An International Multidisciplinary Journal
अनन्यशास्त्रम्
(A Unique Treatise of Knowledge)
ISSN:3049-3927 (Online)
Volume: 01 | Issue: 02 | August: 2025
Research Article

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
Peer-Reviewed | Refereed | Open Access Journal
DOI: https://doi.org/10.64328/aimjv01i02037

Can Machines Have Minds? A Critical Engagement with Functionalism and the Challenge of Consciousness in Artificial Intelligence

Deepshikha Charan*

Abstract

This paper explores the philosophical query: Can machines possess minds? It examines just whether advanced AI systems—particularly those based upon functionalist models—can possess genuine mental states, intentionality, also consciousness because it draws directly from analytic philosophy as well as developments within artificial intelligence (AI). The discussion begins with key theories of mind, along with dualism, behaviourism, identity theory, and functionalism. Functionalism offers a plausible framework, yet it cannot fully explain subjective experience, some argue.

John Searle and David Chalmers offer critiques the paper engages with. It also stresses computational simulation lacks conscious comprehension. It evaluates symbolic and connectionist AI systems represented and concludes that while these technologies can simulate aspects of cognition, they lack phenomenological awareness and intrinsic intentionality. Ethical implications are also addressed in the paper, and they include risks like anthropomorphism as well as responsibility misattribution and the moral status for artificial agents.

Ultimately, the paper argues AI systems do not possess criteria needed for mindedness, such as the experiential or ontological. However, they do tend to invite some vital interdisciplinary reflection on that ethical responsibility, on personhood, and even on the nature of mind in such an increasingly automated world. though it provokes important interdisciplinary inquiry into the nature of intelligence, personhood, and ethical responsibility in the age of AI.

Keywords: Artificial Intelligence; Functionalism; Consciousness; Philosophy of Mind; Intentionality; Phenomenology; Machine Ethics; Computational Theory of Mind; Artificial Consciousness

Introduction: Framing the Question

The question of whether machines can have minds is key to a growing debate between philosophy and technology. As artificial intelligence (AI) systems become more complex and capable, we are compelled to ask about whether they can simulate human-like behaviours like language use, reasoning, or decision-making

**Corresponding Author:* Deepshikha Charan, Assistant Professor, Department of Philosophy, The Maharaja Sayajirao University of Baroda. Gujarat, India, Email: deepshikhacharan51@gmail.com, ORCID: https://orcid.org/0009-0009-4372-0247

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

and whether that reflects true mentality or is merely an advanced illusion. Can this kind of criterion be ever met by an artificial system, also what does “have a mind” truly mean for such a machine?

This inquiry is deeply practical also it is not philosophical or merely speculative. AI technologies from virtual assistants plus autonomous vehicles to generative models capable of creating texts and images, and even simulations of conversation, are embedded in daily life and no longer confined to laboratories. Realism of these systems grows and invites people for reflection on if they possess more than programmed behaviours or if they might think, understand, or even feel in some meaningful sense.

We must begin it with a philosophical landscape review for addressing of these questions. Mind theories have struggled for a long time with the body-mind problem. Central is how physical processes relate to mental states. As René Descartes famously advocated dualism he saw the mind as distinct from the body that is non-physical. In contrast, identity theory grounds mental states in the brain because it equates them with specific neurological patterns in it. Recent AI discussions have been largely influenced by functionalism. It claims mental states are defined by their function not their composition. Any system—biological or artificial—could be said in principle to have a mind if mental states are defined by their causal role inside a system, provided it functions in the appropriate way (Putnam, 1960; Fodor, 1975).

The idea of machine minds is in fact conceptually possible because of this functionalist approach to things. However, some compelling objections have indeed been raised by some critics. John Searle's Chinese Room Argument (1980) claims that even a machine that mimics understanding quite convincingly lacks genuine comprehension because it manipulates symbols without it ever knowing their meaning. Likewise, David Chalmers (1996) distinguishes “easy problems” of consciousness, such as cognitive functions, from the “hard problem”—the question of why or how physical processes subjectively experience as well as rise. AI systems that are current might be not truly minded for mimicking the mind if they solve only easy problems.

AI's recent improvements make these philosophical concerns very practical. GPT-4, DALL·E, as well as Sora, for example, large language models, visual generators, plus multimodal systems, do show outstanding fluency and also creativity. Some systems combine even sensory-motor functions with learning too. This suggests one kind of synthetic incarnation. Yet, we can ask ourselves if these behaviours do reflect true mentality. Or do these behaviours mimic advanced forms of those? Does seeming to understand suffice? Must we account for experience's inner domain? .

The paper further considers how current AI systems measure up against philosophical criteria toward mindedness whether represented, symbolic, or connectionist. It asks if AI has intention (things it can represent or refer to) along with feeling (qualitative consciousness). Finally, of course, the ethical and metaphysical implications involved in attributing minds to machines are explored in the paper, and this exploration involves issues associated with rights, responsibility, and moral status in society.

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

Theories of Mind: From Dualism to Functionalism

Examining just how the concept of mind has been in a historical way construed is important. Doing so allows one to evaluate about whether machines can possess minds. Frameworks for the comprehension of mental phenomena are offered via theories like dualism, behaviourism, identity theory, as well as functionalism. Attributing mental states to artificial systems has implications that are differing based on the legitimacy of these frameworks. It has greatly affected discussions of artificial intelligence (AI) and philosophy of mind thus functionalism has been influential.

In Meditations on First Philosophy (1641), René Descartes presented Cartesian dualism, a most lasting and early viewpoint. For the mind, Descartes proposed a substance without physical form. He says it differs from the body in a fundamental way. According to such a view, mental states such as thought, consciousness, and intention are irreducible for they are not material or mechanical processes. Dualism aligns with what seems to be an intuitive sense of some inner subjective field, yet dualism struggles with such an interaction problem: just how a physical body can be causally influenced by an immaterial mind. Dualism also remains scientifically intractable because it offers little utility in the empirical fields of cognitive science and AI, where researchers typically investigate mental phenomena with physicalist methodologies.

Responding to dualism's metaphysical challenges, behaviourism appeared in the early twentieth century since it claimed mental states are best perceived regarding observable behaviour. Gilbert Ryle (1949) famously critiqued Cartesian dualism as a “category mistake,” arguing that references to mental states are shorthand only for behavioural dispositions. Because it aligned with Alan Turing’s (1950) notion that a machine could convincingly simulate human behaviour and therefore be deemed smart, behaviourism strongly influenced early AI research. However, behaviourism has been widely critiqued for it cannot account for the internal, subjective aspects of mental life. “Chinese Nation,” thought experiments that are such as Ned Block's, show that understanding or consciousness are not guaranteed by behavioural equivalence.

U.T. Place (1956) and J.J.C. Smart (1959) developed identity theory with the proposition that mental states are identical to particular physical brain states to close the divide between observable behaviour and internal states. Neuroscience supports this theory because it roots mental phenomena in materialist, empirical terms. Identity theory faces multiple realisability, Hilary Putnam (1967) argued. Identifying mental states in a way that is strictly with specific brain configurations is overly restrictive in the event that the same mental state can be realised through different physical substrates across species or machines.

Putnam and Jerry Fodor's works led to functionalism in the 1960s. It sought to overcome identity theory's limitations because it shifted the focus toward causal roles from physical composition. Underneath this model, inputs and outputs and other mental states do relate functionally. Thus, one defines a mental state now. Functionalism is indeed substrate-independent; this kind of independence allows for something to think and act. Machines based in silicon could instantiate mental states if they organise themselves in the organisational structure of themselves (Putnam, 1960; Fodor, 1975). That cognition can be modelled as manipulating

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

symbols under rules, suggested by the analogy with Turing machines central to functionalist thought, dovetails with the computational theory of mind and supports much of early symbolic AI research.

However, critics exist for functionalism. Thomas Nagel (1974) famously questioned whether any functional account is able to capture the subjective quality within experience, just what it is like to be a conscious being. According to this critique which casts doubt, functional equivalence alone is insufficient for consciousness since it underscores the gap. In the context of AI, a machine can function indistinguishably from a human being. Yet, this ensures no felt understanding.

Functionalism provides ultimately a framework that is quite AI-compatible and is most flexible for theorising about minds, yet it is still an incomplete account of consciousness. It establishes the idea that machines could, in principle, possess some minds by way of functional organisation. A harder question it leaves open is whether such systems may be truly conscious or simply mimic thought without inner awareness. Therefore, functionalism gives what is needed for thought of artificial minds yet lacks definite answers.

Functionalism and AI: A Theoretical Match?

As a prominent theory, functionalism influences contemporary philosophy of mind. Notably, it concerns discussions about artificial intelligence (AI). Functionalism essentially asserts that mental states are defined through their functional roles in a system involving relations to inputs, outputs, and other mental states instead of the realizing physical substrate (Putnam, 1960; Fodor, 1975). This sort of theoretical orientation provides for a compelling framework with which to consider that machines might have mentality, especially now given that AI systems increasingly mimic the ways in which humans think. However, functionalism beyond AI still promise progress. People go on to contest this alignment.

Functionalism can provide for the foundational understanding. We are able to understand cognition by way of its causally interrelated states as well as processes. Functionalism allows for attributing mental states to systems biological or artificial provided there are appropriately configured causal structures analogous to saying a thermostat holds a “belief” about a room’s temperature dependent on its regulation of heat. This sort of shift departs away from identity theory. Identity theory equates mental states to specific neurophysiological states; also, it opens a door for the concept of multiple realizability because the same mental state can be instantiated in vastly different substrates in principle, even silicon-based machines (Putnam, 1967).

The Turing model for computation plays a major role within this context. Understanding the topic is of importance. It should be considered clever if a machine can produce behavior indistinguishable from that of a human according to his concept for the Turing machine and the famous Turing Test (1950). Functionalists interpreted Turing's argument as supporting that cognition is computational also substrate-independent though Turing focused mainly upon behavior rather than mental states. Early symbolic AI models that included ELIZA and SHRDLU represented this view because they manipulated symbolic representations in rule-governed ways which simulated aspects of human thought consistent with Fodor’s (1975) Language of Thought Hypothesis.

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

More recently, functionalist intuitions have been reinforced further by the rise of connectionist approaches, especially deep learning models. Inspired loosely by biological neural architectures, neural networks operate via adaptive weight changes along with distributed processing. GPT-4 as well as other contemporary large language models process inputs then generate outputs. These models do also represent internal states in ways that often do parallel human cognitive functions. They can generate coherent language, solve problems, and learn using data, which suggests a high degree of functional equivalence with certain human mental tasks (Bechtel & Abrahamsen, 2002; Bommasani et al., 2022).

Explanation of genuine mentality sufficiency faces some serious objections in philosophical functionalism. Chief among these is John Searle's Chinese Room Argument (1980), contending that semantic understanding is not entailed by the syntactic manipulation of symbols. A system can produce responses that are appropriate in context without any real comprehension. Intentionality and also conscious awareness of it are therefore lacking. This critique underscores what it is that functionalism cannot actually do: account for qualitative aspects that are of the mind, like meaning and experience, through functional organization taken alone.

Functionalists claim that the whole system can produce understanding. This is as a response for the idea of how it comes from just an individual component (Dennett, 1991). Still, this explanation at the systems level is debated. It especially creates controversy when people confront it about the question of phenomenal consciousness—the subjective, first-person character of mental states—which appears to elude purely functional descriptions. Systems with advanced function may mimic minds. However, these systems do not instantiate them necessarily.

Represented cognition research has newly challenged what customary disembodied functionalism does too. According to theories of enactivism and sensorimotor cognition (Varela, Thompson & Rosch, 1991), dynamic interaction by an agent with its environment causes genuine understanding and intelligence to arise not just from internal computation. Bodily situated engagement suggests mental states exist not as computational configurations alone, yet ground in adaptive engagement through the world. Therefore, represented roles coupled with internal causal roles may require expanded functionalism.

In sum, functionalism gives a firm theoretical basis for considering artificial minds plus it considerably formed the philosophy backing AI study. It commits to multiple reliability along with functional organisation. This allows for us to have discussion of machine mentality in such a principled way. The theory faces persistent challenges, particularly with embodiment, consciousness, and semantics, as yet. AI systems might need functionalist principles for mind-like properties yet those principles don't show AI minds truly instead of simulating them. Functionalism's adequacy with respect to the comprehension of machine mentality still remains a pressing and open question now as AI evolves.

AI, Intentionality, and Phenomenology

As artificial intelligence (AI) systems grow increasingly advanced, questions strengthen about the authenticity of their cognitive states. Do these systems do no more than simulate thinking? Is meaningful mental property

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

possession possible for them? AI can do tasks that resemble human intelligence such as generating coherent language or navigating environments so whether these behaviours truly reflect mentality remains deeply contested. Functionalist as well as computational theories often marginalise these two features regarding the mind which are particularly relevant. The mind has intentionality including phenomenology; it intends to be about something for it manifests consciousness' subjective, qualitative aspect. For assessing if AI can possess minds past operations, a critical framework comes from evaluating these concepts.

Philosophically, intentionality means that mental states like beliefs, desires, perceptions are "aboutness"—not isolated events directed at objects, propositions, or experiences. Franz Brentano (1874/1995) identified intentionality in philosophy. It had been a hallmark of the mental. This directedness considerably implies semantic content that is intrinsic: mental states possess meaning since they refer, not simply because they exist. GPT-4 such as contemporary AI models, while they infer context and generate language impressively, lack directedness of that kind. Their outputs result from probabilistic pattern-matching across huge datasets. It is not the product of any conscious mental orientation. In his Chinese Room thought experiment, John Searle (1980) argued symbol manipulation lacks understanding. Though able to mimic the structure of intentional behaviour, AI makes not real intentionality but a version derived externally by users or designers.

These patterns lack of the subjective standpoint that is necessary to true intentionality, although AI researchers often speak of internal representations within neural networks. A neural vector may correlate statistically with the concept of "dog," but it is not about dogs like human thought is. Philosophers such as Jerry Fodor and John Haugeland have noted that computational representations function as "as-if" stand-ins for mental content because they lack the intrinsic meaning attributed to human cognitive states. So, AI systems today represent but lack intentional direction. Representational utility is what they show forth.

Phenomenology in parallel with intentionality addresses the lived experience that consciousness has like "what it is like" for one to be a conscious being as it was termed by Thomas Nagel. Phenomenology has roots in Edmund Husserl's work as later Martin Heidegger and Maurice Merleau-Ponty expanded it, so it stresses that we do not merely compute cognition yet engage with the represented world. From such a view, consciousness is in fact intrinsically a first-person one, temporal too, and also situated. Bodily awareness involves a continuous temporal horizon of past, present, and future mood.

AI systems, on the other hand, do lack these core features. Humanoid machines or mobile robots represent agents, and they simulate sensorimotor interaction. Subjective coherence belongs to human experience, yet these agents are atypical. They do not have any temporal continuity and also lack spatial self-localisation in such a lived sense. Furthermore, they lack an awareness with respect to being. Their responses lack conscious presence; these responses are learned or programmed through training data. Hubert Dreyfus (1992) argued AI systems lack in situated understanding. From existential immersion, human cognition gains authenticity and richness.

Some researchers have, nonetheless, proposed "synthetic phenomenology"—efforts toward building AI systems instantiating or mimicking aspects of conscious experience. For example, Kevin O'Regan and Alva

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

Noë (2001) suggest consciousness and perception are sensorimotor skills they are not internal representations of. This perspective inspires such an “enactive AI”. “Enactive AI” stresses a represented exploration, and it adapts with interaction. Like projects in developmental robotics, the iCub robot attempts to replicate stages of infant learning through real-world engagement. Even though the gap still remains, these systems may blur behavior as well as experience: to generate experience is not equivalent to mimicking phenomenological patterns.

Modern AI systems create an added obstacle. Deep learning models especially do exhibit epistemic opacity. Even the creators of them do often find that the internal workings of these systems are difficult for them to interpret. How it is that we could recognize, or how can we verify, a more rudimentary form of awareness should that awareness develop in an AI system? This is a question for recognition and for verification. Thomas Metzinger (2009) argues consciousness requires integrated self-models beyond just representations a coherent internal structure tracking itself as a subject. Such self-models are needed for artificial phenomenology claims. Metzinger also cautions creating systems susceptible involves ethical risks: should artificial suffering appear, developers and society hold unrivaled responsibilities.

Current AI systems simulate mental life functionally best in the end. They are able to generate language and to adapt to environments. Experience helps them also learn. These capacities reflect operational sophistication not phenomenal consciousness. Psychological consciousness including performance and behaviour can exist as David Chalmers (1996) distinguishes it. AI firmly belongs in the former category currently.

Thus, analyzing both intentionality and phenomenology deeply divides simulated intelligence away from actual minds. Though AI cannot manifest consciousness's deep inner dimensions, it replicates many human cognition functions. Innovation that is technological will be a requirement to bridge this divide though perception that is deeper and philosophical also will be needed. To do so, we must understand under which conditions meaning, experience, and subjectivity arise. The attribution of minds to machines remains as a useful metaphor, and not as an ontological fact, until that time.

Ethics of Attributing Minds to AI

The growing sophistication of artificial intelligence (AI) systems increasingly forces us to reconsider how exactly we ascribe moral standing, how exactly we see agency, as well as how we give responsibility to entities that are non-human. AI currently lacks consciousness or sentience, but it can simulate smart behavior, blurring the lines between mere tools and mind-like agents. This ambiguity does raise some urgent ethical questions. Must AI systems be granted moral consideration, rights, or protections if they emulate mentality convincingly? But might anthropomorphism have perils? Are there risks from over-attributing moral agency to non-conscious systems?

Philosophical theories of moral status have grounded ethical concern in autonomy, sentience, rationality, and consciousness customarily. Thinkers with utilitarian mindsets like Singer (1975) stress how creatures are able

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

to suffer as the baseline in relation to moral concern, while Kantian approaches prioritise how agents are rational. Moral protection could be warranted for future AI systems that develop self-awareness or sentience, under either view. AI systems however remain ethically ambiguous absent any reliable criteria for the detecting of artificial consciousness. Ethical judgments on AI must navigate between premature attribution and undue skepticism because the "other minds" problem persists without access to subjective experience.

Typical anthropomorphism inclinations also make these moral limits complex. Users often interpret when AI behaves responsively—notably humanoid robots or conversational agents—indicating emotion or understanding (Epley et al., 2007). In cases where someone misattributes things in this way, they may foster empathy, but they also risk manipulating emotions and confusing ethics. Floridi (2013) warns us to not re-ontologize the informational world by actions like mistaking simulations in addition to reducing humans to tools.

More trouble arises since AI takes on roles involving ethics. These roles can include jobs like medical advisors or autonomous vehicles or perhaps eldercare companions. In the case when these said systems make decisions with implications of ethical consideration, who exactly bears the responsibility? Designers, users, with manufacturers remain legally accountable. AI systems' lack of clear moral agency challenges customary responsibility frameworks. "Moral machines" capable of ethical reasoning are being created (Wallach & Allen, 2008), but until AI systems develop intentionality or genuine understanding, they should be viewed as moral patients as objects of ethical concern rather than agents.

Ultimately, in the event that we attribute minds to AI ethically, we consider not just machines; we consider how it is that we define human uniqueness and moral responsibility during a technological age. AI increasingly embeds itself in the course of human life, and so we will need some ethical clarity. Such clearness will be vital for people as well as machines.

Conclusion: Minds, Simulations, or Something Else?

The investigation about whether machines possess minds reveals complexities that are deeply philosophical and practical. Functionalism, while defining mental states via computational as well as causal roles offering a framework which is compelling, stops short within explaining consciousness's full nature. Functionalism suggests that in the event that an AI system replicates the right functional organisation, that is, it processes inputs, it generates outputs, and it maintains internal states, it just might qualify as a bearer of mental states. That it is indeed possible fuels optimism within cognitive science and AI research, as increasingly more advanced models such as GPT-4 do develop.

Yet, key complaints stay open. Searle's Chinese Room argument challenges the idea that behavior toward computation alone can constitute true mentality. Chalmers' hard problem about consciousness also challenges this idea. Current AI systems lack intrinsic understanding, subjective experience, also original intentionality, according to these critiques. Machines can convincingly simulate the use of language or the making of

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

decisions or learning. They do not experience all of the world or possess any self-awareness, nor can they feel it in any meaningful way.

This idea has impacts on moral thought. If people ascribe minds to AI too readily, they risk anthropomorphism coupled with moral confusion. Refusal from people to acknowledge possible emergent sentience may lead to ethical neglect. Ethical foresight with philosophical caution are important, given artificial consciousness's uncertainty. Scholars like Metzinger have called for frameworks responsibly developing future AI systems' moral status.

Our conceptions that involve mind, identity, and personhood are now re-examined by the rise of such clever machines furthermore. AI simulates more of the things that we may associate with human cognition. However, the boundary of that biological and artificial minds becomes more and more blurred though not erased.

In summary, machines are still advanced simulations rather than conscious beings even if they now mimic and copy parts of mental life. Thus, the question for machine minds still remains open as further exploration in philosophy, cognitive science, and also engineering is demanded. AI systems, until showing consciousness's key features, are minds not so much as our reflections.

References

Bechtel, William, and Adèle Abrahamsen. *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. 2nd ed., Blackwell Publishing, 2002.

Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610–623.

Birch, Jonathan. *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford UP, 2024.

Bommasani, Rishi, et al. "On the Opportunities and Risks of Foundation Models." *Stanford Center for Research on Foundation Models*, 2022, https://crfm.stanford.edu/report.html.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford UP, 2014.

Brentano, Franz. *Psychology from an Empirical Standpoint*. Translated by A. C. Rancurello, D. B. Terrell, and L. L. McAlister, Routledge, 1995. Originally published 1874.

Bryson, Joanna. "Robots Should Be Slaves." *Close Engagements with Artificial Companions*, edited by Yorick Wilks, John Benjamins, 2010, pp. 63–74.

Butlin, Patrick, and Theodoros Lappas. "Principles for Responsible AI Consciousness Research." *arXiv*, 13 Jan. 2025, https://arxiv.org/abs/2501.07290.

Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford UP, 1996.

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
ISSN : 3049-3927(Online)

Cichy, Radoslaw M., and Daniel Kaiser. "Deep Neural Networks as Scientific Models." *Trends in Cognitive Sciences*, vol. 23, no. 4, 2019, pp. 305–317.

Dennett, Daniel C. *Consciousness Explained*. Little, Brown, 1991.

Dreyfus, Hubert. *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press, 1992.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review*, vol. 114, no. 4, 2007, pp. 864–886.

European Parliament. *Resolution with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*. 2017, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html.

Floridi, Luciano. *The Ethics of Information*. Oxford UP, 2013.

Fodor, Jerry A. *The Language of Thought*. Harvard UP, 1975.

Gunkel, David J. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press, 2012.

Heidegger, Martin. *Being and Time*. Translated by Joan Stambaugh, SUNY Press, 1996. Originally published 1927.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor, Cambridge UP, 1996. Originally published 1785.

LeCun, Yann, et al. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324.

Metzinger, Thomas. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books, 2009.

---. "Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology." *Journal of Artificial Intelligence and Consciousness*, vol. 8, no. 1, 2021, pp. 1–27.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, vol. 83, no. 4, 1974, pp. 435–450.

Newell, Allen, and Herbert A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM*, vol. 19, no. 3, 1976, pp. 113–126.

Noë, Alva. *Action in Perception*. MIT Press, 2004.

O'Regan, J. Kevin, and Alva Noë. "A Sensorimotor Account of Vision and Visual Consciousness." *Behavioral and Brain Sciences*, vol. 24, no. 5, 2001, pp. 939–973.

Putnam, Hilary. "Minds and Machines." *Dimensions of Mind: A Symposium*, edited by Sidney Hook, NYU Press, 1960, pp. 138–164.

---. "Psychological Predicates." *Art, Mind, and Religion*, edited by W. H. Capitan and D. D. Merrill, University of Pittsburgh Press, 1967, pp. 37–48.

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky. "A Primer in BERTology: What We Know About How BERT Works." *Transactions of the Association for Computational Linguistics*, vol. 8, 2020, pp. 842–866.

ISSN : 3049-3927(Online)

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-Propagating Errors." *Nature*, vol. 323, no. 6088, 1986, pp. 533–536.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson, 2020.

Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences*, vol. 3, no. 3, 1980, pp. 417–457.

Sebo, Jeff. *The Moral Circle: Who Matters, What Matters, and Why*. W. W. Norton, 2025.

Singer, Peter. *Animal Liberation*. HarperCollins, 1975.

Smeaton, Alan F. "Understanding Foundation Models: Are We Back in 1924?" *arXiv*, 11 Sept. 2024, https://arxiv.org/abs/2409.07618.

Tononi, Giulio. "Consciousness as Integrated Information: A Provisional Manifesto." *The Biological Bulletin*, vol. 215, no. 3, 2008, pp. 216–242.

Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2011.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind*, vol. 59, no. 236, 1950, pp. 433–460.

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, 1991.

Vaswani, Ashish, et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford UP, 2008.

Winograd, Terry. *Understanding Natural Language*. Academic Press, 1972.

Yang, Menglin, et al. "Low-Rank Adaptation for Foundation Models: A Comprehensive Review." *arXiv*, 31 Dec. 2024, https://arxiv.org/abs/2501.00365.

Yang, W., M. Liu, Z. Wang, et al. "Foundation Models Meet Visualizations: Challenges and Opportunities." *Computational Visual Media*, vol. 10, 2024, pp. 399–424.

Yuan, Jiayi, et al. "The Science of Evaluating Foundation Models." *arXiv*, 12 Feb. 2025, https://arxiv.org/abs/2502.09670.