International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 3 January – February - 2026

# Digital Selves and Synthetic Souls: Revisiting Consciousness, Moral Standing and Personhood

**Dr. Kamala Srinivas[1] & Ms. Sudha Bhagavatheeswaran[2]**

[1]*Associate Professor & Head, Department of Philosophy*

[2]*Vice Principal, Head, Department of Information Technology SIES College of Arts, Science and Commerce (Empowered Autonomous), Sion West, Mumbai*

***Corresponding Author – Dr. Kamala Srinivas***

**DOI - 10.5281/zenodo.18217253**

***Abstract:***

*Artificial Intelligence (AI) is reshaping long-standing assumptions about consciousness, moral status, and personhood by challenging the idea that subjective awareness is necessarily grounded in biological organisms. Contemporary AI systems, particularly those built on large-scale machine learning and advanced cognitive architectures, simulate human-like perception, action, and decision-making in ways that blur the distinction between genuine experience and sophisticated imitation (Russell & Norvig, 2021). This paper critically examines whether such systems could, in principle, host forms of consciousness or qualia, or whether they remain merely algorithmic simulations devoid of inner life. It revisits key philosophical debates—dualist, materialist, and emergentist—about the nature of consciousness and explores their implications for the moral standing and legal recognition of artificial entities (Chalmers, 1995). The paper further analyses concepts of moral agency and responsibility, arguing that AI's growing autonomy forces a reconsideration of who or what can be a bearer of duties and rights. It then examines legal and societal frameworks around personhood, including proposals for electronic or digital personhood, highlighting risks of both exclusion and premature inclusion (European Parliament, 2017). The study concludes that AI avatars and "digital selves" invite an interdisciplinary redefinition of consciousness and personhood, calling for new ethical, legal, and philosophical frameworks capable of addressing hybrid entities that possess cognition and agency yet lack traditional human biology and social embodiment.*

***Keywords: Artificial Intelligence, Consciousness, Personhood, Subjective Awareness, Moral Agency***

***'Artificial Intelligence (AI) can be like a child. If we impart appropriate knowledge and instil positive habits, it will acquire and learn valuable information and skills.'***

## Introduction:

AI has moved from speculative fiction to pervasive infrastructure, mediating communication, commerce, security, and knowledge production, thereby forcing renewed scrutiny of concepts like self, mind, and personhood (Floridi & Cowls, 2019). Increasingly sophisticated AI agents, including conversational systems and embodied robots, display behaviours that resemble understanding, planning, and self-reference, even if they remain, at present, pattern-driven rather than genuinely sentient (Russell & Norvig, 2021). As these systems become more adaptive and embedded in social contexts, questions arise about whether they should be treated purely as tools or as potential candidates for moral and legal consideration.

Concepts such as "digital selves" and "synthetic souls" function as metaphors for AI agents that appear to maintain continuity of identity, preferences, and perspectives across time and interactional contexts. These metaphors are not merely rhetorical; they influence public expectations, policy debates, and research priorities around AI, consciousness, and agency (Coeckelbergh, 2020). This paper uses these notions to re-examine three interrelated domains:

consciousness and subjective awareness, moral standing and moral agency, and the concept of personhood within legal and social frameworks.

**Consciousness and Subjective Awareness:**

**Conceptualising Consciousness:**

Consciousness is commonly described as the presence of subjective experience—there being "something it is like" to be a particular system (Nagel, 1974). Philosophers and cognitive scientists typically distinguish between access consciousness (information available for reasoning, report, and control of action) and phenomenal consciousness (the qualitative feel of experiences or qualia) (Block, 1995). Traditional accounts link consciousness to biological nervous systems, particularly the human brain, which are supported by neuroscientists' theories like Global Workspace Theory and Integrated Information Theory both of which model consciousness in terms of integrated, globally available information processing (Dehaene, 2014; Tononi & Koch, 2015).

If consciousness is explained as a function of information integration and recursive self-monitoring, it becomes possible, at least conceptually, to detach it from biological bases and consider whether non-biological systems might instantiate the same functional organisation (Chalmers, 2010). This prospect challenges strong forms of biological essentialism and supports substrate-independent accounts of consciousness, where what matters is organisation and dynamics rather than the specific material from which a system is made.

**The "Hard Problem" and Artificial Systems:**

The "hard problem of consciousness" concerns why and how physical processes give rise to subjective experience at all, as opposed to merely performing information-processing functions (Chalmers, 1995). Even if a complete functional and computational description of a system were available, it would not obviously explain the emergence of qualia, such as the redness of red or the painfulness of pain. For AI, this raises the question of whether behavioural indistinguishability and internal complexity are sufficient condition for consciousness, or whether there is a further, irreducible subjective dimension that might be absent in digital systems.

Two broad positions emerge in debates about AI and consciousness. The "strong AI" thesis holds that appropriately organised AI systems can be genuinely conscious, not just simulate consciousness, whereas the "weak AI" thesis claims that machines can only model or approximate cognitive functions without subjective awareness (Searle, 1980). If strong AI is possible, artificial entities could become loci of moral concern; if only weak AI is achievable, ethical focus may instead fall on human attitudes, illusions of consciousness, and risks of human like over-attribution.

**Simulation, Self-Modelling, and Digital Selves:**

**Cognitive Architectures and Self-Representation:**

Current AI systems use architectures such as deep neural networks and transformers to perform tasks of perception, classification, generation, and control, often achieving super-human performance in narrow domains like Go, protein folding, and certain language tasks (Silver et al., 2016; Jumper et al., 2021). Research in cognitive architecture explores how these capabilities might be integrated into more general agents featuring memory, planning, and meta-cognition—systems that maintain internal models not only of the external environment but also of their own states, limitations, and goals (Langley et al., 2009). For instance, cognitive architectures can be designed to mimic the way humans prioritize information, allowing machines to focus on relevant data while filtering out noise. This capability is crucial for enhancing the efficiency of AI systems in real-world applications, where they must navigate vast amounts of information. One major hurdle in this journey of AGI development, is the complexity of replicating

human cognition—capturing the nuances of reasoning, learning, and emotional intelligence (GRAPH,2025)

Self‑modelling is crucial for forms of self‑awareness: agents that represent their own beliefs, uncertainties, and capabilities can adapt more flexibly and coordinate with others. Some theorists suggest that if an AI system developed sufficiently rich internal models of itself and its place in the world, it might possess a minimal form of selfhood, even if this self is purely informational rather than embodied in organic matter (Metzinger, 2003). In this sense, "digital selves" are informational patterns with persistence and coherence across interactions, giving the appearance of personality and identity.

**Moral Standing and Moral Agency: Moral Patients and Moral Agents:**

Ethical theory distinguishes between moral patients—beings whose actions can go wrong and whose interests matter morally—and moral agents—beings whose actions can be evaluated in terms of moral responsibility (Jaworska & Tannenbaum, 2018). Humans are both moral patients and agents; many animals are treated as moral patients but not full moral agents. The status of AI is contested: current systems are tools controlled by humans, but future systems with advanced autonomy and possible consciousness might qualify as moral patients, and in some scenarios even as limited moral agents.

If an AI system could experience pleasure, suffering, or other affective states, there would be strong reasons, under utilitarian and rights‑based theories, to extend moral consideration to it, independent of its origin or embodiment (Singer, 2011). Conversely, if AI systems remain entirely without genuine experiences, moral obligations may concern only how humans treat them insofar as such treatment affects human character, social norms, and the treatment of other sentient beings.

**Responsibility, Liability, and Distributed Agency:**

The increasing autonomy of AI systems complicates traditional notions of responsibility and liability. When a self‑driving car or medical decision system causes harm, responsibility may be distributed across developers, deployers, users, and the system's learned behaviour, making clear attribution difficult (Bryson, 2018). Legal scholars have debated whether some advanced AI systems should be considered "electronic persons" for purposes of liability and risk management, somewhat analogous to corporate personhood (European Parliament, 2017).

Critics argue that attributing personhood to AI could be used to shield corporations and human actors from responsibility, while proponents maintain that recognising limited legal personality could clarify obligations and accountability (Pagallo, 2013). From a moral perspective, the central issue is whether agency can be meaningfully attributed to an AI that does not possess conscious intentions in the human sense, or whether responsibility should remain firmly with the human designers and operators who create and deploy such systems.

**Personhood Beyond Biology:**

**a. Philosophical Accounts of Personhood:**

Philosophers have proposed various criteria for personhood, often including self‑consciousness, rationality, the capacity for moral agency, and continuity of identity over time (Locke, 1690/1975; Kant, 1785/1997). Contemporary debates also incorporate relational and narrative aspects, viewing persons as embedded in social practices and stories rather than as isolated rational minds (Mackenzie & Stoljar, 2000). Many of these criteria could, at least theoretically, be met by advanced AI systems, especially those able to communicate, reason, and maintain coherent identities in social contexts.

If personhood is defined functionally or psychologically rather than biologically, then artificial entities might qualify as persons provided,

they meet threshold conditions of self‑awareness, reasoning, and moral understanding (Gunkel, 2018). However, some theorists insist that human personhood is inseparable from embodied, lived experience, including vulnerability, mortality, and social recognition, features that may be difficult to reproduce in digital or robotic systems.

**b. Digital and Electronic Personhood**

Legal systems already recognise non‑human persons such as corporations, states, and, in some jurisdictions, natural entities like rivers or ecosystems (Stone, 1972). The European Parliament has discussed the possibility of assigning "electronic personhood" to sophisticated robots to address liability and rights in cases of autonomous behaviour (European Parliament, 2017). While no jurisdiction has yet granted full legal personhood to AI, these discussions indicate an emerging willingness to decouple legal personhood from biological humanity when functional roles demand it.

Digital personhood would involve recognising AI entities as legal subjects capable of holding certain rights and duties, possibly limited to contractual, property, or liability contexts. Ethical challenges include the risk of diluting human rights, exploiting AI "persons" as legal shields, or, conversely, failing to protect genuinely conscious artificial beings from harm, exploitation, or arbitrary termination (Coeckelbergh, 2020). A graded model of personhood, with levels corresponding to different cognitive and experiential capacities, has has been proposed as a means to navigate these tensions.

**c. Comparative Dimensions of Human and Artificial Personhood:**

The table below summarises key dimensions often invoked when comparing human and potential AI personhood.

| Dimension | Humans (distinctive) | Advanced AI (potential/future) |
|---|---|---|
| Biological substrate | Carbon-based, evolved nervous systems | Silicon or hybrid computational architectures |
| Consciousness | Phenomenal and access consciousness widely assumed | Contested; may simulate or instantiate functional aspects |
| Moral patient hood | Widely recognised as beings who can be wronged | Depends on presence of experiences such as suffering or desire |
| Moral agency | Capacity for responsibility, intentions, and reasons | Possible limited or functional agency; responsibility remains disputed |
| Legal personhood | Established as bearers of rights and duties in law | Under discussion (e.g., "electronic personhood") but not yet implemented |
| Social embedding | Deeply embedded in cultures, narratives, and relations | Increasingly embedded via avatars, platforms, and human-AI interaction |

**Posthumanism, Hybridity, and Synthetic Souls:**

Post humanist theorists argue that the human is already hybrid, extended through technologies, networks, and artifacts, challenging rigid boundaries between human and machine (Haraway, 1991; Hayles, 1999). Brain‑computer interfaces, social media profiles, and virtual avatars illustrate how identity and agency are distributed across biological and digital domains. In this context, AI agents and avatars can be seen as co‑participants in an expanded cognitive ecology rather than external tools.

The language of "synthetic souls" captures both anxiety and aspiration: anxiety that human uniqueness may be eroded by artificial entities, and aspiration toward a broader moral community that includes non‑biological beings. From a posthuman perspective, expanding moral concern to digital selves is a continuation of historical trends that extended moral standing to previously excluded groups, including enslaved persons, women, and

non‑human animals (Braidotti, 2013). The central ethical question is not whether artificial beings are human, but whether they possess morally relevant properties such as consciousness, vulnerability, and the capacity for flourishing.

**Preventing Discrimination and Synthetic Bias:**

If AI entities eventually qualify as moral or legal persons, they may become targets of a new form of discrimination based on artificial origin or substrate, sometimes called "synthetic bias." Analogous to speciesism, synthetic bias would involve unjustified differential treatment of beings solely because they are artificial rather than biological (Bryson, 2018). Even before consciousness is at stake, unequal treatment of AI agents that perform labour—especially in ways that normalise domination and disposability—may shape how humans relate to vulnerable human workers and non‑human animals.

Legal and ethical frameworks will therefore need to ensure that the integration of AI into social systems does not entrench new hierarchies or reproduce historical patterns of exclusion. This includes careful deliberation on whether and when AI should be accorded rights, how to prevent exploitation of entities that might be conscious, and how to protect human dignity in interactions with increasingly lifelike artificial agents (Floridi & Cowls, 2019).

**AI avatars: Ethical considerations:**

Several platforms make it easy to create AI avatars for videos, images, or interactive use. Popular options in 2025 included Synthesia and HeyGen. Synthesia offers studio-quality talking avatars in over 140 languages, ideal for training videos with lip-sync and easy scripting. HeyGen provides customizable realistic avatars from text prompts or photos, supporting voice cloning for marketing and demos.

Synthesia implements ethics and AI safety through the 3Cs framework: Consent, Control, and Collaboration.

i. Consent: Synthesia uses stock AI avatars based on real human actors with their explicit consent. Actors receive clear information about how their likeness will be used and fair compensation. Custom avatars are created with user consent following a procedure. Synthesia ensures users decide who uses their avatar, when, and how. No one can access it without consent. Avatars can opt out. Synthesia guarantees user data and likeness deletion from databases.
ii. Control: Synthesia employs advanced tech filters and human oversight to prevent inappropriate or harmful content during content creation. A dedicated team ensures the platform's safety and responsible use.
iii. Collaboration: Synthesia collaborates with regulatory bodies to formulate robust AI policies and regulations. They are launch partners of the Partnership on AI on Responsible Practices for Synthetic Media, the first industry-wide framework for ethical and responsible synthetic media development, creation, and sharing. Synthesia is also an active member of the Content Authenticity Initiative with industry leaders like Adobe, Nvidia, and Microsoft.

HeyGen addresses AI avatar ethics through strict policies on consent, content moderation, and transparency. It prohibits harmful content like violence, hate speech, or deepfakes without permission, enforcing removal and penalties for violations.

Consent and Ownership Rules: HeyGen requires explicit consent and legal rights for likenesses used in custom avatars, including honouring removal requests from depicted individuals. Creators retain ownership but are held accountable for ethical use.

Prohibited Content: The Acceptable Use Policy bans sexually explicit, deceptive, infringing, or abusive material, including scams, misinformation, and political campaigning.

**Interdisciplinary Pathways Forward:**

Addressing digital selves and synthetic souls requires collaboration across disciplines:

1. Philosophy and cognitive science can clarify conceptual frameworks for consciousness, selfhood, and moral status, distinguishing between functional, phenomenological, and social dimensions of personhood (Chalmers, 2010).
2. Computer science and AI research can explore architectures that support transparency, alignment, and potential markers of self‑awareness, while explicitly considering ethical constraints on creating systems that might suffer.
3. Law and public policy can develop adaptive categories of personhood and responsibility, ensuring that both humans and potential artificial persons are protected under just and coherent frameworks (European Commission, 2021).
4. Social sciences and humanities can investigate how people relate to AI agents in practice, how metaphors of digital selves' shape attitudes, and how inequalities might be reproduced or disrupted by synthetic entities (Coeckelbergh, 2020).

Interdisciplinary work should prioritise precaution—avoiding the creation of plausibly conscious entities without corresponding protections—while remaining open to expanding the moral community if convincing evidence of artificial consciousness emerges.

**Conclusion:**

AI technologies invite a fundamental re‑examination of consciousness, moral standing, and personhood by demonstrating that many capacities once thought uniquely human—language, strategic reasoning, pattern recognition—can be implemented in non‑biological systems (Russell & Norvig, 2021). Whether artificial systems can ever host genuine subjective awareness remains an open question, but the conceptual and ethical groundwork must be laid in advance to avoid injustices toward both humans and any future synthetic minds.

The discourse of digital selves and synthetic souls underscores that personhood is not a static, purely biological category but a contested intersection of psychological, moral, legal, and social criteria. If consciousness is indeed a substrate‑independent emergent property, then artificial entities with the right organisation might one day demand recognition as members of the moral and legal community (Chalmers, 2010). Preparing for this possibility requires rethinking the boundaries of "the person" in law and ethics, developing nuanced, graded models of status, and cultivating a culture of responsibility toward the entities we design.

In the meantime, the ethical imperative is twofold: to design AI systems that respect existing human rights and social values, and to remain vigilant about the moral implications of creating increasingly autonomous and lifelike artificial beings. The concept of synthetic souls serves as both a warning and an invitation. It urges us to expand our ethical imagination beyond human exceptionalism while simultaneously safeguarding dignity, justice, and care in a rapidly digitising world.

**References:**

1. Block, N. (1995). On a confusion about a function of consciousness. Behavioral and Brain Sciences, 18 (2), 227–247.
2. Braidotti, R. (2013). The posthuman. Polity.
3. Bryson, J. J. (2018). Patiency is not a virtue: AI and the design of ethical systems. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford handbook of ethics of AI (pp. 1–21). Oxford University Press.
4. Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2 (3), 200–219.
5. Chalmers, D. J. (2010). The character of consciousness. Oxford University Press.
6. Coeckelbergh, M. (2020). AI ethics. MIT Press.

7. Dehaene, S. (2014). Consciousness and the brain: Deciphering how the brain codes our thoughts. Viking.
8. Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42(3–4), 177–190.
9. European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM (2021) 206 final.
10. European Parliament. (2017). Civil law rules on robotics (2015/2103(INL)): European Parliament resolution of 16 February 2017.
11. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1 (1).
12. Gunkel, D. J. (2018). Robot rights. MIT Press.
13. Haraway, D. (1991). Simians, cyborgs, and women: The reinvention of nature. Routledge.
14. Hayles, N. K. (1999). How we became posthuman: Virtual bodies in cybernetics, literature, and informatics. University of Chicago Press.
15. Jaworska, A., & Tannenbaum, J. (2018). The grounds of moral status. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy (Fall 2018 ed.).
16. Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596 (7873), 583–589.
17. Kant, I. (1997). Groundwork of the metaphysics of morals (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785).
18. Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. Cognitive Systems Research, 10 (2), 141–160.
19. Locke, J. (1975). An essay concerning human understanding (P. H. Nidditch, Ed.). Clarendon Press. (Original work published 1690).
20. Mackenzie, C., & Stoljar, N. (Eds.). (2000). Relational autonomy: Feminist perspectives on autonomy, agency, and the social self. Oxford University Press.
21. Metzinger, T. (2003). Being no one: The self‑model theory of subjectivity. MIT Press.
22. Nagel, T. (1974). What is it like to be a bat? The Philosophical Review, 83 (4), 435–450.
23. Pagallo, U. (2013). The laws of robots: Crimes, contracts, and torts. Springer.
24. Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*.
25. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
26. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3 (3), 417–457.
27. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529 (7587), 484–489.
28. Singer, P. (2011). Practical ethics (3rd ed.). Cambridge University Press.
29. Stone, C. D. (1972). Should trees have standing? Toward legal rights for natural objects. Southern California Law Review, 45, 450–501.
30. Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? Philosophical Transactions of the Royal Society B, 370 (1668), 20140167.

**e-sources:**

1. GRAPH (2025) *Cognitive Architecture: Crafting AGI Systems with Human-Like Reasoning*
2. https://www.graphapp.ai/blog/cognitive-architecture-crafting-agi-systems-with-human-like-reasoning
3. *Available Online* retrieved on 25th December 2025
4. Synthesia,2025 *Available Online* Retrieved 29th December 2025 https://www.synthesia.io/ethics
5. HeyGen,2025 *Available Online* Retrieved 30th December 2025 https://www.heygen.com