ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 18, Issue, 10, pp.38386-38389, October, 2026
DOI: https://doi.org/10.24941/ijcr.50969.10.2026

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

INDIAN PHILOSOPHY AND AI ETHICS: EPISTEMIC RESPONSIBILITY, NON-HARM, AND HUMAN AGENCY

Adrija Acharyya

Guest Teacher, Department of Philosophy, West Bengal State University, West Bengal, India

ARTICLE INFO

Article History:
Received 04th July, 2026
Received in revised form
16th August, 2026
Accepted 20th September, 2026
Published online 30th October, 2026

Key words:
Artificial Intelligence Ethics; Indian Philosophy; Nyāya Epistemology; Epistemic Responsibility; Machine Consciousness; Human Agency; Ahiṃsā; Buddhist Ethics; Jain Anekāntavāda; AI Governance.

***Corresponding author:**
Adrija Acharyya

ABSTRACT

Artificial intelligence (AI) increasingly performs tasks associated with human cognition, including linguistic production, classification, prediction, and decision-making. This development intensifies philosophical questions about knowledge, agency, responsibility, consciousness, and harm. This paper argues that Indian philosophy can contribute to AI ethics not by supplying an ancient "AI ethics" but by offering conceptual resources for critically evaluating technologically mediated cognition and action. Nyāya provides a rigorous distinction between a knower, means of knowledge, object, and valid cognition, thereby challenging the automatic epistemic authority of AI output. Sāṃkhya and Advaita Vedānta complicate the identification of computational intelligence with consciousness, while Viśiṣṭādvaita raises questions about the distinction between conscious selves and insentient matter. Buddhist ethics foregrounds the reduction of suffering; Jain anekāntavāda and syādvāda caution against one-sided representations and overconfident classification. Concepts such as ahiṃsā, satya, aparigraha, karma, and loka-saṃgraha can be reconstructed as principles of non-harm, information integrity, data restraint, responsibility, and social welfare. The resulting framework places human responsibility, epistemic humility, and protection of sentient beings at the centre of AI governance while remaining open to future arguments about machine consciousness and moral status.

Copyright©2026, Adrija Acharyya. *This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Citation: ***Adrija Acharyya,*** **2026.** "Indian Philosophy and AI Ethics: Epistemic Responsibility, Non-Harm, and Human Agency*". International Journal of Current Research*, 18, (10), 38386-38389.

INTRODUCTION

Artificial intelligence is no longer confined to mechanical automation. Generative models can produce fluent language, summarize and transform information, write code, recognize patterns, and participate in decision-support. The philosophical difficulty is that successful performance can be mistaken for knowledge, understanding, or agency. A system may generate a medically plausible answer without possessing the epistemic standing of a physician; it may produce a first-person statement without thereby becoming a conscious subject. Contemporary frameworks similarly stress human oversight, accountability, transparency, privacy, fairness, and the prevention of harm.1 The central claim of this paper is that Indian philosophy can deepen these debates by distinguishing capacities that are often collapsed under the word "intelligence." Its traditions offer resources for three connected tasks: evaluating the epistemic status of machine outputs, distinguishing cognition from consciousness, and assigning responsibility for technological consequences. The proposal is reconstructive rather than antiquarian: no Indian school anticipated machine learning, so its concepts cannot simply be converted into modern policy rules.

Nyāya: From Computational Output to Epistemic Authority: Nyāya is especially valuable because it does not identify every cognition with valid knowledge. Its classical analysis distinguishes *pramātṛ* (knower), *pramāṇa* (means of knowledge), *prameya* (object), and *pramā* (valid cognition). The four principal *pramāṇas* are perception, inference, comparison, and testimony.2 This structure exposes a basic problem in treating AI output as knowledge: producing a correct-looking proposition is not by itself sufficient to establish its epistemic status. For Nyāya, cognition is successful when it is veridical; the important issue is therefore the relation between cognition and its object.3 Applied to generative AI, linguistic fluency, statistical confidence, or resemblance to expert discourse cannot themselves confer epistemic authority. An AI-generated proposition must be checked against appropriate evidence and, where relevant, against independent *pramāṇas*. The issue is therefore not merely fact-checking, but the grounds on which AI may legitimately function as a source of belief. The distinction also clarifies responsibility. If an AI system identifies a disease from an image, the relevant question is not simply whether "the AI knows" but who occupies the epistemic roles involved: the system, its designers, the physician interpreting its output, or

1 UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2022), principles on accountability, transparency, human oversight, and non-discrimination.

2 Gautama, *Nyāya-sūtra* 1.1.3; see Matilal, *Perception* (1986), chs. 1 and 4.

3Matilal, *Perception* (1986), on the Nyāya relation between cognition and object.

the institution deploying it. At present, the safer philosophical position is to treat AI as an instrument within a human epistemic practice rather than automatically as a *pramātṛ*. Future systems may warrant a different account; present performance does not settle the question.

Intelligence, Consciousness, and Agency: The problem becomes sharper when AI is described as intelligent or autonomous. Sāṃkhya distinguishes *puruṣa*, pure consciousness, from *prakṛti* and its evolutes, including *buddhi*, *ahaṃkāra*, and *manas*.4 The philosophical lesson is that sophisticated cognitive functioning need not be identical with consciousness. A system's capacity to classify, predict, or generate language therefore does not by itself establish subjective awareness. Advaita Vedānta gives a different but related reason for caution. Consciousness (*cit*) is not treated as an accidental computational property; mind is an instrument through which consciousness is manifested in embodied cognition.5 Thus, an AI system's first-person linguistic reports cannot alone establish that there is an experiencing subject behind them. Viśiṣṭādvaita likewise distinguishes conscious selves (*cit*) from insentient reality (*acit*), making the metaphysical status of an artificial system a substantive question rather than a conclusion to be drawn from functional sophistication.6 These traditions do not prove that machines can never be conscious; they prevent an inference from functional similarity to metaphysical identity. The ethical consequence is equally important. If an AI system is not independently established as a moral subject, assigning responsibility to "the algorithm" can obscure the agency of developers, owners, deployers, and human decision-makers. UNESCO's contemporary principle that AI should not displace ultimate human responsibility converges with this concern, although the philosophical grounds are different.7

Ahiṃsā, Satya, Aparigraha, and the Moral Limits of AI: Indian ethical traditions provide further resources for assessing AI's effects. *Ahiṃsā*, prominently articulated in Jain, Buddhist, and Yoga traditions, can be reconstructed as a principle of avoiding foreseeable harm. Patañjali lists *ahiṃsā* among the *yamas* (*Yoga Sūtra* 2.30).8 In AI ethics, non-harm must extend beyond physical injury to discrimination, manipulation, privacy violations, reputational damage, economic exploitation, and systems that predictably intensify social vulnerability. Ahiṃsā is not identical with modern non-maleficence, but supplies a broader vocabulary for asking who can be harmed and how.

Satya is equally relevant to generative AI. Fabricated citations, synthetic misinformation, deceptive deepfakes, and the presentation of uncertainty as fact threaten the conditions of responsible inquiry. A *satya*-oriented reconstruction would require systems and users to distinguish established information from inference and uncertainty. This complements Nyāya's demand for epistemic warrant: truthfulness concerns how information is represented, while *pramāṇa* concerns the grounds on which cognition is justified.

Aparigraha, traditionally associated with non-hoarding, offers a critical question for data-intensive AI: does the technical ability to collect and retain information establish a moral entitlement to do so? It does not. The concept can be reconstructed as a principle of restraint against unnecessary accumulation and exploitation of personal data.9 Such a reconstruction is especially pertinent where surveillance or behavioural profiling exceeds what is necessary for a legitimate purpose. The concept of *karma* should not be reduced to a simple doctrine of technological causation. More cautiously, its emphasis on action and consequence can illuminate distributed responsibility. Data collection, model design, deployment, and human reliance form a chain through which consequences are produced. Moral responsibility therefore cannot be discharged merely by pointing to an autonomous system. *Loka-saṃgraha*, the *Bhagavad Gītā*'s concern with sustaining social order and welfare, similarly invites evaluation of AI beyond private efficiency or profit.10

Buddhist and Jain Contributions: Suffering, Plurality, and Epistemic Humility: Buddhist ethics centres moral evaluation on the reduction of *duḥkha*. This provides a consequentially sensitive question for AI: does a particular use reduce or increase suffering among sentient beings? The question applies both to beneficial uses—such as health, accessibility, and disaster response—and to harmful uses, including manipulation, discriminatory profiling, surveillance, and autonomous violence. Buddhist epistemology also distinguishes perceptual cognition from conceptual construction and inference. Dignāga's account of perception is therefore relevant to the warning that an AI representation should not be confused with direct access to reality.11

Jain epistemology contributes a different corrective. *Anekāntavāda* resists the assumption that a complex object can be adequately captured from a single standpoint, while *syādvāda* qualifies predication according to the conditions under which a claim is made.12 In algorithmic decision-making, this suggests epistemic humility. A data profile is not the person; a probabilistic classification is not an exhaustive description of the individual. In high-stakes domains such as employment, education, credit, medicine, and criminal justice, an algorithmic output should therefore be presented as conditioned by its data, model, and evidential limits rather than as an unqualified statement about a person. Jain logic does not supply a ready-made algorithmic method; its significance is to resist reducing plural, context-dependent realities to one computational description.

Toward a Reconstructed Indian Framework for AI Ethics: The preceding arguments support a compact framework of six principles. First, epistemic responsibility: AI outputs require proportionate verification and should not gain authority merely from computational sophistication. Second, human responsibility: responsibility for consequential decisions must remain traceable to identifiable human and institutional actors. Third, non-harm: foreseeable physical, psychological, social, economic, and environmental harms must be considered.

4 Gerald James Larson, *Classical Sāṃkhya* (Delhi: MotilalBanarsidass, 1969).
5Śaṅkara, *Brahma-sūtra-bhāṣya*, trans. Swami Gambhirananda (Kolkata: AdvaitaAshrama, 1965).
6Rāmānuja, *Śrībhāṣya*, trans. George Thibaut, SBE 48 (Oxford: Clarendon Press, 1904).
7 UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2022), principle on human oversight.
8Patañjali, *Yoga-sūtra* 2.30; see Edwin F. Bryant, *The Yoga Sūtras of Patañjali* (2009).
9Patañjali, *Yoga-sūtra* 2.30, 2.39; the present use of *aparigraha* is a contemporary reconstruction of restraint.
10*Bhagavad Gītā* 3.20–25; see S. Radhakrishnan, trans., *The Bhagavadgītā* (1948).
11 Jonathan Stoltz, *Illuminating the Mind* (Oxford: Oxford University Press, 2021), 51–73.
12Padmanabh S. Jaini, *The Jaina Path of Purification* (1979); see also Matilal, *The Character of Logic in India* (1998).

Fourth, truthful representation: systems should distinguish knowledge, inference, uncertainty, and fabrication. Fifth, restraint: data collection and behavioural influence require ethical limits rather than being justified solely by technical possibility. Sixth, plurality and humility: algorithmic classifications should remain open to contextual information, alternative perspectives, and correction. These principles intersect with contemporary AI governance without being reducible to it. NIST's AI Risk Management Framework, for example, identifies validity and reliability, safety, security, accountability, transparency, explainability, privacy, and fairness as dimensions of trustworthy AI.13 UNESCO similarly places human dignity, human rights, responsibility, transparency, human oversight, sustainability, and non-discrimination at the centre of its Recommendation on the Ethics of Artificial Intelligence.14 Indian philosophy can complement these frameworks by asking what counts as knowledge, what constitutes a subject, how action becomes attributable, and how human flourishing should constrain technological power.

Critical Qualifications and Conclusion: A philosophically responsible use of Indian traditions must avoid two errors. The first is homogenization: Nyāya, Buddhism, Jainism, Sāṃkhya, Advaita, and Viśiṣṭādvaita disagree on fundamental questions of self, knowledge, and reality. The second is anachronism: none of these traditions addressed neural networks, algorithmic bias, generative models, or digital surveillance. The proper method is therefore critical reconstruction rather than the discovery of supposed ancient predictions of AI. The resulting perspective is modest but important. AI should not be granted epistemic authority simply because it produces persuasive outputs; functional intelligence should not be equated with consciousness; and apparent machine autonomy should not become a device for evading human responsibility. Nyāya supplies epistemic discipline, Sāṃkhya and Vedānta sharpen the consciousness question, Buddhism foregrounds suffering, Jainism encourages plural and qualified judgment, and ethical concepts such as *ahiṃsā*, *satya*, *aparigraha*, *karma*, and *loka-saṃgraha* widen the field of moral evaluation. Together they support a human-centred, epistemically responsible approach to AI. The decisive question is therefore not merely how intelligent machines may become, but what forms of human knowledge, agency, and social life their development ought to sustain.

REFERENCES

1. Bhagavad Gītā. 3.20–25. In S. Radhakrishnan, trans., *The Bhagavadgītā*. London: George Allen & Unwin, 1948.
2. Dignāga. *Pramāṇasamuccaya*. See also Jonathan Stoltz, *Illuminating the Mind: An Introduction to Buddhist Epistemology*. Oxford: Oxford University Press, 2021.
3. Jaini, Padmanabh S. *The Jaina Path of Purification*. Berkeley: University of California Press, 1979.
4. Larson, Gerald James. *Classical Sāṃkhya: An Interpretation of Its History and Meaning*. Delhi: Motilal Banarsidass, 1969.
5. Matilal, Bimal Krishna. *Perception: An Essay on Classical Indian Theories of Knowledge*. Oxford: Clarendon Press, 1986.
6. Matilal, Bimal Krishna. *The Character of Logic in India*. Edited by Jonardon Ganeri and Heeraman Tiwari. Albany: State University of New York Press, 1998.
7. Rāmānuja. *Śrībhāṣya*. Translated by George Thibaut. Sacred Books of the East, vol. 48. Oxford: Clarendon Press, 1904.
8. Radhakrishnan, S., and Charles A. Moore, eds. *A Source Book in Indian Philosophy*. Princeton: Princeton University Press, 1957.
9. Śaṅkara. *Brahma-Sūtra-Bhāṣya*. Translated by Swami Gambhirananda. Kolkata: Advaita Ashrama, 1965.
10. Stoltz, Jonathan. *Illuminating the Mind: An Introduction to Buddhist Epistemology*. Oxford: Oxford University Press, 2021.
11. Tabassi, Elham. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology, 2023. https://doi.org/10.6028/NIST.AI.100-1.
12. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, 2022.
13. Yoga Sūtra of Patañjali. 2.30, 2.39. See Edwin F. Bryant, *The Yoga Sūtras of Patañjali: A New Edition, Translation, and Commentary*. New York: North Point Press, 2009.

REFERENCES

1. Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press, 2014.
2. Bryant, Edwin F. *The Yoga Sūtras of Patañjali: A New Edition*, Translation, and Commentary. New York: North Point Press, 2009.
3. Buddhaghosa. *Visuddhimagga: The Path of Purification.* Translated by BhikkhuÑāṇamoli. Seattle: Buddhist Publication Society, 2011.
4. Cappelen, Herman, and Rachel Sterken, eds. *Communicating with AI: Philosophical Perspectives.* Oxford: Oxford University Press, 2026.
5. Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory.* Oxford: Oxford University Press, 1996.
6. Coeckelbergh, Mark. *AI Ethics*. Cambridge, MA: MIT Press, 2020.
7. Coeckelbergh, Mark. *Introduction to Philosophy of Technology*. New York: Oxford University Press, 2019.
8. Dennett, Daniel C. *Consciousness Explained.* Boston: Little, Brown and Company, 1991.
9. Dharmakīrti. *Nyāyabindu*. Translated by F. Th. Stcherbatsky. In Buddhist Logic, vol. 2. Leningrad: Izdatel'stvoAkademiiNauk SSSR, 1930. Reprint, Delhi: MotilalBanarsidass, 1992.
10. Dignāga. *Pramāṇasamuccaya*. Edited and translated selections by Masaaki Hattori. Harvard Oriental Series, vol. 47. Cambridge, MA: Harvard University Press, 1968.
11. Dubber, Markus D., Frank Pasquale, and Sunit Das, eds. *The Oxford Handbook of Ethics of AI.* Oxford: Oxford University Press, 2020.
12. Floridi, Luciano. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities.* Oxford: Oxford University Press, 2023.
13. Floridi, Luciano, and Massimo Chiriatti. *GPT-3: Its Nature, Scope, Limits, and Consequences. Minds and Machines* 30 (2020): 681–694.

13ElhamTabassi, *AI Risk Management Framework 1.0*, NIST AI 100-1 (2023).
14 UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2022).

14. Gangeśa. *Tattvacintāmaṇi: The Jewel of Reflection on the Truth.* Translated selections from the Pratyakṣa, Anumāna, and related sections. New Delhi: MotilalBanarsidass.
15. Gautama (Akṣapāda). *The NyāyaSūtras of Gautama: With the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara.* Translated by GanganathaJha. 4 vols. Delhi: MotilalBanarsidass, 1984.
16. Gunkel, David J. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics.* Cambridge, MA: MIT Press, 2012.
17. Haugeland, John. *Artificial Intelligence: The Very Idea.* Cambridge, MA: MIT Press, 1985.
18. Īśvarakṛṣṇa. *SāṃkhyaKārikā of ĪśvaraKṛṣṇa.* Translated and annotated by Swami Virupakshananda. Chennai: Sri Ramakrishna Math, 2023.
19. Jaini, Padmanabh S. *The Jaina Path of Purification.* Berkeley: University of California Press, 1979.
20. Keśavamiśra. *Tarkabhāṣā*. Edited and translated editions. Varanasi: Chowkhamba Sanskrit Series.
21. Larson, Gerald James. *Classical Sāṃkhya: An Interpretation of Its History and Meaning.* Delhi: MotilalBanarsidass, 1969.
22. Liao, S. Matthew, ed. *Ethics of Artificial Intelligence.* Oxford: Oxford University Press, 2020.
23. Matilal, Bimal Krishna. *Perception: An Essay on Classical Indian Theories of Knowledge.* Oxford: Clarendon Press, 1986.
24. Matilal, Bimal Krishna. *The Character of Logic in India.* Edited by JonardonGaneri and Heeraman Tiwari. Albany: State University of New York Press, 1998.
25. McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* 1955.
26. Moor, James H. *The Nature, Importance, and Difficulty of Machine Ethics.* IEEE Intelligent Systems 21, no. 4 (2006): 18–21.
27. Patañjali. *The Yoga Sūtras of Patañjali.* Translated and edited by Edwin F. Bryant. New York: North Point Press, 2009.
28. Radhakrishnan, S., *The Bhagavadgītā.* (Trans) London: George Allen & Unwin, 1948.
29. Radhakrishnan, S., and Charles A. Moore, eds. *A Source Book in Indian Philosophy.* Princeton: Princeton University Press, 1957.
30. Rāmānuja. *Śrībhāṣya*. Translated by George Thibaut. Sacred Books of the East. Oxford: Clarendon Press, 1904–1907.
31. Russell, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control.* New York: Viking, 2019.
32. Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach.* 4th ed. Harlow: Pearson, 2021.
33. Śaṅkara. *Brahma-Sūtra-Bhāṣya.* Translated by Swami Gambhirananda. Kolkata: AdvaitaAshrama, 1965.
34. Śaṅkarācārya. *Brahma-sūtra-bhāṣya of ŚrīŚaṅkarācārya.* Translated by Swami Gambhirananda. 2nd ed. Kolkata: AdvaitaAshrama, 1972.
35. Searle, John R. *Minds, Brains, and Programs.* Behavioral and Brain Sciences 3, no. 3 (1980): 417–457.
36. Searle, John R. *The Rediscovery of the Mind.* Cambridge, MA: MIT Press, 1992.
37. Stoltz, Jonathan. *Illuminating the Mind: An Introduction to Buddhist Epistemology.* Oxford: Oxford University Press, 2021.
38. Tabassi, Elham. *Artificial Intelligence Risk Management Framework* (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology, 2023.
39. Turing, Alan M. *Computing Machinery and Intelligence.* Mind 59, no. 236 (1950): 433–460.
40. Umāsvāti (Umāsvāmin). *TattvārthaSūtra.* Translated by K. K. Dixit. Ahmedabad: L. D. Institute of Indology, 2000.
41. VācaspatiMiśra. *Tattvakaumudī*. Commentary on Īśvarakṛṣṇa's*Sāṃkhyakārikā*. Translated editions. Varanasi: Chowkhamba.
42. Vātsyāyana. *Nyāyabhāṣya*. *In The NyāyaSūtras of Gautama*, translated by GanganathaJha. Delhi: MotilalBanarsidass.
43. ViśvanāthaNyāyapañcānana. *Bhāṣāpariccheda with Siddhāntamuktāvalī*. Edited editions. Varanasi: Chaukhamba Sanskrit Pratishthan.
44. Vyāsa. *Mahābhārata*. Translated by BibekDebroy. 10 vols. New Delhi: Penguin Random House India, 2015–2016.
45. Winfield, Alan F. T. *Robotics: A Very Short Introduction.* Oxford: Oxford University Press, 2012.

****