American Journal of Science, Engineering and Technology
2026, Vol. 11, No. 3, pp. 195–204
https://doi.org/10.11648/j.ajset.20261103.19

SciencePG
Science Publishing Group

Research Article

Artificial Intelligence Thinking, Learning, and Generation: Cognitive and Neural-Inspired Foundations of Machine Learning and Deep Learning

Rajan Thapaliya*

College of Engineering and Computer Science, Accord University, Mogadishu, Somalia

Abstract

Artificial intelligence (AI) systems are often discussed as if they think and learn, and are capable of generating, but these claims are typically presented as metaphors rather than analyses. The development of machine learning and deep learning, particularly neural and generative models, has sparked debate over whether artificial systems truly resemble human cognition in any meaningful way or merely reproduce its features through statistical computation. In this paper, the working principles of modern AI systems are examined by situating machine learning and deep learning within the intellectual traditions of neuroscience and cognitive science. Based on recent surveys, theoretical studies, and critical views, the paper discusses how learning processes in artificial systems are motivated by and differ radically from biological thinking. It compares supervised, unsupervised, and reinforcement learning paradigms, the formation of representations in deep neural networks, and how generative models can produce outputs that seem creative but lack comprehension or intention. The paper critiques neural metaphors, cognitive analogies, and assertions of machine intelligence, arguing that interpreting what AI systems can and cannot do requires careful, specific analysis. Finally, the paper offers an interdisciplinary synthesis that helps clarify the conceptual underpinnings of contemporary AI, where anthropomorphic interpretations are flawed. It underscores the need to synthesize insights from cognitive science, neuroscience, and philosophy.

Keywords

Artificial Intelligence, Machine Learning, Deep Learning, Cognitive Science, Neuroscience-Inspired AI, Generative Models, Representation Learning

*Correspondence: Rajan Thapaliya (rajan@accord.edu.so)

Received: 9 July 2026; **Accepted:** 24 July 2026; **Published:** 27 August 2026

1. Introduction: Framing Intelligence in Machines

1.1. The Resurgence of Interest in "Thinking Machines"

The topic of artificial intelligence has come into the limelight in both the mainstream and academia. The methods that have recently advanced in deep learning, grand-scale neural networks, and generative models can solve tasks once considered the stronghold of human intelligence, such as language production, visual perception, and tactical decision-making. Such advances have revived the controversies surrounding whether machines really think and have understanding, or can be described by metaphor in that way. The widespread use of large language models has exacerbated this ambiguity.

Results analogous to reasoning or creativity have even raised concerns that AI is on the brink of general intelligence [3, 5]. Simultaneously, they state that a lack of depth in abstraction, grounding, and understanding can be concealed beneath superficial competence [14, 16]. Such tension indicates a disparity in the description of the AI systems and their functioning. The field has historically been influenced by cycles of optimism and skepticism: symbolic AI promised human-like reasoning based on explicitly defined rules but struggled with the complexities of the real world. At the same time, recent neural networks have placed greater emphasis on data-driven learning and distributed representations, raising radically different conceptual issues.

1.2. Clarifying "Thinking," "Learning," and "Generation"

Words mainly influence arguments over AI. Thinking, learning, and generation are anthropocentric connotations of human cognitive processes, but in artificial systems, these are computational functions. Human thinking is deliberate and meaningful; learning is conditioned by embodiment, motivation, and context. In this context, machine learning typically refers to the process of optimizing an objective function by adjusting parameters. This could be a very efficient process, but it does not entail knowledge concerning what is learned and why, and should not be parallel with human cognition [10, 20].

Generative AI is also likely to be misunderstood. Even though models are sometimes said to create content, generation can be better regarded as probabilistic prediction conditioned on large datasets. Texts generated by language models approximate what is likely to come next, not beliefs or intentions [3]. Philosophy warns that the way philosophers assign mental states to such systems distorts both machine outputs and human cognition [15].

1.3. Cognitive and Neural Inspirations in AI Research

Cognitive science and neuroscience have influenced the direction of modern AI. The initial computational methods aimed to formalize phenomena such as perception, memory, and reasoning, and to provide a conceptual basis for artificial systems [22]. Neural networks are motivated by biological learning, but only one is abstract [2]. Artificially inspired neuroscience is currently a significant interdisciplinary field, and recent surveys have suggested potential advantages in connecting artificial models with biological insights [8, 28]. Artificial networks are not copies of brains; however, they can shed new light on the concept of distributed representation, hierarchy, and dynamics of learning. Meanwhile, criticisms of entirely data-driven learning highlight the importance of structure, causality, and prior knowledge, which are also main aspects of human cognition and may be poorly represented in existing AI systems [14, 27].

1.4. Aim and Structure of the Paper

The paper elucidates how AI systems think, learn, and generate by analyzing machine learning and deep learning through cognitive and neural-inspired frameworks. Instead of asking whether AI is cognitively intelligent like humans, it examines the operation of artificial systems and why their outputs may seem cognitively rich despite underlying constraints.

Section 2 analyzes the cognitive bases; Section 3 analyzes neural inspiration; Section 4 studies the fundamental paradigms of learning; Section 5 examines the formation of representations; Section 6 analyzes generative models; Section 7 analyzes conceptual and ethical constraints, and the conclusion draws on major conclusions.

2. Cognitive Foundations: Human Learning as Inspiration

2.1. Human Cognition and Learning Processes

Human thinking is a complex interplay of perception, memory, abstraction, and reasoning. Human learning is not merely a collection of data; it is conditioned by prior knowledge, context, intentions, and lived experience. Cognitive science highlights frontiers that way humans can learn with sparse information, generalize swiftly, and with the help of structured representations that allow them to be flexible in their reasoning- capacities that many machine learning systems fail to match with due to their need to process massive amounts of data and to optimize them to reach comparable performance [10].

Perception and action are also closely related to human learning. Humans engage in the discovery of environments,

hypothesis testing, and the development of internal representations of the real world. This causal-based learning contributes to intelligence development through interaction among perception, memory, and reasoning, rather than individual pattern recognition [22]. Another benefit is the reliance on prior knowledge: new data are interpreted in the light of an existing conceptual apparatus, enabling transfer and compositional learning. The issue is that many AI systems cannot generalize beyond their training distributions, indicating a persistent gap between human and machine learning [14, 16].

2.2. Early Cognitive Models and Symbolic Artificial Intelligence

First AI was based on cognitive ontologies of reason. Symbolic AI (good old-fashioned AI) was the assumption that explicit representations and logical rules could represent cognition. These systems succeeded in limited domains, such as proving theorems and handicraft systems, yet they were weak and difficult to develop. They were challenged by the ambiguity, perception, and complexity of the real world, and relied more on hand-coded knowledge than on sound learning. Despite these restrictions, symbolic AI elucidated the value of the representational, structural, and interpretability in intelligent acts.

Modern criticisms of data-only learning often lead to a resurgence of these issues, insisting that intelligence cannot be reduced to statistical pattern matching [14]. This has given rise to new neural-symbolic models that may represent renewed interest aligned with neural flexibility [12].

2.3. From Cognitive Theories to Machine Learning Paradigms

As AI moves towards more data-oriented approaches, cognitive inspiration is being redefined rather than discarded. Machine learning focuses on experience-based learning, which is loosely equivalent to human adaptation. All the different forms of supervised, unsupervised, and reinforcement learning are simplified analogs of human learning, but they differ in meaningful ways. Supervised learning is similar to instruction and feedback, except that the models do not necessarily appreciate the meaning or relevance of tasks, as humans do. Unsupervised learning is closer to the human process of pattern discovery, reflecting cognitive propositions about forming concepts through exposure to data structure [10].

The most similar behavioral analog is reinforcement learning, which is learning by trial and error with rewards, but human learning incorporates much deeper world models, social constraints, and long-term objectives in contrast to artificial agents that usually provide optimization of specific rewarding functions in an inherent domain [4, 20].

2.4. Fundamental Differences Between Human and Machine Cognition

There are limits to cognitive analogies. Human cognition is embodied, socially situated, and environmentally constrained by evolutionary limits in general, which have no analog in artificial systems. Human beings study in environments filled with meaning and purpose, whereas machines study abstract representations devoid of lived-in experience. Bereft of the sensory-motor interaction or intrinsic motivation basis, AI can be highly pattern-recognizing and poorly causal and semantic [16, 27]. This is one reason why systems can work well on benchmarks but fail when there are slight shifts in distribution. There is also human cognition, which involves metacognitive abilities, self-monitoring, reflection, and goal revision, aiding flexible adaptation [11]. The existing machine learning systems do not possess such capabilities and instead act as optimization engines, and statements that people believe they think antagonistically need to be approached carefully.

2.5. Implications for Cognitively Inspired AI Research

Cognitive roots help shape the future of AI, not by recreating the human mind, but by selectively incorporating principles that address well-known drawbacks.

These are learning representations, causal reasoning, and inductive biases that enable learning with limited data [10, 14]. Recent evaluations also emphasize greater overlap between cognitive science and machine learning, and interdisciplinary work might enhance efficiency, robustness, and interpretability [8, 28]. Meanwhile, conceptual clarity is imperative: the cognitive inspiration one may be confused with cognitive equivalence.

3. Neural Inspiration: From Biological Neurons to Artificial Networks

3.1. Biological Neurons and Learning Mechanisms

The field of neuroscience has been a source of motivation for AI research, particularly in understanding how biological neurons process information and adapt. Cognition is a consequence of large networks of neurons connected by synapses, whose strengths are modulated by synaptic plasticity in the brain. Learning is distributed across interacting areas supporting perception, memory, and action. One of the generalist principles is Hebbian learning (cells that fire together wire together), which describes how co-activation can promote synaptic strengthening and how learning and memory are formed. Though simplified, it still affects how adaptive behavior can be developed from local learning rules [2].

Biological learning is also energy-efficient, robust to noise,

and prone to rapid generalization, which the brain-inspired AI aims to emulate. Learning has timescales (short-term adaptability and long-term memory), is modulated by reward, attention, and motivation, and brings in some flexibility and context sensitivity that most existing algorithms do not [27].

3.2. Artificial Neurons and the Abstraction of Biology

Artificial neural networks are loosely based on biological systems but applied in an abstract form. Artificial neurons add weighted sums, provide nonlinear activations, and forward afferent signals; they only simulate a tiny example of biological complexity (i.e., temporal dynamics, chemical signals, morphology). Nonetheless, when used in deep networks, multilayer architectures are very efficient: hierarchical feature learning can be conducted by multilayer networks, and they tend to produce increasingly abstract representations, which are regularly compared with processing stages in the sensory cortex [9, 26]. More importantly, success does not mean that it is biologically faithful. The neural networks can be more readily described as inspired than as models of the brain; brain-like plots are functional, not mechanical [7]. Maintaining this difference can ensure that claims about machine cognition should not go too far with neural metaphors.

3.3. Backpropagation and Biological Plausibility

Backpropagation allows for easy optimization of the error gradient, enabling performance-centric deep learning and being fundamental to deep learning's growth. However, its biological plausibility has been compromised because it requires mechanisms such as symmetric weight transport and global error signals, which any neural equivalent cannot neutralize. This has motivated the exploration of other learning rules that are more likely to simulate gradient learning, but are more consistent with neural properties [13, 24].

Predictive-coding and related local-learning frameworks propose that cortical circuits minimize prediction error through local computations, offering a possible bridge between gradient-based learning and biological constraints [2]. Although these practices may not be complete, they strengthen neuroscience as an inspiration and not a blueprint.

3.4. Neuroscience as a Two-Way Exchange

The connection between neuroscience and AI is now two-way: neuroscience may guide AI design, and deep networks have been used to study the brain. Neural responses in the sensory cortex can be predicted using goal-driven models that provide computational predictions of neural coding [19, 26].

Reviews interdisciplinary Reviews suggest that further broader convergence can occur faster between AI, neuroscience, and cognitive science, but interpretations must be used since it is possible that model-brain alignment can be due to non-biological factors [8, 28].

3.5. Limits of the Neural Analogy

It is good to be inspired by the brain, but brain metaphors should not be overused, as they can hide significant differences. Brains were developed within the limits of survival, efficiency, and adaptation, but neural networks are designed to output designer-defined goals.

According to the critics, most AI systems do not have functions that are key to biological intelligence, like innate form, learning through development, or embodiment, and, therefore, they may be no wonder why systems can be effective but fragile with changes in distribution and powerless at causal and semantic comprehension [14, 27]. Understanding such boundaries helps us see that neuroscience can provide AI with direction and boundaries, but not a direct replication object.

4. How Machines Learn: Core Paradigms in Machine Learning

4.1. Learning as Optimization Rather than Understanding

Learning in machine learning is a way of enhancing the performance of a task with experience. Machine learning, compared with human knowledge, follows more or less goal-oriented principles, and optimization, where error, reward, or likelihood are quantified functions, is the result. Systems learn the parameter adjustments to these measures over time. This distinction matters. Although machine learning has shown impressive performance, it is still mechanical; models lack goals and do not identify the tasks they perform. The effectiveness of deep learning, according to [20], does not ask whether machines are also intelligent but rather what existing paradigms are really solving.

The ability and the limitation become understandable when learning is viewed as a disadvantage rather than a thought. The parallels between ML paradigms and cognitive metaphors, though sometimes described as supervised learning as instruction, unsupervised learning as pattern discovery, and reinforcement learning as reward-driven adaptation, are loose.

4.2. Supervised Learning: Learning from Labeled Examples

Supervised learning trains models using input-output pairs by minimizing a loss function that quantifies prediction error. It can be used effectively for classification, speech recognition, and language processing. Supervised learning can be similar to teaching and feedback. However, the comparison is limited: human beings tend to extrapolate from a small number of ex-

amples, and supervised models typically need substantial labeled examples [10]. Designer-defined goals are also coded in labels rather than their intrinsic meaning. Supervised learning models may be vulnerable to changes in distribution. [6] demonstrated that learning systems, in some cases, are highly accurate through shortcut learning, i.e., using superficial cues rather than learning structure, which shows the failures of statistical fitting and human-like abstraction.

4.3. Unsupervised Learning: Discovering Structure Without Labels

Unsupervised learning reduces reliance on labels and enables learning directly from data, such as clusters, latent variables, and representations. This is comparable to the development of concepts in human beings, as they learn to deduce recurrent patterns without being explicitly taught to do so. Deep representation learning can also learn features useful for downstream tasks and produce hierarchical, distributed representations [9].

However, unless anchored in goals, actions, or causal thinking, unsupervised representations can capture correlations that are not semantically coherent. According to recent surveys, inductive biases and structure are vital to leading learning toward more meaningful abstractions [8, 28].

4.4. Reinforcement Learning: Reward-Driven Adaptation

In reinforcement learning, learning occurs between an agent and the environment: different actions yield rewards, and the agent adapts its behavior to maximize total reward. The RL has achieved significant breakthroughs in games, robotics, and control. It is frequently associated with behavioral and decision theories, and signals of reward are analogized to learning processes in animals [4].

These are mere parallels, however. There are typically strong simplifying assumptions in the environment in which artificial agents act, engineered reward functions, mis-specified rewards may lead to unintended behavior, and agents do not usually possess rich internal models and social context. The RL thus only frames a minimal part of human adaptive learning [20].

4.5. Learning Without Awareness or Intent

Machine learning lacks awareness, intent, and reflection across paradigms. Parameter updates maximize goals rather than describe knowledge. The key concept of philosophical critiques pledges that meaning or belief is not possible and required in predicting tokens: it is possible to, and the vision model measures well, despite lacking any fundamental understanding of language, whereas it is possible to classify products without perceptual experience, despite having a consequential language model [15]. The separation of learning-as-optimization and learning-as-cognition provides a better basis for assessing progress.

4.6. Implications for Machine Learning Research

Limiting factors encourage the adoption of methods that are more closely tied to structure and incorporate more abstract strategies, such as hybrid statistical learning models with symbolic algorithms, causal accents, and symbolic representations [12, 21].

It has been noted that further development will not rely solely on scaling but also on reevaluating learning goals and building on previous knowledge more efficiently [18, 28]. This change can facilitate intelligence measures beyond performance measures.

5. Deep Learning and Representation Formation

5.1. Hierarchical Feature Learning in Deep Networks

Hierarchical representation learning is a characteristic of deep learning. Deep neural networks perform a series of transformations on their inputs across layers; earlier layers extract low-level features, and later layers encode task-specific patterns. This stratifying arrangement is similarly likened to hierarchical processing within biological sense organs (particularly vision), but away from the retina [9, 26]. On cognitive grounds, this can be likened to how perception constructs ideas from sensory data.

The mechanisms, however, differ: in artificial systems, hierarchical representations are created through architectural design and optimization goals, rather than through developmental or evolutionary processes. An analogous representational structure, thus, does not mean cognitive similarity. Reports indicate that hierarchy is an inductive bias that facilitates generalization; however, it also relies on extensive datasets and computation, which widens the gap between machine efficiency and human learning [8, 28].

5.2. Internal Representations and Latent Spaces

Deep learning represents knowledge using high-dimensional latent representations: distributed patterns of activations and not symbolic notions. This allows complex relationships that are not easily hand-engineered to be captured in the model and has spurred advancements in vision, speech, and language. However, latent representations may be opaque, and it can be hard to know whether models are based on meaningful structure or spurious correlations- and that is the concern of interpretability and trust [1]. In cognitive science, there exists a contradiction: human depictions are usually organized in

ways that enable exposition, reflection, and discourse.

Deep networks are not able to perform meta-representational tasks and instead serve as function approximators, making it impossible to compare their representations with conceptual understanding [15].

5.3. Representation Learning and Abstraction

Deep learning holds the promise of learning directly from data, enabling transfer and generalization. As practice has shown, abstraction is often frail and task-specific; models are vulnerable to the types of variation that humans can easily handle. Research on shortcut learning demonstrates that extreme accuracy can be achieved by exploiting superficial statistical cues rather than the underlying structure [6].

This implies that acquired abstractions might be causally and semantically ungrounded. To overcome such limits, it has recently been suggested that structure, compositionality, and more aggressive inductive biases be added to achieve more robust abstraction and systematic generalization [12].

5.4. Interpretability and Understanding Learned Representations

With the application of deep learning in high-stakes fields, interpretability is emerging as a key issue of concern. Attribution techniques, probing, and visualization are methods for identifying what drives model behavior, yet interpretability remains challenging, particularly with large models [1].

Processes of human decision-making tend to be explanatory and revisionary; deep networks lack introspection and are not aligned with communicative objectives. Subsequently, good performance with no explanation cannot be regarded as a testament to knowledge. According to some authors, interpretability is not only optional but also increasingly necessary to align with human cognitive and moral expectations [2, 15].

5.5. Generalization, Robustness, and Brittleness

It is believed that generalization as a measure of intelligence is commonly applied, but in training-like distributions, a deep learning system can generalize reliably. When the distribution changes, under adversarial perturbations, or even out-of-domain, performance may suffer. These breakdowns represent dependence on statistical associations rather than on direct models of causality or world form. More recent literature is focusing on combining causal inference with organized knowledge to build learning systems [21]. Brain-inspired AI surveys have also identified the field as the domain where biological systems remain superior to artificial systems, and where further studies are being pursued on how brain systems can learn with stability in the presence of noise and uncertainty [25, 28].

5.6. Implications for Deep Learning Research

The power and limitations of representation formation in deep learning are shown. Hierarchy representations and latent space can achieve remarkable performance, though they tend to be much less rich and flexible in terms of human representation. This disparity encourages blended patterns, increased understandability, and cognitively informed patterns towards portrayal and learning [15, 28].

6. How AI Generates: From Prediction to Creation

6.1. Generative Models in Contemporary Artificial Intelligence

Generative models become the focus of modern AI, especially in massive language models, image generators, and multimodal frameworks. Generative models are the opposite of discriminative models that label the inputs; that is, a generative model can generate new examples similar to the training data. They model a data density by sampling from a learned statistical model, though technically a regression model. Although the initial techniques relied on graphs and variational models, deep learning models such as transformers and diffusion models have since become the most popular methods in modern generative systems [3].

In any case, the general idea is that generation should be based on statistical inference rather than deliberate production. The application of foundation models has advanced the generative AI paradigm by enabling training on large, heterogeneous datasets and cross-task transfer. Nonetheless, the questionnaires indicate that their potential is constrained by training data, architectural biases, and objective functions [25, 28].

6.2. Generation as Probabilistic Prediction

Generative artificial intelligence is innately predictive. Language models consist of text generation, which predicts the probability of the following token given context and manufactures statistically likely continuations in a large corpus. This results in semantic coherence and fluency but does not denote semantic understanding and communicated purpose. This is mainly a question of the difference between prediction and understanding: any model can produce responses that seem thoughtful or creative without knowing what is correct, accurate, or relevant, because those outputs will be determined by optimization objectives, not goals or beliefs [15].

Intention and context influence human generation, linguistic, artistic, or problem-solving. Generative systems do not have this interpretation; instead, they can be considered learned extrapolations of data and not necessarily an expression of the internal mental state [3].

6.3. Creativity and the Illusion of Agency

Generative systems can be found to be creative, creating poetry, images, and novel solutions. However, this imagination is not planned but accidental. High-dimensional representations, along with stochastic sampling, allow models to recombine learned patterns, thereby generating novelty.

However, it is not only novelty that is human creativity; it is also usually guided by a goal, and it involves creative proposals, assessment, and contraction. There is also the promotion of the illusion of agency by conversational interfaces. Fluent behavior may lead people to believe that the user is competent in their language, or, intentionally, meaning, but such judgments should be universal human social tendencies rather than inherent system characteristics [15, 23]. There must be clarity about the concept to avoid treating generative systems as creative agents independent of their environment.

6.4. Large Language Models and Cognitive Limitations

Machine-reasoning arguments have become heated in the wake of the application of large language models. They are capable of doing specific tasks that require inference, analogy, and abstraction, although these are inconsistent and sensitive to prompts and distributional conditions. It is argued that LLMs are effective in completing patterns but poor at grounded reasoning, understanding cause-and-effect, and long-term logical reasoning [14, 16]. Their seemingly rational thought process may be based on superficial regularity rather than a sound conceptual framework, making it unpredictable in the face of novelty or antagonistic input.

Others note that LLMs are limited in perception and perception cannot be reliably defined as truth or plausibility, which makes it challenging to state that programs are moving towards artificial general intelligence [15, 17].

6.5. Generative AI, Cognition, and the Limits of Analogy

The achievements of generative models have brought back the analogy of human cognition. Some argue that large-scale theorizations bring linguistic competence or other elements of world modeling into the picture, however approximative. In contrast, others warn that language modularity ignores the fact that embodiment, learning, and intentionality vary.

Perception, action, motivation, and social context are connected to biological generation, and generative AI operates only on statistical regularities trained on data [25, 27]. Interdisciplinary reviews indicate, therefore, that generative models may be helpful for research on learning and representation but are not typically considered cognitive models of the human mind [8, 28].

6.6. Implications for Evaluating Generative Intelligence

Assessments based solely on the quality of the output can mask fundamental weaknesses in insight, strength, and congruency. There is no such thing as dependable reason, clarification, or customization. Another common way of framing future research is to enhance interpretability and grounding, and to integrate with both causal and symbolic frameworks [12, 21]. Re-conceptualizing generation as a predictive rather than a generative process would justify a less biased evaluation of the contribution of generative AI and non-generative AI.

7. Critical Reflections: Can Artificial Intelligence Really Think

7.1. Performance Versus Cognition

As AI systems are performing exceptionally well at numerous tasks, a question is constantly raised: do they actually think, or do they merely pretend to be intelligent? The solution to this involves the distinction between performance and cognition. Although contemporary systems might be able to execute some task linked to intelligence (e.g., language use, planning), performance per se is not considered thinking in the cognitive sense of the term. Human thinking entails interpreting the meanings, beliefs, and rationales of objectives and outcomes. In contrast, AI systems operate through statistical optimization and are not semantic or intentional. According to [16], when performance is impressive, there is often a tendency to make exaggerated statements about the power of cognition that distort the distinction between task competence and actual knowledge. Such a reduction of intelligence to behavior risks disregarding the internal processes that are the key to human thinking [15].

7.2. The Problem of Anthropomorphism

Anthropomorphic language has a significant impact on perceptions of AI. Artificial intelligence systems are also referred to using short terms such as thinking, reasoning, and understanding, which may be confusingly used to suggest that engineered artworks embody anthropomorphic qualities. Complex behavior is more likely to leave humans with agency and intention beliefs, which may lead to false trust and unrealistic expectations.

Maintains that when AI is viewed as a cognitive agent rather than a tool, it not only misestimates its potential but also its constraints. Anthropomorphism also carries philosophical implications, introducing the notion of a lack of subjective experience and self-awareness in AI [23]. Systems are not conscious of what they create, even when their outputs are similar to reasoning.

7.3. Consciousness, Understanding, and the Limits of Computation

One of these disputed questions is whether machines will ever attain a level of consciousness or actually be understanding. Although speculation, the majority of existing systems are obviously inadequate: they do not provide mechanisms of subjective experience, self-modeling, and reflective awareness. To understand, representations should be compared with the world, and the truth and revision of beliefs should be assessed with evidence. Rather than working on meaningful symbols or vectors, AI systems work on meaningless symbols or vectors, potentially leading to the production of formally coherent but semantically empty outputs [15].

Neuroscience also highlights that cognition is based on the integration of perception, emotion, motivation, and memory processes, which is far more complex than current AI models [2, 27].

7.4. Intelligence Without Understanding

The generative models and deep learning propose that it is possible to be highly competent without any cognition. The systems are neither aware nor can they perform superhumanly in narrow scopes, which contradicts the conventional theories of intelligence. Other scholars prefer operational definitions that are adaptable and focused on problem-solving and might involve highly developed AI. Others also do not think this takes into account the critical dimensions of cognition: intentionality, meaning, and normative reasoning [14]. The stakes involved in practice are obvious: if AI is approached like an intelligent agent, it can be trusted more than it can actually be; if it is approached as mere automation, its value can be undervalued. An intermediate opinion considers AI as contributive smarts, potent, but distinct from Human thinking in specific categories.

7.5. Implications for Ethics, Responsibility, and Trust

When AI is mischaracterized as a thinking entity, ethical implications arise. When systems are considered autonomous actors, they may shift responsibility away from designers, operators, and institutions, thereby diminishing accountability and human decision-making within AI systems. Explainability and transparency, therefore, are important. If AI is perceived as a tool, interpretability becomes the key to responsible exploitation. The concept of explainable AI is based on the need to align systems with values and human expectations, particularly in high-stakes environments [1].

The foundation of trust must be realistic: without excessive trust, there is always an opportunity to misuse it and cause harm; on the other hand, without undertrust, there is always the risk of misuse.

7.6. Toward a More Precise Language of Intelligence

In this paper, the necessity of accuracy is supported. It is rhetorically convenient to refer to AI as the process of thinking or understanding, but this conflates two phenomena. Some of the underlying vocabulary is actively proposed by the authors through learning, optimization, and representation, rather than through cognition or consciousness [15, 20]. Accuracy helps make better judgments and set more realistic expectations. It also shifts the focus of any metaphysical discussions to questions that can be empirically and conceptually addressed: how AIs operate, why they succeed, and why they fail.

7.7. Interdisciplinary Synthesis and Future Research Directions

The current discussions on the topic of artificial intelligence increasingly reveal the boundaries of the approach to AI in the image of engineering performance. Another typical lesson of cognitive science, neuroscience, and philosophy is the realization that intelligence is not a unitary ability but an emergent process produced by various intermediary processes. Human intelligence synthesizes perception, memory, embodiment, social learning, and normative reasoning in a manner that is hardly entirely understandable. Artificial systems, on the contrary, focus on and streamline particular functions with narrow goals. It is important to recognize this distinction when preventing both extreme arguments about machine cognition and extreme definitions of intelligence.

From an interdisciplinary perspective, the further evolution of AI will focus not so much on scaling current architectures but rather on reconsidering long-standing assumptions about learning, representation, and evaluation. Cognitive science also emphasizes structured representations, compositionality, and sparse-data learning, whereas neuroscience focuses on robustness, energy efficiency, and timescale-adaptive learning [2, 27]. In turn, philosophy offers critical means for explaining concepts such as understanding, agency, and responsibility, allowing us to differentiate between metaphor and mechanism [15].

Instead of mimicking human thinking, a more fruitful line of research is to cherry-pick from both fields and use their insights to fill gaps in existing systems. A variety of approaches that integrate statistical learning and causal modeling with symbolic structure and interpretability could offer opportunities to build stronger, more responsible AI [12, 21]. This type of integration makes no move to blur the boundary between artificial and human intelligence; instead, it sharpens it. With the proper conceptual rigor and a willingness to go interdisciplinary, AI studies can develop scientifically and responsibly.

8. Conclusion

The paper discussed the mode of thought, learning, and

generation of artificial intelligence systems by placing machine learning and deep learning in the context of cognitive science and neuroscience. It claimed that human cognition should be distinguished from artificial learning processes and urged avoiding any metaphorical understanding of AI abilities.

Embodiment, intention, prior knowledge, and social context influence human cognition and enable efficient generalization from limited experience. By contrast, machine learning systems evolve through optimization processes, in which their parameters are adjusted based on the data and predetermined goals. Although powerful on particular scales of scope, such learning remains unmatched in understanding and meaning, contributing to weaknesses in robustness and generalization.

Artificial neural networks are not biological brain models, but abstract engineering structures, although they are inspired by neuroscience. Deep learning hierarchical representations are similar to neural processing only on the surface, and neuroscience can serve only as a source of inspiration, not as a template for designing AI.

Generative AI is also an example of the dangers of anthropomorphism. In contrast to intentional creation by design, generative models generate their outputs by predicting, in part, based on statistical regularities, and the perceived creativity reflects these regularities rather than knowledge and agency.

On the whole, AI systems represent potent forms of instrumental intelligence but are not copies of human thought. Advancements in artificial intelligence will not rely solely on scaling methods but will also combine knowledge from cognitive science, neuroscience, and philosophy without losing conceptual clarity.

Abbreviations

AI	Artificial Intelligence
ML	Machine Learning

Author Contributions

Rajan Thapaliya: Conceptualization, Investigation, Resources, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[2] Bengio, Y., Lecun, Y., & Hinton, G. (2021). Towards biologically plausible deep learning. *Neuron, 109*(9), 1359–1374. https://doi.org/10.1016/j.neuron.2021.03.042

[3] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A., Creel, K., Davis, J., Demszky, D., … Liang, P. (2021). *On the opportunities and risks of foundation models*. https://arxiv.org/abs/2108.07258

[4] Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., & Kurth-Nelson, Z. (2020). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences, 24*(5), 408–422. https://doi.org/10.1016/j.tics.2020.02.006

[5] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. https://arxiv.org/abs/2303.12712

[6] Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R. S., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence, 2*(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z

[7] Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

[8] Jiao, L., Ma, M., He, P., Geng, X., Liu, X., & Liu, F. (2024). Brain-inspired learning, perception, and cognition: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*. Advance online publication. https://doi.org/10.1109/TNNLS.2024.3401711

[9] Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science, 1*, 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

[10] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, e253. https://doi.org/10.1017/S0140525X16001837

[11] LeCun, Y., Bengio, Y., & Hinton, G. (2022). A path towards autonomous machine intelligence. *Communications of the ACM, 65*(8), 56–67. https://doi.org/10.1145/3534678

[12] Liang, B., Wang, Y., & Tong, C. (2025). AI reasoning in the deep learning era: From symbolic AI to neural–symbolic AI. *Mathematics, 13*(11), 1707. https://doi.org/10.3390/math13111707

[13] Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience, 21*, 335–346. https://doi.org/10.1038/s41583-020-0277-3

[14] Marcus, G. (2020). *The next decade in AI: Four steps towards robust artificial intelligence*. https://arxiv.org/abs/2002.06177

[15] Millière, R. (2024). Philosophy of cognitive science in the age of deep learning. *Wiley Interdisciplinary Reviews: Cognitive Science, 15*(1), e1670. https://doi.org/10.1002/wcs.1670

[16] Mitchell, M. (2021). Why AI is harder than we think. *AI Magazine, 42*(4), 88–102. https://doi.org/10.1609/aaai.v42i4.7533

[17] Mumuni, A., & Mumuni, F. (2025). *Large language models for artificial general intelligence (AGI): A survey of foundational principles and approaches*. https://arxiv.org/abs/2501.03151

[18] Ren, J., & Xia, F. (2024). *Brain-inspired artificial intelligence: A comprehensive review*. https://arxiv.org/abs/2408.14811

[19] Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Gerstner, W., Giraud-Carrier, C., Gourevitch, B., Izhikevich, E., Jarvis, S., Lucas, C., Lyle, C., Mermillod, M., Morcos, A. S.,… Malinowski, M. (2019). A deep learning framework for neuroscience. *Nature Neuroscience, 22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

[20] Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience, 22*(1), 55–67. https://doi.org/10.1038/s41583-020-00395-8

[21] Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE, 109*(5), 612–634. https://doi.org/10.1109/JPROC.2021.3058954

[22] Sun, R. (2020). Cognitive science and artificial intelligence: Toward cognitive machines. *Wiley Interdisciplinary Reviews: Cognitive Science, 11*(3), e1522. https://doi.org/10.1002/wcs.1522

[23] Tegmark, M. (2020). *Why AI is more like biology than physics*. https://arxiv.org/abs/2010.06894

[24] Whittington, J. C. R., & Bogacz, R. (2019). Theories of error backpropagation in the brain. *Trends in Cognitive Sciences, 23*(3), 235–250. https://doi.org/10.1016/j.tics.2018.12.005

[25] Wang, G. Y., Bao, H. N., Liu, Q., Zhou, T. G., & Wu, S. (2024). Brain-inspired artificial intelligence research: A review. *Science China Technological Sciences, 67*, 2282–2296. https://doi.org/10.1007/s11431-024-2732-9

[26] Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand the sensory cortex. *Nature Neuroscience, 19*(3), 356–365. https://doi.org/10.1038/nn.4244

[27] Zador, A. M. (2019). A critique of pure learning: What artificial neural networks can learn from animal brains. *Nature Communications, 10*, 3770. https://doi.org/10.1038/s41467-019-11786-6

[28] Zhang, Z., Ding, X., Liang, X., Zhou, Y., Qin, B., & Liu, T. (2025). Brain and cognitive science inspired deep learning: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering, 37*(4), 1650–1671. https://doi.org/10.1109/TKDE.2025.3527551