



INDIC KNOWLEDGE AND ARTIFICIAL INTELLIGENCE: DEVELOPING CULTURALLY GROUNDED INTELLIGENT SYSTEMS

Amit Punia¹ and Nandini²

¹Assistant Professor, Department of Management, Galgotias University, Noida.

²Ph.D Scholar, Department of English and Foreign Languages, Central University of Haryana.

Abstract:

The dominant discourse linking Indian intellectual traditions to artificial intelligence oscillates between two unproductive extremes: a revivalist claim that the *śāstras* already contain modern computing, and a dismissive assumption that pre-modern Indian thought can contribute nothing beyond training data in Indian languages. This paper argues for a third position. Indic knowledge systems constitute not a repository of anticipated results but a set of unusually well-developed *formal disciplines* grammatical, epistemological, logical, hermeneutic and aesthetic whose analytical categories bear directly on unresolved problems in contemporary machine intelligence. Working from primary *sūtra* texts, the paper develops five such correspondences: *Pāṇini's Aṣṭādhyāyī* as a conflict-resolving generative rule system with an explicit metalanguage; *kāraka* theory and *Bhartr̥hari's sphoṭa* as a critique of token-additive models of meaning; *Nyāya pramāṇa* theory, and particularly the condition of *āptatva*, as an epistemology adequate to the problem of machine fabrication; the *hetvābhāsa* taxonomy as a ready-made classification of reasoning failure for evaluation design; and *anekāntavāda*, *catuṣkoṭi* and the *vāda* protocols as frameworks for representing contested rather than consensual knowledge. It further argues that *rasa* theory offers a more precise account of aesthetic response than affect classification, and that the *puruṣārtha-dharma* complex yields an ethics of situated obligation distinguishable from principle-based AI governance. These claims are tested against documented computational work in Sanskrit parsing, Indic-language modelling, manuscript recovery, computational musicology and *prakṛti*-based health informatics.

Keywords: Indic Knowledge Systems, Artificial Intelligence, Pāṇini, Pramāṇa, Computational Linguistics, AI Ethics

1. INTRODUCTION

Every technology arrives carrying an implicit theory of knowledge. A system that answers questions has committed itself, whether or not its designers say so, to a position on what counts as knowing, what counts as evidence, who counts as a reliable informant, and what should happen when sources disagree. The remarkable feature of contemporary artificial intelligence is not that it holds such commitments all technologies do but that it holds them almost entirely tacitly, in engineering decisions about tokenization corpus selection, loss functions and reward models, rather than in any articulated epistemology. This is precisely the point at which a tradition that spent two millennia arguing explicitly about the sources and limits of knowledge becomes relevant.

Indian intellectual culture developed, over that long period, a set of disciplines of extraordinary formal rigour: a grammar of some four thousand rules that generates a natural language from a finite ordered system; an epistemology that classifies the means of valid cognition and specifies the failure conditions of each; a logic with a worked taxonomy of fallacies; a hermeneutics devised to extract determinate injunctions from ambiguous text; a theory of debate with procedural rules; and an aesthetics that treats emotional response as a structured, analysable event. *Kauṭilya* called the discipline of systematic reasoning the lamp of all the sciences:

प्रदीपः सर्वविद्यानामुपायः सर्वकर्मणाम् ।

आश्रयः सर्वधर्माणां शश्वदान्वीक्षिकी मता ॥

The lamp of all the sciences, the means of all action, the support of all obligation — such is *ānvīkṣikī*, the discipline of critical inquiry, ever held to be. (Arthaśāstra 1.2.12)

The prevailing discourse connecting these materials to artificial intelligence has nonetheless been poorly served by its own



Cover Page



advocates. One strand, popular and increasingly political, claims that Sanskrit is uniquely suited to machine processing, that the *śāstras* anticipated computation, or that ancient India possessed the substance of modern science in encoded form. Such claims are largely unfalsifiable and have been carefully criticised (Nanda 2003). The opposite strand, common in the engineering literature, treats Indian material as a data problem only: languages to be resourced, scripts to be recognised, corpora to be scraped. Both strands share an assumption worth rejecting that the tradition's possible contribution lies in its *content*.

This paper argues instead that the contribution lies in the tradition's *method*: in the analytical categories by which it formalised language, evidence, inference, disagreement, feeling and obligation. The argument is developed in five stages, each pairing a primary text with a live technical problem, and is then tested against documented computational projects. The paper's overall claim is modest in form and substantial in consequence: that a culturally grounded intelligent system is not one trained on Indian data or decorated with Indian names, but one whose representational and evaluative commitments have been examined using a conceptual vocabulary adequate to the culture it serves.

2. REVIEW OF LITERATURE

The literature relevant to this enquiry has developed in four largely unconnected bodies, and their disconnection is itself part of the problem this paper addresses. The first is the encounter between Sanskrit grammar and computational linguistics. Rick Briggs's much-cited paper in *AI Magazine* (1985) argued that the *śāstric* analysis of sentences into semantic relations resembled the semantic-net formalisms then current in knowledge representation. The paper is frequently over-read as proving Sanskrit's computational superiority, a claim it does not make and cannot sustain; its durable value is the narrower observation that the Indian grammarians had produced a formalised representation of sentence meaning independent of surface word order. That observation was developed rigorously by Bharati, Chaitanya and Sangal (1995), whose *Pāṇinian* dependency framework for Indian-language parsing remains foundational for the field, and it has since been carried forward in the Sanskrit computational linguistics community by Huet's computational toolkit and Kulkarni's *samsādhant* parsers, together with the work presented across the Sanskrit Computational Linguistics symposia (Huet, Kulkarni & Scharf 2009).

The second body of work concerns Indian logic and epistemology as philosophy. Ingalls's study of Navya-Nyāya (1951) opened the technical language of the later logicians to Western readers; Matilal (1986; 1998) established Indian epistemology as a live interlocutor in analytic philosophy; Ganeri (2001; 2011) has shown the *Nyāya* account of testimony and inference to be philosophically serious rather than merely historical; and Mohanty (1992) provided a systematic treatment of *pramāṇa* theory. Almost none of this literature engages artificial intelligence, and almost none of the AI literature on machine testimony and hallucination engages it.

The third body is the critical study of AI in Indian social conditions. Sambasivan and colleagues (2021a) argue that fairness frameworks developed for Western demographic categories misfire in India, where caste, religion, gender and region interact in ways that standard protected-attribute schemas do not capture; their related work on data quality (2021b) documents how the least-valued stage of the pipeline is where most harm originates. Bender and colleagues (2021) supply the general critique of scale-driven language modelling; Gebru and colleagues (2021) supply the documentation practice this paper recommends. The Indian state's own intervention, the *Bhāṣiṇī* mission and the AI4Bharat programme's *IndicNLP* resources (Kakwani et al. 2020; Ramesh et al. 2022), constitutes the largest applied effort at Indic-language machine capability to date.

The fourth body is the discourse of Indic knowledge systems as a curricular and policy category, given institutional force by the National Education Policy 2020 and articulated intellectually by Kapoor (2005) and, in a more polemical register concerned specifically with AI and power, by Malhotra (2021). The gap this paper occupies is in the space between these four literatures: a philosophically informed, technically specific and politically careful account of what Indic knowledge systems can contribute to the design of intelligent systems, and of what they cannot.



3. METHODOLOGY AND SCOPE

The method of this paper is conceptual and comparative rather than experimental. It proceeds by close reading of primary *sūtra* literature in Sanskrit- Aṣṭādhyāyī, Nyāyasūtra, Vaiśeṣikasūtra, Mīmāṃsāsūtra, Vākyapadīya, Nāṭyaśāstra, Madhyamakakārikā with attention to what each text formalises and how, and then by mapping those formal structures onto specified problems in machine intelligence. Where a mapping is genuinely structural it is asserted; where it is merely suggestive it is marked as such. This distinction is the paper's principal methodological safeguard against the analogical overreach that has damaged the field.

Three scope limits should be stated plainly. First, the paper makes no claim of historical anticipation: that *Pāṇini's* system is formally describable in terms later developed by Post, Chomsky and Backus is a fact about formal systems, not evidence that *Pāṇini* was doing computer science.¹ Second, "Indic" is used here to name a plural and internally disputatious field including Buddhist, Jain and materialist positions, not a synonym for a single religious tradition; the disagreements among these schools are, as sections 8 and 9 argue, among the tradition's most usable assets. Third, the paper does not present new empirical results. Its applied section reports and interprets published work, and its recommendations are offered as a research programme to be tested rather than as findings.

Transliteration follows IAST. *Sūtra* numbering follows standard printed editions, with the caveat that numbering varies across recensions and commentarial traditions.² Translations are the authors' working renderings, intended for argumentative clarity rather than literary elegance.

4. WHAT IS MEANT BY INDIC KNOWLEDGE SYSTEMS

The phrase requires definition before it can do any work. Indic knowledge systems are best understood not as a body of doctrine but as an ecology of *śāstra*: disciplines each possessing a defined subject matter, a technical vocabulary, a canonical *sūtra* text, a commentarial apparatus that tests and revises it, and a set of standards for what counts as a valid claim within it. What distinguishes this ecology is the degree of its formalisation. *Vaiśeṣika* opens by declaring an ontological programme:

धर्मविशेष प्रसूतात् द्रव्यगुणकर्मसामान्य विशेषसमवायानां पदार्थानां
 साधर्म्यवैधर्म्याभ्यां तत्त्वज्ञानान्निःश्रेयसम् ॥ १.१.४ ॥

The highest good arises from knowledge of the categories- substance, quality, action, universal, particular and inherence attained through their similarities and differences. (*Vaiśeṣikasūtra* 1.1.4)

The relevance to intelligent systems is direct and unsentimental. This is an *ontology* in the sense the term carries in knowledge engineering: an enumerated set of categories of being, with relations among them, defined by shared and distinguishing properties, and intended to be exhaustive. *Kaṇāda's* inclusion of *samavāya*, inherence, as a category in its own right addresses a problem every ontology designer meets how a quality is bound to its substrate such that the pair is one thing and not two and the several centuries of Indian debate over whether *samavāya* is coherent constitute an argument any serious formal ontologist would recognise.

Three further features of this ecology matter for what follows. It is *metalinguistically self-aware*: *śāstras* routinely specify the conventions by which they are to be read, and *Pāṇini* devotes rules to the interpretation of his own rules. It is *adversarial by construction*: a philosophical position is established only after the strongest form of the opponent's view, the *pūrvapakṣa*, has been stated, so that the tradition preserves its own best objections. And it is *compression-oriented*: the *sūtra* form optimises for memorability and minimal statement, on the principle that a saving of half a syllable is cause for celebration a design constraint that produced techniques of encoding, such as the *kaṭapayādi* and *bhūtasamkhyā* numeral systems, of some interest to anyone who thinks about representation under resource limits.



1. PĀṆINI AND THE ARCHITECTURE OF RULE-GOVERNED GENERATION

The *Aṣṭādhyāyī* is the most sophisticated pre-modern formal system known, and its interest for artificial intelligence lies not in its subject matter but in its architecture. *Pāṇini* constructs a generative device: a finite inventory of roots and affixes, a system of abbreviatory symbols, and an ordered set of rules that derives well-formed Sanskrit utterances. Crucially, he also builds the apparatus needed to manage such a system at scale a metalanguage, a convention that a term names its own form unless otherwise designated, and explicit machinery for rule conflict:

स्वं रूपं शब्दस्याशब्दसंज्ञा।

A word in this grammar denotes its own form, unless it is a technical term of the system. (*Aṣṭādhyāyī* 1.1.68)
 विप्रतिषेधे परं कार्यम्

When two rules conflict, the later one applies. (*Aṣṭādhyāyī* 1.4.2)

These two *sūtras* together solve problems that any large rule-based system must solve and that many fail to solve. The first is the object-language/metalanguage distinction without it, a system that talks about symbols cannot avoid confusing them with what they denote. The second is deterministic conflict resolution: rather than leaving rule interaction to implementation accident, *Pāṇini* declares a global precedence principle, supplemented by the finer devices of *nitya* (obligatory over optional), *antaraṅga* (inner before outer) and *apavāda* (specific overrides general). The last of these is exactly the default-with-exception logic that occupied non-monotonic reasoning research in the 1980s, and the tradition's centuries of commentarial dispute about how the principles interact form a case library of the difficulties.

The lesson for contemporary systems is pointed rather than nostalgic. Large statistical models are, by design, systems without declared precedence: when their training data contains conflicting norms two legal interpretations, two ritual conventions, two dialectal forms the resolution is opaque, unstated and unauditable. *Pāṇini*'s achievement was to make conflict resolution a declared feature of the system, inspectable and arguable. A culturally grounded intelligent system need not be rule-based to learn from this; it needs to make its precedence commitments explicit where its outputs carry normative weight. The practical example is instructive: a system advising on Hindu succession must handle the interaction of codified statute, regional custom and textual injunction, and a design that silently averages these produces confident nonsense, whereas one that names its precedence order produces a contestable and therefore usable answer.

1. KĀRAKA, SPHOṬA AND THE PROBLEM OF MEANING

The Indian grammatical tradition separated two things that natural language processing has repeatedly conflated: the surface morphology of a sentence and the structure of relations it expresses. *Kāraṇa* theory assigns to each participant in an action a semantic role, *kartā*, agent; *karma*, the thing most desired by the agent; *karaṇa*, instrument; *sampradāna*, recipient; *apādāna*, source; *adhikaraṇa*, locus and treats case endings as the (many-to-many, and therefore non-trivial) surface realisation of those roles. Because Sanskrit word order is free, the tradition could not take order as a guide to structure and was obliged to theorise relations directly.

This is the substance of the observation Briggs (1985) made in the AI literature, and it has proved genuinely productive: the *Pāṇinian* dependency framework of Bharati, Chaitanya and Sangal (1995) uses *kāraṇa* relations as the parsing target for Indian languages, and the Hindi and Urdu dependency treebanks built on that scheme perform demonstrably better for these languages than schemes transplanted from English phrase-structure grammar (Bhat et al. 2017). This is an instance of cultural grounding in its strictest and least romantic sense: not Indian content, but an Indian analytical category producing measurably better engineering for Indian languages.

Bharṭṛhari's *sphoṭa* doctrine mounts a deeper challenge. His *Vākyapadīya* opens by identifying language with the ground of reality itself:



Stripped of its metaphysics, *Bhartrhari's* linguistic thesis is that the meaning-bearing unit is the indivisible sentence-whole, and that word- and phoneme-level meanings are analytic conveniences abstracted after the fact. The sequence of sounds is the vehicle; the flash of understanding is unitary. Whatever one concludes about the metaphysics, the thesis constitutes a precise objection to compositional token-additive models: it denies that sentence meaning is recoverable by aggregating sub-sentential units, and it locates comprehension in a holistic event that the parts do not contain. Contemporary transformer architectures, which construct representations from sub-word tokens through contextual attention, sit awkwardly between the two positions they are formally compositional yet functionally holistic, and *Bhartrhari's* framing sharpens the question of what such a system's representation of a sentence actually *is*. The point is not that the tradition solved this; it is that the tradition possesses a well-developed argument about it, which the engineering literature does not. There is a concrete corollary. Sub-word tokenisation trained predominantly on English systematically mishandles the agglutinative morphology and *sandhi* of Indian languages, fragmenting single meaning-units into arbitrary pieces and inflating sequence length a measurable inefficiency and a source of degraded performance across Indic languages. A grammatically informed segmentation is not a cultural courtesy; it is a correction.

1. PRAMĀṆA THEORY AND THE EPISTEMOLOGY OF MACHINE TESTIMONY

The most consequential contribution this paper claims for Indic knowledge systems concerns not language but evidence. *Nyāya* enumerates the means of valid cognition:

प्रत्यक्षानुमानोपमानशब्दाः प्रमाणानि।

Perception, inference, comparison and verbal testimony are the means of valid knowledge. (*Nyāyasūtra* 1.1.3)

What makes this more than a list is that each *pramāṇa* comes with specified conditions of validity and characteristic modes of failure, and that the tradition treats them as heterogeneous and non-interchangeable. A claim's warrant depends on which instrument produced it. Modern machine learning systems, by contrast, collapse this distinction almost totally: perceptual data, inferential chains, analogical extension and reported testimony enter the same corpus, are represented identically, and emerge in outputs of uniform fluency and undifferentiated warrant. The system's confidence carries no information about its epistemic provenance. On *Nyāya's* terms this is not a minor defect but a categorical error.

The tradition's treatment of *śabda* is where the argument becomes sharpest, because testimony is what a language model chiefly produces. *Nyāya* does not hold that verbal report is authoritative as such; it holds that testimony is valid only when it issues from an *āpta* a speaker who has direct knowledge of the matter, is free from the intention to deceive, and is competent to communicate it:

आप्तोपदेशः शब्दः ॥

Verbal testimony is the instruction of a trustworthy person. (*Nyāyasūtra* 1.1.7)

This furnishes an unusually precise diagnosis of what is wrong with machine fabrication. The problem commonly called hallucination is not, on this analysis, a failure of accuracy that better training will fix; it is the structural absence of *āptatva*. A generative model has no perceptual access to the matters it reports, no accountable relation to the truth of its assertions, and no capacity to distinguish its testimony from its fluency. It is, in *Nyāya's* terms, a producer of *śabda* that cannot in principle satisfy the conditions under which *śabda* is a *pramāṇa*. This reframing has design consequences: an epistemically honest system would carry provenance with its assertions, distinguish what it has retrieved from what it has inferred and what it has generated, and mark the difference in the interface rather than smoothing it away. Retrieval-grounded architectures are a partial move in this direction; what the tradition supplies is the principled account of why such grounding is not an optional feature but a condition of the output counting as knowledge at all.

The *Mīmāṃsā* tradition adds a complementary resource. Confronted with a canon of ambiguous, elliptical and apparently contradictory injunctions, and required to extract determinate prescriptions from it, the *Mīmāṃsakas* built what is arguably the world's most elaborate hermeneutic machinery rules for resolving conflicting injunctions, for distinguishing primary



from subordinate statements, for supplying elided elements, and for determining when a passage carries prescriptive force at all. *Jaimini's* definition of the domain is itself a statement about the source of normativity:

चोदनालक्षणोऽर्थो धर्मः॥

Dharma is that good which is characterised by injunction. (*Mīmāṃsāsūtra* 1.1.2)

These *nyāyas* of interpretation were absorbed into Indian juristic practice and are cited in Indian courts to this day. For legal and administrative applications of AI a domain of rapid deployment and considerable risk a hermeneutics developed specifically for the determinate reading of conflicting authoritative text is a more appropriate resource than a statistical model of what similar documents tend to say.

1. HETVĀBHĀSA: A READY-MADE TAXONOMY OF REASONING FAILURE

Evaluation is the weakest link in contemporary AI research: benchmarks measure aggregate accuracy while offering little account of *how* a system's reasoning fails. Indian logic addressed this problem directly, and produced a classification of defective inference:

सव्यभिचार-विरुद्ध-प्रकरणसम-साध्यसम-कालातीता हेत्वाभासाः ॥

The pseudo-reasons are: the inconstant, the contradictory, the counterbalanced, the question-begging, and the mistimed.

(*Nyāyasūtra* 1.2.4)

Each category names a distinct structural defect rather than a wrong answer. *Savyabhicāra* is the reason that does not hold constantly the correlation with counterexamples, which describes with some precision the behaviour of a model that has learned a spurious statistical association. *Viruddha* is the reason that contradicts what is already established. *Prakaraṇasama* is the reason that is exactly counterbalanced by an opposing reason of equal force, the condition of genuine under determination in which a well-calibrated system should decline to conclude. *Sādhyasama* is the reason that presupposes what it sets out to prove circularity, which appears constantly in extended chain-of-thought traces where a model restates its conclusion as a premise. *Kālātīta* is the reason offered at the wrong time, whose relevance conditions have lapsed. The proposal here is concrete: this taxonomy is directly usable as an annotation scheme for the diagnostic evaluation of machine reasoning. Rather than scoring a chain of reasoning as correct or incorrect, an evaluator can classify its failure mode spurious correlation, self-contradiction, under determination misreported as certainty, circularity, stale premise and thereby produce information a developer can act on. That such a scheme was developed for human disputants in the early centuries of the common era does not diminish its utility; it recommends it, since the fallacies were catalogued from observed argumentative practice rather than derived from a theory of error.

2. ANEKĀNTAVĀDA, CATUṢKOṬI AND THE VĀDA PROTOCOLS

Machine learning systems are trained to converge. Given divergent evidence they produce a single most-probable output, and the divergence disappears without record. Cultural knowledge, however, is characteristically contested: ritual practice differs by region and community, historical events carry incompatible authoritative accounts, and religious and philosophical claims are the subject of live disagreement among traditions that have argued for two millennia without resolution. A system that resolves such contests silently is not neutral; it selects a winner and conceals the selection.

Jain *anekāntavāda* was constructed for exactly this situation. Its doctrine of many-sidedness holds that any assertion is true only relative to a standpoint, *naya*, and its *saptabhaṅgī* formalises the resulting predication in seven conditioned forms:

स्यादस्ति नास्ति च तथाऽवक्तव्य इति चोत्तरप्रबन्धो
 सप्तपदा जनितैषा सप्तभङ्गी निरूपिता सुधिया॥

In some respect it is; in some respect it is not; in some respect it both is and is not; in some respect it is inexpressible thus the seven-fold predication. (*Saptabhaṅgī*, standard formulation) The operative element is *syāt*, "in some respect": every assertion is explicitly indexed to the standpoint from which it holds. Translated into system design, this is a representational commitment that a knowledge base should store the claim *together with* its perspective, community and



evidential basis, and should be able to return several indexed answers rather than one unindexed one. The Madhyamaka *catuṣkoṭi* supplies a parallel four-cornered structure:

सर्वं तथ्यं न वा तथ्यं तथ्यं चातथ्यमेव च ।
 नैवातथ्यं नैव तथ्यमेतद् बुद्धानुशासनम् ॥ ८ ॥

All is real; or not real; or both real and unreal; or neither unreal nor real this is the teaching of the Buddha.
 (*Mūlamadhyamakakārikā* 18.8)

Whether these structures constitute paraconsistent logics in the strict modern sense is a matter of ongoing scholarly dispute, and this paper does not need to settle it (Priest 2018; Matilal 1998). The design implication holds under the weaker reading: the tradition provides a vocabulary and a set of formal habits for representing contested knowledge *as contested* something the standard architecture of a single-output model does not do and was never designed to do. The *vāda* tradition adds procedure to representation. Indian philosophy distinguished honest inquiry, *vāda*, from victory-seeking disputation, *jalpa*, and from merely destructive cavil, *vitaṇḍā*, and the *Caraka Saṃhitā* devotes an extended discussion to when a debate should be entered, before which audience, and by what rules. These are protocols for structured disagreement, and they map onto a live technical practice: multi-agent debate and adversarial critique as methods for improving machine reasoning. What the tradition offers that the current literature lacks is the normative distinction the recognition that a procedure optimised for winning produces different and worse epistemic results than one optimised for ascertainment, and that the two must be institutionally separated.

5. RASA AND THE COMPUTATION OF AESTHETIC RESPONSE

Affective computing standardly treats emotion as a classification problem over a small set of labels. *Bharata's* account is structurally different and considerably more sophisticated:

विभावानुभावव्यभिचारिसंयोगाद्रसनिष्पत्तिः ॥

Rasa arises from the conjunction of determinants, consequents and transient states. (*Nāṭyaśāstra* 6.31)

Three features distinguish this from label-based affect models. *Rasa* is *compositional* an emergent product of specified components (causal situation, expressive manifestation, fluctuating accompanying states) rather than an atomic category. It is *relational*: *rasa* is located neither in the work nor in the artist but arises in the cultivated spectator, which is to say that aesthetic effect is a property of a work–audience relation and not of a text in isolation. And it is *conventional*, dependent on a shared repertoire of *vibhāvas* that a competent audience recognises. *Abhinavagupta's* development of the theory adds the crucial condition of *sādhāraṇīkaraṇa*, generalisation the spectator's momentary release from private interest that permits the emotion to be savoured rather than suffered.

For generative systems this framework yields sharper evaluative questions than sentiment scoring can. To ask of a generated poem, image or film sequence which *rasa* it aims at, which *vibhāvas* it deploys, whether the *bhāvas* cohere, and whether a competent audience of the relevant tradition would find the effect realised, is to ask something specific and answerable. The alternative scoring outputs as positive or negative in valence cannot distinguish *karuṇa* from *bhayānaka*, or a successfully achieved *śānta* from flat affect. The tradition also names the failure mode: *rasābhāsa*, the semblance of *rasa*, an effect that displays the outward signs of aesthetic achievement without the coherence that produces it. It is difficult to think of a more exact term for the characteristic output of a competent generative model working from surface pattern.

6. DHARMA, PURUṢĀRTHA AND AN INDIC FRAMEWORK FOR AI ETHICS

Contemporary AI ethics is dominated by principlism: enumerated values fairness, accountability, transparency, beneficence asserted as universal and applied top-down. The approach has known weaknesses: principles conflict without a stated method of resolution, they are stated at a level of abstraction that permits any implementation, and their universality is asserted rather than demonstrated across cultural contexts.



The *dharma* tradition offers a structurally different ethics, and its differences are precisely where its usefulness lies. First, obligation is *situated*: *dharma* is indexed to role, stage of life, capacity and circumstance, so the question is never "what is the right action" in the abstract but "what does this agent, so placed, owe". Applied to AI governance this shifts attention from properties of models to obligations of actors what a deployer owes a user, what a data broker owes a person whose data it holds which is where accountability actually resides. Second, the tradition has an explicit category, *āpaddharma*, for the suspension of ordinary norms under emergency, together with the recognition that this is dangerous and requires justification; this is a more honest apparatus than the unstated exception-making that characterises crisis deployment of surveillance and predictive systems. Third, the *puruṣārtha* scheme relativises means to ends: *artha*, the pursuit of resources and power, is a legitimate human aim but is subordinate to *dharma*, and a technology optimised for *artha* alone is on this analysis defective rather than merely amoral. Fourth, *lokasaṅgraha*, the holding-together of the social world, supplies a criterion of collective rather than individual benefit relevant to a technology whose principal harms are aggregate, appearing as the erosion of a shared informational commons rather than as identifiable injuries to individuals.

Two qualifications are essential, and their omission is what makes much writing in this area untrustworthy. The *dharma* tradition also encodes hierarchy: *varṇāśramadharma* assigns differential obligation by birth, and any uncritical importation of *dharma*-based ethics into a system that allocates opportunity risks laundering caste hierarchy as cultural authenticity. *Ambedkar*'s critique (1936) is not an external objection to be managed but a necessary internal corrective. And the tradition's situationism, without the accountability structures it presupposed, becomes mere discretion. The proposal is therefore not to replace principlism with *dharma* but to use the *dharma* framework's strengths situated obligation, declared exception, subordination of *artha*, collective criterion under constitutional constraint.

7. APPLIED CASES: WHERE GROUNDING HAS BEEN TESTED

The argument so far is conceptual, and conceptual arguments in this field require empirical discipline. Five documented domains show what cultural grounding looks like when it is actually attempted, and each yields a lesson.

Sanskrit computational processing. The *samsādhani* tools developed at Hyderabad and Huet's Sanskrit Heritage platform perform morphological analysis, sandhi splitting and *kāraka* analysis using the grammatical tradition's own categories as the computational model rather than as decoration. The lesson is that the *śāstra* functions here as a formal specification: it tells the system what the correct analysis *is*, which no quantity of unannotated text can supply for a language whose surviving corpus is finite and whose grammatical description is more complete than its data.

Indic-language modelling at scale: AI4Bharat's *IndicNLP* resources, the *IndicBERT* and *IndicTrans* model families, and the *Bhāṣiṇī* mission's translation infrastructure address a real inequity: the majority of Indian speakers were until recently unable to interact with language technology in their own languages. The lesson is a caution. Multilingual coverage is necessary but not sufficient for cultural grounding a system that translates fluently while carrying the norms, referents and assumptions of its dominant training language performs what might be called epistemic dubbing, and users of the twenty-two scheduled languages receive a foreign worldview in a familiar voice.

Manuscript recovery: India holds an estimated thirty million manuscripts, the largest such corpus in the world, much of it uncatalogued and physically deteriorating. Machine transcription of palm-leaf and paper manuscripts across Devanagari, Grantha, Sharada, Modi and Telugu scripts is a task for which machine learning is genuinely well suited and for which no alternative exists at scale.

Here the ethical requirement is custodial rather than technical: digitisation controlled by external institutions reproduces a colonial pattern of extraction, and the communities that preserved these materials for centuries have a claim on the terms of access.



Computational musicology: The CompMusic project's work on Carnatic and Hindustani music demonstrated that Western music information retrieval assumptions equal temperament, the twelve-tone chromatic scale, harmonic verticality, the score as the work actively misrepresent traditions organised around *rāga* as melodic grammar, microtonal *gamaka* ornamentation, cyclic *tāla*, and improvisation as the locus of artistry (Serra 2011). Rebuilding the representational primitives around *rāga* and *tāla* produced systems that could actually describe the music. This is the paper's clearest empirical vindication: the failure was not of data volume but of ontology, and it was corrected by adopting the tradition's own categories.

Health informatics: Āyurvedic *prakṛti* typology classifies constitution in order to individuate treatment, an orientation that converges suggestively with precision medicine's rejection of the average patient. Work on the molecular correlates of *prakṛti* classification exists and is methodologically contested (Prasher et al. 2008). The honest statement of the case is this: the tradition supplies a valuable framework of constitutional individuation and a substantial pharmacopoeia to investigate, and any deployed system must nonetheless be evaluated against clinical outcomes. Cultural grounding cannot be permitted to become an exemption from evidentiary standards, and a system that recommends interventions on textual authority alone would fail the tradition's own test of *āptatva*.

A sixth, smaller case is worth recording for its elegance. The *kolam* drawn daily at South Indian thresholds has been analysed as a formal generative system: Siromoney and colleagues described families of *kolam* patterns with array grammars, showing that a domestic women's practice transmitted without writing encodes recursive rule-governed spatial computation. It is a useful corrective to the assumption that the tradition's formal intelligence resides only in Sanskrit texts and male scholarly lineages.

8. THREE RISKS: DIGITISATION, OVERCLAIM AND DATA INJUSTICE

The first risk is *digitisation without epistemology*. Converting *śāstra* into searchable text while discarding the interpretive apparatus that made it usable produces a corpus stripped of the very thing that constituted it as knowledge. A *sūtra* without its commentary is not a compressed truth but an under-determined string; the tradition never intended it to be read alone. A retrieval system that returns *Manusmṛti* verses as authoritative answers, unmarked by commentarial dispute, historical stratification or the tradition's own contestation, misrepresents the tradition even when it quotes it accurately. Fidelity here requires modelling the commentarial layer-*bhāṣya*, *vārttika*, *ṭīkā* as structure, not merely as more text.

The second risk is *revivalist overclaim*. The assertion that ancient India possessed aviation, genetic engineering or artificial intelligence in substance is both false and self-defeating: it invites the dismissal of the tradition's genuine achievements along with its fabricated ones (Nanda 2003). The *Aṣṭādhyāyī* requires no inflation; it is one of the most impressive intellectual artefacts humanity has produced, and it is diminished rather than honoured by the claim that it is secretly a compiler.

There is also a political dimension that scholarship cannot ignore: the framing of Indic knowledge as a civilisational asset in a contest of national power (Malhotra 2021) tends to produce exactly the essentialised, homogenised and majoritarian account of "Indic" that the tradition's own plurality contradicts.

The third risk is *data injustice*, and it is the most immediate. Sambasivan and colleagues (2021a) show that fairness apparatus built for other societies fails in India, where the salient axes of disadvantage are caste, religion, region, language and gender in interaction, and where the affected populations are frequently absent from the datasets that determine outcomes. A system trained overwhelmingly on English-language, urban, upper-caste-authored digital text will encode that standpoint as the norm and everything else as deviation while a marketing layer describes it as culturally grounded. The corrective is not aspirational but procedural: documented provenance for training corpora, in the manner of Gebru and colleagues (2021); disaggregated evaluation across the axes that matter locally; and participation by the communities whose knowledge is being modelled in decisions about how it is represented. The tradition's own *pūrvapakṣa* discipline supplies the principle



in usable form: a system that cannot state the strongest objection to its own output has not established its position.

9. CONCLUSION

This paper has argued that the relation between Indic knowledge systems and artificial intelligence should be reconstructed at the level of method rather than content. The tradition's usable inheritance is a set of formal disciplines: *Pāṇini's* architecture of ordered rules with declared conflict resolution and an explicit metalanguage; *kāraṇa* theory's separation of semantic relation from surface form, already vindicated in Indian-language parsing; *Bhartr̥hari's* argument against the compositional recovery of sentence meaning; *Nyāya's* heterogeneous *pramāṇas* and its condition of *āptatva*, which diagnoses machine fabrication as a structural rather than incidental defect; the *hetvābhāsa* taxonomy, directly usable as an annotation scheme for reasoning failure; *anekāntavāda* and *catuṣkoṭi* as frames for contested knowledge and *vāda* as protocol for its adjudication; *rasa* theory's compositional and relational account of aesthetic effect, together with its precise name for aesthetic counterfeit; and the *dharma-puruṣārtha* complex as an ethics of situated obligation, to be adopted under constitutional constraint rather than uncritically.

Culturally grounded intelligent systems, on this account, are defined by four commitments rather than by any subject matter. They declare their epistemic provenance, distinguishing what is perceived, inferred, retrieved and generated. They represent contested knowledge as contested, indexed to standpoint. They adopt the representational primitives of the domain they model rather than importing foreign ones *rāga* rather than the chromatic scale, *kāraṇa* rather than phrase structure. And they locate obligation in identifiable actors under stated conditions, including the conditions under which ordinary norms are suspended. None of these commitments is exotic; each is a live requirement of responsible system design, and each is more precisely stateable in the tradition's vocabulary than in the vocabulary currently used.

Two directions for further work follow. The first is empirical: the *hetvābhāsa* annotation scheme proposed in section 8 and the *rasa*-based evaluation of generative output proposed in section 10 are both testable, and should be tested. The second is institutional: the recovery of India's manuscript corpus and the modelling of its commentarial structure are tasks of a scale that requires sustained collaboration between *śāstric* scholars and computational researchers, of a kind that presently exists only in pockets. Neither task will be advanced by the rhetoric of civilisational anticipation, and both are urgent.

It is fitting to close with the injunction that best describes the attitude this programme requires, and which the tradition placed at its own beginning:

आ नो भद्राः क्रतवो यन्तु विश्वतः

Let noble thoughts come to us from every side. (*R̥gveda* 1.89.1)

A tradition confident enough to ask for insight from all directions has no need to claim that it already possessed everything. The strongest case for Indic knowledge systems in the design of intelligent machines is not that they anticipated the machines, but that they thought harder than most traditions about the questions the machines have forced back upon us: what it is to know, who may be believed, how to proceed when the evidence divides, and what we owe to those our judgements bind.

NOTES

1. The formal resemblance between Pāṇini's rule system and later devices of formal language theory is widely noted, and Backus's acknowledgement of Pāṇini in the naming of Panini–Backus form is a matter of record. The resemblance is of interest to the history of formal systems; it establishes nothing about intent.
2. Nyāyasūtra numbering in particular differs across the Nyāyabhāṣya, Nyāyavārttika and modern printed traditions; the *hetvābhāsa* sūtra cited here as 1.2.4 appears as 1.2.6 in some editions.
3. The *saptabhaṅgī* is given in its standard four-limb abbreviated form; the full seven include the two conjunctions



of the inexpressible with affirmation and negation, and their joint form.

4. Estimates of India's surviving manuscript holdings vary between five and forty million depending on definition and survey method; the National Mission for Manuscripts figure of approximately thirty million is used here.
5. No claim of clinical validation is made for prakṛti-based classification in section 12; the literature cited is presented as an open research question.

References:

1. Ambedkar, B.R. (1936). *Annihilation of Caste*. Reprinted (2014), Navayana, New Delhi.
2. Bender, E.M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots.
3. Proceedings of FAccT '21, 610–623.
4. Bhagavad Gītā. Śrīmadbhagavadgītā, with Śaṅkarabhāṣya. Gita Press, Gorakhpur.
5. Bharati, A., Chaitanya, V. & Sangal, R. (1995). *Natural Language Processing: A Pāṇinian Perspective*.
6. Prentice Hall of India.
7. Bhartṛhari. *Vākyapadīya*. Ed. K.A. Subramania Iyer (1966), Deccan College, Poona.
8. Bhat, R.A., Bhatt, R., Farudi, A. et al. (2017). The Hindi/Urdu Treebank Project. In *Handbook of Linguistic Annotation*. Springer.
9. Briggs, R. (1985). Knowledge Representation in Sanskrit and Artificial Intelligence. *AI Magazine*, 6(1), 32–39.
10. Caraka. *Caraka Saṃhitā*, Vimānasthāna 8. Chaukhambha, Varanasi.
11. Ganeri, J. (2001). *Philosophy in Classical India: The Proper Work of Reason*. Routledge. Ganeri, J. (2011). *Artha: Meaning*. Oxford University Press, New Delhi.
12. Gebru, T., Morgenstern, J., Vecchione, B. et al. (2021). Datasheets for Datasets. *Communications of the ACM*, 64(12), 86–92.
13. Gotama. *Nyāyasūtra*, with Vātsyāyana-bhāṣya. Trans. G. Jha (1939), Motilal Banarsidass.
14. Huet, G., Kulkarni, A. & Scharf, P., editors. (2009). *Sanskrit Computational Linguistics*. Springer LNCS 5402.
15. Ingalls, D.H.H. (1951). *Materials for the Study of Navya-Nyāya Logic*. Harvard University Press. Jaimini. *Mīmāṃsāsūtra*, with Śābarabhāṣya. Trans. G. Jha (1933), Oriental Institute, Baroda.
16. Kakwani, D., Kunchukuttan, A., Golla, S. et al. (2020). IndicNLP Suite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Models for Indian Languages. Findings of EMNLP 2020.
17. Kaṇāda. *Vaiśeṣikasūtra*. Trans. N. Sinha (1911), Pāṇini Office, Allahabad. Kapoor, K. (2005). *Indian Knowledge Systems*. D.K. Printworld, New Delhi. Kauṭilya. *Arthaśāstra*. Trans. R.P. Kangle (1972), University of Bombay.
18. Malhotra, R. (2021). *Artificial Intelligence and the Future of Power*. Rupa Publications.
19. Matilal, B.K. (1986). *Perception: An Essay on Classical Indian Theories of Knowledge*. Clarendon Press. Matilal, B.K. (1998). *The Character of Logic in India*. State University of New York Press.
20. Mohanty, J.N. (1992). *Reason and Tradition in Indian Thought*. Clarendon Press. Nāgārjuna. *Mūlamadhyamakakārikā*. Trans. J. Garfield (1995), Oxford University Press. Nanda, M. (2003). *Prophets Facing Backward*. Rutgers University Press.
21. Nāṭyaśāstra of Bharata. Trans. M. Ghosh (1951), Asiatic Society, Calcutta. Pāṇini. *Aṣṭādhyāyī*. Ed. and trans. S.M. Katre (1987), University of Texas Press.
22. Prasher, B., Negi, S., Aggarwal, S. et al. (2008). Whole Genome Expression and Biochemical Correlates of Extreme Constitutional Types Defined in Āyurveda. *Journal of Translational Medicine*, 6, 48.
23. Priest, G. (2018). *The Fifth Corner of Four*. Oxford University Press.
24. Ramesh, G., Doddapaneni, S., Bheemaraj, A. et al. (2022). Samanantar and IndicTrans. *Transactions of the ACL*, 10, 145–162.
25. Sambasivan, N., Arnesen, E., Hutchinson, B. et al. (2021a). Re-imagining Algorithmic Fairness in India and Beyond. *Proceedings of FAccT '21*, 315–328.



Cover Page



-
26. Sambasivan, N., Kapania, S., Highfill, H. et al. (2021b). "Everyone Wants to Do the Model Work, Not the Data Work". Proceedings of CHI '21.
 27. Serra, X. (2011). A Multicultural Approach to Music Information Research. Proceedings of ISMIR 2011, 151–156.
 28. Siromoney, G., Siromoney, R. & Krithivasan, K. (1974). Array Grammars and Kolam. Computer Graphics and Image Processing, 3(1), 63–82