

The Role of Artificial Intelligence in Reshaping Human Consciousness: A Philosophical Psychology Perspective

G. R. Niromi Gunarathne

Department of Philosophy and Psychology, Faculty of Humanities and Social Sciences, University of Sri Jayewardenepura, Sri Lanka

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150700124>

Received: 05 August 2026; Accepted: 10 August 2026; Published: 20 August 2026

ABSTRACT

The very notion of consciousness has begun to transform from an abstract computation tool, Artificial Intelligence (AI) has morphed into a defining transformative agent that is remodelling the cognitive framework in which identity and decision-making occur at all human levels. Approaching both from the methodologically less selective angle of a Philosophical Psychology that attends separately to certain aspects and then weaves them into a fused together understanding, this study raises questions on how current AI-related advances influence our insights into consciousness, selfhood, agency, intentionality and cognition.

Employing a qualitative conceptual research design based on philosophical inquiry together with a systematic interdisciplinary literature search (2015-2025 within Scopus, Web of Science, PsycINFO, JSTOR, IEEE Xplore and PhilPapers), evidence from cognitive science, neuroscience and AI ethics was synthesised. The analysis revealed six significant emergent themes:

- AI as cognitive augmentation and "second brain"
- There is re-definition of consciousness and phenomenological experience
- Selfhood and Privacy: Changing identities
- Re-evaluations of human agency and moral accountability
- Psychological effects of prolonged human - machine dialogue; and
- The importance of a harmonised philosophical-psychological paradigm

The results suggest that, although modern AI does not yet possess the subjective consciousness experience of humans, it is radically altering how people think, remember, decide and construct identity. AI operates more as a cognitive partner, which leads to urgent ethical questions of autonomy, authenticity and accountability. We find that AI is no loutelligenceew technology but active overwriting of the cognitive and psychocultural context in which people encounter new data modalities through new "intelligences". Key implications include the urgent requirement for adaptive curricula, ethically sound policy-making and anticipatory psychological practices to maintain human agency and critical thinking in the longer term.

Keywords: Artificial Intelligence; Human Consciousness; Philosophical Psychology; Cognitive Psychology; Philosophy of Mind; Selfhood; Human Agency; Extended Mind Theory; Human–AI Interaction.

INTRODUCTION

Artificial Intelligence (AI) is one of these technologies that have transcended an applied computational method and has become a massive social force, augmenting and mirroring the basic features of existence, such as communication, education transparency, healthcare efficacy, creative expression etc. Recent developments in

machine learning, deep learning, and generative AI have largely increased the ability of synthetic systems to perform sophisticated cognitive tasks that once were considered uniquely human natural language processing, complex pattern identification or recognition; adaptive problem-solving; and artistic synthesis. As a result, researchers in various fields have revisited central questions about human self-awareness, agency, intentionality and personal identity.

Philosophical psychology goes deeper than just consciousness as processing raw data. It includes phenomenology, intentionality, self-awareness, and meaning attribution of lived experience. Classical Philosophy asserts that conscious awareness is essentially not computationally executable, pointing to qualia—the qualitative feel of experience (Nagel, 1974; Searle, 1980). In contrast, modern cognitive science increasingly analyses mental phenomena because of sophisticated computational and informational architectures (e.g. Chalmers, 1996; Clark, 1997). Theories friction presents one existential question: is AI just a simulation of an intelligent machine, or advanced architectures have the potential to model properties akin to conscious subjective experience?

Contrary to this, psychological research shows that in our daily interactions with digital assistants, recommender algorithms and generative models are designed to interact with us so as to continuously modify human memory, attention, and decision making. The Extended Mind Theory (Clark & Chalmers, 1998) asserts that cognitive processes go outside the (biological) brain and out into external technological systems. This strategy of offloading certain cognitive tasks changes how information is stored, retrieved, and solved - eroding the distinctions of biological cognition versus technological support.

In bringing together this cluster of concepts, this paper proposes a philosophical psychology that draws on multiple disciplines. It addresses two overarching questions:

How do intelligent technologies change traditional philosophical understandings of consciousness, agency and selfhood?

As human lives are eventually becoming more intimately intertwined with AI systems, what might the implicational bite on wide perceptions go beyond simple cognitive processes through emotional development and social behavior?

Theoretical Background and Review of Literature

The Philosophy of Mind and the "Hard Problem"

Since the history of philosophy about consciousness, attempts to categorize cognitive operations and quality of subjective experience have been dissociated. Nagel (1974) famously drew attention to the irreducibility of subjective experience, arguing that physicalist reductions should not be able to capture everything about "what it is like" to be a conscious organism. Chalmers (1996) advanced this by distinguishing the "easy problems" of cognition (e.g., discrimination, integration, attention) from the "hard problem" which asks why and how physical processing translates to phenomenal experience, or qualia.

In contrast, Dennett (1991, 2017) undermined Cartesian intuitions and argued for a functionalist "Multiple Drafts" model, which treats subjective consciousness as a constructed narrative provisionally assembled by asynchronous distributed implementations. Around the same time, Searle (1980) carried out a thought experiment in which he demonstrated that syntactic (symbol processing) operations do not equal semantic (understanding/intention) ones. Current discussions of machine consciousness are thus highly dependent on rival contemporary theories:

Global Workspace Theory (GWT): Describes consciousness as a central processor that disseminates information to specialized sub-systems (Baars, 1988). **Integrated Information Theory (IIT)** | Defines consciousness mathematically as a physical system applying the capacity of integration information () (Tononi, 2008; Tononi et al., 2016). **Predictive Processing Models:** Describe conscious experience as a continuous hierarchical minimization of prediction error (Friston, 2010).

Cognitive Offloading and the Extended Mind

In the case of cognitive psychology, the idea of a solitary brain as an isolated processing unit has been challenged by externalist paradigms. In 1998, Clark and Chalmers unveiled Extended Mind Theory to show that external artifacts like notebooks or digital algorithms can become literal components of human cognition. Now that generative AI and personal recommendation algorithms abound, cognitive offloading becomes active co-creative reasoning rather than passive retention of storage. Though that enhances problem-solving capabilities, it also creates risks surrounding cognitive habituation and lost powers of critical reflection.

Information Identity & the Person

In the digital age, personal identity is historically mediated by information infrastructures. Floridi (2014, 2019) proposes that human beings are essentially inforgs (informational organisms) living in an infosphere where the online/offline reality is constantly being eroded. In line with this idea, Metzinger (2009) postulates that the human self is not a stable object but instead a "phenomenological self-model" (PSM). This self-model is transformed by deep interaction with AI agents and algorithmic curation, altering narrative identity, self-determination, and autonomy.

Comparative Analysis of Theoretical Perspectives

Theoretical Perspective	Core Premise on Consciousness / Cognition	View on AI Systems	Implications for Human Selfhood & Agency
Biological Naturalism (Searle)	Consciousness requires specific biological/biochemical substrates.	AI merely manipulates syntax; cannot achieve genuine intentionality or subjective awareness.	Human identity and consciousness remain strictly distinct from technological tools.
Functionalism & Computationalism (Dennett, Turing)	Mental states are defined by functional organizational roles, regardless of physical material.	Sufficiently complex algorithms can realize genuine cognition and functional consciousness.	Dissolves rigid boundaries between biological and synthetic intelligence.
Extended / Embodied Cognition (Clark, Chalmers, Varela)	Cognition spans brain, body, and external environmental tools.	AI acts as a constitutive, external node in human cognitive circuits ("Second Brain").	Blurs self-boundaries; agency becomes distributed across human-technology networks.
Information Ethics & PSM (Floridi, Metzinger)	Identity is an informationally constructed, dynamic self-model.	AI continuously rewrites the informational environment (infosphere).	Alters narrative identity, making selfhood vulnerable to algorithmic manipulation.

METHODOLOGY

Design and Search Parameters

This qualitative conceptual study employed systematic interdisciplinary review and hermeneutic synthesis. A comprehensive literature search was conducted across seven major databases: Scopus, Web of Science, PsycINFO, Google Scholar, JSTOR, IEEE Xplore, and PhilPapers.

- **Search Period:** Primary window 2015 - 2025; seminal theoretical texts (e.g., Turing 1950; Nagel 1974; Searle 1980) were included for foundational grounding.
- **Search Strings:** Applied using Boolean operators
- **Inclusion Criteria:** Peer-reviewed journal articles, scholarly monographs, and proceedings from established academic conferences focused on the theoretical, psychological, or ethical intersections of AI and human cognition/consciousness; published in English.
- **Exclusion Criteria:** Non-peer-reviewed commentaries, trade publications, news media, duplicate studies, and papers exclusively technical without psychological or philosophical depth

Quality Assessment and Analysis Procedure

Selected sources underwent quality screening based on theoretical rigors, clarity of conceptual framework, and peer-review standing. Data analysis combined **Thematic Analysis** (categorizing recurring empirical/theoretical patterns) with **Philosophical Conceptual Analysis** (dissecting logical consistency, assumptions, and semantic definitions).

Conceptual Framework: The Human - AI Consciousness Matrix

To illustrate theoretical integration across emergent themes, Figure 1 presents a conceptual framework mapping the interactions between AI technological drivers, human psychological dimensions, and the overarching philosophical outcomes.

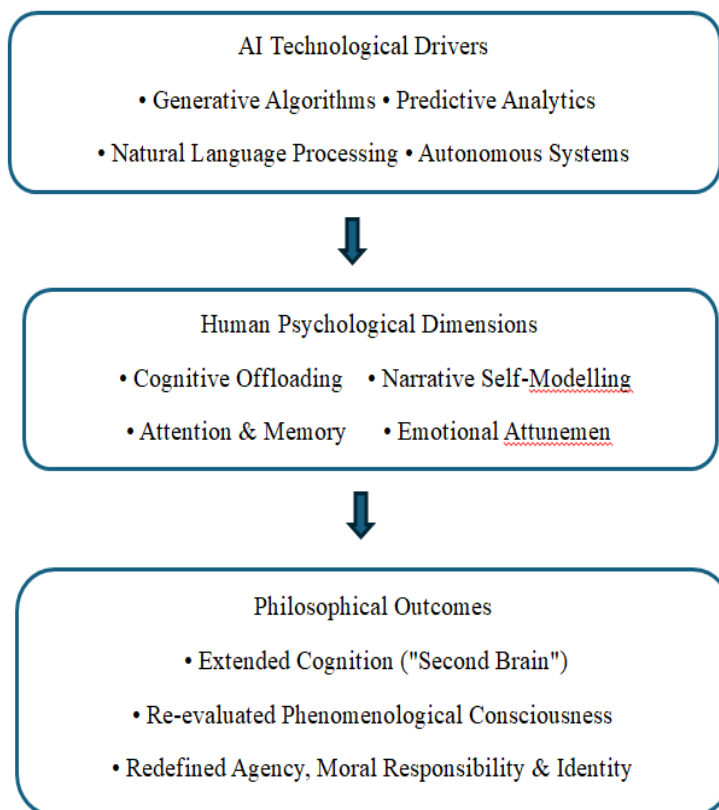


Figure 1: The Human–AI Consciousness Matrix—illustrating how algorithmic inputs re-structure core psychological mechanisms to redefine philosophical paradigms of mind.

FINDINGS AND CRITICAL DISCUSSION

The thematic synthesis yielded six core themes at the intersection of AI and human consciousness.

AI as a "Second Brain" and Extended Cognition

Modern AI tools act as functional extensions of the human nervous system. Offloading memory storage, analytical computation, and draft synthesis to generative models alters internal neural demand. While this expands immediate working memory capacities, uncritical reliance introduces risks of cognitive atrophy - where baseline capacities for independent synthesis and deep critical engagement diminish.

Re-imagining Consciousness and Subjective Experience

Despite advanced natural language generation and heuristic performance, current large language models and neural networks display no empirical or theoretical evidence of *phenomenal consciousness* or intrinsic *aboutness* (intentionality). Modern computational systems execute sophisticated syntax without semantic understanding (Searle, 1980). However, AI forces a conceptual decoupling: intelligence and high-level problem-solving no longer strictly imply subjective awareness.

AI and the Re-definition of Human Identity

Identity formation is increasingly collaborative. As algorithms curate news, suggest interpersonal responses, and generate creative works, human self-narratives become co-authored by automated systems. This ongoing algorithmic feedback loop reshapes user values, preferences, and aesthetic sensibilities, complicating traditional metrics of psychological authenticity.

Human Agency, Moral Accountability, and Autonomous Action

As automated systems assume diagnostic, financial, and legal evaluation roles, human agency faces structural diffusion. Distributed decision-making creates "responsibility gaps," wherein moral accountability becomes ambiguous. Maintaining human oversight (*human-in-the-loop*) is essential to prevent the erosion of moral autonomy and ethical decision-making.

Psychological Consequences of Human-AI Interaction

Prolonged interactions with humanlike AI interfaces introduce unique psychological dynamics. Parasocial attachment to virtual companions can reduce interpersonal tolerance for complex human relationships, while constant exposure to predictive algorithms alters attention spans and emotional regulation strategies.

Toward an Integrated Philosophical-Psychological Paradigm

Addressing these changes requires transcending disciplinary silos. Philosophical psychology offers a unifying lens that pairs philosophical inquiry with empirical cognitive research, establishing a holistic framework to evaluate how human cognitive environments evolve alongside artificial agents.

Critical Perspectives and Counter-Arguments

To ensure analytical balance, this paper engages with counter-arguments challenging prevailing assumptions regarding technology, consciousness, and identity:

1. **The Technological Determinism Fallacy:** Critics argue that framing AI as an active "rewriter" of human consciousness overemphasizes technological agency while ignoring human adaptability, societal regulation, and historic resilience to past technological disruptions (e.g., the printing press, personal computing).
2. **The Illusion of Extended Mind:** Eliminationist frameworks suggest that external tools remain mere passive instruments; calling an algorithm a "constituent of mind" conflates external informational inputs with internal cognitive processes.

3. **Biological Chauvinism vs. Functionalist Overreach:** While biological naturalists assert that machine consciousness is impossible without carbon-based biology, radical computationalists risk reducing subjective human experience to mere information processing, ignoring embodied, affective, and visceral dimensions of human life.

Cultural and Individual Moderating Factors

Psychological and philosophical effects of AI vary on a human population scale. Key moderating factors include:

Cultural differences: The fast-track view for these members of society is individualistic, such as Western societies, where AI may be seen through the lens of personal autonomy or privacy and agency. On the other hand, collectivistic cultures may see integrated AI tools as social instruments to achieve harmony or distributed responsibility.

Age and Developmental Stage: Digital natives who interact with generative AI as part of their cognitive development have higher baseline rates of cognitive offloading compared to older cohorts, which may develop alternative long-term memory structures, attention habits.

Socioeconomic Access: The unequal ability to obtain cutting-edge AI tools leads to a cognitive dividend where privileged demographics magnify creatively and analytically while marginalized groups run the risk of algorithmic surveillance or automated prejudice.

Practical Implications and Recommendations

To mitigate psychological risks and maximize cognitive enhancement, concrete interventions are recommended across three key sectors:

For Educators

- **Metacognitive Curricula:** Integrate explicitly designed coursework evaluating *when* and *how* to offload tasks to AI without compromising baseline critical thinking and problem-solving skills.
- **Assessment Redesign:** Shift from product-oriented evaluation (e.g., standard essays) to process-oriented viva voces, oral defenses, and real-time conceptual synthesis.

For Policymakers

- **Algorithmic Transparency Standards:** Mandate clear disclosure when content, decisions, or emotional interfaces are AI-generated (explicability principles).
- **Agency-Preserving Legislation:** Establish regulatory safeguards that require meaningful human oversight in high-stakes domains (healthcare, employment, judicial sentencing).

For Mental Health Practitioners

- **Clinical Diagnostic Frameworks:** Develop screening protocols for digital parasocial reliance, algorithmic identity dissociation, and technology-induced cognitive passivity.
- **Therapeutic Interventions:** Promote digital mindfulness and intentional cognitive training to restore internal reflective capacities.

Limitations and Directions for Future Research

Conceptual and Methodological Limitations

This paper is based on qualitative synthesis of existing theoretical and pre-empirical literature. As technology changes fast, empirical validations of psychological models remain in progress.

Future Research Agenda

To build on these basic insights, future research should focus on:

Longitudinal Empirical Studies: Assessing cognitive offloading over multi-year developmental cohorts and its long-term effects on neural plasticity and memory consolidation

Cross-Cultural Comparative Analyses: Empirical studies across multiple nations exploring how national culture shapes trust, parasocial attachment and moral responsibility in human–AI interactions

Neuroimaging Questions: Mapping functional neural connectivity over time and across various experiential states during sustained co-creativity with generative AI methods through fMRI/EEG.

CONCLUSION

Artificial Intelligence is reshaping the landscape of Human Life on cognitive psychological and existential level. Though neither system currently are still computational entities without phenomenal consciousness, the integration of AI into everyday life changes everything from human memory and agency to self-image and moral reasoning. An integrated philosophical-psychological framework shall then support scholars and practitioners to orient themselves in this changing landscape - safeguarding that as technology augments machine abilities human agency, rigorous reflection and fundamental selfhood do not dissolve.

REFERENCES

1. Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
2. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
3. Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
4. Chalmers, D. J. (2023). *Reality+: Virtual worlds and the problems of philosophy*. W. W. Norton.
5. Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press.
6. Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.
7. Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
8. Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Harcourt.
9. Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.
10. Dennett, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. W. W. Norton.
11. Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
12. Floridi, L. (2019). *The logic of information*. Oxford University Press.
13. Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
14. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
15. Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.
16. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
17. Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self*. Basic Books.
18. Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450.
19. Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126.
20. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
21. Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457.
22. Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.
23. Shanahan, M. (2015). *The technological singularity*. MIT Press.

24. Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin*, 215(3), 216–242.
25. Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461.
26. Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
27. Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
28. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization.
29. Yuste, R., Goering, S., Arcas, B. A. Y., Bi, G., Carmena, J. M., Carter, A., ... & Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551(7679), 159–163.
30. Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.