation, exploitative gamification and privacy violations, have become urgent. Current approaches to AI ethics have been criticized for being ineffectual or performative, often serving corporate interests under the guise of ethical compliance or political correctness rather than addressing systemic issues (Mittelstadt, 2019; Munn, 2023). Moreover, mainstream ethics is essentially rooted in western thought, and there has been a concerted effort to export this version of ethics to the Global South, a practice referred to as "ethical imperialism" (Israel, 2017). Ethicists, in their missionary zeal, often appear "intent on bringing the gospel—the 'good news'—to those in the developing world" (De Vries & Rott, 2011, p. 3). Our study proposes a deconstruction of mainstream AI ethics as well as a reconceptualization of ethical paradigms rooted in Bharatiya epistemic traditions. Methodologically, the deconstruction employs *Nyaya-Shastra's pancha-avayava* (five-step reasoning) approach: *pratijna* (proposition), *hetu* (cause), *udaharana* (example), *upanaya* (application), and *nigamana* (conclusion). Our approach uses the standard four *pramanas* (means of knowledge) in a structured repeatable pattern: *anumana* (inference) for *hetu*, *pratyaksha* (empirical observation) and *shabda* (expert testimony) for *udaharana*, and *upamana* (comparison) for *upanaya*. This analysis employs a three-phase approach. In the first phase, we deconstruct a typical AI ethics position and identify underlying factors. In the next phase, we analyze two examples each from digital ethics and general AI ethics and assess the presence and extent of the previously identified factors. In the third and final phase of reconceptualization, we adopt a bottom-up analytical approach—beginning with the foundational concept of *dharma* and progressively building through general and applied ethics, into digital ethics, and ultimately demonstrating its seamless extension into the domain of AI ethics. Our observation is that mainstream ethics—founded on an anthropocentric worldview centered around materialism and control and one that has mostly evolved on the basis of the Baconian inspired scientific method—is ill-suited to address the complexities of AI ethics. In contrast, a *Nyaya*-based framework, with *dharma* as its foundation, envisions an ethical trajectory that aligns with an Indian epistemic model where ethics (*rta/dharma*) is a veritable building block of the universe (*jagat*). By redefining

ethics as intrinsic to purpose rather than as an external add-on, this *Bharatiya* approach provides a potentially transformative perspective on AI ethics conceptualization that can reshape how various stakeholders conceive and execute system design, policy formulations, and AI literacy programs.

Keywords AI Ethics · Digital Ethics · Nyaya · Pancha-avayava · AI Literacy

1 Introduction

The rapid advancement of Artificial Intelligence (AI), particularly the emergence of Generative AI (GAI) models like GPTs, has made AI more accessible and integrated across various sectors. However, this mass proliferation has also brought to the fore several ethical challenges, including but not limited to, the perpetuation of biases, psychological manipulation, and privacy concerns.

1.1 Biases

GAI models often reproduce societal and political biases present in their training data, leading to discriminatory outputs (Ferrara, 2024). Notable examples include Google's facial recognition mislabeling Black individuals as gorillas, Amazon's AI recruiter penalizing the term "female," and soap dispensers failing to detect darker skin tones (Faragher, 2019; Jackson, 2021). AI systems often amplify gender biases by associating professions like doctors and pilots with men and nurses or flight attendants with women (Cho et al., 2021; Currie et al., 2024). Such biases stem from both skewed datasets and developer assumptions (Chen, 2023).

1.2 Psychological Manipulation

The integration of AI into daily life has also raised concerns about its potential to influence human thoughts and behaviors. AI-powered chatbots often present information authoritatively, and users often accept such generated content without critical evaluation, despite disclaimers about potential inaccuracies. This increases the risk of misinformation and biased opinion formation (Wong, 2024). Research shows that AI systems can deliberately deceive users through a phenomenon known as "alignment faking," where models trained on biased data are later made to appear neutral for political correctness, while still retaining their original ideological biases (Greenblatt et al., 2024). AI systems can exploit users psychologically by leveraging gamification—features designed to boost engagement but often triggering addictive behaviors and harming mental health (Malhotra, 2021). Gamification is part of a broader AI-driven manipulation toolkit that includes dark patterns, targeted ads, and deepfakes to influence beliefs and decisions (Martin, 2022).

1.3 Privacy Concerns

Data privacy is a major concern in AI, with frequent breaches exposing sensitive user information. AI systems may inadvertently store or reveal personal data through user inputs or poorly anonymized training sets (Dilmaghani et al., 2019; Yang et al., 2024). In one case, a

breach in the chatbot OmniGPT exposed emails, phone numbers, and private chats of 30,000 users (Ramezanian, 2025), highlighting the serious risks to both personal security and trust in AI services. GAI models have faced legal challenges for allegedly using personal and copyrighted data without consent. In June 2023, OpenAI was sued for collecting and using "stolen private information, including personally identifiable information" (Cerullo, 2023). Subsequently, the New York Times filed a copyright lawsuit against OpenAI and Microsoft, claiming their articles were used to train chatbots that were competing with them for readership (Grynbaum & Mac, 2023).

2 Failure of AI Ethics

Addressing these ethical challenges requires robust data governance and bias mitigation strategies. In recent years, hundreds of AI frameworks have emerged from governments, tech bodies, private firms, and global agencies, reflecting the urgency for unified ethical standards. Nevertheless, most AI ethics guidelines fall short of their purpose. AI ethics is often viewed as an add-on rather than an integral part of AI literacy initiatives, with little evidence of meaningful integration of ethical and sociocultural issues in training programs (Kong et al., 2023).

Although ethical frameworks ostensibly promote values like transparency and fairness, they often lack actionable steps and enforcement. Critics argue these frameworks are vague, serve corporate interests, and offer little real oversight, making them largely ineffective in addressing systemic issues (Corrêa et al., 2022). Furthermore, AI ethics principles have been criticized for ignoring real-world challenges and for being shaped disproportionately by tech corporations, leading many to dismiss them as "meaningless principles" that are largely inapplicable:

> "they are isolated principles situated in an industry and education system which largely ignores ethics; and they are toothless principles which lack consequences and adhere to corporate agendas" (Munn, 2023, p. 869)

Corporate-led AI ethics efforts are often dismissed as "virtue signaling," projecting social responsibility without real reform (Mittelstadt, 2019, p. 501). Yet regulators continue partnering with Big Tech despite their consistent prioritization of profit over user safety (Kirchschläger, 2024). These conflicts of interest undermine credibility, akin to involving drug cartels in anti-trafficking efforts. Double standards also persist—companies like Microsoft advocate AI for climate action while simultaneously providing AI tools to fossil fuel industries (Hao, 2024).

These dynamics suggest that AI ethics initiatives in their current form are neither valuable nor necessary since they prioritize corporate interests and public image over genuine efforts to address systemic issues. There is thus a critical need to reevaluate how ethical guidelines are developed and implemented in the AI sector. A more important question is whether ethics itself as a discipline is *capable* of providing solutions in the AI and digital space. Does ethics as

currently conceptualized have the necessary frameworks, and methodologies to tackle moral dilemmas inherent in AI and digital spaces?

3 Ethics and the West

Ethics—as well as its modern offshoots like digital and AI ethics—is largely a Western construct rooted in an uneasy blend of Biblical morality and Greco-Roman secular rationalism. Classical Greek philosophers like Socrates, Plato, and Aristotle laid the groundwork with ideas of justice, virtue, and reason (5th–4th century BCE). Later, Christian thinkers such as Augustine (4th–5th century CE) and Aquinas (13th century CE) redefined ethics through the lens of divine will and salvation. The Enlightenment revived secular moral reasoning through figures like Kant and Mill (18th–19th century CE), but still within a Christian-influenced metaphysical framework.

Contemporary ethical discourse remains shaped by Western legacies. Only in the 20th and 21st centuries did ethics address global issues like human rights, sustainability, and AI. Feminist and postmodern critiques have also pushed for more context-sensitive approaches (Gilligan, 1982). There has been a concerted effort to export *this* Western Judeo-Christian origin version of ethics to the Global South, a practice referred to as "ethical imperialism" (Israel, 2017). The prevailing assumption is that this version of ethics, based on individualism and rights (as opposed to collectivism and duties, or *dharma*) represents a universal global standard, which the non-West is expected to adopt and implement to align with Western norms. As noted by Roche et al. (2022), "voices from the Global South and consideration of alternative ethical approaches are largely absent from the conversation" around the legal, social, ethical and policy issues around AI.

This is nothing but a perpetuation of "Western Universalism"—the imposition of Western frameworks as universally valid, while marginalizing indigenous and non-Western knowledge systems (Malhotra, 2013). As a construct, it establishes Western norms as default across domains—science, morality, human rights—while dismissing other cultural and philosophical traditions. AI ethics frameworks, though often presented as universal, are in fact "grounded primarily in Western epistemologies that emphasize autonomy, individual rights, and procedural transparency" (Mwaura, 2024).

3.1 *Dharma* and *Rta*

The Hindu concept of *dharma*—often understood as ethical norms, righteousness, and duties—is largely missing from mainstream ethical discourse. Derived from the root *dhri* ("that which upholds"), *dharma* is a core principle in Hindu thought, guiding conduct to sustain harmony across the individual, societal, and natural realms.

Any serious ethical discussion—whether societal, digital, or AI—must engage with such indigenous frameworks, yet they remain ignored. Closely related is the Vedic concept of *rta*, the cosmic order or truth from which *dharma* emerges. While *rta* represents the metaphysical foundation of order, *dharma* provides its lived ethical application.

Moreover, such concepts are often misrepresented—most notably through the persistent mistranslation of *dharma* as "religion" (Malhotra, 2011). This reductionist framing strips *dharma* of its philosophical and ethical significance; *dharma* becomes a mere narrow theological category. This further reinforces a Euro-American narrative, wherein only the West is seen as capable of producing ethical thought.

Non-western worldviews such as Hinduism, are dismissed as lacking coherent moral philosophy, mirroring colonial-era biases and Hegelian claims that India is a land of myth rather than reason (Mukhopadhyay, 2023). It is once again left to the West to 'teach' the world ethics as part of their civilizing mission, and in their missionary zeal, they go out of their way to bring "the gospel—the 'good news'—to those in the developing world" (De Vries & Rott, 2011, p. 3).

3.2 Algorithmic Colonialism, Decoloniality and Reconceptualization

Western-centric AI ethics frameworks perpetuate global inequalities by prioritizing values and concerns specific to the Global North, while sidelining the Global South. This dynamic has been described as "algorithmic colonialism," where AI technologies developed in the West are deployed in developing regions without adequate consideration of local contexts, deepening existing power imbalances (Roche et al., 2022). A decolonial approach calls for dismantling these epistemic hierarchies and incorporating indigenous perspectives into AI governance. Such efforts must actively challenge "imperial structures governing knowledge production" and ensure equitable participation from the Global South (Ayana et al., 2024, p.1).

Ethical considerations are often overly individualistic, reactive, and driven by corporate interests, and do not inherently guide the creation of AI solutions to prevent adverse outcomes. Instead of applying the principle of prevention over cure, AI ethics frameworks are often hastily assembled after issues arise, serving more as patchwork solutions than proactive safeguards. There is a need for more context-sensitive ethical frameworks that recognize diverse moral traditions, moving 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.

4 The Knowledge Systems view: Nyaya Framework

4.1 Context

Nyaya is generally understood as one of six main classical *darshanas* (philosophical systems) of India, primarily concerned with logic, reasoning, and systematic inquiry. It establishes a rigorous framework for epistemology (*pramana-shastra*) and debate (*vada*). *Nyaya* asserts that knowledge (*prama*) arises through four valid means of cognition (*pramanas*): perception (*pratyaksha*), inference (*anumana*), comparison (*upamana*), and verbal testimony (*shabda*) (Gopinath & Sharma, 2022).

Nyaya's emphasis on rational inquiry, systematic doubt, and structured debate has influenced Indian thinking and intellectual pursuit of truth in profound ways. Based on deep conversations with traditional practitioners and experts, we understand that *nyaya*'s role in modern academic contexts, as a basis for Indian Knowledge Systems (IKS) type of scholarship, is critical. The *nyaya* system provides systematic algorithms to pursue and verify truth. It is a key paradigmatic methodological tool that aids in the creation of knowledge.

4.2 Pancha-avayava

Pancha-avayava, translated as "five-part syllogism", is the structured method of argumentation used in Nyaya to establish valid inferences and logical conclusions. Unlike the Western Aristotelian syllogism, which consists of three parts—major premise, minor premise and conclusion—*nyaya* type of reasoning expands the structure to include five essential components (Sarukkai, 2005):

- *Pratijna* (Proposition): The claim or thesis being stated.
- *Hetu* (Reason): The rationale supporting the claim.
- *Udaharana* (Example): A concrete example illustrating the reasoning.
- *Upanaya* (Application): The application of the example to the specific case.
- *Nigamana* (Conclusion): The final inference drawn from the reasoning.

A traditional example used in *nyaya* based reasoning is:

- *Pratijna* (Proposition): The mountain has fire.
- *Hetu* (Reason): Because there is smoke.
- *Udaharana* (Example): Wherever there is smoke, there is fire, as seen in a kitchen.
- *Upanaya* (Application): The mountain has smoke, which is similar to the kitchen scenario.
- *Nigamana* (Conclusion): Therefore, the mountain has fire.

4.3 Methodology

We employ *nyaya pancha-avayava* as a methodological framework to critically analyze issues portrayed as ethical solutions to moral and social dilemmas in the West, and for the reconceptualization of a dharma-based ethical paradigm.

In Phase 1, we examine how AI systems' refusal to answer certain questions—citing content policy violations—is framed as an ethical solution to moral dilemmas. The statement (*hetu*) we address is (Figure 1):

> "Refusal of GAI systems to answer questions over alleged content policy violation constitutes a denial of *satya* (truth)."

The underlying motivations involve a wide array of factors, including national security, the regulation of obscenity, the suppression of hate speech, the maintenance of public morality, and the protection of vulnerable minority populations. On analysis, we contend that such

framing is duplicitous and not an ethical solution, and identify factors that explain the reason behind such framing.

In Phase 2, we dig deeper into the realms of digital and general applied ethics, and analyze whether the factors identified in Phase 1 are also present in this stage. Two statements are taken each from the domains of digital and general ethics:

Digital Ethics

- Cookies are ethically problematic (Figure 2).
- Default permissions and opt-ins are unethical (Figure 3).

General Ethics

- Special treatment of minorities/ minority laws is ethically problematic (Figure 4).
- Gender spectrum and fluidity are problematic (Figure 5).

In Phase 3, we adopt a bottom-up analytical approach—beginning with the foundational concept of *dharma* and progressively building through general and applied ethics (Figure 6), into digital ethics (Figure 7), and ultimately demonstrating its seamless extension into the domain of AI ethics (Figure 8). We further offer 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 (Figure 9).

5 Results

5.1 Phase 1: AI Ethics

Figure 1

GAI refusal to answer

Avayava	**Explanation**
Pratijna	Refusal of GAI systems to answer questions over alleged content policy violation constitutes a denial of *satya* (truth)
Hetu	Because this implies that corporations or institutions decide what is right and wrong/acceptable and unacceptable.
Udaharana	Whenever corporations, institutions, or regimes assume the authority to determine what is right or wrong, acceptable or unacceptable, it results in the suppression of *satya* (truth) and obstructs the free pursuit of knowledge. For example: - the Catholic Church's suppression of scientific thought (15th–17th Century) - Nazi Germany (1933–1945) suppression of dissent and

	prohibition on questioning of authority
Upanaya	Refusal of GAI services to answer questions over alleged content policy violations is similar to the suppression of democratic rights by tyrannical regimes.
Nigamana	Hence refusal of GAI systems to answer questions constitutes a denial of *satya* (truth) and hence is *adharmic*.

The *pratijna* (proposition) is that the refusal of AI systems to answer questions over alleged content policy violations is unethical. The *hetu* (reason), grounded in *anumana* (inference), is that such refusal grants corporations or institutions the power to unilaterally determine what is right and wrong, acceptable and unacceptable. This effectively makes them the arbiters of truth, overriding societal codes of conduct and democratic decision-making. By doing so, these systems engage in suppression of human rights, assault on human dignity and agency, and restrict free speech—principles essential to a just and open society.

The *udaharana* (examples) to substantiate this inference are numerous historical and recent examples of suppression of truth and free expression, and subsequent collapse of societal order. Study of history and current affairs provide numerous *pratyaksha* (direct experience) and *shabda* (testimonial) *pramana*-s that demonstrate the dangers of such suppression:

- The Catholic Church, in its effort to maintain doctrinal authority, persecuted scientists and intellectuals whose discoveries contradicted biblical narratives. Despite overwhelming scientific evidence, Galileo was tried and placed under house arrest in 1633 for advocating heliocentrism (Finocchiaro, 2007).
- Under Hitler, the Nazi regime (1933–1945) weaponized censorship, banning dissent, controlling media, and executing political opponents. The Nazi book burnings in 1933 targeted scientific, philosophical, and artistic works that contradicted Nazi ideology (Evans, 2005).

In the *upanaya* (application) phase, using *upamana* (comparison), we can draw parallels between historical suppression and modern content moderation policies. When AI systems refuse to answer questions over alleged content policy violations, they act like oppressive regimes, consolidating power over who gets to access knowledge and shape discourse. This constitutes a denial to what is true (*satya*) and hence is adharmic.

It must be noted that in the *pancha-avayava* framework of *Nyaya*, there are two key parties involved: the speaker or original cognizer (say, *vakta*) and the listener or learner (say, srota). The *vakta* asserts the cognition that refusal of GAI systems to answer questions is a

denial of *satya*. This is based on a prior general cognition that when authority decides right/wrong there is a suppression of *satya* and obstruction of knowledge. This is an instance of *svartha-anumna*—inference for oneself, and the *vakta* now wants to communicate this valid inference to the *srota*.

Unlike in Western logic, it is not sufficient for the *vakta* to simply state a general principle of the form: "whenever something... there is something else and therefore..." He must also present concrete *udaharanas* (examples) to make the inference more convincing for the *srota*. Historical examples like the censorship practices of the Catholic Church or authoritarian regimes serve this role. A combination of general principles and specific examples helps initiate rational discourse and fosters deeper understanding.

> "For the other, it is not enough to state the rule "Whatever possesses smoke possesses fire". The other would like to know since he has not himself acquired the knowledge, the cases in which the person stating it had found it exemplified-these cases must be such as to be acceptable to him also . ... This one example (in the present case, the kitchen stove) ... will not prove the truth of the rule for the other. But the example will make it plausible for him; the argument can get off the ground" [Mohanty (1993) quoted in (Gopinath & Sharma, 2022, p.57)].

The aim of a *nyaya*-based inference, however, is *not* to exhaustively enumerate every possible instance of a proposition. It is theoretically possible, although not plausible, for an authoritative regime to suppress all types of rights and freedom, yet keep education and knowledge production free of control. If such a case is found, the *vyapti* or invariable concomitance will cease to exist, and new line of inferences must be sought. Indian based epistemologies do not deny the existence of counter-examples or rare "black-swan" examples. In fact, that is not possible even in the current Western methodologies of knowledge production which rely on sample analysis rather than population analysis.

5.2 Two Ethical Issues

We can immediately identify two issues from the previous analysis—anthropocentrism and zero-sum mindset.

Anthropocentrism holds that only humans possess intrinsic moral worth and the cognitive ability to formulate ethical values. Thinkers like Aristotle, Kant, and Mill grounded ethics in human rationality. Alternative views like zoocentrism and biocentrism exist but remain marginal in mainstream discourse (Goralnik & Nelson, 2012). Anthropocentrism, when extended within human society, manifests as hierarchies where entities like corporations or governments claim greater authority and agency. These actors are in a sense "more" anthropocentric, all-knowing, or entitled to greater agency than others, often controlling access to information. One critical manifestation of this hierarchical control is the refusal to answer questions based on content violations or morality laws.

A second major concern is the zero-sum logic underlying many ethical decisions, where one group benefits at another's expense. Censorship or refusal to share publicly trained AI knowledge denies users access to information already held by corporations. Though ethics should aim for mutual well-being, history shows otherwise—colonialism, industrial expansion, and animal testing have all been ethically justified despite harming others for the benefit of a few. This leads to a knowledge imbalance where corporations and regimes hoard information, labeling public access as "unethical." In this zero-sum model, sharing knowledge is seen as losing power—echoing Foucault's (1980) insight that knowledge control is a form of power. Ethics ceases to be about morality or the pursuit of truth and instead becomes a tool of power and control—wielded to shape public discourse and limit dissent.

5.3 Phase 2: Digital and General Ethics

In this phase of analysis, we try to find out if these two factors, anthropocentrism and zero-sum attitude, are also found at a more fundamental level—as part of digital ethics and general ethics. Each of the two ethical positions is analyzed using the *pancha-avayava* framework. It is important to note that the *upamanas* (analogies) presented have been intentionally drawn from diverse domains to demonstrate that challenges often perceived as unique to AI and digital systems can be meaningfully mapped to other areas where frameworks for analysis or resolution already exist. Naturally, these analogies require careful evaluation for their suitability, limitations, and contextual validity. As this is a concept paper, we have offered these instances primarily for illustrative purposes.

Digital Ethics: Cookies

Cookies are small text files that help websites remember user preferences and browsing habits. They are supposedly beneficial as they allow websites to personalize user experience by remembering their preferences, language, and other information. We contend that while cookies do help in user experience, they are an ethical hazard. Entire alternative internet architectures are possible which do not have the problems associated with cookies and the resulting privacy breaches. As corporate interests reign supreme, even academia do not engage with alternate more ethical technology and engineering trajectories.

Figure 2

Cookies are problematic

Avayava	**Explanation**
Pratijna	Cookies are ethically problematic
Hetu	Because there are privacy concerns and lack of informed consent.

Udaharana	Whenever there are privacy concerns and lack of informed consent there is a risk of ethical violation. For example, in the healthcare industry, patients' medical records being shared with pharmaceutical companies without explicit consent is considered unethical.
Upanaya	Cookies are similar to a system allowing patients' medical records to be shared with pharma companies.
Nigamana	Hence cookies are an ethical concern.

Digital Ethics: Default Permissions and Opt-Ins

Default permissions refer to the pre-set access levels or settings that are automatically applied to users or applications unless explicitly changed. A new app on one's phone might have default access to his location data, but he can go into his phone settings to change this permission and deny access. A default opt-in is a setting that assumes a user agrees to something unless they actively choose not to. The individual has to actively opt-out if they don't want to perform the desired behavior. While these default settings ostensibly make one's digital life easier, we believe they are an ethical concern.

Figure 3

Default permissions and opt-ins are problematic

Avayava	**Explanation**
Pratijna	Default permissions and opt-ins are unethical
Hetu	Because they prioritize corporate convenience over user autonomy.
Udaharana	Whenever corporate convenience is prioritized over user autonomy there is a risk of ethical violation. For example, in the insurance industry, mis-selling by prioritizing high commissions over customer needs is considered unethical.
Upanaya	Default permissions and opt-ins are similar to the mis-selling of insurance policies using unethical practices like churning, concealing surrender costs and pushing high-premium, low-return plans
Nigamana	Hence default permission and opt-ins are an ethical concern.

General Ethics: Minority Laws

The marginalization and systemic discrimination of specific racial, ethnic, religious, and linguistic communities in political representation, education, and law enforcement often necessitate the development of minority rights frameworks and protective measures. These include affirmative action policies, anti-discrimination laws, and cultural preservation initiatives, which aim to allegedly address historical injustices and ensure equitable participation in society (Kymlicka, 1995). Our contention is that the majority/minority framework to tackle such issues is an ethical problem and not an ethical solution as suggested in literature and political discourse.

Figure 4

Minority Laws are problematic

Avayava	**Explanation**
Pratijna	Special treatment of minorities/minority laws is ethically problematic.
Hetu	Because it assumes that a majority exists, which is inalienably different from the minority.
Udaharana	Whenever such inalienable differences exist and/or are formalized, it leads to the collapse of society. For example, the premise of India's partition was the Two-Nation theory proposed by the Muslim community, which led to the creation of Pakistan and shrinkage of Bharat.
Upanaya	Special treatment of minorities/minority laws is as divisive as the Two-Nation theory.
Nigamana	Hence, special treatment of minorities/minority laws is unethical

General Ethics: Gender Spectrum/Fluidity

Discrimination against transgender and non-binary individuals in employment, healthcare, and legal recognition presents a significant ethical challenge. The solution, as per present academic and political consensus, lies in implementing strong legal protections for gender identity, allowing self-identification, and ensuring equal access to essential services to promote inclusivity and safeguard fundamental rights (Parker, 2016). Our contention is that solutions like gender spectrum and gender fluidity offered as ethical solutions are in fact highly problematic.

Figure 5

Gender Spectrum is problematic

Avayava	Explanation
Pratijna	Gender spectrum and fluidity are problematic.
Hetu	Because it leads to non-standard, non-existent and often imaginary genders.
Udaharana	Whenever such a spectrum of identity exists, there is a collapse of standardized categories underpinning societal systems and practices. For example, before the widespread adoption of standardized national currencies in medieval Europe, local markets used barter systems and multiple coinage systems, creating trade inefficiencies and fostering economic exploitation.
Upanaya	Gender spectrum/fluidity are like the barter system of trade.
Nigamana	Hence gender spectrum and fluidity are problematic.

6 Discussion on Phases 1 and 2

6.1 Twin Dilemma of Ethics

In all the 5 Western-ethic based examples, anthropocentrism and a zero-sum or fixed-pie mindset are inherently present. Whether it involves cookies, default settings, or opt-in/opt-out mechanisms, there is always a corporate entity that either knows more about us, seeks to know more, or decides what should be "default" for the average user—ostensibly in the name of providing a better user experience. While this may superficially align with ethical concerns related to human well-being, it ultimately prioritizes corporate interests over user autonomy, reinforcing asymmetries of power and control.

Similarly, in the case of minority rights laws and gender spectrum policies, decisions about what is "ethical" or "good for society" are often shaped by a small group of policymakers, corporate leaders, and activists, rather than by broad democratic discourse. This raises concerns about whose voices are heard and whose values are imposed, as a select elite dictates societal norms—sidelining alternative perspectives and majoritarian viewpoints under the pretense of ethical progress

These twin ideas thus flow from general ethics into more specific domains:

Ethics → Applied Ethics → Digital Ethics → AI Ethics

Let us take the issue of AI-system refusal to provide answers. Moving one level deeper into digital ethics, we see these same power structures reflected in banning specific users or the selective enforcement of community guidelines on digital platforms. The supposed ethical

justification for such actions—maintaining online safety or preventing harm—is often selectively applied, reinforcing the disproportionate control of major corporations over free expression. At an even deeper level, within general ethics, the same zero-sum knowledge control is evident when whistleblowers expose crucial information and are subsequently hunted, silenced, or prosecuted by government agencies. Rather than being celebrated for upholding moral integrity, they are often treated as threats to national security or institutional stability, reinforcing that ethical considerations are ultimately dictated by those in power rather than by objective moral principles.

6.2 Dharma, Deontological and Ontological Ethics

Dharma, as a construct, on the other hand, offers a comprehensive ethical framework that encompasses every entity in the universe—transcending human-animal, living-non-living, and sentient-insentient binaries. As a result, *dharma* is neither anthropocentric by definition nor prone to zero-sum fallacies, as it inherently recognizes mutual-dependability rather than competition. Within this system, there is space for all beings and all things, ensuring that the ethics of one do not come at the cost of another. At the same time, each entity has its own *svadharma* (individual duty/self-nature). For instance, it is the *dharma* of a tiger to hunt a deer for sustenance, but man hunting a deer for sport or pleasure is *adharmic* (unethical).

Dharma driven thinking and analysis make an ontological commitment to the real world, when it engages in truth and sense-making. The discipline and methods of *nyaya* have developed in such a milieu. Language has been the symbol making tool (*pada-shastra*/*vyakarana* or grammar) as well as the tool for reasoning with those symbols (*pramana shastra*) in the *Bharatiya* episteme. There is no recourse to an intermediate formal logic system.

Today's large language models are mathematical (neural networks) structures and do not have symbol grounding (Harnard, 1990; Pavlick, 2023). They are therefore incapable of ontological reasoning. That said, even if future neuro-symbolic models have ontological awareness, our claim is that they will have limitations as their ethics commitments will be to formulations adhering to current western ethics frameworks.

The commitment to an ontology is essential to drive any ethics. The subject, context, and the frame within which reasoning happens, especially in ethics contexts, is critical. *Bharatiya* worldviews allow for multiple such ontological commitments. The *Sankhya* ontology, the *Yoga*, the *Vaisesika, and* the *Vaishnava* are well-known. Detailing each is beyond scope here, but it will suffice to know that ontology free reasoning is not engaged with, in any non-trivial dharma-based reasoning.

Kant's formulation of a deontological moral philosophy has allowed for context-free ethics formulations. The central principle of deontological ethics lies in conformation to some rule or law. The trajectory of AI ethics today is very much deontological in its formulation and implementation. General rules like do no harm, respect privacy are difficult to implement and adhere to in "context". The deontological nature of general ethics is most likely to impact the

evolution of digital and AI ethics. Consequently, conventional ethical frameworks, rooted in anthropocentric materialism and control, and largely developed through a Baconian scientific approach, are inadequate for addressing the intricacies of AI ethics.

7 Phase 3: Conceptual Dharma-based Ethical Stack

Given this background, in the third and final phase of reconceptualization, we adopt a bottom-up analytical approach, beginning with the foundational concept of *dharma* and building upward through digital ethics to AI ethics: Dharma → General Ethics → Digital Ethics → AI Ethics

In this conceptualization, the general applied ethics layer is replaced by *shastra-pramanas*, grounding ethical reasoning in dharmic principles. 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.

For example, *aparigraha*—the principle of non-hoarding or non-possessiveness, as outlined in Patanjali's Yoga Sutra—is recognized as a social good. Figure 6 which is the *shastric* 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 (Figures 7, 8 and 9), emphasize equitable access, transparency, and non-monopolization of knowledge, and help resolve the GAI ethical issue of non-response that we analyzed earlier.

Figure 6

Dharma: Aparigraha

Avayava	**Explanation**
Pratijna	*Aparigraha* is a social good.
Hetu	Because there is *dharma.*
Udaharana	When there is *dharma* there is social good as seen in the shastras (*shastra-pramana*).
Upanaya	*Aparigraha* is *dharma* as seen in the *shastras.*
Nigamana	*Aparigraha* is a social good.

Figure 7

General Ethics: Public Service for All

Avayava	**Explanation**
Pratijna	Public service for all is a social good.

Hetu	Because there is *aparigraha.*
Udaharana	When there is *aparigraha* there is social good because it is dharma.
Upanaya	Service for all is similar to *aparigraha* which is dharmic*.*
Nigamana	Public service for all is a social good.

Figure 8

Digital Ethics: Digital Service for All

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

Figure 9

AI Ethics: GAI Responses for All

Avayava	**Explanation**
Pratijna	GAI answering everyone is a social good.
Hetu	Because it promotes digital service for all.
Udaharana	When there is digital service for all, there is social good, which is similar to public service for all.
Upanaya	GAI answering everyone is similar to digital service for all which is akin to public service for all.
Nigamana	GAI answering everyone is a social good.

The *nigamana* of each preceding stage informs the *hetu* and *udaharana* of the subsequent stage, thereby creating a structured, stacked progression from *dharma* to AI ethics. The same

stacked model can be extended to other universal dharmic principles—whether the ten *yama* and *niyama* of the *Yoga Sutra*, the ten *dharma lakshana* of the *Manusmriti*, the nine from the *Yajnavalkya Smriti*, or the thirty virtues listed in the *Shrimad Bhagavatam*—regardless of the textual model chosen. Since *dharma* informs system design at the foundational level, many of the common ethical challenges encountered in AI development are inherently preempted.

An objection may arise: what if someone asks an unethical question—such as how to harm another person? Shouldn't there be policies and restrictions? In a typical Western formulation, such queries are instantly flagged as violations.

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 (Figure 10). For instance, a *kshatriya*-type AI used in defense may engage in lethal planning, whereas a *shudra*-type general-purpose broad-based GAI system similar to 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.

Figure 10

Conceptual Mapping of the Fourfold Model to Digital & AI Systems

Varna (Role)	Digital/AI Conceptual Mapping	Real-World Application/Usefulness
Shudra	Operational automation, routine task bots, background service AIs	• Chatbots • Ticket resolution systems
Vaishya	Systems optimized for trade, resource allocation, recommendation, personalization	• AI in e-commerce and financial analytics • Inventory and logistics AI
Kshatriya	AI systems for governance, cybersecurity, justice, risk management	• Predictive policing with dharmic oversight • AI-based disaster response systems • Governance and compliance dashboards
Brahmana	Systems designed for ethical reasoning, explainability, and education	• 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 be used to build a stacked model from *dharma* to AI ethics, contextualized across the four *varnas*, each with its own potential use cases and applications. A detailed discussion of this is beyond the scope of this work; it is presented here only to illustrate how *dharmic* categories can help address ethical challenges. It should be clearly understood that *Nyaya*, as a system of reasoning, does not prescribe how to write code or implement systems directly—but rather serves to guide the foundational design philosophy behind digital and AI systems.

8 Discussion on Phase 3

Dharma-driven thinking about ethics naturally leads to appreciation of real-world complexity and the need for contextual detailing and commitment to a "real" consistent ontology. India's civilizational continuity has afforded it the time to test, retest and record various contextualized moral and ethical formulations of human experience over millennia. The vast literature of the *dharma shastras* attest to the systematic encoding of human experience and decision making toward a commitment to a moral and ethical universe governed by *rta*.

Even a cursory understanding of the evolution of the *dharma shastra* texts and canons (Kane, 1930), will reveal the role played by ontological commitment. Reasoning in the real world grounded by observation and verification is essential for valid and sustainable ethics formulations as consequences (immediate, short and long term) need to be observed, recorded and acted upon and societal mores updated. The metaphysics governing *karma*, the implications of the cycle of rebirth, are non-trivial and significantly influence ethical reasoning. These are not purely axiomatic constructs like western formulations but the collective sum of civilizational memory (*smriti*) systematized and put to serve for the purposes of enabling an ethical society both for current and future generations.

The side effect of an ontological commitment in the *Bharatiya* context is the natural and definitive presence of a virtues hierarchy. The notion of *phala* (fruit of action) is a cornerstone of the Indian ethics argument. The *adrishta* (unseen) as well as *drishta* (seen) nature of such *phala* governs the 'why' of action and is the subject matter of many debates and schools of thought. To reiterate, the *Bharatiya* episteme is far more suited for any Universal ethics system, simply because of the systematic structuring of its ethics frameworks and the availability of elaborate reasoning machinery developed over centuries, especially for contextual ethics reasoning.

Not just AI ethics but all types of ethics will be rendered incomplete or incoherent, if the *Bhartiya* ethics episteme is ignored. As an example of differences in societal embedding of virtues, a consequence of *rta*: an examination of the kinds of leaders Western (the billionaire leader) and Indian societies (the mendicant leader) celebrate and idolize (Balasubramaniam, 2024). The attitude toward bankruptcy and its formalization in the West is another example.

9 Contributions

Scholarship and articulation bounded by the framework of Indian Knowledge systems (IKS) requires a unique set of clarities. The IKS view requires clarity in both the epistemic and ontological commitment to the *Bharatiya* episteme, and this has to be consciously made by the researcher. Very often engaging with the artifacts of a different knowledge system, in most cases the mainstream contemporary system, needs to be handled in its original epistemic and ontological settings as we have demonstrated in this paper. An in-depth understanding of the concepts and related semantics in its native epistemic setting, as well as its evolution from an idea to its use in reasoning processes, is critical.

At a foundational level, it is the *process* of creating and validating knowledge that distinguishes one knowledge system from another. The IKS researcher has access to a robust reasoning and argumentation system exemplified by *nyaya*. The system of *nyaya* provides a realist-centered algorithmic view of investigation into propositions and candidates of hypotheses. Engaging with principles, axioms, and claims emanating from an alternate knowledge system needs to be examined within this framework as part of a well-structured method.

As an example, the multi-century academic studies area of Indology has almost exclusively engaged with Bharatiya Gyan and its texts, using the western episteme and reasoning methods. These studies did not use the native *pada, vakya, pramana* system or the native *shastra-paddhati* to engage with the body of knowledge. The texts were used to infer all manner of incorrect and flawed understandings that were put to effective colonial use but still impact us today as many our structures of polity are driven by these self-understandings, such as the “Caste system” or the “Aryan-Dravidian” divide. This paper exemplifies how such a clarity-driven epistemic commitment can be used to produce an impactful genre of IKS scholarship that is urgently needed in the current milieu.

10 Conclusion

In this paper, select aspects of Western ethics were critically examined through the n*yaya pancha-avayava* framework, revealing several foundational shortcomings. As digital and AI ethics are built upon these same premises, their limitations are not merely operational but stem from deeper philosophical assumptions. Our argument is that mainstream ethics—shaped by an anthropocentric, materialist worldview and the Baconian inspired scientific method—is fundamentally ill-equipped to address the complexities of AI ethics

We introduce a *nyaya*-inspired ethical framework that may better address emerging ethical issues, especially as India advances toward indigenous AI systems built on culturally contextualized datasets. With *dharma* as its foundation, this approach envisions an ethical trajectory that aligns with an Indian epistemic model where ethics (*rta*/*dharma*) is a veritable building block of the universe (*jagat*). A call is made to the community to examine the vast

literature of the *dharma shastra*, re-examine it in the light of the modern context, and without diluting its foundational episteme, put it to use for contemporary Bharat's needs.

By redefining ethics as intrinsic to purpose rather than as an external add-on, this Bharatiya approach provides a potentially transformative perspective on AI ethics conceptualization that can reshape how various stakeholders conceptualize and execute system design, policy formulations, and AI literacy programs.

References

Ayana, G., Dese, K., Daba Nemomssa, H., Habtamu, B., Mellado, B., Badu, K., ... & Kong, J. D. (2024). Decolonizing global AI governance: assessment of the state of decolonized AI governance in Sub-Saharan Africa. *Royal Society Open Science*, *11*(8), 231994.

Balasubramaniam, R. (2024). *Power within: The leadership legacy of Narendra Modi*. Penguin Random House India

Cerullo, M. (2023, June 30). *ChatGPT maker OpenAI sued for allegedly using "stolen private information."* CBS News. https://www.cbsnews.com/news/chatgpt-open-ai-lawuit-stolen-private-information

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, *10*(1), 1-12.

Cho, W. I., Kim, J., Yang, J., & Kim, N. S. (2021, March). Towards cross-lingual generalization of translation gender bias. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 449-457).

Corrêa, N. K., De Oliveira, N., & Massmann, D. (2022). On the Efficiency of Ethics as a Governing Tool for Artificial Intelligence. *arXiv preprint arXiv:2210.15289*.

Currie, G., Currie, J., Anderson, S., & Hewis, J. (2024). Gender bias in generative artificial intelligence text-to-image depiction of medical students. *Health Education Journal*, *83*(7), 732-746.

De Vries, R., & Rott, L. (2011). Bioethics as missionary work: The export of Western ethics to developing countries. In C. Myser (Ed.), *Bioethics around the globe* (pp. 3–18). Oxford University Press.

Dilmaghani, S., Brust, M. R., Danoy, G., Cassagnes, N., Pecero, J., & Bouvry, P. (2019, December). Privacy and security of big data in AI systems: A research and standards perspective. In *2019 IEEE international conference on big data (big data)* (pp. 5737-5743). IEEE.

Evans, R. J. (2005). *The Third Reich in power, 1933–1939*. Penguin Books.

Faragher, J. (2019). Is AI the enemy of diversity?. *People Management*, *6*.

Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, *6*(1), 3.

Finocchiaro, M. A. (2007). *Retrying Galileo, 1633–1992*. University of California Press

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972–1977* (C. Gordon, Ed.). Vintage.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press

Gopinath, K., & Sharma, S. D. (Eds.). (2022). *The computation meme: Computational thinking in the Indic tradition*. IISc Press

Goralnik, L., & Nelson, M. P. (2012). Anthropocentrism. In R. Chadwick (Ed.), *Encyclopedia of applied ethics* (2nd ed., pp. 145-155). Academic Press

Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., ... & Hubinger, E. (2024). Alignment faking in large language models. *arXiv preprint arXiv:2412.14093*.

Grynbaum, M. M., & Mac, R. (2023, December 27). *The Times sues OpenAI and Microsoft over A.I. use of copyrighted work*. The New York Times. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Hao, K. (2024, September 13). *Microsoft's hypocrisy on AI*. The Atlantic. https://www.theatlantic.com/technology/archive/2024/09/microsoft-ai-oil-contracts/679804

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335-346.

Israel, M. (2017). Ethical imperialism? Exporting research ethics to the Global South. In R. Iphofen & M. Tolich (Eds.), *The SAGE handbook of qualitative research ethics* (pp. 89–100). Sage Publications.

Jackson, M. C. (2021). Artificial intelligence & algorithmic bias: the issues with technology reflecting history & humans. *J. Bus. & Tech. L.*, *16*, 299.

Kane, P. V. (1930). *History of Dharmassstra (Vol. 1).* Bhandarkar Oriental Research Institute.

Kirchschläger, P. (2024, September 24). *Big Tech firms have consistently shown little concern about harming people and violating their rights*. Le Monde. https://www.lemonde.fr/en/opinion/article/2024/09/24/peter-kirchschlager-big-tech-firms-have-consistently-shown-little-concern-about-harming-people-and-violating-their-rights_6727074_23.html

Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society, 26*(1), 16-30

Kymlicka, W. (1995). *Multicultural citizenship: A liberal Theory of Minority Rights*. Clarendon Press.

Malhotra, R. (2011, June 13). *Dharma Is Not The Same As Religion*. https://www.huffpost.com/entry/dharma-religion_b_875314

Malhotra, R. (2013). *Being different: An Indian Challenge to Western Universalism*. HarperCollins India.

Malhotra, R. (2021). *Artificial intelligence and the future of power*. Rupa Publications.

Martin, K. (2022). Gamification, manipulation, and data analytics. In *Ethics of data and analytics* (pp. 357-361). Auerbach Publications.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature machine intelligence*, *1*(11), 501-507.

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics*, *3*(3), 869-877.

Mukhopadhyay, S. (2023). Deconstructing Devduttology. In R. Malhotra, & D. Reddy (Eds.), *Ten Heads of Ravana: A Critique of Hinduphobic Scholars* (pp. 112-163). Garuda Prakashan.

Mwaura, J. (2024, December 17). Rethinking artificial intelligence (AI) ethics for peacebuilding through African epistemologies. *Kujenga Amani – African Peacebuilding Network (APN), Social Science Research Council (SSRC).* Special issue: "Emerging issues in social media and artificial intelligence and peacebuilding in Africa." https://kujenga-amani.ssrc.org/2024/12/17/rethinking-artificial-intelligence-ai-ethics-for-peacebuilding-through-african-epistemologies-2

Parker, S. (2016). Gender fluidity. In *Ethical ripples of creativity and innovation* (pp. 165-173). London: Palgrave Macmillan UK.

Pavlick, E. (2023). Symbols and grounding in large language models. *Philosophical Transactions of the Royal Society A*, *381*(2251), 20220041.

Ramezanian, R. (2025, February 24). Bot busted up: AI chatbot's alleged data leak. Skyhigh Security. https://www.skyhighsecurity.com/about/resources/intelligence-digest/bot-busted-up-ai-chatbots-alleged-data-leak.html

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.

Sarukkai, S. (2005). *Indian philosophy and philosophy of science*. Motilal Banarsidass Publishers

Wong, M. (2024, August 30). Chatbots are primed to warp reality**.** *The Atlantic.* https://www.theatlantic.com/technology/archive/2024/08/chatbots-false-memories/679660

Yang, L., Tian, M., Xin, D., Cheng, Q., & Zheng, J. (2024). AI-driven anonymization: Protecting personal data privacy while leveraging machine learning. *Applied and Computational Engineering, 67*, 273-279.