FUNCTIONAL AND AETIOLOGICAL CORRESPONDENCES BETWEEN NATURAL AND ARTIFICIAL INTELLIGENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

In this article, we present a descriptive framework for analysing natural and artificial intelligence across three dimensions: functional characterisation, developmental aetiology, and inference aetiology. We argue that, despite significant differences in implementation, both forms of intelligence exhibit functional comparability, understood as the capacity to map inputs to outputs through internally mediated transformations.

At the same time, we identify fundamental aetiological differences. In terms of development, natural intelligence emerges from endogenous, evolutionarily shaped processes, whereas artificial systems are trained under explicitly defined optimisation regimes. In terms of inference, natural intelligence can be activated both exogenously and endogenously, while artificial systems remain predominantly dependent on externally triggered initiation mechanisms.

The proposed framework integrates competing perspectives on the uniqueness of human cognition while preserving its functional analogy to artificial systems. By distinguishing between functional similarity and aetiological difference, it provides a principled basis for comparing natural and artificial intelligence without reducing one to the other.

1 INTRODUCTION

A widely cited observation, often attributed to Dobzhansky Dobzhansky (1962), holds that although all organisms are unique, humans appear distinctive in the scope and flexibility of their cognitive capacities. This perceived uniqueness has traditionally been grounded in human intelligence, which is typically characterised as domain-general and adaptable, in contrast to the more specialised cognitive abilities observed in non-human animals Premack (2004). While the similarities and differences between human and animal intelligence remain an active topic of scientific inquiry, recent advances in artificial intelligence (AI) have introduced a new and pressing dimension to this debate Placani (2024).

In particular, the rapid development and widespread deployment of large-scale machine learning models has prompted renewed reflection on the nature and limits of intelligence. Contemporary AI systems exhibit increasingly sophisticated behaviours, often perceived as indicative of general cognitive competence. This has led to fundamental questions concerning the status of artificial systems: are they comparable to natural intelligence, or merely sophisticated artefacts that simulate it Shneiderman (2020)? More broadly, do such developments challenge the long-standing assumption of human cognitive uniqueness?

This paper addresses these questions by examining the relationship between natural and artificial intelligence through a unified analytical framework. We argue that, although these forms of intelligence can be understood as functionally similar — both capable of performing complex input–output transformations — they differ fundamentally in their aetiology. In particular, we distinguish between (i) the processes through which these systems are developed, and (ii) the conditions under which their inferential mechanisms are initiated and controlled.

In the developmental dimension, natural intelligence arises through evolutionary and developmental processes that lack explicit, system-accessible optimisation objectives, whereas artificial systems are

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

trained under explicit or implicit optimisation regimes, such as loss minimisation or reward maximisation. In the inferential dimension, natural intelligence exhibits endogenous forms of control, whereby cognitive processes may be initiated internally, whereas artificial systems are typically activated exogenously, for example through user inputs, automated processes, or interactions with other systems.

To this end, we propose a threefold framework for analysing intelligence, encompassing (i) a functional perspective, (ii) an aetiology of development, and (iii) an aetiology of inference. This framework enables a more precise articulation of both the similarities and the differences between natural and artificial systems, and provides a basis for reassessing their philosophical and practical implications.

2 The Cattell–Horn–Carroll Framework of Human Intelligence

Scientific inquiry into the structure of human intelligence can be traced back to the work of Spearman Spearman (1927), who, through early applications of factor analysis, identified a general factor of intelligence (g) alongside a set of task-specific factors (s). Although his work has been subject to extensive debate and refinement, it remains foundational in contemporary psychometrics.

The currently dominant model of human intelligence is the Cattell–Horn–Carroll (CHC) framework, which integrates theoretical contributions from Cattell, Horn, and Carroll into a hierarchical taxonomy of cognitive abilities Schneider & McGrew (2018). Within this framework, human intelligence is typically understood as comprising three strata:

Stratum I: narrow abilities (specific cognitive skills),

Stratum II: broad abilities (clusters of related skills),

Stratum III: a general intelligence factor (g).

While there is substantial empirical support for the existence of a general intelligence factor across populations and cultures Warne & Burningham (2019); de la Fuente et al. (2021), it is no longer commonly interpreted as an isolated or unitary cognitive trait. Rather, g is often understood as reflecting the aggregate outcome of multiple interacting cognitive processes shaped by both biological and environmental influences Schneider & McGrew (2012); Laland & Seed (2021); Wilcox et al. (2026).

At the intermediate level, the CHC framework identifies a number of broad cognitive abilities (Stratum II), including, among others, fluid intelligence (Gf), crystallised knowledge (Gc), quantitative knowledge (Gq), short-term memory (Gsm), and long-term storage and retrieval (Glr). Additional domains include sensory processing capabilities (e.g. visual and auditory processing), psychomotor abilities, and processing speed. Some recent studies have proposed extending the framework to incorporate further domains, such as financial literacy or face-processing ability, although their status as independent broad abilities remains under discussion Callis et al. (2023); Walker et al. (2023).

The CHC framework provides a structured account of the organisation of human cognitive abilities, and, importantly for present purposes, highlights the layered and functionally differentiated nature of natural intelligence. This makes it a useful point of reference for evaluating whether and to what extent artificial systems exhibit comparable functional organisation, an issue to which we now turn.

3 On Progressive Artificial Intelligence

Artificial intelligence systems are typically understood as tools designed to perform tasks that would otherwise require human cognitive abilities. Among the various ways of classifying such systems, one influential taxonomy distinguishes between different stages of AI development, often described in progressive terms IBM Topics (2024).

This classification commonly identifies three categories:

Weak AI (narrow AI): systems designed to perform specific tasks, such as image recognition or natural language processing, without possessing general cognitive abilities;

Artificial General Intelligence (AGI): a hypothetical form of AI characterised by human-level, domain-general cognitive competence;

Superintelligence: a speculative stage in which artificial systems surpass human intelligence across all relevant domains.

Narrow AI is already widely realised and deployed across a range of domains and industries. In contrast, AGI remains largely aspirational. While some researchers argue that recent developments in large-scale, multimodal foundation models or large language models may constitute steps towards more general forms of intelligence Fei et al. (2022); Ananthaswamy (2024), others remain sceptical about the feasibility of such systems Fjelland (2020); Muñoz et al. (2025), including concerns regarding the limits of scaling current solutions aaa (2025). A central issue in this debate is the role of intersubjectivity, inter-corporeality Abbate (2023), and embodiment, which has been argued to be essential for achieving genuinely general intelligence aaa (2025). At the same time, ongoing research in biomimetic and embodied systems suggests that progress in this direction may be, at least to some extent, possible Hong et al. (2026).

Despite the impressive capabilities demonstrated by contemporary AI models, there is currently no broad scientific consensus that AGI has been achieved, or that its realisation is imminent Ananthaswamy (2024). This lack of agreement is partly due to the absence of a precise and widely accepted definition of artificial general intelligence Chen et al. (2026). From the perspective of the CHC framework, AGI may be interpreted as the integrated realisation of multiple narrow and broad cognitive capabilities. In this sense, it aligns with subsumptive or layered architectures, such as those proposed by Brooks Brooks (1990), and reflects challenges highlighted by Moravec's paradox Moravec (1988), according to which tasks that are simple for humans may remain difficult for artificial systems, and vice versa.

Superintelligence represents a further, largely speculative stage in this taxonomy. It is often discussed in both technical and philosophical contexts, although it is frequently conflated with AGI in public discourse. From an analytical standpoint, however, it is important to distinguish clearly between human-level general intelligence and forms of intelligence that would exceed it.

The progressive classification of artificial intelligence provides a useful conceptual backdrop for the present discussion. However, rather than focusing on whether current systems qualify as AGI, the analysis that follows shifts attention to a different question: to what extent can artificial systems, regardless of their position within this taxonomy, be considered functionally comparable to natural intelligence, while remaining distinct in their aetiology of development and inference?

4 Functional Perspective

The idea that intelligence can be characterised in functional terms, independently of its physical realisation, dates back at least to Turing. On this view, intelligence is not defined by its biological substrate, but by the role it plays as a system capable of transforming inputs into appropriate outputs. This position is closely aligned with functionalist approaches in the philosophy of mind Levin (2023), which emphasise behavioural and computational equivalence over material constitution.

Although the biological mechanisms underlying natural intelligence remain only partially understood, it is widely recognised that cognition emerges from the activity of large-scale networks of interconnected neurons, involving both electrical and chemical signalling Harari (2019). Contemporary developments in artificial neural networks, including spiking neural networks, attempt (tough in simplified form) to capture aspects of this distributed and dynamic processing Wang et al. (2020); Fang et al. (2023); Shi et al. (2025).

From a functional perspective, natural intelligence can be modelled as a system that processes stimuli — both external (e.g. sensory input) and internal (e.g. memory or internal states) — by encoding them into latent representations and mapping them onto actions or decisions Mathis et al. (2024); Brincat et al. (2018); Gallivan et al. (2018). This process can be expressed abstractly as a time-dependent mapping:

$$y_t = f_t(x_t, C_t, f_{t-1}, \Theta_t), \tag{1}$$

where x_t denotes input stimuli, C_t represents contextual factors (environmental and social), f_{t-1} reflects the prior functional state of the system, and Θ_t denotes the system's parameters (e.g. neural structure and connectivity). The output y_t corresponds to a behavioural action or decision. This formulation highlights the dynamic and potentially self-modifying character of natural intelligence, consistent with views of cognition as an adaptive, recurrent process Christensen & Hooker (2000).

While this formulation is necessarily an abstraction, given the immense complexity and partially unknown structure of biological systems, it captures a key feature: natural intelligence operates as a mapping from inputs and internal states to outputs. Computational neuroscience can be understood, in part, as an attempt to approximate or characterise this mapping Khilkevich et al. (2024); Kriegeskorte & Douglas (2018).

A similar functional description can be applied to artificial intelligence systems, including both supervised, unsupervised, and reinforcement learning. Consider, for instance, the latter one, in which an agent interacts with an environment in order to maximise a reward signal Sutton et al. (1998); Stuart & Peter (2010). Such systems can be described by a mapping of the form:

$$y_t = f(x_t, C_t, r_t, y_{t-1}, \Theta), \tag{2}$$

where x_t represents input data (e.g. sensory input or symbolic representations), C_t denotes the environment, r_t is a reward signal guiding learning (how the undertaken action is good or bad), y_{t-1} captures prior outputs or states, and Θ denotes model parameters. The output y_t corresponds to an action or decision taken by the system.

Despite significant differences in implementation, the structural similarity between these formulations is evident. Both natural and artificial systems can be understood as performing transformations over inputs, contextual information, and internal states to generate outputs. In this sense, they exhibit functional comparability.

This does not imply that the underlying processes are identical, nor that the resulting mappings are equivalent in complexity or scope. Rather, the claim is more modest: at an abstract level, both natural and artificial forms of intelligence instantiate systems of input–output transformation mediated by internal representations and state-dependent dynamics. The functional perspective thus provides a principled basis for comparison, while leaving open the possibility of deeper differences in their aetiology, to which we now turn.

5 Developmental Aetiology

Despite their functional comparability, natural and artificial intelligence differ fundamentally in the processes through which they are developed. These differences are not reducible to a simple contrast between biological and computational substrates; rather, they reflect distinct forms of developmental aetiology.

Accounts of the development of human intelligence have long emphasised the role of interaction with the environment. In particular, Piaget's influential theory of cognitive development characterises intelligence as emerging through a staged process driven by the continuous interaction between the organism and its surroundings Easley Jr (1978). Central to this view is the notion of cognitive equilibrium: discrepancies between existing cognitive structures and new experiences generate disequilibrium, which in turn drives the adaptation and reorganisation of cognitive structures. Although Piaget's framework has been substantially revised and extended in contemporary research Rochat (2024), it remains indicative of an important feature of natural intelligence — namely, its development is guided by internally mediated processes of adjustment and self-organisation.

From an evolutionary perspective, the development of intelligence can be understood in terms of adaptive pressures acting over long time scales. Cognitive systems are shaped to support survival and efficient interaction within complex environments and social systems Dunbar et al. (2014). One important aspect of this process is the optimisation of energy expenditure: well-practised behaviours tend to become automatised, reducing cognitive load and conserving metabolic resources Shine & Shine (2014); Saberi et al. (2024). In this sense, natural intelligence reflects a form of adaptive organisation emerging from distributed, often implicit pressures, rather than from explicitly specified objectives.

Artificial intelligence systems, by contrast, are typically developed under explicit optimisation regimes. Learning processes are guided by objective functions, such as loss minimisation in supervised learning or reward maximisation in reinforcement learning, which are defined externally and imposed deliberately Sun et al. (2019). These objectives may remain fixed or evolve during training, for example in adaptive or curriculum-based learning settings Sivan et al. (2024); Luo et al. (2025). In all cases, however, the optimisation criteria are, in principle, accessible and specifiable independently of the system itself.

This contrast can be captured in terms of differing forms of teleology. The development of natural intelligence may be described as exhibiting a form of *endogenous* or internally mediated goal-directedness, arising from processes such as self-regulation, homeostasis, and evolution: in short: adaptation. By contrast, artificial systems exhibit *exogenous* teleology, in which goal-directed behaviour is defined by externally specified objectives and optimisation criteria.

The distinction is not absolute: evolutionary processes can be interpreted as a form of optimisation, and some artificial systems incorporate adaptive or self-modifying mechanisms. Nevertheless, the difference in how objectives are specified and realised remains significant. Natural intelligence develops under diffuse, multi-level pressures that are not directly encoded as explicit objective functions, whereas artificial systems are, at least in principle, constructed to optimise clearly defined goals.

This divergence in developmental aetiology complements the functional perspective outlined above. While both natural and artificial systems may be described in terms of input–output mappings, the processes through which those mappings are formed differ in systematic and philosophically relevant ways.

6 Inference Aetiology

Having examined how natural and artificial intelligence are developed, we now turn to a related but distinct question: what initiates and governs their use? In other words, what triggers the activation of intelligence in both systems?

Both natural and artificial intelligence can be activated by external stimuli. In the case of human cognition, environmental interactions, such as visual or tactile stimuli, are well-known triggers of neural activity and cognitive processing Stenfors et al. (2019); Dehaghani & Zarei (2025); Steinmetz et al. (2005). Similarly, artificial systems may be activated by external inputs, including sensor data (e.g. motion detection, temperature thresholds), user prompts, or signals from other systems Serra-Simón et al. (2025); Xie et al. (2025).

External triggering is not limited to physical stimuli. In humans, social interactions play a central role in activating and shaping cognitive processes. Encounters with other individuals, recognition of familiar faces, and participation in conversation all serve as powerful triggers for cognitive activity Kasai et al. (2015); Naci et al. (2018); Gobbini et al. (2004). Analogously, artificial systems may be activated through inter-agent communication in multi-agent environments, where the output of one system serves as input to another Chen & Decker (2002); Dastani et al. (2004); Baldoni et al. (2015). Such interactions can be interpreted as a functional counterpart to socially mediated cognition.

In addition, both natural and artificial systems may be activated by temporally structured processes. Periodic or scheduled triggers, such as circadian rhythms in biological systems or timer-based execution in computational systems, can initiate processing independently of immediate external input.

Despite these similarities, a crucial difference emerges in the presence of internally generated triggers. Human cognition is not solely reactive to external stimuli; it can also be initiated endogenously. Processes such as reflection, imagination, and introspection enable cognitive activity to arise without direct external input Gupta et al. (2025); Liang et al. (2026); Jack & Roepstorff (2002); Schwitzgebel (2012). These internally driven processes play a central role in deliberation, planning, and self-regulation.

Artificial systems, by contrast, typically lack genuinely endogenous initiation mechanisms. One might argue that autonomous agents exhibit forms of self-initiated activity; however, such processes remain grounded in externally specified architectures, objectives, or execution schedules. After all, the concepts such as self-reflection or meta-cognition, these are generally implemented as higher-

order processing stages or feedback loops within predefined architectures Qin et al. (2025); Nisa et al. (2026) without strong self-initialisation. Crucially, such mechanisms are still ultimately dependent on external triggering, whether through human input, system scheduling, or interaction with other agents Salmani & Lewis (2025).

The contrast can thus be framed in terms of the aetiology of inference. Natural intelligence exhibits both exogenous and endogenous modes of activation, whereas artificial systems are predominantly, if not exclusively, exogenously triggered. This results in a fundamental difference in how cognitive processes are initiated and controlled.

This distinction complements the developmental analysis presented earlier. While both natural and artificial intelligence can be described functionally as systems that map inputs to outputs, they differ not only in how these mappings are formed, but also in how and when they are activated.

7 Conclusions

In this paper, we proposed a threefold framework for comparing forms of intelligence, encompassing natural intelligence, as exhibited by humans, and artificial intelligence. The framework integrates three dimensions of analysis: functional characterisation, developmental aetiology, and inference aetiology. The distinction developed throughout is descriptive rather than normative, aiming to characterise both the functional similarities and the aetiological differences between these systems, particularly in terms of their formation and deployment.

The proposed framework reconciles opposing perspectives in the debate on natural versus artificial intelligence. On the one hand, it acknowledges functional similarity, understood as the capacity of both systems to perform input–output mappings mediated by internal representations. On the other hand, it highlights fundamental differences in developmental aetiology. Natural intelligence does not develop according to explicitly defined or formally specified objective functions. Instead, it emerges through endogenous processes shaped by evolutionary pressures, self-regulation, and the pursuit of adaptive equilibrium under energetic and environmental constraints. In contrast, artificial systems are developed under explicitly defined optimisation regimes, guided by objective functions or policies imposed externally.

A further distinction arises at the level of inference aetiology. While both natural and artificial systems can be activated by exogenous stimuli, only natural intelligence exhibits robust endogenous modes of activation, such as imagination, reflection, and introspection. Although analogous mechanisms are increasingly discussed within the AI literature, current artificial systems implement these processes within externally defined architectures and remain dependent on exogenous parameters, training procedures, or execution schedules. As a result, artificial systems do not presently demonstrate genuinely endogenous initiation mechanisms comparable to those observed in natural intelligence.

Taken together, these considerations suggest that natural and artificial intelligence are best understood as functionally comparable yet aetiologically distinct. This distinction provides a principled basis for analysing their similarities and differences, while avoiding both reductive equivalence and unwarranted claims of incomparability. More broadly, the framework developed here contributes to a more nuanced understanding of intelligence across biological and artificial domains, and offers a foundation for future philosophical and empirical investigations.

AI use statement

In this work, we used generative AI tools for language correction, identify relevant literature, and provide feedback on research methodology. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

Nothing to declare.

Reproducibility statement

The research presented in this paper does not come with the code.

Author Contributions

This section is anonymized.

Acknowledgments

This section is anonymized.

References

AAAI 2025 presidential panel on the future of AI research. Technical report, Association for the Advancement of Artificial Intelligence (AAAI), March 2025. URL https://aaai.org/wp-content/uploads/2025/03/AAAI-2025-PresPanel-Report-Digital-3.7.25.pdf. Presidential Panel Report.

Francesco Abbate. Natural and artificial intelligence: a comparative analysis of cognitive aspects. *Minds and Machines*, 33(4):791–815, 2023.

Anil Ananthaswamy. How close is ai to human-level intelligence. *Nature*, 636(8041):22–25, 2024.

Matteo Baldoni, Cristina Baroglio, Federico Capuzzimati, and Roberto Micalizio. Empowering agent coordination with social engagement. In *Congress of the Italian Association for Artificial Intelligence*, pp. 89–101. Springer, 2015.

Scott L Brincat, Markus Siegel, Constantin von Nicolai, and Earl K Miller. Gradual progression from sensory to task-related processing in cerebral cortex. *Proceedings of the National Academy of Sciences*, 115(30):E7202–E7211, 2018.

Rodney A Brooks. Elephants don't play chess. *Robotics and autonomous systems*, 6(1-2):3–15, 1990.

Zoe Callis, Paul Gerrans, Dana L Walker, and Gilles E Gignac. The association between intelligence and financial literacy: A conceptual and meta-analytic review. *Intelligence*, 100:101781, 2023.

Eddy Keming Chen, Mikhail Belkin, Leon Bergen, and David Danks. Does ai already have human-level intelligence? the evidence is clear. *Nature*, 650(8100):36–40, 2026.

Wei Chen and Keith Decker. Developing alternative mechanisms for multiagent coordination. In *Pacific Rim International Workshop on Multi-Agents*, pp. 63–76. Springer, 2002.

Wayne D Christensen and Clifford Alan Hooker. An interactivist-constructivist approach to intelligence: self-directed anticipative learning. *Philosophical Psychology*, 13(1):5–45, 2000.

Mehdi Dastani, M Birna Van Riemsdijk, Joris Hulstijn, Frank Dignum, and John-Jules Ch Meyer. Enacting and deacting roles in agent programming. In *International Workshop on Agent-Oriented Software Engineering*, pp. 189–204. Springer, 2004.

Javier de la Fuente, Gail Davies, Andrew D Grotzinger, Elliot M Tucker-Drob, and Ian J Deary. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nature human behaviour*, 5(1):49–58, 2021.

Narjes Soltani Dehaghani and Mojtaba Zarei. Pre-stimulus activities affect subsequent visual processing: Empirical evidence and potential neural mechanisms. *Brain and Behavior*, 15(2):e3654, 2025.

Theodosius Dobzhansky. *Mankind evolving*. JSTOR, 1962.

Robin Dunbar, Clive Gamble, and John Gowlett. *Thinking big: How the evolution of social life shaped the human mind*. Thames & Hudson, 2014.

JA Easley Jr. The development of thought: Equilibration of cognitive structures, 1978.

Wei Fang, Yanqi Chen, Jianhao Ding, Zhaofei Yu, Timothée Masquelier, Ding Chen, Liwei Huang, Huihui Zhou, Guoqi Li, and Yonghong Tian. Spikingjelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40):eadi1480, 2023.

Nanyi Fei, Zhiwu Lu, Yizhao Gao, Guoxing Yang, Yuqi Huo, Jingyuan Wen, Haoyu Lu, Ruihua Song, Xin Gao, Tao Xiang, et al. Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, 13(1):3094, 2022.

Ragnar Fjelland. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(1):10, 2020.

Jason P Gallivan, Craig S Chapman, Daniel M Wolpert, and J Randall Flanagan. Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, 19(9):519–534, 2018.

M Ida Gobbini, Ellen Leibenluft, Neil Santiago, and James V Haxby. Social and emotional attachment in the neural representation of faces. *Neuroimage*, 22(4):1628–1635, 2004.

Kurusetti Vinay Gupta, Jyotiranjan Beuria, Lokeswara Kumar Vijanapalli, Amit Sethi, and Laxmidhar Behera. Self-reflection, sense of agency, and underlying neural correlates: A pilot study. *PloS one*, 20(12):e0335276, 2025.

Yuval Noah Harari. *21 Lessons for the 21st Century*. Random House, 2019.

Jianlong Hong, Yukun Xiao, Yuqi Chen, Shengshun Duan, Shengxin Xiang, Xiao Wei, Huiyun Zhang, Lei Liu, Jun Xia, Wei Lei, et al. Biomimetic hairy affective-touch sensory ai interface. *Nature Communications*, 2026.

IBM Topics. What is artificial intelligence (ai)? types of ai. `https://www.ibm.com/think/topics/artificial-intelligence-types`, 2024. Accessed: 2026-05-28.

Anthony I. Jack and Andreas Roepstorff. Introspection and cognitive brain mapping: from stimulus–response to script–report. *Trends in Cognitive Sciences*, 6(8):333–339, 2002. doi: 10.1016/s1364-6613(02)01941-1. URL `https://www.sciencedirect.com/science/article/abs/pii/S1364661302019411`.

Kiyoto Kasai, Masato Fukuda, Noriaki Yahata, Kentaro Morita, and Naotaka Fujii. The future of real-world neuroscience: imaging techniques to assess active brains in social environments. *Neuroscience Research*, 90:65–71, 2015.

Andrei Khilkevich, Michael Lohse, Ryan Low, Ivana Orsolic, Tadej Bozic, Paige Windmill, and Thomas D Mrsic-Flogel. Brain-wide dynamics linking sensation to action during decision-making. *Nature*, 634(8035):890–900, 2024.

Nikolaus Kriegeskorte and Pamela K Douglas. Cognitive computational neuroscience. *Nature neuroscience*, 21(9):1148–1160, 2018.

Kevin Laland and Amanda Seed. Understanding human cognitive uniqueness. *Annual Review of Psychology*, 72(1):689–716, 2021.

Janet Levin. Functionalism. In Edward N. Zalta and Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2023 edition, 2023. URL `https://plato.stanford.edu/entries/functionalism/`. First published Aug 24, 2004; substantive revision Apr 4, 2023.

Belle Liang, Timothy Klein, Zoë Ting, and Gabe Avakian Orona. Reflective intelligence: A neuroscience-informed framework for purpose and belonging. *Trends in Neuroscience and Education*, pp. 100292, 2026. ISSN 2211-9493. doi: https://doi.org/10.1016/j.tine.2026.100292. URL `https://www.sciencedirect.com/science/article/pii/S2211949326000177`.

Mingyue Luo, Hui Li, Wanbo Luo, Hewei Li, and Jianan Li. Goal-oriented autonomous decision-making for social robots via collaborative interactive inverse reinforcement learning approach. *Scientific Reports*, 15(1):27724, 2025.

Mackenzie Weygandt Mathis, Adriana Perez Rotondo, Edward F Chang, Andreas S Tolias, and Alexander Mathis. Decoding the brain: From neural representations to mechanistic models. *Cell*, 187(21):5814–5832, 2024.

Hans Moravec. *Mind children: The future of robot and human intelligence*. Harvard University Press, 1988.

José M Muñoz, Javier Bernacer, Alva Noë, and Evan Thompson. Why ai will never be able to acquire human-level intelligence. *Nature*, 637(8047):794–794, 2025.

Lorina Naci, Amelie Haugg, Alex MacDonald, Mimma Anello, Evan Houldin, Shakib Naqshbandi, Laura E Gonzalez-Lara, Miguel Arango, Christopher Harle, Rhodri Cusack, et al. Functional diversity of brain networks supports consciousness and verbal intelligence. *Scientific reports*, 8 (1):13259, 2018.

Ume Nisa, Muhammad Faizan Shirazi, Mohamed Ali Saip, and Muhammad Syafiq Mohd Pozi. Agentic AI: The age of reasoning—a review. *Journal of Automation and Intelligence*, 5(1): 69–89, 2026. doi: 10.1016/j.jai.2025.08.003. URL https://www.sciencedirect.com/science/article/pii/S2949855425000516.

Adriana Placani. Anthropomorphism in ai: hype and fallacy. *AI and Ethics*, 4(3):691–698, 2024.

David Premack. Is language the key to human intelligence? *Science*, 303(5656):318–320, 2004.

Ruilin Qin, Changle Zhou, and Mengjie He. A comprehensive taxonomy of machine consciousness. *Information Fusion*, 113:102994, 2025. ISSN 1566-2535. doi: 10.1016/j.inffus.2025.102994.

Philippe Rochat. The evolution of developmental theories since piaget: A metaview. *Perspectives on Psychological Science*, 19(6):921–930, 2024.

Majid Saberi, Jenny R Rieck, Shamim Golafshan, Cheryl L Grady, Bratislav Misic, Benjamin T Dunkley, and Ali Khatibi. The brain selectively allocates energy to functional brain networks under cognitive control. *Scientific Reports*, 14(1):32032, 2024.

Parisa Salmani and Peter R. Lewis. A reflective architecture for llm-based systems. In *2025 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, pp. 61–68. IEEE, 2025. doi: 10.1109/ACSOS-C66519.2025.00029. URL https://doi.org/10.1109/ACSOS-C66519.2025.00029.

W Joel Schneider and Kevin S McGrew. The cattell-horn-carroll model of intelligence. 2012.

W Joel Schneider and Kevin S McGrew. The cattell-horn-carroll theory of cognitive abilities. *Contemporary intellectual assessment: Theories, tests, and issues*, 733:163, 2018.

Eric Schwitzgebel. Introspection, what? *New Waves in Philosophy of Mind*, pp. 29–43, 2012.

Jordi Serra-Simón, Mònica Puntí-Brun, Sílvia Espinosa-Mirabet, Maria Alice de Faria Nogueira, Ramón Martín-Guart, and Sandro Tôrres de Azevedo. Generative artificial intelligence in advertising. field applications in rio de janeiro and catalonia. *Telecommunications Policy*, pp. 103033, 2025.

Tuo Shi, Lili Gao, Ruixi Zhou, Yang Tian, Pei Chen, Yanting Ding, Shuangzhu Tang, Huiqin Ma, Jian Lu, Hui Zhang, et al. Fully memristive spiking neural network for energy-efficient graph learning. *Science Advances*, 11(19):eadv2312, 2025.

JM Shine and Richard Shine. Delegation to automaticity: the driving force for cognitive evolution? *Frontiers in Neuroscience*, 8:90, 2014.

Ben Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6):495–504, 2020.

Dawn Sivan, K Satheesh Kumar, Veena Raj, and Rajan Jose. 7 reinforcement learning from human feedback (rlhf). *Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks*, pp. 135, 2024.

Charles Edward Spearman. *The Abilities of Man: Their Nature and Measurement*. Macmillan, New York, 1927.

PN Steinmetz, SS Hsiao, KO Johnson, and E Niebur. Attention to tactile stimuli increases neural synchrony in somatosensory cortex. In *Neurobiology of Attention*, pp. 534–537. Elsevier, 2005.

Cecilia UD Stenfors, Stephen C Van Hedger, Kathryn E Schertz, Francisco AC Meyer, Karen EL Smith, Greg J Norman, Stefan C Bourrier, James T Enns, Omid Kardan, John Jonides, et al. Positive effects of nature on cognitive performance across multiple experiments: Test order but not affect modulates the cognitive effects. *Frontiers in psychology*, 10:438004, 2019.

Russell Stuart and Norvig Peter. Artificial intelligence a modern approach third edition, 2010.

Shiliang Sun, Zehui Cao, Han Zhu, and Jing Zhao. A survey of optimization methods from a machine learning perspective. *IEEE transactions on cybernetics*, 50(8):3668–3681, 2019.

Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Dana L Walker, Romina Palermo, Zoe Callis, and Gilles E Gignac. The association between intelligence and face processing abilities: A conceptual and meta-analytic review. *Intelligence*, 96: 101718, 2023.

Xiangwen Wang, Xianghong Lin, and Xiaochao Dang. Supervised learning in spiking neural networks: A review of algorithms and evaluations. *Neural Networks*, 125:258–280, 2020.

Russell T Warne and Cassidy Burningham. Spearman'sg found in 31 non-western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin*, 145(3):237, 2019.

Ramsey R Wilcox, Babak Hemmatian, Lav R Varshney, and Aron K Barbey. The network architecture of general intelligence in the human connectome. *Nature Communications*, 2026.

Lingyun Xie, Kui Shan, and Shengwei Wang. A generic framework and strategies for integrating ai into building automation systems for field-level optimization of hvac systems. *Energy*, 333: 137405, 2025.