World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

A review of findings from neuroscience and cognitive psychology as possible inspiration for the path to artificial general intelligence

Ismatova Khilola Shukhratovna *

D'Overbroeck's Sixth Form Oxford, Oxfordshire, United Kingdom.

World Journal of Advanced Research and Reviews, 2025, 28(01), 599-604

Publication history: Received on 27 August 2025; revised on 06 October 2025; accepted on 08 October 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3481

Abstract

This paper examines how insights from neuroscience and cognitive psychology can inform the development of Artificial General Intelligence (AGI). It highlights neural mechanisms such as plasticity, predictive coding, and global integration, alongside cognitive functions like working memory, attention, and metacognition. The study argues that AGI should combine biological adaptability with cognitive intentionality through hybrid and embodied architecture. Integrating these interdisciplinary principles can guide the creation of self-regulating, context-aware, and ethically aligned intelligent systems.

Keywords: AGI; Neuroscience; Cognitive Psychology; Neural Plasticity; Metacognition; Embodied Cognition

1. Introduction

The quest for Artificial General Intelligence (AGI), a system capable of reasoning, understanding, and learning across domains with the same adaptability as human cognition has evolved from a speculative ambition into one of the most sophisticated research challenges of the 21st century. While the last two decades have witnessed dramatic progress in Artificial Narrow Intelligence (ANI) through deep learning, reinforcement learning, and generative models, the leap from specialized competence to generalized understanding remains elusive. AGI requires not only computational power but a deeper comprehension of how intelligence arises, evolves, and interacts with its environment questions that have occupied neuroscience and cognitive psychology for over a century.

Traditional AI systems are largely data-driven, relying on massive datasets and optimization of parameters to achieve proficiency in narrowly defined tasks. Despite their impressive performance, such systems lack contextual flexibility, causal reasoning, and self-reflective awareness, which are hallmarks of human cognition. For instance, while a language model can process vast linguistic corpora, it still lacks genuine comprehension of semantics and intentionality. This discrepancy between performance and understanding sometimes referred to as the "symbol grounding problem" underscores the limitations of purely statistical approaches. Neuroscience and cognitive psychology, by contrast, focus on mechanisms of understanding rather than outcomes, offering insights into how perception, memory, emotion, and consciousness co-evolve to produce intelligent behavior.

In neuroscience, intelligence is not merely an emergent property of neuron quantity but of dynamic interconnectivity and adaptive reconfiguration—what is known as neural plasticity. The human brain continuously reshapes its synaptic pathways based on experience, a feature that grants it resilience and creativity. This principle stands in stark contrast to the rigid architecture of current artificial neural networks, which, once trained, often exhibit limited adaptability to new or unstructured inputs. Moreover, the energy efficiency and robustness of biological neural systems far exceed

* Corresponding author: Ismatova Khilola Shukhratovna

those of their artificial counterparts, inviting questions about whether computational models can emulate the brain's hierarchical, predictive, and feedback-driven organization.

From the perspective of cognitive psychology, intelligence is a multifaceted construct encompassing attention, working memory, reasoning, problem-solving, and metacognition. Unlike the reductionist view that treats cognition as mere computation, modern theories emphasize context-dependence, embodiment, and interaction. Human cognition emerges through continuous engagement with the environment what Varela, Thompson, and Rosch (1991) termed "enactive cognition." Such findings challenge the notion that intelligence can be abstracted from physical and social experience. This insight has profound implications for AGI, suggesting that true general intelligence may require not only computational simulation but also embodied experience and goal-directed intentionality.

Furthermore, the integration of neuroscience and cognitive psychology offers a promising interdisciplinary framework for understanding intelligence as both a biological and computational phenomenon. Cognitive psychology provides models of high-level functions such as reasoning, decision-making, and self-regulation while neuroscience uncovers the micro- and meso-level mechanisms that implement these functions within neural substrates. By aligning these domains, researchers can move beyond algorithmic replication toward architectural inspiration: designing artificial systems that reflect the layered complexity of human cognition rather than imitating its superficial outputs.

2. Literature Review

The concept of Artificial General Intelligence (AGI), often described as the "holy grail" of computer science, traces its intellectual origins to Alan Turing's (1950) seminal question, "Can machines think?" Turing's vision of universal computation laid the theoretical groundwork for machines capable of general reasoning. Early efforts in symbolic AI during the 1950s–1980s (Newell and Simon, 1976) were largely inspired by human problem-solving models in cognitive psychology. However, these approaches failed to replicate the adaptability and context sensitivity inherent to human cognition, prompting a paradigm shift toward connectionist and later deep learning frameworks (Rumelhart and McClelland, 1986; Schmid Huber, 2015).

The emergence of Artificial Neural Networks (ANNs) and subsequent deep learning architectures reinvigorated optimism toward AGI. Nonetheless, despite their statistical prowess, modern neural networks are criticized for their lack of interpretability, transferability, and cognitive flexibility (Marcus, 2020; Bengio et al., 2021). These limitations reveal a fundamental gap between pattern recognition and genuine understanding of a gap that neuroscience and cognitive psychology may help to close by elucidating the mechanisms underlying biological intelligence.

Recent developments in neuroscience have illuminated several biological principles that could inform AGI design. One of the most influential is the predictive coding framework, which conceptualizes the brain as a hierarchical inference machine that minimizes prediction error through feedback loops (Friston, 2010; Clark, 2013). This theory resonates with deep learning models that adjust weights based on backpropagation to reduce loss; however, the biological brain achieves this with vastly greater efficiency and adaptability.

Another critical insight is neural plasticity, the brain's capacity to rewire itself in response to experience (Yuste, 2015). Artificial systems, by contrast, typically operate on fixed architectures post-training. Introducing dynamic structural plasticity into machine learning models could potentially lead to more autonomous forms of lifelong learning, a key feature of AGI (Hassabis et al., 2017).

Research into neural synchronization and global broadcasting mechanisms (Baars, 1988; Dehaene and Changeux, 2011) has also provided inspiration for cognitive architectures that model conscious awareness and attention distribution. The Global Workspace Theory (GWT) suggests that consciousness arises from the integration of information across specialized modules. Translating this concept into AI could yield architectures capable of prioritizing and coordinating distributed processes, enhancing generalization and decision-making in complex environments.

Complementary to GWT, the Integrated Information Theory (IIT) proposed by Tononi et al. (2016) describes consciousness as a measurable quantity of integrated information within a system. Although its direct computational implementation remains debated, IIT introduces the idea that intelligence may be linked to information integration density, a property that could inspire future AI architectures with emergent self-awareness.

While neuroscience elucidates the brain's physical substrate, cognitive psychology offers abstract models of how cognition operates functionally. Foundational frameworks such as Baddeley's (1992) model of working memory and Anderson's (2007) ACT-R cognitive architecture have been instrumental in understanding executive control, decision-making, and goal management. These mechanisms correspond to computational challenges in AI, including task switching, contextual learning, and multi-objective optimization.

The development of attention mechanisms in deep learning, especially the Transformer architecture, was directly inspired by psychological theories of selective attention (Posner and Petersen, 1990). Yet, unlike humans, AI lacks the intentional control that governs attention allocation based on motivation and goals. Integrating motivation-driven attention systems could thus enhance AI autonomy.

Equally central to AGI research is metacognition of the ability to monitor and regulate one's cognitive processes (Flavell, 1979). Humans use metacognition to detect errors, plan strategies, and reflect on outcomes. Replicating these abilities in artificial agents would enable them to evaluate uncertainty, improve self-learning, and develop rudimentary self-awareness (O'Reilly et al., 2016).

Moreover, the embodied cognition framework (Varela et al., 1991) emphasizes that intelligence arises from dynamic interaction between mind, body, and environment. This challenges purely computational notions of intelligence, suggesting that AGI must incorporate sensorimotor experiences to achieve grounded understanding. Modern research in robotic learning and embodied AI (Brooks, 1999) reflects this transition, highlighting the importance of physical and social contexts in intelligent behavior.

3. Analysis and Results

The analytical phase of this research builds upon the theoretical and empirical insights reviewed in the preceding sections, aiming to uncover the functional parallels between human cognition and artificial intelligence architectures. While neuroscience provides a biological account of how intelligence emerges from complex neural dynamics, cognitive psychology elucidates the structural and procedural organization of human thought. Together, these domains offer a multidimensional framework for understanding the mechanisms that may guide the evolution from narrow machine intelligence toward genuine Artificial General Intelligence (AGI).

Table 1 Neuroscientific principles and their implications for AGI development

Neuroscientific Mechanism	Empirical Evidence and Function	Potential Application in AGI Design
Neural Plasticity	Brain circuits reorganize in response to new stimuli; supports lifelong learning (Yuste, 2015).	Implementation of *continual learning* systems that dynamically adapt to novel environments without catastrophic forgetting.
Predictive Coding	Cortical networks minimize prediction error via hierarchical inference (Friston, 2010; Clark, 2013).	Design of *hierarchical predictive architectures* capable of real-time learning and adaptation to uncertainty.
Neural Synchronization	Consciousness and attention emerge from synchronized neural oscillations (Dehaene and Changeux, 2011).	Development of *modular coordination networks* that integrate distributed AI subsystems through temporal coherence.
Energy Efficiency	The human brain uses ~20 W to achieve massive parallel processing (Hassabis et al., 2017).	Creation of *neuromorphic computing architectures* using event-driven spiking networks for efficient processing.
Global Workspace Integration	Information becomes conscious when globally broadcast to specialized regions (Baars, 1988).	Development of *meta-controller systems* that integrate information across modules to support reasoning and awareness.

Source: Compiled by the author

The analysis of neuroscientific mechanisms demonstrates that biological efficiency and adaptability are central to human-level intelligence. Unlike artificial networks, which are optimized for single-task performance, the human brain exhibits structural flexibility, contextual prediction, and global integration.

The principle of neural plasticity directly challenges the rigidity of deep neural networks by suggesting architectures that continuously evolve rather than being statically trained. Similarly, the predictive coding model provides a mathematical foundation for the brain's ability to infer and anticipate environmental stimuli, a property that could enhance machine adaptability under uncertainty.

Moreover, neural synchronization and global workspace integration underscore that intelligence is not merely computational but coordinative—a product of harmonized communication among specialized subsystems. This insight motivates the design of meta-cognitive controllers in AGI that dynamically allocate computational attention and integrate multi-modal data streams.

Finally, the energy efficiency of the brain highlights a critical engineering challenge. While current large language models demand gigawatts of computational power, the biological brain achieves superior efficiency through spike-based signaling and selective activation. Replicating these mechanisms in neuromorphic chips could significantly advance sustainable AGI development.

Table 2 Cognitive psychology constructs and their relevance for AGI architecture

Cognitive Function / Model	**Psychological Framework / Evidence**	**Analogous Implementation in AGI**
Working Memory (Baddeley, 1992)	Integration of sensory, episodic, and executive subsystems for temporary information manipulation.	Design of *contextual memory modules* for dynamic reasoning and task sequencing.
Selective Attention (Posner and Petersen, 1990)	Allocation of cognitive resources to relevant stimuli; goal-directed information prioritization.	*Transformer-based attention models* that emulate human focus and context weighting.
Metacognition (Flavell, 1979)	Self-monitoring and regulation of thought processes; error detection and learning from feedback.	Creation of *self-evaluation layers* that monitor model confidence, uncertainty, and ethical decision-making.
Embodied Cognition (Varela et al., 1991)	Intelligence as interaction between body, mind, and environment; sensorimotor grounding.	Integration of *sensorimotor loops* in embodied AI and robotics for situated learning.
Executive Control (Anderson, 2007; Laird, 2012)	Coordination of cognitive modules for planning, inhibition, and decision-making.	Implementation of *cognitive control architectures* that manage submodules via top-down goal hierarchies.

Source: Compiled by the author

The findings from cognitive psychology emphasize that human intelligence is modular yet integrated, reflective yet goal-directed, and contextually grounded. The working memory model suggests that flexible reasoning in AGI requires the ability to retain and manipulate multiple information streams simultaneously an aspect underrepresented in current deep learning models. Implementing episodic buffers and short-term contextual layers may bridge this gap.

The development of attention mechanisms in modern AI already demonstrates partial success in mimicking human focus. However, cognitive theories highlight that attention is not purely reactive but intentional, guided by internal goals and motivational states. Thus, AGI systems must evolve beyond passive attention models toward purpose-driven attentional control.

The incorporation of metacognitive processes allowing a system to reflect on its performance marks a critical step toward self-awareness. By embedding self-evaluation layers that assess uncertainty and ethical boundaries, AGI could demonstrate adaptive self-correction and accountability, reducing risks of misalignment.

Lastly, the embodied cognition perspective shifts the AGI debate from purely symbolic reasoning to sensorimotor interaction. This underscores the notion that genuine understanding cannot be achieved without grounding perception in physical and social experience. As a result, embodied AGI agents could better interpret contextual cues, emotions, and causality central to human intelligence.

4. Conclusion and Recommendations

The findings of this study underscore that the pursuit of Artificial General Intelligence (AGI) must move beyond purely algorithmic or data-driven paradigms toward architectures inspired by the adaptive, integrative, and reflective nature of human cognition. The convergence of neuroscience and cognitive psychology reveals that intelligence is not a static computational state, but an evolving, context-sensitive process characterized by continuous learning, prediction, and self-regulation. Thus, the following recommendations emerge as essential strategic directions for the next phase of AGI research.

First, AGI development should incorporate neurobiologically inspired architectures that emulate plasticity, hierarchical inference, and energy-efficient computation. Implementing mechanisms analogous to neural plasticity and predictive coding could enable systems capable of continuous adaptation and error minimization under uncertain environments. Moreover, embedding synchronization and global workspace dynamics may allow artificial agents to integrate distributed processes, forming coherent awareness and decision-making like human cognition.

Second, the psychological dimension of intelligence must be operationalized through cognitive-level modeling. Principles derived from working memory, attention, and metacognition should inform AGI's executive control systems, allowing them to manage multiple goals, reflect on internal states, and adjust learning strategies autonomously. Introducing metacognitive self-assessment modules would enable artificial systems to evaluate their confidence, uncertainty, and ethical boundaries crucial for maintaining interpretability and safety in complex real-world interactions.

Third, AGI research should prioritize embodied and context-aware learning frameworks. Cognitive science demonstrates that human understanding emerges from sensorimotor engagement with the environment. Therefore, integrating embodied cognition principles through robotics, multimodal perception, and reinforcement-based learning could ground AI understanding in real physical and social contexts, enhancing its capacity for abstraction and empathy.

Finally, the pathway toward AGI requires a unified interdisciplinary methodology, combining computational modeling, experimental neuroscience, and cognitive simulation. Collaboration between neuroscientists, psychologists, and computer scientists should form the foundation for hybrid models that merge symbolic reasoning with neural computation. Such integration will not only advance technical innovation but also promote ethical, transparent, and human-aligned intelligence.

In conclusion, this research demonstrates that the architecture of AGI must reflect the dual essence of intelligence, biological adaptability and cognitive intentionality. By synthesizing the structural insights of neuroscience with the functional frameworks of cognitive psychology, it becomes possible to design artificial systems capable of reasoning, learning, and reflecting in ways that approximate human-level understanding. The future of AGI lies not in replicating the brain's complexity, but in grasping the principles of integration, self-regulation, and contextual awareness that make human cognition truly general. Only through such interdisciplinary synthesis can the path to Artificial General Intelligence become scientifically grounded, ethically guided, and evolutionarily sustainable.

References

[1] Anderson, J. R. (2007). How can the human mind occur in the physical universe? Oxford University Press.

[2] Baars, B. J. (1988). A cognitive theory of consciousness. Cambridge University Press.

[3] Baddeley, A. D. (1992). Working memory. Science, 255(5044), 556–559.

[4] Bengio, Y., LeCun, Y., and Hinton, G. (2021). Deep learning for AI. Communications of the ACM, 64(7), 58–65.

[5] Brooks, R. A. (1999). Cambrian intelligence: The early history of the new AI. MIT Press.

[6] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.

[7] Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200–227.

[8] Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. Science, 338(6111), 1202–1205.

[9] Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911.

[10] Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138.

[11] Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. Neuron, 95(2), 245–258.

[12] Laird, J. E. (2012). The SOAR cognitive architecture. MIT Press.

[13] Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. arXiv preprint arXiv:2002.06177.

[14] Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. Communications of the ACM, 19(3), 113–126.

[15] O'Reilly, R. C., Hazy, T. E., and Herd, S. A. (2016). The Leabra cognitive architecture: How to play 20 principles with nature and win! In S. O. Murray and E. Margolis (Eds.), The Oxford handbook of cognitive science (pp. 91–120). Oxford University Press.

[16] Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. Annual Review of Neuroscience, 13(1), 25–42.

[17] Rumelhart, D. E., and McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. MIT Press.

[18] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117.

[19] Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. Nature Reviews Neuroscience, 17(7), 450–461.

[20] Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.

[21] Varela, F. J., Thompson, E., and Rosch, E. (1991). The embodied mind: Cognitive science and human experience. MIT Press.

[22] Yuste, R. (2015). From the neuron doctrine to neural networks. Nature Reviews Neuroscience, 16(8), 487–497.