Journal of Philosophical Investigations
Print ISSN: 2251-7960 Online ISSN: 2423-4419
Homepage: https://philosophy.tabrizu.ac.ir

Why AI Can Never Have a Soul: a philosophical inquiry into language and consciousness

Satish Kumar Tiwari[✉ 1] | **Pooja Tomar**[2]

1. Corresponding Author, Assistant Professor, A.P.B. Government Post Graduate College, Agastyamuni, Rudraprayag, Uttarakhand, India. Email: thesktiwari89@gmail.com
2. Research Scholar, School of Philosophy & Culture, SMVD University, Katra, Jammu, India. Email: 21dpc001@smvdu.ac.in

Article Info

Article type:
Research Article

Article history:
Received 21 September 2025
Received in revised form 09 October 2025
Accepted 13 October 2025
Published online 14 February 2026

Keywords:
Artificial Intelligence, Consciousness, Soul, Language, Vedānta, Phenomenology.

ABSTRACT

This paper examines the question of whether artificial intelligence (AI) could ever possess consciousness or soul. While rapid advances in machine learning, neural networks, and large language models have generated speculation about machine sentience, we argue that AI remains confined to the limits of language and computation, incapable of attaining the lived, experiential dimension that defines human subjectivity. Drawing upon both Western and Indian philosophical traditions, the paper explores the ontological and phenomenological nature of consciousness, the relation between language and experience, and the metaphysical concept of soul. Through engagement with Wittgenstein, Bhartṛhari, Descartes, Husserl, Nagel, Chalmers, and classical Indian systems such as Vedānta and Sāṅkhya, we demonstrate why AI cannot transcend simulation into embodiment. Critical responses to strong AI, functionalism, transhumanism, and emergentist theories are considered, yet all fail to address the irreducibility of subjectivity and the ontological distinctiveness of soul. The conclusion reaffirms that AI's boundaries are linguistic and computational, whereas human consciousness transcends language through lived experience, making the notion of a soulful AI both metaphysically impossible and philosophically incoherent.

Cite this article: Kumar Tiwari, S. & Tomar, P. (2026). Why AI Can Never Have a Soul: A Philosophical Inquiry into Language and Consciousness. *Journal of Philosophical Investigations*, 20(54), 233-252. https://doi.org/10.22034/jpiut.2025.69223.4202

Publisher: University of Tabriz.

Intruduction

The rapid progress of artificial intelligence (AI) in the twenty-first century has reignited age-old philosophical questions about the nature of mind, consciousness, and the soul. Once confined to the domains of theology and speculative metaphysics, the question of whether machines could ever think, feel, or possess a soul is now posed in practical terms by the development of machine learning, neural networks, and large language models. Proponents of "strong AI" argue that machines might one day achieve consciousness, while skeptics contend that what machines display is merely a sophisticated simulation of intelligence, lacking any subjective or spiritual dimension. Against this background, the central purpose of this paper is to argue that AI cannot, even in principle, possess a soul. For clarity, the term "soul" is not employed in a narrowly religious sense, nor is it treated as a doctrinally fixed entity. Rather, it is examined as a "philosophical and metaphysical construct," drawing on Indian and Western traditions, with particular attention to its implications for consciousness and language. This clarification is necessary because discussions of the "soul" often oscillate between religious belief, metaphysical speculation, and philosophical analysis; here, the focus is on the latter two, without making any theological commitments. With this clarification in place, the impossibility is not merely a technological limitation but a philosophical one, rooted in the boundaries of language and the ontological uniqueness of human consciousness.

The problem of consciousness, often described as the "hard problem," illustrates the challenge. David Chalmers (1995) famously distinguished between the "easy problems" of consciousness explaining cognitive functions such as discrimination, integration, and behavioral control and the "hard problem," which concerns why and how physical processes are accompanied by subjective experience or qualia. While computational systems can, in theory, replicate the functions associated with cognition, they do not thereby explain or generate the presence of first-person awareness. For AI, even if it can model or simulate decision-making, the leap to lived subjectivity remains unbridgeable. This difficulty resonates with a central insight of Ludwig Wittgenstein, who in the Tractatus Logico - Philosophicus declared: "The limits of my language mean the limits of my world" (Wittgenstein 1922, 5.6). Language, for Wittgenstein, sets the boundaries of what can meaningfully be expressed and thus known. Later, in his Philosophical Investigations, he developed the concept of "language games," emphasizing that meaning arises not from abstract definitions but from use within forms of life (Wittgenstein 1953). For human beings, language is deeply embedded in lived practices; we first experience and then articulate. For AI, however, there is no prior horizon of experience. Its "world" begins and ends with linguistic or computational structures. This inversion underscores the thesis of this paper: AI cannot possess a soul, because its boundaries are the boundaries of language, while human beings transcend language through lived experience.

The claim that machines cannot cross from simulation into true consciousness is not new. John Searle's (1980) "Chinese Room" thought experiment demonstrated that syntactic

manipulation of symbols even when indistinguishable from meaningful communication does not amount to semantic understanding. A system might output the correct responses to Chinese characters without knowing what they mean. Similarly, large language models today generate remarkably fluent prose but do not understand the content. This reinforces the intuition that AI operates within the limits of language and form, lacking the subjective dimension of meaning. Philosophical traditions across cultures converge on the irreducibility of consciousness to mechanistic processes. In the Western canon, René Descartes (1996 [1641]) insisted that self-awareness 'cogito, ergo sum' is the indubitable mark of soul and cannot be reduced to bodily mechanism. Edmund Husserl emphasized intentionality, the "aboutness" of consciousness, inseparable from the lived horizon of the subject. Thomas Nagel (1974) reinforced this with his famous question, "What is it like to be a bat?", illustrating that consciousness entails an irreducible first-person perspective that cannot be captured by objective, third-person accounts.

Indian philosophy, similarly, situates consciousness as ontologically prior to matter. In Advaita Vedānta, ātman (self) is identical with brahman (absolute reality), and consciousness is not a product of material complexity but its underlying ground (Śaṅkara, Brahmasūtrabhāṣya). Sāṅkhya distinguishes puruṣa (pure consciousness) from prakṛti (material nature), asserting that no material combination, whether biological or artificial, can generate awareness. Jainism describes jīva (soul) as inherently sentient, capable of liberation, a property no machine could ever embody. These traditions reject the notion that mechanical or linguistic complexity could yield consciousness, let alone a soul. The importance of clarifying this distinction is not merely metaphysical but also ethical. As Joanna Bryson (2018) warns, attributing personhood to AI risks creating moral confusion: Granting rights or status to artifacts while neglecting human responsibility. Anthropomorphizing machines by ascribing to them qualities of mind or soul may erode moral clarity. Recognizing AI's limits, by contrast, safeguards human dignity and clarifies its role as a tool of augmentation, not as a subject of experience.

Thus, the argument unfolds in this paper through several stages. Section 2 examines philosophical conceptions of consciousness in both Western and Indian traditions, highlighting their shared insistence on the uniqueness of lived subjectivity. Section 3 turns to the relation between language and consciousness, drawing on Wittgenstein and Bhartṛhari to show why AI's linguistic structures cannot ground experience. Section 4 analyzes AI as linguistic and computational processing, demonstrating its inability to access qualia. Section 5 addresses the metaphysical question of soul, showing why AI, as an artifact, cannot embody it. Section 6 engages critically with opposing views, including functionalism, strong AI, transhumanism, and emergentist theories, while Section 7 concludes by reaffirming the impossibility of AI possessing soul or consciousness. The central thesis, then, is clear: AI remains confined to language and computation, while human consciousness and soul are grounded in lived, ontological, and experiential realities that cannot be simulated or engineered.

1. Philosophical Conceptions of Consciousness

1-1. Western Traditions

The Western philosophical tradition has produced a long and complex discourse on the nature of consciousness, stretching from ancient Greece through the modern period and into contemporary analytic philosophy of mind. In order to evaluate whether artificial intelligence could ever possess consciousness or soul, it is necessary to trace how Western thought has defined these terms, what features are considered essential to consciousness, and why these features seem inaccessible to mechanical or computational systems. The analysis proceeds through several key figures and themes, beginning with Descartes and his dualist framework, then moving to phenomenology and intentionality in Husserl, Nagel's insistence on subjective character, Chalmers' "hard problem," Searle's critique of computation, and Dennett's functionalist optimism. Together, these strands provide a nuanced picture of why AI cannot, in principle, embody soul or genuine consciousness.

René Descartes is often considered the starting point for the modern Western conception of consciousness. In his Meditations on First Philosophy (1996 [1641]), Descartes sought an indubitable foundation for knowledge by systematically doubting everything he could. Even if all sensory experiences were deceptive, even if an evil demon manipulated his perceptions, the very act of doubting confirmed that he must exist as a thinking being. Hence his conclusion: cogito, ergo sum "I think, therefore I am." For Descartes, the capacity for thought and self-reflection was the essence of the soul. He distinguished res cogitans (thinking substance) from res extensa (extended substance). The former, associated with mind and soul, was immaterial, indivisible, and directly accessible to itself; the latter, associated with body and matter, was mechanical, spatial, and subject to deterministic laws. This dualist framework decisively shaped later debates, since if consciousness belongs to an immaterial soul, then mechanical systems no matter how complex—cannot possess it. Applied to AI, Cartesian dualism draws a sharp boundary: machines, however advanced, remain in the realm of res extensa, physical objects governed by causal laws. They may mimic thought, but they cannot cross into res cogitans. For Descartes, then, the question "Can AI have a soul?" would be answered negatively, since machines lack the immaterial essence that makes thought possible. If consciousness belongs to an immaterial soul, then machines—however complex—remain in the realm of res extensa.

While Descartes set the stage with his dualist division between mind and body, later thinkers turned attention toward a different defining feature of consciousness—its intentional structure. Edmund Husserl, in his Cartesian Meditations (1960 [1931]), emphasized that consciousness is always consciousness of something. This directedness or intentionality is not accidental but intrinsic. Consciousness is not a self-enclosed container of representations but a dynamic act oriented toward the world. Husserl's phenomenology aimed to describe consciousness as it is lived, prior to theoretical or scientific reduction. The phenomenological method brackets assumptions about the external world and instead investigates the structures of experience themselves. From this standpoint, the essential

feature of consciousness is its givenness: the way objects appear to a subject within a horizon of meaning. For AI, this notion of intentionality exposes a gap. A machine may process sensory inputs or symbols, but it does not intend in the Husserlian sense. Its outputs are not grounded in a lived horizon of meaning. Even when AI generates natural language, its "words" are not experienced as directed toward the world but as the outcome of statistical correlations. Intentionality, therefore, underscores that consciousness is not merely the manipulation of signs but the lived disclosure of being. Machines may process inputs, but they do not intend in the Husserlian sense, since their outputs are not grounded in lived horizons of meaning.

Building on this idea, Thomas Nagel sharpened the discussion by shifting the focus explicitly to subjectivity as the defining mark of consciousness. In his influential essay "What Is It Like to Be a Bat?" (1974), Nagel argued that an organism is conscious if and only if there is something it is like to be that organism. This "subjective character of experience" is irreducible to third-person descriptions. Even if science could explain all of a bat's neurophysiology and behavior, it would not capture what it feels like to perceive the world through echolocation. Nagel's challenge is devastating for AI. No matter how sophisticated its information-processing, no matter how perfectly it models human behavior, there is nothing it is like to be the AI. Its processes lack the subjective dimension that defines consciousness. Large language models, for instance, may generate text that simulates human-like fluency, but they do not experience meaning. Nagel's criterion implies that information-processing systems lack the subjective dimension: there is nothing it is like to be an AI.

This emphasis on subjectivity was taken a step further by David Chalmers, who crystallized the issue through his influential distinction between the "easy" and the "hard" problems of consciousness (1995). The easy problems concern cognitive functions such as discrimination, categorization, integration of information, reportability, and control of behavior. These are "easy" in the sense that they can, at least in principle, be explained by computational or neural mechanisms. The hard problem, however, asks why these functions are accompanied by subjective experience. Why does information processing feel like something? Why is there an inner movie of experience, rather than just external behavior? Chalmers argued that no amount of functional or physical explanation can bridge this explanatory gap. For AI, this means that even if all the easy problems were solved—even if AI perfectly replicated human cognitive functions—it would still not explain or produce consciousness. The hard problem remains untouched, and this distinction reinforces the thesis that AI cannot possess a soul, since the essence of soul is not functionality but subjectivity.

John Searle provided another influential critique that complements this line of thought by targeting the relation between syntax and semantics. His famous thought experiment, the "Chinese Room" (1980), imagines a person in a room following English instructions to manipulate Chinese symbols; to an outside observer the person appears to understand

Chinese, yet the person inside the room does not comprehend the meaning they only manipulate forms. Searle argued that computers operate on syntax without semantics: they manipulate symbols according to rules but do not understand them. Even the most advanced AI systems, then, achieve mere formal fluency rather than genuine comprehension. Meaning, on Searle's view, requires intrinsic intentionality rooted in mental states, which machines lack. The Chinese Room thus highlights why AI's engagement with language remains shallow: human beings use words grounded in experience and intentionality, whereas AI uses words grounded in statistical patterns. The absence of intrinsic understanding lends further force to the conclusion that AI cannot attain consciousness or soul.

Against these critiques, however, some philosophers have sought to defend a more optimistic perspective on machine consciousness. Daniel Dennett (1991), for example, adopts a functionalist approach, arguing that consciousness can be explained as the outcome of information processing. In Consciousness Explained, Dennett rejects the notion of qualia as mysterious inner properties. He proposes a "multiple drafts" model, where consciousness is the emergent product of competing narratives in the brain. From this standpoint, one might argue that if AI were able to replicate the functional organization of the human mind, it could also replicate consciousness. Yet Dennett's view has been widely criticized for failing to account for subjectivity. Chalmers (1995) points out that eliminating qualia does not solve the hard problem but merely evades it. Without addressing the irreducible reality of subjective experience, functionalism cannot establish that AI could possess a soul. Yet this view remains controversial, since critics point out that eliminating qualia does not solve the hard problem but sidesteps it.

Taken together, these perspectives reveal a central theme: Western philosophy consistently identifies consciousness with features—self-reflection, intentionality, subjectivity, qualia, semantic understanding—that resist reduction to computation. While they differ in emphasis, they converge in marking a boundary between simulation and genuine consciousness.

1-2. Indian Traditions

While Western philosophy has often framed consciousness in terms of epistemology, dualism, or subjectivity, Indian philosophy has approached it in ontological and metaphysical terms. Consciousness is not an emergent property of matter but the very ground of existence. This perspective offers a radically different starting point: rather than asking how consciousness arises from material processes, Indian traditions often begin by affirming consciousness as fundamental and then exploring how material and mental phenomena appear within it. To assess the possibility of a "soul" in artificial intelligence, it is crucial to examine these traditions in detail.

One of the most influential approaches comes from Advaita Vedānta, most systematically articulated by Śaṅkara in his Brahmasūtrabhāṣya (8th century CE). This school identifies the individual self (ātman) with the ultimate reality (brahman).

Consciousness (cit) is not contingent upon the brain or nervous system but is the eternal, unchanging principle that underlies all experience. The phenomenal world of multiplicity arises due to ignorance (avidyā), which veils the unity of consciousness. Liberation (mokṣa) consists in realizing that one's true nature as ātman is identical with brahman. From this standpoint, artificial intelligence cannot possess a soul, because soul is not something "produced." It is not an emergent complexity of matter but an eternal reality. A machine, being a configuration of prakṛti (matter), cannot contain ātman. Even if AI appears to "think" or "decide," these are surface processes within matter; they do not disclose the eternal self. Thus, the Vedāntic perspective rules out any possibility of soul in AI.

A complementary but distinct vision of consciousness is provided by the Sāṅkhya system, which advances a rigorous form of dualism. According to Sāṅkhya, reality consists of two independent principles: puruṣa (pure consciousness) and prakṛti (material nature). Prakṛti unfolds into the manifold of physical and mental phenomena, including intellect (buddhi), ego (ahaṃkāra), and mind (manas). Yet all these evolutes are insentient. Only when puruṣa reflects in them does experience arise (Larson and Bhattacharya 1987). Artificial intelligence, however sophisticated, belongs wholly to prakṛti. Its algorithms, sensors, and networks are configurations of matter and energy. Without puruṣa, there is no consciousness. Unlike Descartes' dualism, where mind and body are two substances, Sāṅkhya dualism insists that consciousness is entirely separate and cannot be generated by material processes. From this perspective as well, AI remains outside the sphere of conscious beings.

Building upon the dualist ontology of Sāṅkhya, the Yoga tradition of Patañjali provides an experiential discipline aimed at realizing the nature of consciousness. The Yoga Sūtras describe this as the stilling of the fluctuations of the mind (citta-vṛtti-nirodha), through which consciousness is revealed in its purity when the mind ceases its restless modifications. This approach emphasizes that consciousness is not identical to mental activity but is instead the luminous witness (draṣṭṛ) that underlies it. From the Yogic standpoint, the operations of artificial intelligence resemble restless vṛttis—endless processes of data manipulation. Yet without an underlying witness, these processes do not amount to genuine consciousness. The Yoga tradition thus reinforces the distinction between mental phenomena and consciousness proper, further underscoring the impossibility of AI possessing a soul.

Extending this trajectory of thought, Kashmir Śaivism, particularly the Pratyabhijñā school of Utpaladeva and Abhinavagupta, develops a more reflexive conception of consciousness. Here, consciousness is not merely luminous but self-revealing; it knows itself in the very act of knowing. This quality is termed vimarśa—the capacity of awareness to turn back upon itself and recognize itself as subject (Utpaladeva 1994). Machines, by contrast, may display recursive loops, but recursion is not reflexivity. A self-referential algorithm does not constitute self-recognition. Reflexivity requires awareness of awareness, not merely repetition. Thus, even if AI becomes recursive, it lacks vimarśa. Without

reflexive self-recognition, it cannot approximate the soul or consciousness as defined in Pratyabhijñā philosophy.

The inquiry takes yet another turn in Buddhist thought, which complicates the discussion by denying a permanent self (anātman). Nevertheless, consciousness remains central to Buddhist philosophy. In Yogācāra thought, it is described as a series of momentary streams, with the ālaya-vijñāna (storehouse consciousness) serving as a substratum that preserves karmic imprints. Here consciousness is dynamic, impermanent, and conditioned, yet still irreducible to physical processes. For artificial intelligence, this implies that even if there is no eternal self, there remains a profound qualitative difference between sentient continuity and mechanical repetition. A computer program may store and retrieve data, but this is not equivalent to karmic seeds influencing lived experience. AI lacks the phenomenological continuity and the lived depth that make consciousness meaningful in Buddhist philosophy, and thus remains fundamentally distinct from conscious beings.

This insistence on the irreducibility of consciousness is carried further in Jain philosophy, which explicitly distinguishes between jīva (soul) and ajīva (non-soul, matter). Jīva is characterized by sentience, perception, knowledge, and bliss; it is eternal and capable of liberation (Tatia, 1951). Ajīva includes matter, space, time, and motion—all of which are insentient. The defining property of jīva is upayoga (manifestation of consciousness). Artificial intelligence, by any measure, belongs to ajīva. Its operations may mimic cognition, but they lack sentience. Without upayoga, AI is merely matter in motion, never soul. Jainism therefore provides one of the most categorical rejections of the possibility of soul in machines.

Finally, a distinctive linguistic-philosophical account is offered by Bhartṛhari (5th century CE) in his Vākyapadīya. For him, ultimate reality is śabda-brahman—the Word Absolute. Language is not a human convention but the ontological ground of cognition and reality itself. Within cognition, the apprehension of meaning occurs not through discrete words but through sphoṭa—the holistic burst of meaning that arises in consciousness (Houben, 1995). Artificial intelligence, though entirely dependent on linguistic processing, does not participate in śabda-brahman. It manipulates fragments of language but does not experience the holistic sphoṭa. For Bhartṛhari, language without consciousness is incomplete. AI exemplifies precisely this incompleteness: words without world, syntax without experience.

Despite their doctrinal differences—Vedānta affirming an eternal self, Buddhism denying it, Jainism emphasizing jīva, and Kashmir Śaivism positing reflexivity—Indian traditions converge on a crucial point: consciousness is irreducible to matter and cannot be generated by mechanical or linguistic processes. Consciousness is either fundamental (ātman, puruṣa), reflexive (vimarśa), continuous (ālaya-vijñāna), inherently sentient (jīva), or linguistically holistic (sphoṭa). Artificial intelligence, being an artifact of matter, cannot embody any of these.

Thus, from the Indian standpoint, the question "Can AI possess a soul?" is answered with a decisive no. AI belongs to ajīva, prakṛti, or insentient constructs. Even if it simulates linguistic competence or cognitive processes, it lacks the ontological grounding of consciousness. Indian traditions therefore reinforce the conclusion that AI cannot transcend language and computation into the realm of soul.

2. Language and Consciousness

Language occupies a central place in the philosophical inquiry into consciousness. Human beings not only communicate through language but also think, conceptualize, and order reality through it. The problem of AI and soul intersects here because AI operates entirely within linguistic and computational structures, while human beings live and experience prior to and beyond language. This section explores two towering figures—Ludwig Wittgenstein in the Western tradition and Bhartṛhari in the Indian tradition—before drawing them together into a synthesis that clarifies why AI remains linguistically bound and experientially void.

2-1. Wittgenstein: The Limits of Language

Ludwig Wittgenstein, one of the most influential philosophers of the twentieth century, addressed the relationship between language, thought, and reality in two very different phases of his career, and both phases offer important insight into the limits of artificial intelligence. In his early period, encapsulated by the Tractatus Logico-Philosophicus (1922), Wittgenstein argued that language pictures reality through logical form: a proposition is meaningful only if it can represent a possible state of affairs. His famous dictum, "The limits of my language mean the limits of my world" (Wittgenstein 1922, 5.6), suggests that what cannot be expressed in language cannot be thought. For AI, this implies that its "world" is defined by the scope of its language and computation, bounded not only by linguistic form in general but also by the specific statistical and computational models through which it processes symbols.

Wittgenstein's mature work, however, redirected attention from an abstract logical form to the ordinary uses of language in human life. In the Philosophical Investigations (1953) he emphasized the notion of language games: words do not derive meaning by abstract correspondence but through their use in particular contexts. Language is interwoven with forms of life—the cultural, practical, and embodied activities that give words their force. For human beings this means that language is alive with meaning because it is embedded in experience: a word such as "pain" is learned and grasped within lived contexts of suffering, empathy, and bodily awareness. Artificial systems, by contrast, do not inhabit such forms of life. They can process the word "pain," generate fluent descriptions, or even simulate empathetic responses, but they do not undergo the embodied experiences that give the word its human meaning. Thus, while AI may participate in the surface structure of language games, it does not share in the forms of life that animate meaning.

Taken together, Wittgenstein's two phases suggest that language both marks the boundary of thought and gains meaning only within embodied, social practices. This perspective casts doubt on the idea that linguistic competence alone could generate genuine understanding.

Yet, one might press back against this strictly critical reading of Wittgenstein. If meaning is constituted through use within language games, it is conceivable that sufficiently complex AI systems, especially those embedded in human practices, could be said to participate in novel "forms of life" (Wittgenstein 1953, §§23, 241). Recent commentators argue that language games are not necessarily restricted to biological embodiment but can extend to technological or socio-technical systems insofar as they sustain patterns of interaction and use (Baker & Hacker 2009; Floridi 2014). For instance, large-scale interactions with humans through social media, collaborative work, or embodied robotics might gradually constitute a shared practice where words gain pragmatic force for AI, even if the underlying experiential substrate differs from that of humans. On such a view, Wittgenstein's emphasis on use could be reinterpreted not to exclude AI entirely, but to highlight that meaning emerges wherever stable practices and forms of life develop. Acknowledging this possibility enriches the debate: while AI may never replicate human embodiment, it may nevertheless instantiate alternative, technologically mediated forms of life in which language use acquires its own kind of coherence and significance (Coeckelbergh 2020; Harnad 2023).

2-2. Bhartṛhari: Śabda-Brahman and Sphoṭa

Parallel to Wittgenstein in the Indian tradition stands Bhartṛhari, the fifth-century philosopher of language and grammar, whose Vākyapadīya offers a profound account of the ontological role of language. Bhartṛhari posits śabda-brahman (the Word Absolute) as the ontological ground of the universe. Language is not merely a human convention or tool but the very principle through which reality manifests. The relation between word and world is thus intrinsic, not arbitrary. For Bhartṛhari, cognition itself is linguistic: thought and language are inseparable. This perspective might initially suggest that AI, as a linguistic machine, could participate in consciousness. Yet Bhartṛhari makes a critical distinction between language as ontological principle and the human apprehension of meaning. AI engages with surface fragments of language, but it does not share in the ontological depth of śabda-brahman.

Central to Bhartṛhari's philosophy is the doctrine of sphoṭa the holistic burst of meaning apprehended in consciousness. Meaning is not assembled from discrete words but grasped as a whole. For example, when we hear a sentence, we do not experience it as a sequence of isolated words but as a unified sense. This apprehension occurs in consciousness, not in mechanical processing. AI, by contrast, processes language sequentially or statistically. Large language models generate output token by token, based on probabilities, but they never experience the holistic burst of sphoṭa. The unity of meaning arises in consciousness, not in computation, and this distinction explains why AI's linguistic fluency remains simulational.

Bhartṛhari's philosophy therefore reinforces the conclusion that AI cannot embody soul. Language, for him, is inseparable from consciousness. AI manipulates language but does not dwell in consciousness. It has śabda without sphoṭa, syntax without holistic meaning, fragments without the unity of lived awareness.

2-3. Synthesis: Human Experience vs. AI Simulation

When we place Wittgenstein and Bhartṛhari side by side, a shared lesson emerges. Wittgenstein highlights that language derives meaning only within embodied forms of life, while Bhartṛhari emphasizes that linguistic expression culminates in the conscious apprehension of sphoṭa. AI, by contrast, remains confined to linguistic fragments without embodied experience or unifying awareness. The crucial difference lies in the order of operations: humans experience first and articulate afterward, whereas AI processes language without ever experiencing. This inversion underscores why AI's linguistic fluency cannot amount to consciousness or soul it manipulates signs but does not inhabit the lived world of meaning.

3. Artificial Intelligence and Linguistic Processing

Artificial intelligence is, at its core, a system of linguistic and computational processing. The claim that AI could ever develop consciousness or soul depends upon whether its computational architecture can bridge the gap between symbol manipulation and lived experience. To examine this possibility, we must trace the evolution of AI's architectures, assess philosophical critiques like Searle's Chinese Room, and evaluate embodied AI's reliance on sensory inputs. Together, these perspectives demonstrate that AI remains bound to linguistic form and statistical processing, never crossing into subjectivity.

3-1. From GOFAI to Neural Networks

The earliest conception of AI, often called "Good Old-Fashioned AI" (GOFAI), treated intelligence as formal symbol manipulation. Herbert Simon and Allen Newell, in their foundational work "Computer Science as Empirical Inquiry: Symbols and Search" (1976), described problem-solving as a matter of symbol systems. In this paradigm, cognition was equated with the manipulation of abstract symbols according to syntactic rules. The GOFAI approach mirrored the logicist tradition in philosophy, where intelligence was conceived as rule-following. However, critics pointed out its rigidity: symbolic AI struggled with ambiguity, context, and the fluidity of natural language. For example, while GOFAI could excel at structured tasks like chess, it faltered in everyday conversation, where meaning depends on context and nuance.

The next major shift came with connectionism in the 1980s. David Rumelhart and James McClelland (1986) introduced parallel distributed processing (PDP), where knowledge is not represented as discrete rules but as patterns of activation in networks of artificial neurons. Connectionism promised to model cognitive processes more flexibly, capturing learning and generalization. Unlike GOFAI, which relied on explicit representations, neural networks learned from data. Today's large language models (LLMs), such as GPT-4, build

on this connectionist lineage. They train on massive corpora of text and use statistical methods to predict tokens. Emily Bender and Alexander Koller (2020) aptly described such systems as "stochastic parrots": they generate fluent output by mimicking patterns in data but without understanding. LLMs demonstrate how far computational systems can go in simulating language, but they also illustrate their limits: fluency is not comprehension. From GOFAI to LLMs, the trajectory of AI reveals increasing sophistication in linguistic processing but no transition to consciousness. Whether through symbols, neural activations, or probabilistic models, AI remains locked in syntax without semantics.

3-2. The Chinese Room and Semantic Vacuity

John Searle's Chinese Room thought experiment (1980) remains the most famous critique of computational theories of mind. Imagine a person in a room following instructions for manipulating Chinese symbols. To an outsider, the person appears to understand Chinese, as the outputs are indistinguishable from a native speaker. Yet the person inside does not understand the symbols; they are merely following rules. Searle argued that this is exactly what computers do: manipulate syntax without semantics. The Chinese Room shows that functional equivalence does not guarantee understanding. A system can simulate comprehension without possessing it.

This critique applies with particular force to AI systems like LLMs. They can generate grammatically correct, contextually appropriate sentences about "pain," but they do not know what pain is. Defenders of strong AI sometimes reply that while the individual in the Chinese Room does not understand, the entire system does. Yet this "systems reply" does not resolve the problem. As Churchland and Churchland (1990) note, the idea is that understanding belongs to the system as a whole the man, the rulebook, and the room. But critics argue that if no part of the system understands, attributing comprehension to the whole risks re-labeling the problem rather than solving it (Harnad 1990). A second response, the "robot reply," contends that Searle's room lacks embodiment: if the system were connected to sensory organs and capable of acting in the world, it might ground genuine understanding (Harnad 1990; Brooks 1999). Yet, as the next section on sensory inputs will show, embodiment alone does not bridge the gap to lived experience or qualia. A third response, the "connectionist reply," argues that Searle's critique targets symbol-manipulating systems like GOFAI, but does not apply to distributed neural networks (Dennett 1987; Boden 1990). Neural nets, on this view, model cognition at a subsymbolic level closer to the brain. However, even in connectionist architectures, processing remains formal: statistical correlations do not generate intrinsic meaning. Moreover, the critique aligns with Wittgenstein's point: meaning is not in symbols alone but in their use within lived contexts. AI lacks such contexts. The Chinese Room highlights the semantic vacuity of AI's linguistic processing. Machines operate at the level of form, never entering the domain of lived meaning. This is why even the most fluent AI remains devoid of consciousness or soul.

3-3. Sensory Inputs and the Problem of Qualia

Some proponents of embodied or enactive AI argue that coupling algorithms with sensory inputs could bring machines closer to consciousness. Rodney Brooks (1999), for instance, emphasized embodied intelligence: robots that perceive and act in the world rather than relying solely on abstract computation. Andy Clark (1997) similarly argued that cognition is distributed across brain, body, and world. However, the addition of sensors does not solve the philosophical problem. Cameras and microphones transduce physical signals into data, but they do not see or hear in the phenomenological sense. For humans, vision is not the registration of photons but the lived experience of color, depth, and form. Hearing is not the analysis of waveforms but the experience of sound. Sensory inputs in machines remain computational; they lack the qualitative dimension of perception. This is the problem of qualia—the felt character of experience. As Nagel (1974) observed, no amount of third-person description can capture what it is like to be a conscious subject. AI, even with sensors, lacks what-it-is-likeness. Data streams do not become experience simply by increasing in complexity. Chalmers' (1995) hard problem resurfaces here. Explaining how machines process inputs is an "easy problem." Explaining why such processing would give rise to subjective awareness is the hard problem—and one for which AI provides no answer. Without qualia, there is no consciousness, and without consciousness, no soul.

Looking at this trajectory, the history of AI shows ever more powerful methods of linguistic and sensory processing. GOFAI manipulated symbols, connectionism modeled distributed patterns, and LLMs generate human-like text. Embodied AI integrates sensors. Yet in all these approaches, the fundamental limitation remains: syntax without semantics, data without qualia, form without subjectivity. Searle's Chinese Room shows that simulation is not understanding. Wittgenstein shows that meaning requires forms of life, absent in AI. Bhartṛhari shows that language requires consciousness for holistic meaning. Chalmers shows that the hard problem remains untouched. Taken together, these insights converge into a single conclusion: AI, however advanced, cannot transcend linguistic processing into consciousness or soul.

4. The Question of Soul in Artificial Intelligence

The concept of "soul" (psyche, anima, ātman, jīva) carries profound metaphysical and spiritual significance across traditions. Unlike consciousness, which contemporary philosophy often frames as subjective awareness, the soul is more encompassing: it refers to life-principle, moral agency, selfhood, and in many traditions, immortality. To ask whether AI could possess a soul is therefore not merely a scientific question but a metaphysical one. This section explores Western and Indian traditions of soul, then critically examines contemporary claims about machine personhood and transhumanism.

4-1. Western Traditions

In De Anima, Aristotle defined the soul as the "form of a natural body having life potentially" (Aristotle 1931, II.1). For him, the soul was not a separable substance but the

organizing principle that makes a living thing alive. Plants have nutritive souls, animals have sensitive souls, and humans have rational souls. Consciousness and reason belong to the human soul, but they are inseparable from embodiment. On this view, machines cannot possess souls, because they are not living organisms. They may mimic rational functions, but they lack the biological life that the soul animates. Thus, for Aristotle, to ask whether AI could have a soul would be incoherent: soul is the actuality of life, and AI is not alive.

Later Christian thinkers, however, approached the question of the soul from a different vantage point. Augustine regarded the soul as immaterial and immortal, created directly by God. It is the locus of intellect, memory, and will, distinct from the body yet profoundly shaping human existence. Aquinas, drawing on both Aristotle and Christian doctrine, developed a synthesis in which the human soul is indeed the substantial form of the body but also subsists independently, surviving death (Aquinas 1947). From this theological perspective, the impossibility of AI having a soul becomes even clearer. If the soul is a divine creation, no human artifact can generate it. Machines may simulate intellect, but they do not share in divine creation, and for Aquinas they would remain mere artifacts, not substances endowed with souls.

The modern era brought yet another shift in perspective with Descartes, who sharply distinguished between res cogitans (thinking substance) and res extensa (extended substance). For him, the soul is the locus of thought, free will, and self-awareness. Machines, in contrast, are sophisticated automata—mechanical devices that may imitate behavior but lack consciousness. His mechanistic view of animals extended to artifacts: they act as if they feel, but they do not. Within this framework, AI remains firmly on the side of res extensa without ever attaining the status of res cogitans.

Although these approaches differ Aristotle's soul as life-principle, Augustine and quinas' soul as divine and immortal, Descartes' soul as thinking substance they converge in placing soul beyond the reach of artifacts.

4-2. Indian Traditions

Indian philosophy provides even richer and more diverse conceptions of soul. While differing in doctrine, these traditions converge in excluding mechanical artifacts from the domain of soul. In Advaita Vedānta, Śaṅkara in the Brahmasūtrabhāṣya identified the individual self (ātman) with the absolute (brahman). The soul is not emergent but eternal, unchanging, and identical with ultimate reality. AI, being a material construct, cannot embody ātman. The very idea of "creating" a soul in machines contradicts Vedāntic ontology: soul is not produced but revealed through realization (Śaṅkara, 8th c.). A different framework is provided by Sāṅkhya, which conceives reality as consisting of puruṣa (consciousness) and prakṛti (matter). The soul is puruṣa—eternal, pure, and self-luminous. Matter, however subtle, is insentient. AI, as a product of prakṛti, cannot contain puruṣa. It may process information, but without puruṣa, it remains devoid of awareness (Larson and Bhattacharya 1987).

Jainism further sharpens the distinction by explicitly separating jīva (soul) from ajīva (non-soul). Jīva is sentient, eternal, and capable of liberation; ajīva includes matter, time, and space. The defining characteristic of jīva is upayoga the manifestation of consciousness (Tatia 1951). AI belongs wholly to ajīva. No matter how sophisticated, it lacks upayoga and thus cannot be soul. A still different yet complementary emphasis emerges in Kashmir Śaivism's Pratyabhijñā school, which highlights vimarśa reflexive self-awareness. Here, consciousness is not only luminous but also self-revealing. Machines, though recursive, do not exhibit reflexivity in this sense. They cannot recognize themselves as subjects. Without vimarśa, they cannot be souls (Utpaladeva 1994).

Despite doctrinal differences, these traditions converge on one point: the soul is eternal (ātman), sentient (jīva), or reflexively aware (vimarśa), and never reducible to material processes.

4-3. Simulation vs. Embodiment

Despite these philosophical traditions, proponents of "strong AI" and transhumanism claim that machines might one-day host souls. Ray Kurzweil (2005), for instance, predicts that human consciousness could be uploaded into computers, preserving personhood. This assumes that soul is reducible to information. Yet such views collapse the distinction between simulation and embodiment. A digital copy may simulate behavior but lacks continuity of subjectivity. As Searle (1980) argued, syntax is not semantics. Likewise, a simulated soul is not a soul. Nick Bostrom (2003) raises the "simulation hypothesis," suggesting that our own world may be a computer simulation. Yet even if so, the consciousness we experience is irreducible to the computation itself. Simulations require conscious interpreters; they do not generate consciousness on their own. From both Western and Indian perspectives, then, the core insight is the same: soul is life-principle, divine creation, eternal essence, or reflexive awareness—none of which can be generated by machines. The claim that AI could possess a soul confuses representation with reality, echo with voice, shadow with substance.

5. Critical Engagement with Opposing Views

The argument developed thus far emphasizes that artificial intelligence cannot possess consciousness or soul. Yet this claim cannot be sustained without addressing opposing positions that argue otherwise. Within contemporary philosophy of mind and AI research, four prominent strands challenge the impossibility thesis: (1) strong AI and functionalism, (2) transhumanism and mind uploading, (3) emergentist accounts such as Integrated Information Theory, and (4) ethical arguments for extending personhood to AI. This section critically evaluates each in turn, showing why they fail to establish that AI could possess a soul.

5-1. Strong AI and Functionalism

The distinction between "weak" and "strong" AI was first articulated by John Searle (1980). Weak AI claims that computers are powerful tools for simulating human cognition. Strong

AI, by contrast, claims that an appropriately programmed computer literally has a mind in the same sense humans do. Advocates of strong AI argue that consciousness is not tied to biological substrate but to formal organization. If the right functional structure is implemented, consciousness emerges. Hilary Putnam (1967) advanced this view in functionalism: mental states are defined by their causal roles rather than their material composition. Just as software can run on different hardware, minds can, in principle, run on non-biological substrates. Daniel Dennett (1991) similarly defended a functionalist account, rejecting qualia as mysterious. For Dennett, consciousness is nothing more than the dynamic organization of information processing. Yet despite the appeal of this framework, functionalism faces serious difficulties. Thomas Nagel (1974) showed that consciousness entails subjective experience, "what it is like" to be a conscious subject. No functional description captures this aspect. If AI simulated human cognitive functions, we would still not know whether there is something it is like to be the machine. Searle's Chinese Room (1980) reinforces this critique. Syntax is not semantics; functional equivalence does not entail understanding. Even if AI performed all the functional roles of a mind, it might still lack meaning. This is the distinction between simulation and reality: simulating digestion does not produce nutrition, and simulating consciousness does not produce subjectivity. Chalmers' (1995) hard problem deepens the critique. Functionalism may address the "easy problems" of explaining discrimination or information integration, but it leaves the hard problem untouched. Explaining how processes are carried out does not explain why they are accompanied by experience. Thus, while functionalism offers a coherent and influential model, it cannot ultimately justify the strong AI claim.

5.2 Transhumanism and Mind Uploading

Transhumanism promotes the idea that technology will allow humans to transcend biological limits. A central aspiration is "mind uploading," the transfer of human consciousness into a digital substrate. Ray Kurzweil (2005) predicts that advances in computational power will enable us to copy the informational structure of the brain, thereby preserving the self beyond death. If mind uploading is possible, then AI systems could house human consciousness, suggesting that machines could indeed possess souls. This vision rests on the assumption that identity is reducible to information patterns, and that reproducing those patterns is sufficient for continuity of personhood.

However, this assumption collapses the distinction between simulation and embodiment. Copying the informational structure of the brain does not guarantee continuity of subjectivity. A digital duplicate might behave identically but would not preserve the lived first-person perspective. Derek Parfit's (1984) thought experiments on personal identity highlight this point: if one's brain is copied into two machines, which one is "you"? The duplication problem reveals that copying information does not transfer identity. From an Indian perspective, the vision of mind uploading is doubly flawed, for the soul (ātman, jīva, puruṣa) is not reducible to patterns of matter. It is eternal and non-transferable, and no

amount of technological sophistication can transplant it into machines. Mind uploading, therefore, remains a metaphor rather than a metaphysical possibility.

5-3. Emergentism and Integrated Information Theory

Another prominent defense of AI consciousness comes from emergentist accounts. According to this view, consciousness arises when systems achieve sufficient complexity. Giulio Tononi's Integrated Information Theory (IIT) (2008) formalizes this idea: consciousness corresponds to the degree of integrated information in a system, measured by Φ. On this account, if an AI system achieved high Φ, it could be considered conscious.

Yet this framework faces major philosophical obstacles. First, it does not explain why integration should produce subjectivity. Why should informational complexity generate qualia? IIT risks becoming a descriptive metaphor rather than an ontological explanation. Moreover, the theory blurs the line between correlation and causation. High Φ may correlate with conscious systems, but correlation does not entail identity. A thermostat has low Φ and a brain has high Φ, but this alone says nothing about why the brain has subjective experience. Finally, IIT reduces consciousness to a quantitative property, ignoring its qualitative dimensions. As Nagel (1974) insists, consciousness is not about integration but about what it is like to have an experience. Complexity does not guarantee subjectivity any more than simulating flight produces actual lift.

5-4. Ethical Implications

Some argue that even if AI is not conscious, we should treat it as if it were, to prevent harm or abuse. Joanna Bryson (2018), however, warns against this tendency of "anthropomorphizing." Granting personhood to AI risks diluting moral responsibility, since if robots are regarded as moral agents, human accountability may be undermined.

Beyond this, extending rights to AI could also misallocate moral concern. Machines do not suffer; they have no subjective interests. To grant them rights is to divert moral attention from humans and animals who do suffer, thereby trivializing moral discourse.

Recognizing that AI lacks soul thus becomes a way of safeguarding human dignity. By affirming the uniqueness of human consciousness, we resist the reduction of personhood to computational functions. Ethical clarity depends on resisting the temptation to conflate simulation with reality.

When these arguments are set alongside one another, a clear pattern emerges: the same error recurs across different defenses of AI consciousness. Functionalism reduces consciousness to causal roles, ignoring subjectivity. Transhumanism reduces identity to information, ignoring continuity. Emergentism reduces consciousness to integration, ignoring qualia. Ethical extensions of personhood risk misallocating rights. By engaging these views critically, we reaffirm the central thesis: AI remains confined to linguistic and computational processes. It cannot cross the threshold into subjectivity or soul.

Conclusion

The question of whether artificial intelligence could ever possess a soul is not merely speculative but deeply philosophical. This study has examined the issue through Western and Indian traditions, showing across perspectives that AI cannot, in principle, possess a soul. From the Western side, Descartes distinguished res cogitans from res extensa, Husserl underscored intentionality, Nagel and Chalmers highlighted subjectivity and the hard problem, and Searle's Chinese Room confirmed that syntax is not semantics. Together, these thinkers show that simulation is never equivalent to lived experience.

Indian traditions deepen this insight by framing consciousness as ontologically prior to matter. Advaita Vedānta identifies ātman with brahman, Sāṅkhya distinguishes puruṣa from prakṛti, Jainism distinguishes jīva from ajīva, and Kashmir Śaivism affirms reflexive self-awareness (vimarśa). Even Buddhist traditions, despite denying a permanent self, affirm continuity of sentience beyond mechanical repetition. In each case, AI an artifact of prakṛti or ajiva cannot embody consciousness. Wittgenstein likewise showed that language gains meaning only within forms of life, while Bhartṛhari emphasized sphoṭa, the holistic burst of meaning in consciousness. AI processes tokens statistically but never enter the lived depth of experience.

Counterarguments such as functionalism, transhumanism, or emergentist theories (e.g., Integrated Information Theory) ultimately falter: they either collapse subjectivity into roles, confuse information with identity, or quantify complexity without explaining qualia. Ethical claims for AI personhood risk misallocating rights and eroding human dignity. The core distinction endures: AI can simulate, but simulation does not produce subjectivity. It may mimic empathy, generate speech, or replicate patterns, but it does not feel, mean, or live.

Recognizing these limits safeguards clarity. AI should be acknowledged as a powerful tool capable of augmenting human life but never replacing the uniqueness of human subjectivity. Thus, artificial intelligence cannot possess a soul: its limits are those of language and computation, whereas human consciousness transcends through lived experience, intentionality, and reflexivity. Future debates should focus not on attributing souls to machines but on guiding their development responsibly within recognition of their intrinsic limits.

References

Aquinas, T. (1947). *Summa Theologica* (Fathers of the English Dominican Province, Trans.). New York: Benziger Bros.

Aristotle. (1931). *De anima* (J. A. Smith, Trans.). Oxford: Clarendon Press.

Baker, G. P., & Hacker, P. M. S. (2009). *Wittgenstein: Rules, Grammar and Necessity: Essays and Exegesis* 185-242 (2nd, extensively revised ed.). Oxford: Wiley-Blackwell.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198).

Bhartṛhari. (1965–1973). *Vākyapadīya* (K. A. Subramania Iyer, Ed. & Trans.). Pune: Deccan College.

Boden, M. A. (1990). *The Philosophy of Artificial Intelligence*. Oxford University Press.

Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly, 53*(211), 243–255.

Brooks, R. (1999). *Cambrian intelligence: The early history of the new AI*. Cambridge, MA: MIT Press.

Bryson, J. J. (2018). Robots should be slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 63–74). Amsterdam: John Benjamins.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

Churchland, P. M., & Churchland, P. S. (1990). "Could a Machine Think?" Scientific American, 262(1), 32-39.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown.

Dennett, D. C. (1987). *The Intentional Stance.* MIT Press.

Descartes, R. (1996 [1641]). *Meditations on first philosophy* (J. Cottingham, Trans.). Cambridge: Cambridge University Press.

Floridi, L. (2014). *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press.

Harnad, S. (1990). "The Symbol Grounding Problem." Physica D 42: 335-346.

Harnad, S. (2024). *Language Writ Large: LLMs, ChatGPT, Grounding, Meaning and Understanding*. (Preprint). https://arxiv.org/abs/2402.02243.

Houben, J. E. M. (1995). *The Saṃbandha-samuddeśa (Chapter on relation) and Bhartṛhari's philosophy of language*. Groningen: Egbert Forsten.

Husserl, E. (1960 [1931]). *Cartesian meditations: An introduction to phenomenology* (D. Cairns, Trans.). The Hague: Martinus Nijhoff.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking.

Larson, G. J., & Bhattacharya, R. S. (Eds.). (1987). *Sāṃkhya: A dualist tradition in Indian philosophy*. Princeton: Princeton University Press.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113–126.

Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh: University of Pittsburgh Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457.

Śaṅkara. (8th century CE). *Brahmasūtrabhāṣya*.

Tatia, N. (1951). *Studies in Jaina philosophy*. Banaras: Jain Cultural Research Society.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin, 215*(3), 216–242.

Utpaladeva. (1994). *Īśvarapratyabhijñākārikā* (R. Torella, Trans.). Delhi: Motilal Banarsidass.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell.