AI and Brain
https://www.sciltp.com/journals/aib



Editorial

Understanding Intelligence across Biological and Artificial Systems

Tommaso Dorigo

Istituto Nazionale di Fisica Nucleare, Sezione di Padova, 35131 Padova, Italy; tommaso.dorigo@gmail.com
Received: 19 October 2026; Accepted: 27 October 2026; Published: 6 November 2026

How To Cite: Dorigo, T. Understanding Intelligence across Biological and Artificial Systems. *AI and Brain* **2026**, *1*(1), 1.

1. Intelligence in Nature and Artefacts: Questions and Routes

What is intelligence, and how can it arise in a physical system? Intelligence has been associated with a variety of capacities: learning, adaptation, perception, memory, communication, empathy, problem solving, planning, and the ability to act successfully in changing environments. No single definition captures all of these aspects, nor is intelligence an exclusively human phenomenon: many of these capacities are found, to different degrees and in different forms, throughout the natural world. The observed diversity of systems capable of intelligent behaviour also raises the issue of which physical organizations and computational principles are sufficient to produce it. The scientific investigation of intelligence has therefore followed several routes in parallel: understanding the human mind and brain, comparing cognition across species, and, more recently, attempting to reproduce intelligent behaviour in artificial systems.

Studies in comparative cognition have identified sophisticated behaviour in species with very different evolutionary histories and nervous systems. Complex problem solving, use of specialized tools, memory, planning, and social cognition have, for example, been documented in both primates and corvids, despite differences in their neural architectures [1]: Nature appears to have conducted more than one experiment in the construction of intelligent behaviour. This invites a deeper question: which properties of intelligent systems are essential, and which instead reflect the particular biological implementation through which they evolved?

Long before artificial intelligence (AI) emerged as a discipline, intelligence was studied from several different viewpoints. Psychology studied learning, memory, perception, reasoning, and behaviour; comparative and behavioural sciences examined how such capacities arise and vary across species and environments; and neuroscience sought the biological mechanisms that support them. These approaches differed greatly in methods and levels of explanation, but they largely shared a common starting point: intelligent behaviour already existed in a natural system, and the task was to understand how it arose. The emergence of AI in the second half of the twentieth century added a qualitatively different route, introducing the possibility of building it from scratch. Although the early developments of AI were closely connected with ideas about biological computation, the new discipline soon acquired its own methods and objectives, and for much of its history it developed partly independently of the biological sciences. AI may provide a synthetic branch of the broader search for the principles of intelligence: instead of beginning with an existing brain and working backward toward its mechanisms, one can begin with computational mechanisms and ask whether intelligent behaviour can emerge from them.

Today these different routes are converging. AI provides powerful tools for analysing neural and behavioural data and offers computational models with which to investigate biological representations and processes. In the opposite direction, neuroscience inspires new approaches to learning, memory, adaptation, perception, and efficient computation. In addition, brain-computer interfaces now allow biological and artificial systems to interact directly, and the complexity of modern AI has made artificial networks themselves objects of scientific investigation. It is this growing scientific exchange that motivates the creation of *AI and Brain*.

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2. Constructing Intelligence: From Logical Neurons to Foundation Models

Early attempts to model computation and learning in the brain helped lay the foundations of artificial intelligence. In 1943, McCulloch and Pitts showed that simplified neurons could be treated as logical units capable, in principle, of performing computation [2]. A few years later, Hebb proposed that learning could result from changes in the strength of neuronal connections, an idea that would profoundly influence both neuroscience and artificial neural networks [3]. Turing then reframed machine intelligence in operational terms, asking under what circumstances a machine could display behaviour indistinguishable from that of a human interlocutor. Turing is also credited with devoting substantial attention to the possibility of constructing machines that could learn independently rather than having every aspect of their functioning directly programmed [4]. These early contributions already contained themes that remain central today: distributed computation, adaptation through experience, and the possibility that complex behaviour may emerge from relatively simple mechanisms.

The Dartmouth proposal of 1955, and the workshop held the following year, transformed the above ideas into an explicit research programme encompassing learning, language, abstraction, problem solving, and self-improvement [5]. Different approaches soon emerged: one emphasized symbolic representations, logical reasoning, and search; another sought inspiration in networks of interacting units, building on Rosenblatt's perceptron [6]. The subsequent history is sometimes narrated as a competition in which symbolic artificial intelligence eventually gave way to neural networks; such a description is oversimplified. In fact, search, planning, probabilistic inference, optimization, and symbolic manipulation remain important components of contemporary systems. What did change dramatically was the ability of machines to learn useful internal representations from data.

This shift became particularly visible with the revival of connectionist approaches in the 1980s, when Hopfield networks showed how collective dynamics among simple units could implement associative memory, linking neural computation to concepts from dynamical systems and statistical physics [7]. Shortly afterward, backpropagation was recognized as a general method for adjusting internal parameters in multilayer networks, reducing errors at their outputs [8]. Hidden units could acquire representations that had not been specified by the programmer but emerged through training because they were useful for solving the task at hand.

A parallel intellectual lineage concerned learning through interaction with the environment. Long before reinforcement learning became a computational framework, experimental psychology had investigated how behaviour is shaped by experience, association, reward, and expectation, from the work of Thorndike and Pavlov through Skinner and later formal models such as Rescorla and Wagner's account of associative learning [9]. These studies provided part of the conceptual background from which modern reinforcement-learning theories developed, linking behavioural adaptation to the consequences of actions and the prediction of future events.

The practical impact of neural networks and reinforcement learning expanded dramatically with the availability of large datasets, greater computing power, improved optimization strategies, and deeper architectures. The success of AlexNet in the 2012 ImageNet competition became a visible marker of this transition [10]; deep learning thereafter rapidly spread across vision, speech, language, and other domains [11]. Deep reinforcement learning further combined learned perception with computational models of learning through interaction [12]. AlphaGo and its successor AlphaZero demonstrated how neural networks, reinforcement learning, and search algorithms could be integrated to achieve superhuman performance in Go, chess, and shogi [13,14].

The latest major transition came with the introduction of the transformer architecture [15]. Combined with self-supervised learning, growing datasets, and increasingly powerful computational resources, transformers enabled the development of models with billions of trainable parameters, and eventually more than a trillion in some architectures. They form the architectural basis of today's large language models, which display broad linguistic capabilities and can perform a wide range of tasks with little or no task-specific training, including few-shot and in-context learning [16]. These developments contribute to the broader concept of foundation models: systems trained on diverse data whose representations can support many downstream tasks [17]. The same paradigm now extends beyond language to multimodal systems integrating text, images, audio, video, and action, further blurring the distinction between narrowly specialized models and more general-purpose systems.

This progression from hand-crafted to large-scale learned representations has also made modern AI systems objects of scientific study in their own right. The architectures, training objectives, and data of modern AI systems are chosen by their designers, but their detailed internal organization develops through optimization and cannot simply be read from the code. This makes it interesting to ask whether different systems arrive at similar computational solutions, and whether any of these resemble those found in biological brains.

The investigation of internal representations is only one aspect of contemporary AI research. Work on multimodal learning, generative models, autonomous agents, and continual learning also raises questions about how artificial systems acquire knowledge, adapt, and act. AI research can therefore contribute to a broader science of

intelligence, and produce concepts and tools whose relevance to biological cognition may become apparent only later.

3. Understanding Natural Intelligence: From Cells to Systems

The biological investigation of intelligence took an almost opposite route to the one followed in artificial intelligence: neuroscience began with a functioning nervous system and worked backward toward the principles underlying perception, cognition, and behaviour.

The neuron doctrine, established through the pioneering work of Ramón y Cajal in the late nineteenth century, identified the neuron as the basic structural unit of the nervous system. This provided a foundation for progressively linking cellular activity to circuits, representations, and computation. Subsequent developments in neurophysiology extended this anatomical understanding toward quantitative descriptions of neuronal function. Hodgkin and Huxley showed that the electrical behaviour of the neuronal membrane could be described through a mathematical model of ionic conductances [18]. At a larger scale, Hubel and Wiesel demonstrated that neurons in the visual cortex respond selectively to particular features and are organized across successive stages of processing [19]. Such developments encouraged the view that perception and behaviour arise from transformations distributed across interacting neural populations, not from isolated cells.

The focus of explanatory attempts consequently shifted from individual neurons toward circuits, populations, and computations. An influential framework developed by David Marr distinguished among the computational problem being solved, the algorithms and representations used to solve it, and the physical mechanisms in which these processes are implemented [20]. This distinction remains relevant today because similar computational goals may be achieved through different algorithms and physical substrates. Work on neural populations has also shown that information can be represented through distributed activity, while the temporal evolution of neural populations may itself carry computational structure. Studies of motor cortex, for example, have described neural activity during movement in terms of trajectories through high-dimensional state spaces, rather than simple one-to-one mappings between neurons and behavioural variables [21].

Recent technological advances have also brought a dramatic expansion of our ability to observe the working brain, through EEG and MEG, high-density electrophysiology, structural and functional neuroimaging, calcium imaging, connectomics, and increasingly rich combinations of behavioural, genetic, physiological, and clinical data. Neuroscience has moved from relatively sparse and localized measurements toward high-dimensional, multimodal datasets spanning multiple scales. Researchers now seek not only to collect these data, but to identify the representations and dynamics they reveal. The study of individual neurons has thus grown into an investigation of neural systems in which experimental neuroscience, mathematical modelling, and artificial intelligence increasingly work together.

4. What Kind of Computation Does a Brain Perform?

Our increasing ability to observe neural activity has not so far produced a consensus on a computational theory of the brain. Contemporary neuroscience instead offers partly complementary and partly competing principles for explaining how neural systems transform sensory information into perception, memory, decisions, and action. Efficient-coding theories, for example, emphasize adaptation to the statistical structure of the environment, while hierarchical accounts describe successive transformations of information; population-based approaches focus on the collective dynamics of interacting neurons. These views need not be mutually exclusive. We need to understand which mechanisms operate at the different spatial and temporal scales involved in cognition, and how they interact.

One influential family of theories treats perception as a process of inference. Predictive-coding models propose that higher levels of a neural hierarchy generate predictions about activity at lower levels, while mismatches between these predictions and incoming information drive updating [22]. Bayesian formulations similarly describe perception as inference under uncertainty, while active-inference and free-energy approaches extend this logic to action [23]. The broad explanatory scope of these attempts is attractive, but also creates a methodological challenge: competing theories must ultimately generate predictions that allow them to be distinguished experimentally. Determining how prediction and uncertainty are represented in neural systems, and under which conditions predictive, feed-forward, or other mechanisms dominate, remains an active area in which AI can help implement and compare competing computational hypotheses.

The study of how organisms learn from experience and use that information to guide action provides another connection between computational theory and neuroscience. Reinforcement-learning models formalize how behaviour can improve through rewards and prediction errors; dopaminergic activity has been related to reward-prediction signals resembling those produced by temporal-difference algorithms [24]. Yet biological learning is considerably richer than any single reinforcement-learning formulation: organisms learn through observation, remembered experience, social

interaction, and internally generated goals, not only from externally supplied rewards.

The learning processes described above operate through neural circuits whose activity changes from moment to moment. Dynamical-systems approaches study how activity evolves across neuronal populations, treating these collective patterns as part of the computation itself. Learning also alters the circuits that generate those patterns, raising the problem of how the brain can adapt while preserving previously acquired abilities.

Many forms of intelligent behaviour require an organism to act on information that is no longer present in its immediate sensory input. Intelligent organisms maintain goals, manipulate memories, suppress habitual responses, and evaluate alternative actions. The prefrontal cortex has long been associated with such cognitive control; in an influential account, Miller and Cohen proposed that prefrontal activity maintains goal-related representations that bias processing elsewhere in the brain [25]. How goals are represented, updated, and transformed into action remains an open question. Modern generative models and AI agents provide an interesting point of comparison, as they can use contextual information to generate possible actions and plans. Whether the mechanisms through which artificial systems organize goal-directed behaviour have anything in common with those of the prefrontal cortex is an issue worth investigating.

Another difficult unresolved problem in brain science concerns the relationship between intelligence and consciousness. Although consciousness is closely associated with many cognitive functions in humans, intelligent behaviour does not by itself establish the presence of subjective experience. Despite substantial theoretical and experimental effort, there is still no consensus on the neural basis of consciousness or on how competing accounts should be decisively distinguished experimentally [26]. A large study published in 2025 directly tested predictions from two leading theories of consciousness and challenged important predictions of both, illustrating how much remains unresolved [27]. Artificial intelligence adds a new dimension to this debate: the most capable systems can display sophisticated performance in perception, language, reasoning, and planning, yet such performance alone does not establish whether subjective experience is present. Artificial systems thus offer a new perspective on an old question: which components of intelligence actually require consciousness, and which do not?

5. AI as an Instrument for Understanding the Brain

The most immediate contribution of today's artificial intelligence to neuroscience is its ability to extract structure from complex data. Machine-learning methods are now applied across electrophysiology, EEG and MEG, neuroimaging, behavioural measurements, and combinations of these modalities, supporting tasks ranging from classification and prediction to the identification of latent neural and behavioural states. Recent approaches model brain activity and behaviour jointly, using discriminative, generative, and contrastive methods to uncover relationships that may be difficult to identify from either domain alone [28].

Artificial networks can also serve as candidate models of biological computation. A model trained to solve a behaviourally relevant task may develop intermediate representations that were not specified in advance; these can then be compared with neural recordings. Representational similarity analysis provides a quantitative framework for comparing representations in computational models with those measured in behaviour and the brain [29]. Goal-driven deep networks optimized for sensory tasks have, for example, developed internal representations predictive of responses in the visual cortex [30]. Yet models that fit the same experimental data may rely on different computational mechanisms. Distinguishing them requires experiments or perturbations designed to elicit different predictions.

Foundation models seek to learn reusable representations from large and heterogeneous neural datasets, with the aim of transferring them across experiments, individuals, or tasks. A recent foundation model trained on visual-cortical activity from a sample of mice was shown to generalize to new animals and previously unseen stimulus domains, while also predicting aspects of neuronal anatomy and connectivity [31]. Results like this one suggest the possibility of developing models that integrate information across experiments and perhaps eventually across recording modalities, behavioural contexts, or species. However, variability between subjects, limitations and heterogeneity of available datasets, and difficulties of interpretation remain substantial obstacles. The hope is that, in addition to offering better predictions of brain activity, larger and more complex neural models may reveal transferable principles of neural computation.

The relevance of AI-based modelling extends beyond basic neuroscience to medical applications, where predicting a patient's outcome is not the same as identifying an intervention that could change it. Computational psychiatry combines data-driven approaches with mechanistic models to investigate clinical heterogeneity, disease progression, and treatment response [32]. Digital-twin approaches pursue a related ambition at the level of individual patients: constructing computational models that might predict how a patient's physiological or neural state would evolve under different conditions or treatments. Before such models can reliably guide clinical decisions, their predictions, and especially their estimates of treatment effects, must be validated in independent and prospective studies.

6. Brain–Computer Interfaces: Closing the Loop

Brain–computer interfaces (BCIs) provide a particularly concrete example of how neural activity can be coupled to artificial computational systems. Early BCIs focused largely on decoding motor intentions to control cursors or external devices. More recent systems have moved toward increasingly complex forms of communication and action. For example, in 2021 an intracortical interface decoded attempted handwriting from motor-cortex activity and translated it into text at 90 characters per minute [33]. Only two years later, speech neuroprostheses demonstrated rapid decoding of attempted speech from cortical activity into large-vocabulary text; in another effort cortical surface recordings were successfully used to generate text, synthesized speech, and facial-avatar movements [34,35]. These advances illustrate a transition from simple command decoding toward restoring richer and more natural forms of interaction.

When a person learns to control a BCI, the decoder may be adapting at the same time. The system translates neural signals into actions or communication in real time, and the feedback received by the user can influence subsequent neural activity. Some interfaces also deliver sensory information directly through neural stimulation. The two forms of learning can interfere: a decoder that changes its mapping too quickly may undermine the user's efforts to master it. Understanding how these adaptations interact is important both for designing reliable interfaces and for studying neural plasticity and human–machine co-learning.

A decoder that learns alongside its user must also remain reliable when neural signals change over days, months, or years. Implanted devices add requirements of long-term stability and safety, alongside the technical demands of accurate decoding. Privacy and consent are particularly sensitive when neural recordings contain information about intended movements or attempted communication. Adaptation also raises questions of agency: if the decoder changes how it interprets a user's signals, the user must retain meaningful control over the resulting actions. I believe that this combination of scientific opportunities and practical challenges makes BCI developments a particularly interesting subject for our new journal.

7. From Brains to AI and Back

Biological intelligence also offers examples of capabilities that remain difficult to reproduce artificially. Brains learn continually, adapt rapidly to changing environments, integrate multiple sensory streams, operate robustly in the presence of noise, and accomplish all of this under severe energy and computational constraints. They also learn through active interaction with the world. NeuroAI investigates the relationships between neural and artificial intelligence, including the possibility of using principles of biological computation to improve artificial systems. One proposed benchmark is the "embodied Turing test," which shifts attention from language and game-playing toward the flexible sensorimotor abilities shared by animals, asking whether artificial agents can interact with the physical world as effectively as their biological counterparts [36]. The practical value of borrowing from biology must ultimately be demonstrated through measurable improvements in artificial systems.

Neuromorphic computing makes this search for alternative principles concrete at the level of hardware. Neuromorphic systems explore some of the distinguishing features of the biological brain, such as event-driven computation, spiking communication, distributed memory and processing, asynchronous operation, and local adaptation. These systems may become particularly valuable where conventional architectures are constrained by power consumption, latency, or the need for continuous operation at the edge. However, realizing their potential requires more than new devices. Algorithms, learning rules, sensors, hardware, and applications must be developed together, making full-stack co-design an important direction for the field [37]. Major open questions concern how neuromorphic systems should learn, how robustness can be maintained in the presence of device variability, which applications genuinely benefit from spiking or event-driven computation, and how such systems should be compared fairly with conventional hardware.

A different approach uses living neural tissue itself as part of the computational substrate. Biological computing includes organoid intelligence, in which three-dimensional neural organoids are interfaced with electronic systems, as well as platforms based on cultured neuronal networks coupled to multielectrode arrays. Such systems can receive structured inputs, generate measurable outputs, and adapt through closed-loop feedback. Brain-organoid reservoir computing has been explored for tasks such as speech recognition and nonlinear prediction [38], while cultured neuronal networks have been shown to change their behaviour during interaction with a simulated environment [39]. This reverses the approach of neuromorphic engineering: instead of reproducing selected principles of neural computation in artificial hardware, biological computing incorporates living neural networks into engineered computational systems.

The parallels between biological and artificial systems extend to the methods we use to investigate them. Artificial systems can be studied at the level of behaviour, independently of assumptions about biological similarity or internal mechanisms. Systematic characterization of their capabilities, errors, strategies, adaptation, and interactions can reveal regularities that themselves require explanation. Their internal organization offers

another level of investigation: researchers probe activations, remove or perturb components, compare internal representations, and attempt to identify circuits responsible for particular behaviours. Early approaches such as Network Dissection quantified associations between individual hidden units and interpretable visual concepts [40]. More recent interpretability research seeks systematic ways to connect internal components to model behaviour. The methodological resemblance to experimental neuroscience is striking, and raises the prospect that tools may migrate in both directions: methods developed to understand neural populations may help interrogate artificial networks; meanwhile, the much greater observability and manipulability of artificial systems may provide new test beds for theories of representation, learning, and distributed computation.

The methodological resemblance between neuroscience and AI interpretability should not be mistaken for biological equivalence. Similar behaviour, and even similar internal representations, may arise from very different mechanisms; the distinction between reproducing observations and explaining them remains contested even within neuroscience [41]. The value of comparing brains and artificial networks lies in testing specific computational principles across different architectures, physical substrates, and experimental conditions. Where the comparison breaks down may be as revealing as where it succeeds.

8. A New Cognitive Ecology and a New Journal: *AI and Brain*

Artificial systems are becoming part of the environment in which human cognition itself takes place: they assist decisions, retrieve and organize information, support learning and creativity, and participate in communication and problem solving. How does reliance on these systems alter attention, memory, reasoning, or judgement? When do humans and machines complement one another, and when does automation encourage dependency or misplaced trust? The consequences may also extend beyond cognition, affecting social relationships and mental health. For example, in recent experiments on human–AI feedback loops, Glickman and Sharot [42] found that biases present in human judgements could be amplified by AI systems and subsequently reinforced in the people interacting with them.

Scientific research is also becoming part of this changing cognitive environment. AI systems can generate predictions, hypotheses, and candidate solutions without necessarily making the underlying phenomena easier to understand. Messeri and Crockett [43] have called attention to the resulting 'illusion of understanding'. A related challenge has recently emerged in mathematics. in October 2026 OpenAI released more than 700 AI-generated mathematical preprints, many addressing open problems [44]. The speed at which these mathematical claims were produced creates difficulties for independent verification, attribution, and scientific interpretation, concerns already articulated a few months earlier in the Leiden Declaration [45]. Formal verification can establish the correctness of suitably formalized arguments, but does not by itself settle questions of originality, significance, or scientific understanding.

The longer-term consequences of AI are much harder to anticipate. Artificial systems may profoundly reshape education, the work environment, scientific research, political institutions, and social relationships, with consequences that will not necessarily be uniformly beneficial or harmful. Some possibilities are still more radical. Progress in brain–computer interfaces may eventually lead beyond the restoration of lost functions toward closer forms of integration between human cognition and artificial systems. And artificial general intelligence may be closer than we expect, although its definition, feasibility, and timescale remain debated. A further possibility is that the development of more capable artificial systems may help us investigate whether consciousness is necessarily a biological phenomenon. These prospects differ greatly in their scientific maturity, but they deserve serious consideration, even where experimental evidence is still limited and predictions remain speculative.

AI and Brain is intended to provide a forum for research across these disciplinary boundaries. Its scope includes computational neuroscience and psychiatry, behavioural science, comparative cognition, NeuroAI, cognitive modelling, brain foundation models, imaging genetics, digital twins, brain–computer interfaces, neuromorphic and biological computing, mechanistic interpretability, and human–AI interaction. The journal encourages research that connects different levels of description, develops computational methods applicable across fields, or devises experiments capable of distinguishing competing explanations of intelligent behaviour.

It is my hope that *AI and Brain* will become the default meeting place for researchers studying intelligence from different starting points. The journal will welcome contributions that directly investigate the relationship between biological and artificial systems, as well as research that advances AI or deepens our understanding of cognition on its own terms. In addition, developments in language and multimodal models, learning and reasoning, interpretability, evaluation and benchmarking, and trustworthy AI will be valuable additions to the journal even when their relevance to biological cognition is not immediately apparent. The converse applies to discoveries in neuroscience and cognitive science whose implications for artificial systems may emerge only later. I would also like the journal to welcome scientifically grounded speculation about the future of artificial and biological

intelligence, including their possible integration, their effects on human societies, and questions about the nature of consciousness. Such contributions need not offer definitive answers, but they should distinguish established findings from hypotheses and predictions, and remain open to critical scrutiny.

Funding

This research received no external funding.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The author declares no conflict of interest. Given his role as Editor in Chief, Tommaso Dorigo had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the author used CGPT5.6Sol for literature searches, consistency checks, and proofreading. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

1. Emery, N.J.; Clayton, N.S. The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science* **2004**, *306*, 1903–1907. https://doi.org/10.1126/science.1098410.
2. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133.
3. Hebb, D.O. *The Organization of Behavior: A Neuropsychological Theory*; Wiley: New York, NY, USA, 1949.
4. Turing, A.M. Computing Machinery and Intelligence. *Mind* **1950**, *LIX*, 433–460. https://doi.org/10.1093/mind/lix.236.433.
5. McCarthy, J.; Minsky, M.L.; Rochester, N.; et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Proposal Dated 31 August 1955. Available online: https://www-formal.stanford.edu/jmc/history/dartmouth.pdf (accessed on 1 November 2026).
6. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. https://doi.org/10.1037/h0042519.
7. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities.. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. https://doi.org/10.1073/pnas.79.8.2554.
8. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. https://doi.org/10.1038/323533a0.
9. Rescorla, R.A.; Wagner, A.R. A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In *Classical Conditioning II: Current Research and Theory*; Appleton-Century-Crofts: New York, NY, USA, 1972; pp. 64–99.
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105.
11. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. https://doi.org/10.1038/nature14539.
12. Mnih, V.; Kavukcuoglu, K.; Silver, D.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. https://doi.org/10.1038/nature14236.
13. Silver, D.; Huang, A.; Maddison, C.J.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. https://doi.org/10.1038/nature16961.
14. Silver, D.; Hubert, T.; Schrittwieser, J.; et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **2018**, *362*, 1140–1144. https://doi.org/10.1126/science.aar6404.
15. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
16. Brown, T.B.; Mann, B.; Ryder, N.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33, pp. 1877–1901.
17. Bommasani, R.; Hudson, D.A.; Adeli, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2021**, arXiv:2108.07258.

18. Hodgkin, A.L.; Huxley, A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **1952**, *117*, 500–544. https://doi.org/10.1113/jphysiol.1952.sp004764.
19. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837.
20. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*; W. H. Freeman: San Francisco, CA, USA, 1982.
21. Churchland, M.M.; Cunningham, J.P.; Kaufman, M.T.; et al. Neural population dynamics during reaching. *Nature* **2012**, *487*, 51–56. https://doi.org/10.1038/nature11129.
22. Rao, R.P.N.; Ballard, D.H. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **1999**, *2*, 79–87. https://doi.org/10.1038/4580.
23. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138.
24. Schultz, W.; Dayan, P.; Montague, P.R. A Neural Substrate of Prediction and Reward. *Science* **1997**, *275*, 1593–1599. https://doi.org/10.1126/science.275.5306.1593.
25. Miller, E.K.; Cohen, J.D. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* **2001**, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.
26. Seth, A.K.; Bayne, T. Theories of consciousness. *Nat. Rev. Neurosci.* **2022**, *23*, 439–452. https://doi.org/10.1038/s41583-022-00587-4.
27. Ferrante, O.; Gorska-Klimowska, U.; Henin, S.; et al. Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* **2025**, *642*, 133–142. https://doi.org/10.1038/s41586-025-08888-1.
28. Mathis, M.W.; Mathis, A. Joint modelling of brain and behaviour dynamics with artificial intelligence. *Nat. Rev. Neurosci.* **2026**, *27*, 87–100. https://doi.org/10.1038/s41583-025-00996-1.
29. Kriegeskorte, N.; Mur, M.; Bandettini, P.A. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2008**, *2*, 249. https://doi.org/10.3389/neuro.06.004.2008.
30. Yamins, D.L.K.; DiCarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **2016**, *19*, 356–365. https://doi.org/10.1038/nn.4244.
31. Wang, E.Y.; Fahey, P.G.; Ding, Z.; et al. Foundation model of neural activity predicts response to new stimulus types. *Nature* **2025**, *640*, 470–477. https://doi.org/10.1038/s41586-025-08829-y.
32. Huys, Q.J.M.; Maia, T.V.; Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **2016**, *19*, 404–413. https://doi.org/10.1038/nn.4238.
33. Willett, F.R.; Avansino, D.T.; Hochberg, L.R.; et al. High-performance brain-to-text communication via handwriting. *Nature* **2021**, *593*, 249–254. https://doi.org/10.1038/s41586-021-03506-2.
34. Willett, F.R.; Kunz, E.M.; Fan, C.; et al. A high-performance speech neuroprosthesis. *Nature* **2023**, *620*, 1031–1036. https://doi.org/10.1038/s41586-023-06377-x.
35. Metzger, S.L.; Littlejohn, K.T.; Silva, A.B.; et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **2023**, *620*, 1037–1046. https://doi.org/10.1038/s41586-023-06443-4.
36. Zador, A.; Escola, S.; Richards, B.; et al. Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nat. Commun.* **2023**, *14*, 1597. https://doi.org/10.1038/s41467-023-37180-x.
37. Schuman, C.D.; Kulkarni, S.R.; Parsa, M.; et al. Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput. Sci.* **2022**, *2*, 10–19. https://doi.org/10.1038/s43588-021-00184-y.
38. Cai, H.; Ao, Z.; Tian, C.; et al. Brain organoid reservoir computing for artificial intelligence. *Nat. Electron.* **2023**, *6*, 1032–1039. https://doi.org/10.1038/s41928-023-01069-w.
39. Kagan, B.J.; Kitchen, A.C.; Tran, N.T.; et al. In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron* **2022**, *110*, 3952–3969.e8. https://doi.org/10.1016/j.neuron.2022.09.001.
40. Bau, D.; Zhou, B.; Khosla, A.; et al. Network Dissection: Quantifying Interpretability of Deep Visual Representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6541–6549. https://doi.org/10.1109/cvpr.2017.354.
41. Ross, L.N.; Bassett, D.S. Causation in neuroscience: Keeping mechanism meaningful. *Nat. Rev. Neurosci.* **2024**, *25*, 81–90. https://doi.org/10.1038/s41583-023-00778-7.
42. Glickman, M.; Sharot, T. How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nat. Hum. Behav.* **2024**, *9*, 345–359. https://doi.org/10.1038/s41562-024-02077-2.
43. Messeri, L.; Crockett, M.J. Artificial intelligence and illusions of understanding in scientific research. *Nature* **2024**, *627*, 49–58. https://doi.org/10.1038/s41586-024-07146-0.
44. Castelvecchi, D. OpenAI posts 700 maths preprints online: Mathematicians are up in arms. *Nature* **2026**. https://doi.org/10.1038/d41586-026-03196-8.
45. Alper, J.; Barany, M.J.; et al. Leiden Declaration on Artificial Intelligence and Mathematics. Available online: https://leidendeclaration.ai/ (accessed on 1 November 2026).