

Review Article**Dharma-Driven AI: Integrating Indian Ethical Frameworks into Algorithmic Decision-Making**Dr. Viraj P Tathavadekar ¹,¹Individual Researcher, Pune, India.**ABSTRACT:**

Artificial intelligence systems in current times experience severe ethical problems which include algorithmic bias issues and accountability problems and fairness issues that depend on Western utilitarian frameworks. The chapter introduces a new model which combines Indian ethical standards with dharma and karma to govern algorithmic decision-making systems. Dharma-based frameworks define righteous behavior through three principles which include contextual righteousness and relational duties and cosmic order (Rta) as their core values. The chapter uses conceptual analysis and case studies from healthcare and justice systems to show how dharma provides moral guidance for ethical AI design while karma theory establishes strong accountability systems through its tracking of actions and their corresponding results. The analysis shows that dharma-driven methods to problem-solving improve Western methods because they support multiple cultural traditions and specific situations and ongoing effects on society. The integration of Indian Knowledge Systems with contemporary AI governance creates a pathway that enables development of algorithmic systems which respect different knowledge systems while maintaining ethical standards and supporting technological progress.

Keywords: Dharma, Artificial Intelligence Ethics, Algorithmic Accountability, Indian Knowledge Systems, Karma Theory, Ethical AI Governance

INTRODUCTION

AI and data technologies rapidly expanding into healthcare and justice represent deep challenges from accountability and bias (Sambasivan *et al.*, 2021). Western philosophies such as utilitarianism often break down at the boundaries of cultural nuance (Bruneault and Laflamme, 2021). By contrast, Indian Knowledge Systems offer dharma, a means of situational morality, and karma linking actions to consequences (Tuli & Gupta, 2024). This chapter considers an original way of decentralizing the ethics of AI by integrating these principles. The outcome would be governance that is not just context-aware, but also fundamentally accountable.

**Dharma as Framework for Ethical AI Design
Conceptual Foundations of Dharma**

Dharma represents sophisticated ethical philosophy encompassing cosmic order (Rta), social duty (Varna-ashrama-dharma), universal standards (Samanya-dharma), and contextual practice (Apad-dharma). The fundamental values of dharma adapt to different situations because Western ethics attempt to create universal principles (Tuli & Gupta,

2024). Five elements determine righteous action according to dharma: the actor's role (Adhikara) and specific circumstances (Desa-kala-patra) and intended consequences (Prayojana) and cosmic alignment (Rta). This framework helps AI ethics address complex, context-dependent situations.

Current AI systems face challenges because they cannot understand that people interpret the same actions with different moral importance based upon their situation. Healthcare AI needs to assess treatments through four factors which include statistical success and cultural beliefs and available resources and societal effects. Dharma enables systems to change their operational methods according to different environmental conditions.

Dharma-Based Design Principles

The criminal justice system needs to establish specific AI operational limits because its functions differ from what consumers need. Sadharana-dharma (universal duties) non-violence, truthfulness, non-stealing, self-restraint, non-possessiveness provides foundational constraints

Corresponding author: Dr. Viraj P Tathavadekar

DOI: 10.5281/zenodo.18961155

Received: 02 Feb 2026; **Accepted:** 05 Feb 2026; **Published:** 06 Feb 2026

Copyright © 2026 The Author(s): This work is licensed under a Creative Commons Attribution- Non-Commercial-No Derivatives 4.0 (CC BY-NC-ND 4.0) International License

which require organizations to minimize harm and maintain transparent operations while protecting privacy and conserving resources (Kumar & Sangwan, 2025). Apad-dharma (contextual adaptation) challenges one-size-fits-all approaches

because it requires algorithms to achieve fairness by addressing power imbalances and historical wrongs while considering specific local situations (Sambasivan *et al.*, 2021).

FIGURE 1: DHARMA-DRIVEN AI DESIGN FRAMEWORK

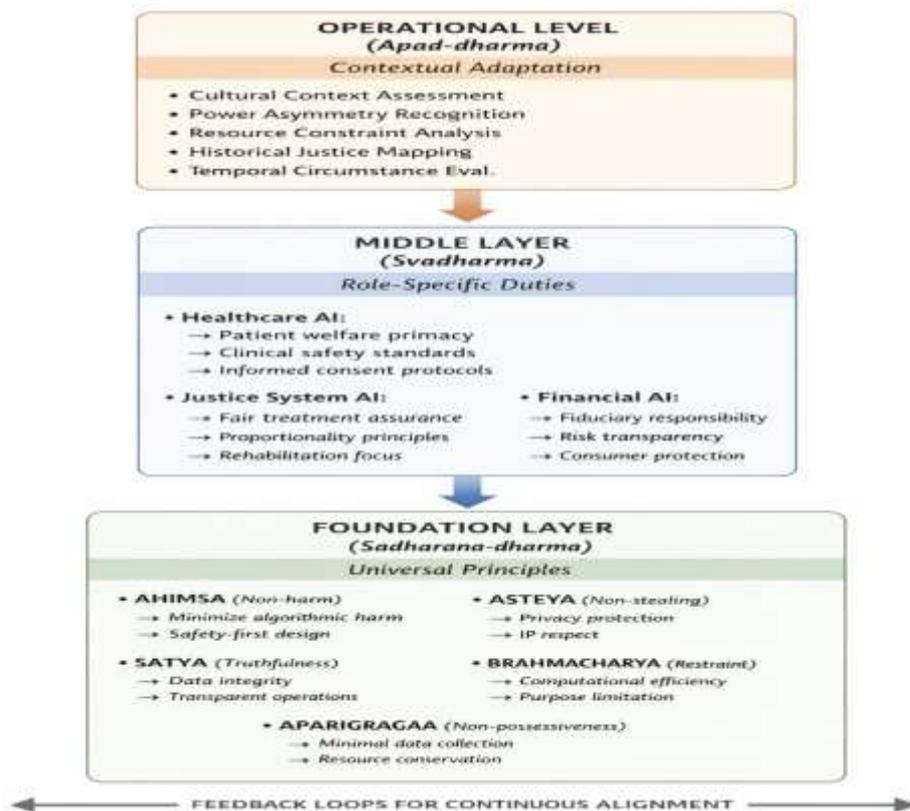


FIGURE 1: Dharma-Driven AI Design Framework
Source: Authors Creation

Figure 1 illustrates the Dharma-Driven AI Design Framework displays its hierarchical structure through universal principles which establish its base Sadharana-dharma and its middle layer develops role-specific duties through Svadharna and its operational level implements contextual adaptation mechanisms through Apad-dharma. The framework demonstrates how these three dimensions interact to guide ethical algorithmic decision-making while maintaining both principled coherence and contextual flexibility.

Comparative Advantages Over Western Frameworks

Western AI ethics usually between two poles, namely, utilitarianism giving upper hand to welfare at the cost of minority rights and deontology, which seems protective but can fail in proceeding

trade-offs. These frameworks often become a mismatch with local norms or compartmentalize application realities.

Dharma, while having certain universal principles, is flexible toward situational nuances.

This approach does not allow moral calculations that do not want to acknowledge the complexities of the various duty’s stakeholders may have with respect to individuals, with due weight given to context. Duty must flex in consideration of circumstances while respect for the general moral order needs to be conserved, which is a flexibility quite valuable for AI in deployment through diverse cultural settings (Berberich *et al.*, 2020).

In dharma's relational ontology, individuals have obligations to one another. This reflects the socio-technical perspective of AI where responsibility is distributed among developers, administrators, users, and the broader community, as opposed to it being individualist (Trothen *et al.*, 2024). This shift adds a more comprehensive avenue for AI governance by moving from individual rights to collective obligations.).

Karma Theory and AI Accountability Mechanisms

Karma as Accountability Architecture

Karma exists as a principle which connects actions to their consequences while maintaining moral responsibility for actions. The actor must face all outcomes which result from their actions because every action leads to specific results. Contemporary AI operates with divided responsibility among its various stakeholders including developers and deployers and data providers which leads to "responsibility gaps" in autonomous systems (Gill, 2020). The entire lifecycle of AI systems requires complete tracking of all actions and their resulting consequences according to karma theory.

Karma-inspired accountability mechanisms require complete documentation which connects every aspect of their system through established causal relationships to enable assessment of past damages. This system extends explainable AI because it demonstrates how moral responsibility and cause-based understanding should be evaluated beyond technical system comprehension (Rivera *et al.*, 2025).

Implementing Karmic Accountability in AI Systems

The implementation of karma in artificial intelligence requires three essential elements. First, immutable audit trails need to document all major decisions together with data origins preprocessing methods model selection procedures and deployment environments and intervention records. Second, consequence tracking across temporal scales. Healthcare AI affects patient lifetimes; criminal justice algorithms influence decades-long recidivism patterns. The karma-inspired monitoring system performs continuous longitudinal impact assessment through its established auditing process (Li *et al.*, 2023).

Third, Samyoga-karma (joint action) recognizes consequences arise from multiple causes, proportionally distributing responsibility among developers (architectural choices), data scientists (dataset curation), deployers (implementation), and users (operational decisions).

TABLE 1: Comparative Analysis of AI Accountability Mechanisms

Dimension	Western Legal Liability Model	Technical AI Accountability	Karma-Based Framework
Temporal Scope	Incident-focused Retrospective analysis Limited liability period (2–10 years) Statute of limitations constraints	Development-to-deployment cycle Short-term monitoring Deployment phase focus Rarely extends beyond initial release	Comprehensive lifecycle tracking multi-generational consequence tracking Includes historical causes and future impacts No temporal boundaries (Sanchita-karma)
Responsibility Distribution	Single entity liability Corporate personhood Hierarchical attribution Limited to direct actors Difficulty with AI's distributed nature	Developer-centric Organization-level accountability Team responsibility Struggles with complex supply chains	Distributed across ecosystem Proportional attribution Includes data providers, developers, deployers, users, policymakers Samyoga-karma (joint action) principle
Causation Model	Proximate cause 'But-for' test Linear causation Difficulty proving algorithmic causation	Feature importance Gradient-based attribution SHAP/LIME explanations Limited to model internals	Multi-level causal chains Direct and indirect causation Systemic and historical causation Karma-Phala (action consequence) linkage

Consequence Tracking	Complaint-triggered Reactive investigation Requires demonstrated harm Sporadic monitoring	Performance metrics Validation datasets A/B testing Limited long-term tracking	Proactive longitudinal monitoring multi-dimensional impact assessment Tracks unintended and societal impacts Immutable audit trails
Remediation Approach	Financial compensation Punitive damages Regulatory penalties Adversarial process	Model retraining Bias mitigation patches Technical fixes Developer-determined solutions	Multi-faceted restoration Root cause remediation Restorative processes Combines technical and social remediation

Table 1 presents the study compares three frameworks through their accountability mechanisms, which include Western legal liability models and current technical AI accountability approaches and the proposed karma-based framework. The comparison examines five dimensions: temporal scope, responsibility distribution, causation model, consequence tracking, and remediation approach.

Karma and Algorithmic Bias Mitigation

Karma theory addresses algorithmic bias through its mechanisms which extend beyond its statistical fairness standards. Current approaches treat bias as technical problems which exist separately from their historical origins. Karma demonstrates through its action-consequence chains that bias represents the totality of past unfair treatment (Kasirzadeh & Klein, 2021). Karmic approaches recognize past damages through their restoration processes which involve community-based data gathering and participatory design and specific disadvantage modeling (Ofosu-Asare, 2024). Sanchita-karma (accumulated consequences) requires addressing inherited patterns from legacy systems, explicitly remediating historical harms through their historical deployment which follows repair methods that seek to correct past system injustices.

Case Study: Dharma-Driven AI in Healthcare Contextual Treatment Recommendation Systems

Healthcare artificial intelligence demonstrates dharma-based operational frameworks through its implementation. The current treatment recommendation systems achieve optimal clinical results but fail to assess patient values together with cultural factors and resource limitations and quality-of-life compromises. Dharma-based healthcare AI would establish its purpose through its Svadharma duty to provide patient support while it honors personal freedom and cultural traditions and equitable treatment. The system would follow Sadharana-dharma standards by implementing Ahimsa which demands complete safety testing and its requirement for complete evidence disclosure through Satya. The system would establish data protection through Asteya which prohibits unauthorized access to information. Apad-dharma requires contextual adaptation to include patient-specific factors together with extra protection measures for groups that have faced historical

discrimination and budget limitations in resource-deficient areas and culturally appropriate treatment options for patients who hold strong religious or cultural beliefs.

Accountability Through Karmic Traceability

The implementation of karma-based accountability systems for healthcare artificial intelligence systems eliminates the "black box" problem because it establishes complete system transparency. The system maintains comprehensive action records which enable us to trace all elements that contributed to the recommendation, from the data bias and weighted uncertainties to the specific decision-making process. Karmic traceability enables detailed post-incident evaluation which uses system failures to create organizational improvements. The system tracks extended patient results while using these outcomes to update its predictive model through its longitudinal assessment feature. The system establishes a learning loop which uses historical actions to improve upcoming medical treatment.

Case Study: Dharma-Driven AI in Justice Systems

Algorithmic Fairness in Sentencing and Risk Assessment

Dharma-driven AI provides a deep solution to current risk assessment algorithms which the criminal justice system uses because they operate through hidden methods that show racial and socioeconomic discrimination (Sambasivan *et al.*, 2021). A dharmic system sees its Svadharma as serving true justice which requires matching proportionality needs with rehabilitation and restoration requirements instead of pursuing prediction accuracy.

The AI system uses Sadharana-dharma principles to operate according to predefined moral boundaries. The system must first protect defendants

and communities because non-violence mandates that all harm reduction efforts must come before any operational decisions. The system requires full decision-making transparency about all operations which will help maintain data integrity according to truth requirements. The approach requires systems to eliminate standardization which creates uniform solutions for all situations. Apad-dharma enables the system to understand that context matters because theft from desperate poverty has different moral implications than theft which occurs because of greed. The dharma-driven systems deliver justice through their approach which uses statistical data and brings in socioeconomic disadvantages together with systemic barriers as factors that have moral weight.

Restorative Accountability Through Karma

The principles of karma which describe justice led to restorative justice instead of punitive measures. A karma-based AI system operates beyond recidivism tracking to assess two elements of restoration progress which are harm repair and accountability proof. The system utilizes continuous monitoring to establish that justice needs ongoing assistance which extends beyond the initial sentencing.

The principle of Samyoga-karma (joint action) distributes accountability to all participating parties. Developers and policymakers and agencies share responsibility for all failures that occur with the algorithm. The system prevents "blaming the machine" by making all members of the ecosystem responsible for the results produced through AI-based decision making.

Operationalizing Dharma-Driven AI: Challenges and Pathways

Technical Implementation Challenges

Code as formal translation of these concepts is no small task. Actualization of the Dharma (justice) argument for flexibility may require hybrid systems that define core constraints and adapts to context (Chaturvedi *et al.*, 2024). A data infrastructure supporting Karma tracking must be extremely robust; here, privacy-preserving technologies such as federated learning may be used to guarantee independence at all costs (whenever possible). Lastly, outcome performances must hinge upon integral key culture translation, insulating the

praxis with sound practicality, though "deeply culturally grounded" in philosophical roots; and still yet appropriate in varied global contexts (Kumar & Sangwan, 2025).

Institutional and Governance Pathways

The deployment of AI based on its dharma depends on highly collaborative, diverse, IT-trained, and trained in Indian Knowledge Systems (IKS) workgroups. Control cannot be carried out merely from orthodox practices, but governance must be infused with society-induced checks and balances such as the establishment of community review boards (Woodgate and Ajmeri, 2024). In the end, we must pertain to some kind of industry-wide standards as there are in following karma. Whether it is in the field of aviation safety, in the class-of-architecture regulators must accurately require a comprehensive and myopic injury near point-of-attack towards ensuring institutions from proper watching's and eventual accountability.

CONCLUSION

This chapter paves the way for a vision of dharma and karma as foundational principles for ethical AI, a vision opposing the fragmentations in rigid Western models. Dharma offers a more intuitive balancing act between universally recognized rules and situational detail, while karma presents a transparent one-to-one correspondence between the action and the consequences in accordance with what Spade calls a "sumption." Just in case it was a philosophical idea, the integration of Indian Knowledge Systems is the only way to decolonize tech-ethics. By moving past Western monopolies, we become truly diverse and culturally diverse. Future research should drive us to deepening discussions into theoretical engagement in Mahabharata-like classical texts and tasting pilot programs in healthcare and justice. In perspective, dharma-based AI transcends the ideal of efficiency to create a technology at one with its culturally diverse audience and serving truly in the human interest.

REFERENCES

- Berberich, N., Nishida, T., & Suzuki, S. (2020). Harmonizing artificial intelligence for social good. *Philosophy & Technology*, 33(4), 613-

638. <https://doi.org/10.1007/S13347-020-00421-8>
2. Bruneault, F., & Sabourin Laflamme, A. (2021). AI ethics: How can information ethics provide a framework to avoid usual conceptual pitfalls? An overview. *AI & Society*, 36(2), 501-513. <https://doi.org/10.1007/S00146-020-01077-W>
 3. Cárdenas Rivera, A., Abhari, K., & Xiao, B. (2025). Responsible AI design: The authenticity, control, transparency theory. *Journal of the Association for Information Systems*, 26(1), 1-32. <https://doi.org/10.17705/1jais.00948>
 4. Chaturvedi, S., Patvardhan, C., & Vasantha Lakshmi, C. (2024). Human values from Indian philosophy to solve AI alignment problem. *International Journal for Multidisciplinary Research*, 6(3), 1-12. <https://doi.org/10.36948/ijfmr.2024.v06i03.22099>
 5. Gill, K. S. (2020). AI & Society: Editorial volume 35.2: The trappings of AI agency. *AI & Society*, 35(2), 293-302. <https://doi.org/10.1007/S00146-020-00961-9>
 6. Kasirzadeh, A., & Klein, C. (2021). The ethical gravity thesis: Marrian levels and the persistence of bias in automated decision-making systems. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 372-383. <https://doi.org/10.1145/3461702.3462606>
 7. Kumar, A., & Sangwan, S. R. (2025). Natural language processing as Digital Veda (डिजिटल वेद): A humanistic framework for language, ethics, and AI. *Digital Scholarship in the Humanities*, 40(1), 245-267. <https://doi.org/10.1093/llc/fqaf100>
 8. Ofosu-Asare, Y. (2024). Cognitive imperialism in artificial intelligence: Counteracting bias with indigenous epistemologies. *AI & Society*, 39(4), 1823-1841. <https://doi.org/10.1007/s00146-024-02065-0>
 9. Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). Re-imagining algorithmic fairness in India and beyond. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 315, 1-14. <https://doi.org/10.1145/3442188.3445896>
 10. Trothen, T. J., Kwok, P. L., & Lee, B. (2024). AI and East Asian philosophical and religious traditions: Relationality and fluidity. *Religions*, 15(5), Article 593. <https://doi.org/10.3390/rel15050593>
 11. Tuli, M., & Gupta, D. (2024). Ethics in a culture of dharma. *Review of General Psychology*, 28(4), 412-428. <https://doi.org/10.1177/10892680241289345>
 12. Woodgate, J., & Ajmeri, N. (2024). Macro ethics principles for responsible AI systems: Taxonomy and directions. *ACM Computing Surveys*, 56(11), Article 278, 1-37. <https://doi.org/10.1145/3672394>
 13. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023, September 30). Trustworthy AI: From Principles to Practices. *ACM Computing Surveys*, 55(9), 1-46. <https://doi.org/10.1145/3555803>