

Posthumanist Epistemology and Algorithmic Decolonization: Reconstructing Svayam Prakāśa Consciousness in Artificial Intelligence (AI) Ethics

Hendri Syafriadi¹, Siti Fatimah²

^{1,2}Universitas Negeri Padang, Indonesia

Email: hendrisyafriadi22@gmail.com

Abstract

The development of artificial intelligence (AI) raises fundamental philosophical questions concerning the relationship between algorithms, knowledge, consciousness, and human moral responsibility. This study aims to reconstruct the epistemological foundations of AI ethics through a dialogue among the philosophy of technology, decolonial AI, the philosophy of consciousness, and the concept of svayam prakāśa in Advaita Vedānta. The study employs a philosophical inquiry approach through critical conceptual analysis, philosophical hermeneutics, and theoretical reconstruction. The findings indicate that the computational paradigm can explain the functional aspects of intelligence but remains insufficient to account for the subjective experience of consciousness. Furthermore, algorithms are not neutral, as they shape epistemic structures embedded with particular assumptions about values and power. This study proposes a Self-Luminous AI Ethics Framework that places reflective awareness, epistemic humility, epistemic pluralism, ethical human–technology relations, and respect for human dignity at the foundation of responsible AI development.

Keywords: *Artificial Intelligence (AI), AI Ethics, Epistemology, Svayam Prakāśa, Advaita Vedānta, Decolonial AI, Consciousness.*



A. INTRODUCTION

The development of artificial intelligence (AI) has transformed how information is processed, decisions are made, and knowledge is produced through algorithmic systems. Beyond computational advancement, AI raises fundamental philosophical and ethical questions concerning knowledge, responsibility, and human–technology relations, particularly how algorithms shape human judgment, embed specific values, and affect humans as epistemic and moral subjects (Coeckelbergh, 2020; Floridi & Cowls, 2019). AI algorithms are not entirely neutral or value-free. Decisions concerning data selection, categorization, system objectives, and output interpretation embed human assumptions, interests, and values. AI should therefore be understood as a sociotechnical system shaped by the social contexts in which it is designed and deployed (Selbst et al., 2019).

A major development in AI ethics is the emergence of principle-based approaches intended to guide responsible AI development and governance. Jobin, Ienca, and Vayena (2019) identify recurring principles in global AI ethics guidelines, including transparency, justice and fairness, non-maleficence, responsibility, and privacy. However, their interpretation and implementation vary across institutions

and countries, indicating that ethical AI involves not only technical performance but also the representation of human values and moral judgments in technological systems (Jobin et al., 2019). Although these principles have expanded AI discourse beyond technical concerns, their practical application remains challenging. Mittelstadt (2019) highlights the gap between abstract principles and implementation, as concepts such as fairness, transparency, and responsibility require contextual interpretation. This limitation has encouraged critical approaches that examine how algorithms are shaped by human decisions, data, social values, and specific contexts of use (Selbst et al., 2019). In the philosophy of technology, Ihde's *postphenomenology* challenges the conception of technology as a neutral instrument, emphasizing its role in mediating human experience and understanding of the world (Ihde, 1990). Verbeek (2011) further argues that technological mediation shapes human action, experience, and decision-making, situating human–technology relations within complex interactions among humans, technologies, and the world (Rosenberger, 2015).

Critical AI scholarship similarly rejects algorithmic neutrality. Crawford (2021) demonstrates that AI is shaped not only by code, data, and computational infrastructure but also by social, economic, political, and material conditions. This perspective is extended by *decolonial AI*, which examines how power relations, historical coloniality, and epistemological choices influence AI design and implementation. Mohamed, Png, and Isaac (2020) argue that AI may reproduce or intensify injustice when particular knowledge systems dominate technological development. Thus, AI ethics must address not only technical performance but also power, epistemology, and the diversity of knowledge perspectives (Mohamed et al., 2020). Floridi's (2011) *philosophy of information* and concept of the *infosphere* further situate human life within an information environment connecting people, data, and digital technologies. AI consequently influences knowledge formation, decision-making, and social practices rather than functioning merely as a computational tool (Coeckelbergh, 2020). Despite established principles of transparency, fairness, responsibility, and privacy (Jobin et al., 2019), the role of human consciousness as an ontological foundation for moral values and ethical responsibility remains insufficiently examined.

This issue is particularly relevant to the relationship between AI, cognition, and consciousness. *Computationalism* conceptualizes mental processes as information processing explainable through computational mechanisms, a view reflected in classical symbolic AI, where symbol manipulation underpins intelligent problem-solving (Newell & Simon, 1976; Pylyshyn, 1980). However, functional competence does not necessarily imply genuine understanding or conscious experience. Searle's (1980) *Chinese Room* argument demonstrates that symbol manipulation does not automatically produce understanding, while Chalmers (1995) identifies subjective experience as a fundamental problem that remains insufficiently explained by physicalist and computational approaches (Chalmers, 1995; Searle, 1980).

Despite contributions from AI ethics, the philosophy of technology, and *decolonial AI*, the philosophical foundations of humans as conscious subjects remain

underexplored. Existing AI ethics research primarily emphasizes normative principles such as transparency, fairness, accountability, responsibility, and human rights protection (Floridi & Cowls, 2019; Jobin et al., 2019). Although essential for AI governance, these principles leave deeper questions concerning human nature, dignity, and the ontological grounds for recognizing humans as more than data sources or objects of algorithmic processing. Hanna and Kazim (2021) accordingly emphasize the need for stronger philosophical foundations to clarify human dignity, moral agency, and the human position in relation to digital technologies (Hanna & Kazim, 2021).

Decolonial AI exposes coloniality in knowledge production, data practices, and algorithmic decision-making (Mohamed et al., 2020), its scholarship has largely focused on epistemic injustice and power relations. Less attention has been devoted to connecting these critiques with non-Western philosophical traditions that explicitly address the nature of consciousness. This gap becomes increasingly significant as AI systems simulate more aspects of human cognition. One relevant tradition is *Advaita Vedānta*, particularly its concept of *svayam prakāśa* (self-luminous consciousness). According to Fasching (2021), consciousness is not an observable object but the fundamental condition that makes experience possible. Rather than reducing consciousness to computational processes or information representation, this perspective treats it as the basis through which humans experience and understand the world (Fasching, 2021). Fasching (2021) does not directly address AI, *svayam prakāśa* offers a philosophical basis for reconsidering the relationships among AI, consciousness, subjectivity, and human dignity. To address this gap, the present study develops the *Self-Luminous AI Ethics Framework*, extending AI ethics beyond algorithmic regulation toward an approach that recognizes humans as conscious subjects with intrinsic value.

Accordingly, this study aims to: (1) critically analyze how the computational paradigm may reduce the understanding of human consciousness; (2) reconstruct *svayam prakāśa* as an alternative epistemology of consciousness in response to algorithmic dominance; and (3) formulate *Self-Luminous AI Ethics* as a normative framework for AI development and governance that respects humans as conscious subjects. Using a qualitative philosophical approach combining critical conceptual analysis, philosophical hermeneutics, and reconstructive synthesis, the study seeks to contribute to the philosophy of science and technology by offering an alternative perspective on the relationship among consciousness, algorithms, and AI ethics.

B. METHOD

This study adopts a *philosophical inquiry* approach combining *critical conceptual analysis*, *philosophical hermeneutics*, and *philosophical reconstruction*. Rather than testing empirical relationships, it examines and critiques the philosophical assumptions underlying artificial intelligence (AI). Philosophical inquiry involves conceptual clarification, argumentative evaluation, interpretation, and theoretical reconstruction to develop a deeper and more coherent understanding (Daly, 2015; Rescher, 2001).

Accordingly, this study investigates how *computationalism* shapes conceptions of intelligence, how AI influences understandings of knowledge and consciousness, and how *svayam prakāśa* in *Advaita Vedānta* offers an alternative perspective on human consciousness.

The first stage employs *critical conceptual analysis* to examine AI, *computationalism*, algorithms, consciousness, and technology ethics. Beyond defining concepts, this method investigates their underlying assumptions, logical relationships, and philosophical implications (Beaney & Raysmith, 2024). It critically assesses whether understanding intelligence as computation can adequately explain human conscious experience, particularly its subjective dimension.

The second stage uses philosophical hermeneutics to interpret *svayam prakāśa* within *Advaita Vedānta* and engage it with contemporary debates on consciousness and AI. Gadamer (2004) emphasizes that philosophical understanding emerges through the interaction of texts, historical contexts, and the interpreter's horizon. Accordingly, *svayam prakāśa* is reconsidered not merely as a classical metaphysical concept but as a philosophical perspective relevant to contemporary questions of consciousness and technology (Gadamer, 2004). Understood as *self-luminous consciousness*, it refers to consciousness's capacity to reveal or illuminate itself (Ram-Prasad, 2007), providing a basis for questioning reductionist accounts that explain consciousness solely through algorithmic processes.

The final stage reconstructs the *Self-Luminous AI Ethics Framework* by integrating theoretical perspectives to develop a more comprehensive understanding of humans, consciousness, and technology (Daly, 2015; Rescher, 2001). Four perspectives inform this synthesis: *computationalism*, which conceptualizes intelligence as information processing; *postphenomenology*, which explains how technology mediates human experience and understanding; *decolonial AI*, which critiques dominant epistemological assumptions in algorithmic development; and *svayam prakāśa*, which offers an alternative understanding of consciousness that cannot be fully reduced to computational mechanisms. This synthesis argues that AI ethics must extend beyond technical principles such as accuracy, transparency, and efficiency to recognize humans as conscious subjects who experience and interpret the world.

The study draws on primary and secondary academic literature concerning AI philosophy, philosophy of mind, philosophy of technology, AI ethics, and *Advaita Vedānta*. Sources are selected based on their theoretical relevance and contribution to the study's argument. Analysis proceeds through four stages: philosophical interpretation, argumentative analysis, conceptual comparison, and reconstructive synthesis. The rigor of philosophical research is assessed not through empirical testing but through argumentative strength, logical consistency, and conceptual coherence (Daly, 2015; Rescher, 2001).

C. RESULTS AND DISCUSSION

1. The Limitations of the Computational Paradigm in Explaining Consciousness

This study finds that the computational paradigm remains a central foundation of modern AI, assuming that mental processes and intelligence can be explained through information processing, representation, and computational operations. This perspective is rooted in functionalism, which defines mental states through their functional roles within cognitive systems (Putnam, 1967), representational theories that explain thinking through formally analyzable mental structures (Fodor, 1987), and the *physical symbol system* hypothesis, which proposes that manipulating symbols appropriately can produce intelligent behavior (Newell & Simon, 1976). Although this paradigm has enabled AI systems to perform complex cognitive tasks, including pattern recognition, classification, prediction, learning, and decision making, these achievements primarily demonstrate functional capabilities rather than establish the existence of consciousness. Nagel (1974) argues that consciousness involves a subjective, first person perspective, expressed through the question of “what it is like to be” something. Consequently, explaining how a system processes information does not necessarily explain the subjective experience associated with those processes. This limitation reflects Chalmers’s (1995) *hard problem of consciousness*, which distinguishes mechanistically explainable cognitive functions from the fundamental question of why or how such processes generate conscious experience.

Searle’s (1980) *Chinese Room* argument further challenges computationalism by distinguishing syntactic symbol manipulation from semantic understanding. Although a computer may generate appropriate outputs by following formal rules, this does not necessarily mean that it understands their meaning. Similarly, Block (1995) distinguishes *access consciousness*, which makes information available for reasoning, thinking, and action control, from *phenomenal consciousness*, which refers to subjective experience. These distinctions indicate that AI’s capacity to process information and produce apparently intelligent behavior does not necessarily establish genuine understanding or conscious experience. Nevertheless, the relationship between computation and consciousness remains contested. Some functionalist perspectives maintain that sufficiently developed functional structures and organization could, in principle, support mental states (Chalmers, 1995; Putnam, 1967). In contrast, critics of computationalism argue that consciousness cannot be fully reduced to computational functions, particularly because subjective experience remains difficult to explain through functional mechanisms alone (Nagel, 1974; Chalmers, 1995). The distinction between access and phenomenal consciousness further highlights why successful information processing cannot, by itself, demonstrate that AI possesses subjective experience (Block, 1995).

AI can simulate cognitive functions but does not yet provide an adequate ontological basis for explaining consciousness as subjective experience. Overall, the computational paradigm effectively explains the mechanisms and functions of intelligence but remains limited in explaining the nature of human consciousness.

This limitation suggests that AI development should not focus exclusively on computational capabilities and system performance. It must also address philosophical questions concerning understanding, subjective experience, and the relationship between intelligence and consciousness.

2. The Paradigm Shift in AI Ethics from Technical Concerns to Epistemological Questions

The literature reveals a shift in AI ethics from technical principles toward deeper questions about technology, human beings, and values. Jobin et al. (2019) identify transparency, fairness, responsibility, privacy, and non-maleficence as prominent principles in global AI ethics frameworks, reflecting recognition of AI as a technology embedded in social and normative contexts. However, Mittelstadt (2019) argues that implementing these principles remains challenging because of the gap between ethical concepts and algorithmic practice. AI ethics must therefore examine not only algorithmic safety and transparency but also how responsibility, fairness, agency, and moral value are understood in relation to intelligent technologies.

The central crisis in AI ethics concerns not only the limitations of algorithmic regulation but also the epistemological limitations of understanding humans as moral subjects who establish values, responsibilities, and the purposes of AI.

3. Algorithms as Non-Neutral Epistemic Structures

Postphenomenology challenges the view that technology is a neutral instrument, arguing that it mediates how humans experience and interpret reality (Ihde, 1990). Similarly, algorithms do more than process data: they select, rank, and classify information, shaping what becomes visible, relevant, and accessible to users (Gillespie, 2014). Crawford (2021) further demonstrates that AI systems are embedded in social, economic, and political contexts, with data collection, categorization, and system objectives reflecting human decisions about what matters and what counts as measurable knowledge.

Synthesizing postphenomenology, algorithm studies, and critical AI studies, this research conceptualizes algorithms as epistemic structures that shape how reality is perceived and understood. Through data selection, categorization, prediction, and recommendation, algorithms determine which information receives attention and which is marginalized. Algorithmic neutrality is therefore problematic because every AI system embodies assumptions about data, classification, and relevant knowledge. AI design is consequently not merely an engineering activity but also a philosophical undertaking with epistemological consequences.

Algorithms are epistemic mediators that shape how humans experience, organize, and understand reality, embedding assumptions about data, classification, and relevant knowledge.

4. Decolonizing AI Requires an Expanded Epistemology of Consciousness

Decolonial AI scholarship challenges technological neutrality by demonstrating that AI reflects historical conditions, power relations, and particular epistemological assumptions (Mohamed et al., 2020). Data-driven epistemologies may reproduce the coloniality of knowledge when specific understandings of reality are treated as universal (Ricaurte, 2019). AI decolonization therefore requires examining not only social representation but also how AI defines humanity and establishes standards of valid knowledge (Adams, 2021). Although existing approaches address representation, algorithmic injustice, data coloniality, and technological power (Mohamed et al., 2020; Ricaurte, 2019, 2022), the ontology of humans as conscious subjects and sources of moral value remains relatively underdeveloped. This study extends AI decolonization toward the decolonization of the epistemology of consciousness, promoting pluralistic understandings of humanity, subjectivity, and consciousness beyond computational reductionism. This position aligns with Coeckelbergh's (2021, 2023) emphasis on human meaning, experience, and moral responsibility in understanding AI.

AI decolonization requires epistemological pluralism concerning humanity and consciousness, not merely data diversification. AI ethics must examine how humans are understood as conscious subjects and sources of moral value, rather than focusing solely on their algorithmic representation.

5. Reconstructing Svayam Prakāśa as an Alternative Ontology of Consciousness

A hermeneutical analysis of Advaita Vedānta suggests that *svayam prakāśa* offers an alternative to reductionist explanations of consciousness as a product of material or computational processes. In this tradition, consciousness is the fundamental condition that makes experience possible, rather than an object arising within experience. Fasching (2021) describes *prakāśa* as a self-manifesting presence that does not require an external source to reveal its existence.

This reconstruction provides an ontological distinction between computational capabilities and conscious experience. Greater information-processing capacity, linguistic fluency, or algorithmic complexity does not automatically establish *phenomenal consciousness*. This issue reflects Chalmers's (1995) *hard problem of consciousness*, which distinguishes explanations of cognitive mechanisms from explaining how and why subjective experience arises. Similarly, Searle's (1980) *Chinese Room* argument distinguishes formal symbol manipulation from semantic understanding, demonstrating why apparently intelligent AI outputs cannot automatically be equated with conscious understanding.

Whether AI can possess consciousness remains an open question. However, Butlin et al. (2023) argue that assessments should rely on theoretically grounded indicators from consciousness science and conclude that current AI systems provide insufficient evidence of consciousness. Accordingly, *svayam prakāśa* contributes to AI

ethics by emphasizing interiority, subjectivity, and experience as dimensions that cannot be reduced solely to data, symbolic representations, or algorithmic processes.

Svayam prakāśa offers an alternative ontological foundation for understanding consciousness as a fundamental dimension of subjective experience distinct from mental objects, data, and algorithms. It emphasizes that functional intelligence is not equivalent to phenomenal consciousness and that AI evaluation must distinguish computational capabilities from subjective experience.

6. Synthesis of Findings and the Self-Luminous AI Ethics Framework

The findings indicate that AI ethics must extend beyond technical performance, safety, efficiency, and transparency to examine human values, technological limitations, and social consequences. Existing frameworks emphasize transparency, accountability, fairness, privacy, and human autonomy (Jobin et al., 2019), but external regulation alone cannot resolve deeper questions about the relationship between intelligence, consciousness, and technology. This study therefore proposes the *Self-Luminous AI Ethics Framework*, inspired by *svayam prakāśa* in Advaita Vedānta. The term *self-luminous* does not imply that AI possesses human-like consciousness. Rather, it serves as a reflective principle grounded in the idea that consciousness enables experience and knowledge instead of being merely an external object of observation (Fasching, 2021). Applied to AI ethics, it calls for critical reflection on the sources of knowledge, values embedded in algorithmic design, and social consequences of AI-generated decisions.

The framework comprises four dimensions:

- a. **Epistemic Awareness: Awareness of Knowledge Formation**
AI should not be treated as a neutral producer of knowledge because its data, categories, optimization objectives, and performance criteria reflect human decisions. Ethical evaluation must therefore examine how algorithmic knowledge is produced and how epistemic biases emerge. This extends AI ethics beyond system outputs to the processes through which knowledge is constructed. *Value Sensitive Design* similarly advocates integrating human values from the earliest stages of technological design (Umbrello & van de Poel, 2021).
- b. **Reflective Transparency: Transparency Grounded in Understanding**
Transparency must go beyond technical disclosure to explain the reasons, contexts, values, and consequences underlying AI decisions. It should clarify not only how an algorithm operates but also why a decision is made and who is affected. This approach reinforces the importance of explicability and accountability in trustworthy AI (Jobin et al., 2019).
- c. **Human-Centered Ethical Relations: Ethical Relations Grounded in Human Dignity**
AI must remain situated within human social relations because systems are developed, deployed, and directed by people. Moral responsibility cannot be transferred entirely to machines. This dimension aligns with human-centered

AI, which prioritizes human needs, social values, and well-being, and emphasizes translating abstract ethical principles into practical design (Göbels et al., 2026).

- d. **Ontological Humility: Humility Regarding the Ontological Limits of AI**
AI development must recognize that information processing and simulated cognitive functions do not establish subjective experience. *Ontological humility* guards against both computational reductionism, which treats intelligence solely as computation, and premature anthropomorphism, which attributes consciousness to artificial systems without sufficient evidence.

Overall, the *Self-Luminous AI Ethics Framework* shifts AI ethics from a primary focus on risk control toward a reflective approach that examines epistemological foundations, embedded values, human dignity, and the limits of computational knowledge. Rather than replacing existing frameworks, it extends them by incorporating philosophical perspectives on consciousness, reflection, and epistemic limitations.

Table 1. The Self-Luminous AI Ethics Framework

Dimension	Literature Findings	Conceptual Reconstruction	Ethical Principle
Ontology	AI explains intelligence through computation	Consciousness has a non-reductive dimension	Primacy of Consciousness
Epistemology	Data underpin algorithmic decisions	Algorithmic knowledge has limitations	Epistemic Humility
Technology	AI mediates human experience	Technology shapes human–world relations	Relational Technology Ethics
Decolonization	AI embodies particular epistemological assumptions	Epistemological pluralism is necessary	Decolonial Epistemic Plurality
Morality	Humans are often reduced to data profiles	Humans possess conscious interiority	Human Interior Dignity

Source: Authors' analysis and conceptual reconstruction (2026).

This table synthesizes the research findings that underpin the *Self-Luminous AI Ethics Framework*. The model was developed through philosophical analysis and reconstruction of scholarship on the philosophy of mind, AI ethics, the philosophy of technology, AI decolonization, and *svayam prakāśa* in *Advaita Vedānta*. Each dimension begins with a principal finding from the literature analysis and is then reconstructed into a more comprehensive concept addressing the limitations of AI ethics approaches that have predominantly emphasized technical concerns. These reconstructed

concepts are subsequently formulated as the ethical principles that constitute the *Self-Luminous AI Ethics Framework*. The model does not claim that AI possesses consciousness comparable to human consciousness. Instead, it proposes an ethical framework that positions consciousness, epistemic humility, human–technology relations, epistemological pluralism, and respect for human dignity as foundations for responsible AI development and governance.

Thus, the model’s principal contribution is to demonstrate that ethical AI is not merely AI that functions correctly in technical terms, but AI developed through critical reflection on what is known, how knowledge is constructed, which values underpin it, and how technology relates to human life.

This study argues that contemporary artificial intelligence (AI) raises not only computational challenges but also fundamental philosophical questions about the nature of intelligence (Dreyfus, 1992; Haugeland, 1989). Computational paradigms and deep learning have significantly advanced pattern recognition, language processing, representation learning, and data-driven prediction (Goodfellow et al., 2016; LeCun et al., 2015). However, functional success does not explain how subjective experience emerges (Block, 1995; Chalmers, 1995). Chalmers (1995) identifies this issue as the *hard problem of consciousness*: explaining why and how physical processes give rise to conscious experience.

Searle’s (1980) *Chinese Room* argument further distinguishes computational performance from understanding, demonstrating that a system can manipulate symbols according to formal rules without understanding their meaning. Thus, apparently intelligent outputs do not necessarily establish understanding, consciousness, or intentionality (Searle, 1980). Although deep learning enables systems to learn hierarchical representations and perform complex tasks (LeCun et al., 2015), human intelligence remains distinct in its flexibility, capacity to learn from limited examples, causal reasoning, and ability to generalize to unfamiliar situations (Lake et al., 2017). Similarly, language models can learn statistical relationships without necessarily understanding meaning through experience, interaction, and contextual knowledge (Bender & Koller, 2020).

These findings motivate a shift from *functional intelligence* toward *ontological consciousness inquiry*, a theoretical approach that examines the relationship between intelligence, consciousness, and human experience. AI should therefore be evaluated not only by accuracy, efficiency, and predictive performance but also by its implications for human values, responsibility, and social well-being. Responsible AI requires principles of beneficence, non-maleficence, autonomy, justice, and explicability (Floridi & Cowls, 2019). Accordingly, AI must be understood in relation to humans as conscious beings with experience and moral values, rather than merely as an information-processing technology.

Principle-based approaches dominate efforts to promote responsible AI. A global review by Jobin et al. (2019) identifies transparency, justice and fairness, non-maleficence, responsibility, and privacy as recurring principles, although convergence in principles does not guarantee agreement on their interpretation or implementation.

Mittelstadt (2019) similarly argues that abstract moral principles do not automatically produce ethical AI because practical implementation involves substantial complexity. These limitations reveal that AI ethics concerns not only algorithmic performance but also the understanding of humans as moral subjects. Principles such as fairness, transparency, and accountability are essential, yet they do not fully address how technology shapes human self-understanding, knowledge, and perceptions of reality (Coeckelbergh, 2020; Mittelstadt, 2019). This study therefore proposes *Self-Luminous AI Ethics* as a philosophical extension, rather than a replacement, of existing principles. It asks how humans, as the source of ethical values, should be understood before AI systems are designed and governed.

The central concern is the reduction of humans to data, behavioral patterns, or objects of computational prediction. Because subjective experience cannot easily be reduced to information processing or computational functions, AI ethics requires a stronger anthropological and ontological foundation that recognizes humans as central moral subjects (Chalmers, 1995; Coeckelbergh, 2020).

Decolonial AI scholarship challenges the assumption that AI is objective and value-neutral. Mohamed et al. (2020) conceptualize AI as a sociotechnical system shaped by historical, cultural, and power relations, raising questions about who determines the knowledge categories, values, and assumptions embedded in technological systems. Algorithmic bias can therefore arise not only from poor data or modeling errors but also from particular worldviews encoded in classification, decision-making, and digital representation (Mohamed et al., 2020). This study extends decolonial AI debates beyond social representation, power distribution, and dominant perspectives by questioning whether humans can be fully understood through data, information, and algorithmic representations. Drawing on *svayam prakāśa*, it shifts the focus from who is represented by technology to how humanity itself is conceptualized. AI decolonization should therefore involve not only greater cultural diversity but also a critical examination of assumptions that reduce humans to objects of calculation. *Self-Luminous AI Ethics* advances this deeper form of decolonization by questioning the epistemological foundations through which human consciousness is understood.

This study's principal contribution is to reconstruct *svayam prakāśa* as a philosophical foundation for consciousness in AI ethics. Chalmers (1995) demonstrates that subjective experience cannot be adequately explained through functional or mechanistic accounts alone. In Advaita Vedānta, *svayam prakāśa*, or self-luminous consciousness, describes consciousness as the foundation that makes experience and knowledge possible, rather than merely an object known through another process (Fasching, 2021). From this perspective, machine learning, complex responses, and large-scale information processing do not automatically establish human-like consciousness. This distinction does not require treating humans and technology as entirely separate entities; rather, it recognizes that functional capabilities differ from subjective experience. Accordingly, this study positions *svayam prakāśa* as an epistemic boundary principle in AI ethics. AI may extend human

capabilities, but it should not become the sole standard for defining human nature. Through this reconstruction, *Self-Luminous AI Ethics* expands the debate from how to make machines behave ethically to more fundamental questions about consciousness, values, and humanity as the foundations of human–technology relations.

D. CONCLUSION

This study demonstrates that the central challenge of artificial intelligence ethics lies not only in developing accurate, transparent, and safe algorithms, but also in understanding humans as the source of moral values in their relationship with technology. The computational paradigm can explain cognitive functions but remains limited in explaining consciousness as subjective experience. Through the perspective of decolonial AI, this study shows that algorithms constitute non-neutral epistemic structures because they embody assumptions about knowledge, values, and reality. The reconstruction of *svayam prakāśa* offers a philosophical foundation for recognizing that human consciousness has dimensions that cannot be reduced to algorithmic processes. Based on this synthesis, the Self-Luminous AI Ethics Framework argues that responsible AI must be grounded not only in technical regulation but also in epistemological and ontological reflection on technological limitations and human dignity as a conscious subject.

REFERENCES

1. Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46(1–2), 176–197. <https://doi.org/10.1080/03080188.2020.1840225>
2. Beaney, Michael Raysmith, T. (2024). Analysis. In *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2024/entries/analysis/>
3. Bender, E. M., & Koller, A. (2020). Climbing towards {NLU}: {On} Meaning, Form, and Understanding in the Age of Data. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.463>
4. Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247. <https://doi.org/DOI: 10.1017/S0140525X00038188>
5. Butlin, P., Long, R., Elmoznino, E., Birch, J., Constant, A., Deane, G., Fleming, S., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. <https://doi.org/10.48550/arXiv.2308.08708>
6. Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219. <https://philpapers.org/rec/CHAFUT>
7. Coeckelbergh, M. (2020). *AI Ethics*. The MIT Press. <https://mitpress.mit.edu/9780262538190/ai-ethics/?utm>
8. Coeckelbergh, M. (2021). Time Machines: Artificial Intelligence, Process, and Narrative. *Philosophy & Technology*, 34(4), 1623–1638.

- <https://doi.org/10.1007/s13347-021-00479-y>
9. Coeckelbergh, M. (2023). Narrative responsibility and artificial intelligence. *AI & SOCIETY*, 38(6), 2437–2450. <https://doi.org/10.1007/s00146-021-01375-x>
 10. CRAWFORD, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
 11. Daly, C. (2015). *The Palgrave handbook of philosophical methods*. Palgrave Macmillan.
 12. Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.
 13. Fasching, W. (2021). Prakāśa. A few reflections on the Advaitic understanding of consciousness as presence and its relevance for philosophy of mind. *Phenomenology and the Cognitive Sciences*, 20(4), 679–701. <https://doi.org/10.1007/s11097-020-09690-2>
 14. Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199232383.001.0001>
 15. Floridi, L., & Cows, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1), 1–15. <https://doi.org/10.1162/99608f92.8cd550d1>
 16. Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
 17. Gadamer, H. G. (2004). *Truth and Method*. Continuum.
 18. Gillespie, T. (2014). The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press.
 19. Göbels, V. P., Fischer-Preßler, D., Guhl, J., Manca, T., Alischer, N., & Kutz, J. (2026). Aligning AI with human values: Design principles for human-centered AI. *Procedia CIRP*, 142, 486–491. <https://doi.org/10.1016/j.procir.2026.05.296>
 20. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.
 21. Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach. *AI and Ethics*, 1(4), 405–423. <https://doi.org/10.1007/s43681-021-00040-9>
 22. Haugeland, J. (1989). *Artificial Intelligence: The Very Idea*. MIT Press.
 23. Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Indiana University Press.
 24. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
 25. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, e253. <https://doi.org/10.1017/S0140525X16001837>
 26. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
 27. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
 28. Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as

- Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
29. Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450. <https://doi.org/https://doi.org/10.2307/2183914>
30. Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), 113–126. <https://doi.org/10.1145/360018.360022>
31. Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (pp. 37–48). University of Pittsburgh Press.
32. Pylyshyn, Z. W. (1980). Computation and cognition: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(1), 111–132. <https://doi.org/DOI:10.1017/S0140525X00002053>
33. Ram-Prasad, C. (2007). *Indian Philosophy and the Consequences of Knowledge Themes in Ethics, Metaphysics and Soteriology*. Routledge.
34. Rescher, N. (2001). *Philosophical Reasoning A Study in the Methodology of Philosophizing*. ohn Wiley & Sons Ltd.
35. Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, 20(4), 350–365. <https://doi.org/10.1177/1527476419831640>
36. Ricaurte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. *Media, Culture & Society*, 44(4), 726–745. <https://doi.org/10.1177/01634437221099612>
37. Robert Rosenberger, P. P. C. C. V. (Editor). (2015). *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lexington Books.
38. Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
39. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. <https://doi.org/10.1145/3287560.3287598>
40. Umbrello, S., & van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, 1(3), 283–296. <https://doi.org/10.1007/s43681-021-00038-3>
41. Verbeek, P.-P. (2011). *Moralizing Technology: understanding and designing the morality of things*. University of Chicago Press.