International Journal of Emerging Research in Computer Science
Volume 2 | Issue 4 | April 2026 | ISSN: 3139-101X (Online)
Available online at: https://ijercs.com/
DOI: https://doi.org/10.59828/ijercs.v2i4.28

How Machine Learning Systems Simulate the Human Brain Using Artificial Intelligence

Aditya Vijay Shinde
Department of Computer Science, Sarhad College of Arts, Commerce and Science, Pune

Received: 02 April 2026 | **Accepted:** 17 April 2026 | **Published:** 29 April 2026

Abstract

Artificial Intelligence (AI) has emerged as one of the most transformative technology of the twenty-first century. At its core, AI attempts to replicate certain cognitive capabilities of the human brain — the ability to learn, recognize patterns, make decisions, and adapt to new situations. This paper investigates how modern machine learning systems, particularly neural networks and deep learning architectures, are designed with direct inspiration from the biological structures and processes of the human brain. Topics covered include the architecture of biological neurons, how artificial neural networks (ANNs) model synaptic behavior, the role of deep learning in replicating cognitive functions, and the limitations and ethical considerations that arises from this simulation.

The study draws on published literature spanning neuroscience, cognitive science, and computer science. Findings suggest that while machine learning have achieved remarkable brain-like capabilities in specific tasks such as image recognition, language understanding, and decision-making, true general intelligence that matches the adaptability of human brain still remains an unsolved problem. The gap between biological intelligence and machine intelligence continues to narrow, driven by advances in neuromorphic computing, reinforcement learning, and transformer-based architectures.

Keywords: *Artificial Intelligence, Neural Networks, Deep Learning, Cognitive Simulation, Human Brain*

1. INTRODUCTION

The human brain is arguably the most sophisticated information processing system known to exist. Composed of approximately 86 billion neurons connected through trillions of synaptic junctions, it enables perception, reasoning, memory, emotion, and creativity — all simultaneously and with extraordinary energy efficiency. For decades, scientists and engineers have attempted to understand this biological marvel and use those insights to design machines that can think, learn, and behave intelligently.

The discipline of Artificial Intelligence (AI) was formally established in mid-twentieth century, with early researchers imagining that computers could be programmed to simulate human reasoning. However, it quickly became apparent that hard-coded logic rules were insufficient for capturing the richness of human cognition. The field then shifted it's attention toward learning systems — machines that could improve their performance through experience, much like the brain does through neuroplasticity and repeated exposure.

Machine learning (ML), a subfield of AI, provides the computational framework for this kind of learning. At its foundation lies the artificial neural network (ANN), a mathematical model inspired directly by the biological neuron. Over the past two decades, advances in computing power, data availability, and algorithmic innovation have transformed ANNs into deep neural networks capable of performing tasks — from diagnosing cancer to translating languages — that were once considered exclusively human capabilities.

This paper aims to trace the conceptual and architectural connections between the human brain and machine learning systems. It examines how biological principles have shaped the design of artificial intelligence, where the two converge and diverge, and what the future may hold as computational neuroscience and AI increasingly inform each other.

2. LITERATURE REVIEW

The foundational connection between the brain and computing was established by Warren McCulloch and Walter Pitts in 1943, who proposed a simple mathematical model of a neuron. They demonstrated that logical operations could be performed by networks of such artificial neurons, laying the groundwork for neural computation. Frank Rosenblatt extended this concept in 1958 with the Perceptron, a hardware device capable of learning simple pattern classification from labeled examples. Although Minsky and Papert (1969) exposed it's limitations — particularly its inability to solve nonlinearly separable problems — the Perceptron established the conceptual template for all future neural networks.

The resurgence of interest in neural networks came with the introduction of backpropagation algorithm by Rumelhart, Hinton, and Williams (1986). Backpropagation enabled multi-layer networks to learn complex input-output mappings by adjusting weights in proportion to the error signal, closely analogous to how the brain adjusts synaptic strengths through Hebbian learning. This development opened the door to practical applications in speech recognition, image classification, and financial forecasting throughout the 1990s.

The modern era of deep learning began with the work of Hinton et al. (2006), who showed that deep neural networks could be effectively trained using a technique called greedy layer-wise pre-training. This was followed by Krizhevsky, Sutskever, and Hinton (2012), whose AlexNet architecture achieved a dramatic reduction in image classification error on the ImageNet benchmark, signaling a paradigm shift in the field. Convolutional Neural Networks (CNNs), inspired by the hierarchical processing structure of the visual cortex (Hubel and Wiesel, 1962), became the dominant architecture for visual tasks.

Parallel developments in neuroscience further informed AI design. The discovery of place cells and grid cells in the hippocampus (O'Keefe and Dostrovsky, 1971; Moser et al., 2008) inspired spatial memory models in AI. The theory of predictive coding in the brain (Rao and Ballard, 1999) influenced the design of generative models. More recently, the Transformer architecture (Vaswani et al., 2017) and models such as GPT-4 and BERT (Devlin et al., 2018) have shown that attention mechanisms, loosely analogous to selective attention in the brain, are enormously effective for language tasks. These developments together form a rich, bidirectional relationship between neuroscience and AI research.

3. RESEARCH OBJECTIVES

The primary objective of this study is to systematically examine the structural and functional parallels between the human brain and modern machine learning systems. Specifically, the research investigates how biological neural mechanisms have been translated into mathematical and computational models across different generations of AI architectures.

A secondary objective is to identify where current machine learning systems succeed and fail in replicating brain-like behavior. This includes examining performance benchmarks on cognitive tasks such as language understanding, visual perception, and decision-making, and comparing these against known properties of human cognition.

The study also aims to survey emerging research directions — including neuromorphic hardware, spiking neural networks, and neuro-symbolic AI — that attempt to close the remaining gap between machine and biological intelligence. Finally, the paper seeks to discuss the broader implications of brain-like AI, including ethical questions around consciousness, autonomy, and the social impact of increasingly intelligent systems.

4. METHODOLOGY

This study follows a secondary research methodology based on a systematic review of published literature. No new computational experiments or empirical datasets were generated as part of this work. The review draws on peer-reviewed journal articles, conference proceedings from major AI venues including NeurIPS, ICML, ICLR, and CVPR, as well as key publications in computational neuroscience journals.

Sources were selected according to their relevance to the central theme of brain-inspired computation, their citation impact within the field, and their coverage of both historical developments and recent advances up to 2024. Both foundational theoretical papers and applied experimental studies were included. Where quantitative performance comparisons are referenced, the figures are drawn directly from the cited papers.

The methodology for organizing the review followed a thematic structure, grouping findings according to major topic areas: biological neuron architecture and it's computational analog, learning mechanisms, cognitive simulation, and current limitations. This thematic grouping was chosen to highlight conceptual continuity across decades of research rather than treating developments in purely chronological isolation.

4.1 THE BIOLOGICAL NEURON AND ITS ARTIFICIAL ANALOG

A biological neuron consists of a cell body (soma), dendrites that receive input signals, an axon that transmits output, and synaptic terminals that communicate with neighboring neurons. When the integrated input signals exceed a threshold, the neuron fires an action potential — a brief electrical spike that carries information along the axon. The strength of the connection between neurons, called synaptic weight, is modifiable through experience, a property underlying learning and memory.

The artificial neuron mirrors this structure. Each unit in an ANN receives a weighted sum of inputs from connected units, applies a nonlinear activation function (analogous to the firing threshold), and passes the result to subsequent layers. Activation functions such as ReLU (Rectified Linear Unit), sigmoid, and softmax have been developed over time to improve learning dynamics. Weight adjustment through backpropagation acts as a surrogate for biological synaptic plasticity, although the mechanisms are mathematically quite different from Hebbian learning or spike-timing-dependent plasticity (STDP) observed in real neurons.

4.2 DEEP NEURAL NETWORKS AND CORTICAL HIERARCHY

The human brain process information through a hierarchical series of cortical areas. In the visual system, for instance, the primary visual cortex (V1) responds to simple edges and orientations, while higher cortical areas respond to increasingly complex features such as faces and objects. This hierarchical abstraction is the direct inspiration for Convolutional Neural Networks (CNNs).

A CNN applies a series of convolutional layers, each learning to detect progressively more abstract visual features. Early layers detect edges and textures; intermediate layers detect shapes and object parts; and deep layers represent complete objects or scenes. This hierarchical feature learning has made CNNs extraordinarily effective for image classification, achieving superhuman accuracy on benchmarks such as ImageNet. The parallel to cortical hierarchy is structural as well as functional: both systems achieve invariance to transformations like scale and rotation through feature pooling.

4.3 RECURRENT NETWORKS AND MEMORY

Human memory is not a static storage system but a dynamic, context-dependent reconstruction process. The hippocampus plays a central role in forming new episodic memories and in consolidating them into long-term cortical storage during sleep. Recurrent Neural Networks (RNNs), which maintain hidden state across time steps, were designed to model sequential and temporal dependencies in data, loosely analogous to how the brain integrates informations over time.

Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) addressed the vanishing gradient problem that limited early RNNs, enabling learning over longer time horizons. LSTMs use gating mechanisms — input gate, forget gate, and output gate — that control the flow of information into and out of memory cells. While the analogy to biological memory is approximate, LSTM networks achieve strong performance on tasks such as speech recognition, machine translation, and time-series prediction that require sustained memory over many input steps.

4.4 ATTENTION MECHANISMS AND THE TRANSFORMER

The Transformer architecture (Vaswani et al., 2017) introduced self-attention as the primary computational mechanism, replacing recurrence entirely. Self-attention computes pairwise relationships between all elements of an input sequence, allowing the model to focus on the most relevant parts of the input for each output element. This is broadly analogous to selective attention in the brain, where cognitive resources are directed toward task-relevant stimuli while suppressing irrelevant ones.

The impact of Transformers has been extraordinary. BERT (Devlin et al., 2018) demonstrated that bidirectional pretraining on large text corpora followed by fine-tuning on specific tasks achieves state-of-the-art results across nearly all natural language benchmarks. GPT-3 (Brown et al., 2020) showed that scaling a language model to 175 billion parameters produces emergent capabilities — few-shot learning, analogical reasoning, code generation — that were not explicitly trained. These emergent properties draw comparisons to how the brain develops general capabilities from domain-specific experiences.

4.5 REINFORCEMENT LEARNING AND THE REWARD SYSTEM

The brain's reward system, centered on the basal ganglia and dopaminergic pathways, is responsible for learning behaviors that lead to positive outcomes. Dopamine functions as a teaching signal — it's release following an unexpected reward strengthens the neural pathways associated with the action that produced it, a mechanism formalized in neuroscience as temporal difference learning.

Reinforcement Learning (RL) in AI directly implements this principle. An RL agent interacts with an environment, takes actions, and receives reward signals that guide learning toward policies that maximize cumulative reward. The temporal difference (TD) learning rule at the heart of modern RL was developed in direct analogy to dopaminergic reward prediction error signals in the brain (Schultz et al., 1997). Deep Reinforcement Learning, which combines neural networks with RL, achieved remarkable milestones: DeepMind’s AlphaGo defeated world champions in Go (Silver et al., 2016), and AlphaFold solved protein structure prediction (Jumper et al., 2021), demonstrating that brain-inspired learning principles can unlock solutions to problems far beyond the scope of traditional programming.

5. CHALLENGES AND LIMITATIONS

Despite the impressive advances in brain-inspired AI, fundamental differences between biological and artificial intelligence create significant challenges. One major issue is energy efficiency. The human brain operates on approximately 20 watts of power, yet outperforms any current AI system on general reasoning tasks. Training a large language model such as GPT-3 requires millions of watts of compute over weeks, and inference requires substantial hardware. The brain’s efficiency stems from sparse coding, spiking neural activity, and analog computation — properties that are difficult to replicate in current digital hardware.

Sample efficiency is another critical limitation. The brain can learn new concepts from very few examples — a child learns to recognize a dog from seeing a handful of instances. Machine learning systems typically requires thousands or millions of labeled examples to achieve comparable performance. This data hunger is a fundamental limitation of statistical learning, and while few-shot learning research has made progress, the gap with biological learning efficiency remains large.

Generalization and robustness are ongoing concerns. Machine learning models can be brittle: small, carefully crafted perturbations to an input — adversarial examples — can cause confident misclassification that no human would make. This brittleness reveals that ML systems learn statistical correlations in data rather than the deep structural understanding that characterizes human cognition. The brain’s ability to reason causally and counterfactually — to understand not just what happened but why, and what would happen under different conditions — is not yet captured by current ML architectures.

Interpretability is a growing concern. Deep neural networks are often described as black boxes: their internal representations are difficult to decode, making it hard to understand why they produce a particular output. The brain, while also not fully understood, has at least been studied through centuries of neuroscience, and brain damage studies, neuroimaging, and electrophysiology have given researchers rich insight into its functional organization. Equivalent interpretability tools for neural networks are still immature, raising important questions in high-stakes applications such as medical diagnosis and autonomous vehicles.

Finally, there are philosophical and ethical considerations. As AI systems become more brain-like, questions arise about consciousness, agency, and moral status. Current AI has no subjective experience and no inner life, but as systems grow in complexity and capability, the boundaries become philosophically unclear. Issues of bias, fairness, and accountability in AI decisions also become more pressing as systems are deployed in consequential social domains including hiring, criminal justice, and healthcare.

6. FINDINGS

The review of literature confirms that machine learning has achieved genuinely brain-like performance in a range of specific cognitive tasks. Deep neural networks now match or exceed human accuracy in image classification (Krizhevsky et al., 2012; He et al., 2015), and large language models demonstrate sophisticated language understanding and generation that was unimaginable a decade ago. These achievements have been driven primarily by increases in computational scale, larger datasets, and architectural innovations directly inspired by neuroscientific understanding.

It was further found that the most successful AI architectures share structural principles with the brain, including hierarchical feature extraction, distributed representation, and context-sensitive processing. The success of attention mechanisms in Transformers suggests that the brain's attentional selectivity is a crucial computational principle that can be abstracted and implemented in non-biological systems. Similarly, the success of reinforcement learning confirms that the dopaminergic reward signal is a universal learning principle that extends beyond biology.

However, the findings also make clear that current machine learning systems simulate particular functional properties of the brain rather than replicating its full architecture or its general intelligence. The brain integrates vision, language, memory, emotion, and social cognition in a unified, embodied system that interacts dynamically with the world. Current AI systems, by contrast, are modular, task-specific, and trained in isolation from the rich sensorimotor experience that shapes human intelligence from infancy.

Emerging research directions show promise for narrowing this gap. Neuromorphic computing, using hardware architectures modeled on the brain's spiking neural dynamics, offers potential improvements in energy efficiency. Neuro-symbolic AI attempts to combine the pattern recognition strengths of deep learning with the structured, causal reasoning of symbolic AI, better replicating the full range of human cognitive capabilities. Multimodal models that jointly process vision, language, and action are beginning to approach a more integrated form of machine intelligence.

7. CONCLUSION

The relationship between the human brain and artificial intelligence is one of the most profound scientific intersections of our time. Machine learning systems have drawn deeply from neuroscience and cognitive science to develop architectures that achieve remarkable brain-like capabilities in specific domains. From the artificial neuron to the Transformer, each major development in AI has been shaped by our evolving understanding of how the brain process information.

Yet the human brain remains far ahead of any machine in terms of generality, efficiency, robustness, and adaptability. The brain learns from a handful of examples, integrates multiple senses seamlessly, reasons causally about the world, and achieves all of this on twenty watts of biological power. Replicating these properties in silicon will require not only more powerful hardware and larger datasets, but fundamentally new theoretical frameworks that goes beyond statistical learning.

The future of AI likely lies in a deeper synthesis of neuroscience and machine learning — drawing not just on the brain's architecture as a design template, but on its computational principles at a functional level. As these fields continue to advance together, the boundary between biological and artificial intelligence may become increasingly difficult to draw. This convergence brings both enormous opportunity and profound responsibility, demanding careful scientific, ethical, and societal engagement from researchers and policymakers alike.

REFERENCES

[1]. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5(4), 115–133.

[2]. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408.

[3]. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

[4]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[5]. Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79–87.

[6]. Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593–1599.

[7]. Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554.

[8]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. NeurIPS 2012.

[9]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

[10]. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS).

[11]. Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489.

[12]. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. NeurIPS 2020.

[13]. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.

[14]. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv:1512.03385.

[15]. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed. draft). Stanford University.

[16]. Moser, E. I., Kropff, E., & Moser, M. B. (2008). Place cells, grid cells, and the brain's spatial representation system. Annual Review of Neuroscience, 31, 69–89.

Cite this Article:

Shinde, A.V. (2026). How Machine Learning Systems Simulate the Human Brain Using Artificial Intelligence. International Journal of Emerging Research in Computer Science, 2(4), 1-6.

Journal URL: *https://ijercs.com/*

DOI: https://doi.org/10.59828/ijercs.v2i4.28