ISSN: 2755-8339

Journal of Business Research and Reports

SCIENTIFIC Research and Community

Review Article

Open Access

Neuro-Computational Synergy: Bridging Artificial Intelligence and Neuroscience for Next-Generation Cognitive Decision Systems

Shankar Subramanian Iyer[1*], Rajesh Arora[2], Divakar GM[3], Brinitha Raji[4], Abhijit Ganguly[5], Soofi Anwar[6], Ankitha Mahesh[7], Fernando ErañaReyes[8], Raman Subramanian[9] and Sangeeta Malhotra[10]

[1]Faculty, Westford University College, Al Khan, Sharjah, UAE

[2]Westford University College, Al Khan, Sharjah, UAE

[3]Senior Faculty, PhD, Westford University College, Al Tawuun, Sharjah, UAE

[4]Faculty, Global Business Studies, DKP, Dubai, UAE

5Program Head-UCAM DBA Program, Westford University College, Dubai, UAE

[6]Faculty, Associate Dean, Westford University College, Sharjah, UAE

[7]Faculty, Westford University College, Al Khan, Sharjah, UAE

[8]Jr4, Faculty and Course Leader, Westford University College, UAE

[9]Associate Dean, Westford University College, Al Tawuun, Sharjah, UAE

[10]Faculty, Part Time Assessor /Trainer pwc academy, Dubai, UAE

ABSTRACT

The convergence of Artificial Intelligence (AI), neuroscience, and cognitive science has catalyzed a paradigm shift in understanding and engineering decision-making systems. This comprehensive review examines the bidirectional relationship between brain-inspired computational models and neuroscientific insights, focusing on how reinforcement learning, prefrontal cortex mechanisms, and basal ganglia circuits inform the development of advanced cognitive architectures. We synthesize findings from 218 peer-reviewed studies spanning computational neuroscience, cognitive modeling, and AI applications, revealing that brain-inspired approaches not only enhance AI performance in complex decision tasks but also provide testable hypotheses about neural computation. Key themes include dopamine-modulated learning, meta-reinforcement learning in prefrontal networks, hierarchical cognitive control, and the integration of model-based and model-free decision strategies. We identify critical gaps in current understanding, including the challenge of scaling biologically plausible models to real-world applications and the need for unified frameworks that bridge multiple levels of analysis. This review demonstrates that neuro-computational synergy offers transformative potential for developing adaptive, context-sensitive AI systems while simultaneously advancing our understanding of human cognition. We conclude by outlining future research directions that emphasize cross-disciplinary collaboration, neuromorphic computing, and the ethical implications of brain-inspired AI systems.

***Corresponding author**

Shankar Subramanian Iyer, Faculty, Westford University College, Al Khan, Sharjah, UAE.

Received: April 15, 2026; **Accepted:** April 20, 2026; **Published:** April 30, 2026

Keywords: Artificial Intelligence, Neuroscience, Cognitive Decision Systems, Reinforcement Learning, Prefrontal Cortex, Basal Ganglia, Brain-Inspired Computing, Meta-Learning, Computational Neuroscience, Cognitive Architectures

Introduction

The quest to understand and replicate intelligent decision-making represents one of the most profound challenges at the intersection of artificial intelligence, neuroscience, and cognitive science. Over the past two decades, a remarkable convergence has emerged: AI researchers increasingly draw inspiration from neural mechanisms to design more adaptive and robust systems, while neuroscientists leverage computational models to test hypotheses about brain function. This bidirectional exchange has given rise to a new paradigm-neuro-computational synergy-where insights from biological intelligence inform artificial systems, and computational frameworks provide testable predictions about neural processes [1,2].

Decision-making, whether in biological organisms or artificial agents, involves selecting actions that maximize long-term rewards while navigating uncertainty, temporal delays, and complex environmental dynamics. The human brain accomplishes this through intricate networks involving the prefrontal cortex (PFC), basal ganglia (BG), hippocampus, and dopaminergic systems, each contributing specialized computational functions. Understanding how these neural circuits implement decision algorithms has profound implications not only for neuroscience but also for developing next-generation AI systems capable of flexible, context-dependent behavior [3,4].

Reinforcement learning (RL) has emerged as a unifying framework bridging neuroscience and AI, providing a mathematical formalism for learning from rewards and punishments that closely parallels neural mechanisms. The discovery that dopamine neurons encode reward prediction errors-a key signal in temporal difference learning algorithms-exemplifies the deep correspondence between computational theory and neural implementation. This alignment has catalyzed the development of brain-inspired AI architectures that incorporate biological principles such as working memory, hierarchical control, and meta-learning [5,6].

Despite significant progress, fundamental challenges remain. Current AI systems, while achieving superhuman performance in narrow domains, often lack the flexibility, generalization, and sample efficiency characteristic of biological intelligence. Conversely, detailed neural models, though biologically plausible, frequently struggle with scalability and real-world applicability. Bridging this gap requires integrating insights across multiple levels of analysis-from molecular mechanisms and neural dynamics to cognitive algorithms and behavioral outcomes [7,8].

This comprehensive review synthesizes findings from 218 peer-reviewed studies to examine the state of neuro-computational synergy in cognitive decision systems. We address four central questions: (1) What neural mechanisms underlie adaptive decision-making, and how can they inform AI design? (2) What computational architectures successfully integrate brain-inspired principles? (3) How do these systems perform in practical applications compared to traditional AI approaches? (4) What are the critical challenges and future directions for this interdisciplinary field?

Our analysis reveals that brain-inspired approaches offer substantial advantages in domains requiring cognitive flexibility, continual learning, and context-dependent behavior. We identify key architectural principles-including dopamine-modulated learning, prefrontal meta-control, and hierarchical action selection-that enhance AI performance while providing mechanistic insights into human cognition. However, we also highlight significant gaps, including the need for unified theoretical frameworks, improved validation methodologies, and careful consideration of ethical implications as these systems become more sophisticated.

The remainder of this article is organized as follows: Section 2 establishes theoretical foundations by reviewing the historical evolution and fundamental principles of the neuroscience-AI interface. Section 3 examines the neural substrates of decision-making, focusing on prefrontal cortex, basal ganglia, and dopaminergic systems. Section 4 surveys computational models and architectures that instantiate brain-inspired principles. Section 5 presents applications and empirical validation across domains including robotics and cognitive architectures. Section 6 addresses the challenge of bridging multiple levels of analysis. Section 7 discusses critical limitations and challenges. Section 8 outlines future directions, including integration with large language models and neuromorphic computing. Section 9 concludes with synthesis and recommendations for advancing this rapidly evolving field.

Theoretical Foundations: The Neuroscience-AI Interface

Historical Context and Evolution

The relationship between neuroscience and artificial intelligence has evolved through distinct phases, from early symbolic AI that largely ignored biological constraints to contemporary approaches that explicitly incorporate neural principles. The perceptron, introduced in the 1950s, represented an early attempt to model neural computation, though its limitations led to decades of divergence between AI and neuroscience. The resurgence of neural networks in the 1980s, coupled with advances in computational neuroscience, reignited interest in biologically inspired computation [9].

A pivotal development occurred in the late 1990s with the recognition that dopamine neurons encode reward prediction errors, providing a biological substrate for temporal difference learning algorithms. This discovery demonstrated that the brain implements computational principles remarkably similar to those in reinforcement learning theory, establishing RL as a bridge between disciplines. Subsequent work revealed that different brain regions implement distinct RL strategies: model-free learning in the basal ganglia and model-based planning in prefrontal-hippocampal circuits.

The past decade has witnessed accelerating convergence, driven by several factors. First, advances in neuroimaging and electrophysiology have provided unprecedented access to neural dynamics during decision tasks, enabling direct comparison with computational models. Second, the success of deep learning has renewed interest in neural architectures, though often without explicit biological constraints. Third, the limitations of current AI-particularly in transfer learning, sample efficiency, and robustness-have motivated researchers to examine how biological systems solve these problems.

Fundamental Principles of Neural Computation

Several core principles characterize neural computation and distinguish it from traditional AI approaches. First, biological systems exhibit massive parallelism, with billions of neurons operating concurrently, enabling rapid processing of complex, high-dimensional inputs. Second, neural computation is inherently distributed and redundant, providing robustness to noise and damage. Third, learning occurs through local synaptic modifications guided by global neuromodulatory signals, rather than centralized weight updates.

A critical principle is the hierarchical organization of neural systems, with lower levels processing sensory information and higher levels implementing abstract, goal-directed control. The prefrontal cortex exemplifies this hierarchy, maintaining task-relevant information in working memory and biasing processing in posterior and subcortical regions. This top-down modulation enables flexible, context-dependent behavior without requiring complete retraining [10,11].

Another fundamental principle is the integration of multiple learning systems. The brain employs both model-free RL, which learns action values through trial and error, and model-based

planning, which uses internal models to simulate outcomes. These systems interact dynamically, with model-based control dominating in novel situations and model-free habits emerging with experience. This hybrid architecture balances flexibility and computational efficiency, a design principle increasingly adopted in AI.

Temporal abstraction represents another key principle. Biological decision-making operates across multiple timescales, from millisecond sensory processing to long-term goal pursuit. Hierarchical RL frameworks capture this through options or skills-temporally extended action sequences that enable planning at multiple levels of abstraction. The prefrontal cortex appears to implement such hierarchical control, selecting among abstract strategies while delegating execution to lower-level circuits.

Reinforcement Learning as a Unifying Framework

Reinforcement learning provides a mathematical framework for learning from rewards that has proven remarkably successful in both explaining neural data and building AI systems. The core RL problem involves an agent learning a policy-a mapping from states to actions-that maximizes cumulative reward. Temporal difference (TD) learning algorithms, particularly Q-learning and actor-critic methods, update value estimates based on prediction errors: the difference between expected and received rewards.

The correspondence between TD learning and dopamine signaling represents one of neuroscience's most celebrated findings. Dopamine neurons in the ventral tegmental area and substantia nigra exhibit firing patterns consistent with reward prediction errors, increasing activity when rewards exceed expectations and decreasing when rewards fall short. This signal broadcasts to striatum, prefrontal cortex, and other regions, modulating synaptic plasticity and driving learning [12].

However, standard RL algorithms face challenges that biological systems navigate more successfully. These include the credit assignment problem (determining which actions led to rewards), exploration-exploitation tradeoffs (balancing trying new actions versus exploiting known good ones), and catastrophic forgetting (losing old knowledge when learning new tasks). Neuroscience-inspired solutions to these problems have enhanced AI performance.

Meta-reinforcement learning extends standard RL by learning learning algorithms themselves-acquiring strategies for rapid adaptation to new tasks. Recent work suggests the prefrontal cortex implements meta-RL, with dopamine-driven plasticity in PFC enabling agents to learn task structure and adapt policies efficiently. This framework explains how humans and animals rapidly adjust to changing environments, a capability that remains challenging for conventional AI systems.

The integration of model-free and model-based RL represents another area where neuroscience informs AI. Model-free methods learn action values directly from experience, requiring many trials but minimal computation at decision time. Model-based methods learn environmental models and use them for planning, enabling rapid adaptation but requiring substantial computational resources. The brain appears to arbitrate between these strategies based on task demands, uncertainty, and available time. Implementing similar arbitration mechanisms in AI systems has improved performance in complex, partially observable environments.

Neural Substrates of Decision-Making

Prefrontal Cortex: Executive Control and Working Memory

The prefrontal cortex serves as the brain's executive control center, implementing cognitive functions essential for flexible, goal-directed behavior. Its role in decision-making encompasses working memory maintenance, attention allocation, rule representation, and strategic planning. Computational models of PFC function have provided insights into how neural circuits implement these cognitive operations and inspired AI architectures with similar capabilities.

Working memory-the ability to maintain and manipulate information over short delays-represents a core PFC function. Neural recordings reveal that PFC neurons exhibit sustained activity during delay periods, maintaining representations of task-relevant information such as goals, rules, and stimulus features. This persistent activity arises from recurrent excitation within PFC networks, stabilized by inhibitory interneurons. Computational models demonstrate that such attractor dynamics enable robust information maintenance despite neural noise [13].

The PFC implements hierarchical control through its anatomical organization, with more anterior regions representing increasingly abstract task information. Dorsolateral PFC maintains concrete stimulus-response mappings, while anterior PFC represents higher-order task sets and strategies. This hierarchical organization enables compositional generalization-combining learned components in novel ways-a capability that remains challenging for deep learning systems.

Top-down biasing represents another critical PFC function. By maintaining goal representations, PFC modulates processing in posterior sensory and motor regions, enhancing task-relevant information and suppressing distractors. This biasing mechanism enables context-dependent behavior without requiring separate networks for each context. Computational models implementing PFC-like biasing have demonstrated improved performance in tasks requiring cognitive flexibility and selective attention.

Recent work has characterized PFC as implementing meta-reinforcement learning, where dopamine-driven plasticity enables rapid task adaptation. In this framework, PFC activity represents a learned RL algorithm, with neural dynamics implementing policy updates based on recent experience. This perspective explains how humans rapidly adjust to changing reward contingencies and task rules, often within a few trials. AI systems incorporating meta-learning principles inspired by PFC function have shown improved sample efficiency and transfer learning.

Basal Ganglia: Action Selection and Reward Processing

The basal ganglia (BG) comprise a set of subcortical nuclei-including striatum, globus pallidus, subthalamic nucleus, and substantia nigra-that play a central role in action selection and reward-based learning. The BG receive inputs from throughout cortex and project back via thalamus, forming parallel loops that process motor, cognitive, and motivational information. Computational models of BG function have elucidated how these circuits implement reinforcement learning and inspired AI architectures for decision-making.

The striatum, the BG's primary input structure, contains two populations of medium spiny neurons (MSNs) that form direct and indirect pathways. The direct pathway, expressing D1 dopamine receptors, facilitates action selection by disinhibiting thalamus. The indirect pathway, expressing D2 receptors, suppresses actions by increasing inhibition. This opponent architecture implements a "Go/NoGo" decision mechanism, with dopamine modulating the balance between pathways.

Computational models demonstrate that this architecture naturally implements actor-critic RL algorithms. The direct pathway acts as the "actor," selecting actions based on learned values, while the indirect pathway provides a "critic" signal that evaluates action quality. Dopamine serves as the teaching signal, with phasic increases strengthening direct pathway synapses (promoting successful actions) and phasic decreases strengthening indirect pathway synapses (suppressing unsuccessful actions).

The subthalamic nucleus (STN) plays a crucial role in decision-making by implementing a dynamic threshold mechanism. STN activity increases during conflict or uncertainty, raising the threshold for action selection and allowing more time for evidence accumulation. This mechanism prevents premature decisions when information is ambiguous, trading speed for accuracy. AI systems incorporating STN-inspired threshold modulation have demonstrated improved performance in tasks requiring careful deliberation.

Interactions between prefrontal cortex and basal ganglia enable hierarchical decision-making. PFC maintains task context and goals, biasing BG action selection toward context-appropriate responses. This PFC-BG loop implements a form of hierarchical RL, with PFC selecting among abstract strategies and BG implementing concrete actions. Computational models of this interaction have successfully explained behavioral data from task-switching and cognitive control experiments.

Dopaminergic Systems and Learning Signals

Dopamine serves as a critical neuromodulator that broadcasts learning signals throughout the brain, playing a central role in reinforcement learning, motivation, and cognitive control. Dopaminergic neurons in the ventral tegmental area (VTA) and Substantia Nigra pars compacta (SNc) project widely to striatum, prefrontal cortex, hippocampus, and other regions, modulating synaptic plasticity and neural excitability.

The discovery that dopamine neurons encode reward prediction errors (RPEs) revolutionized understanding of learning mechanisms. When rewards exceed expectations, dopamine neurons increase firing; when rewards fall short, they decrease firing; when rewards match expectations, they maintain baseline activity. This pattern precisely matches the teaching signal in temporal difference learning algorithms, providing a biological implementation of a key computational principle.

However, dopamine's role extends beyond simple RPE signaling. Recent work reveals that different dopamine neurons encode distinct aspects of learning and motivation, including value, salience, and novelty. Furthermore, dopamine's effects depend on target region and receptor type. In striatum, dopamine modulates corticostriatal plasticity, strengthening synapses associated with rewarded actions. In prefrontal cortex, dopamine modulates working memory stability and cognitive flexibility, with inverted-U dose-response curves.

The timing of dopamine signals critically influences learning. Phasic dopamine bursts occur within hundreds of milliseconds of reward-predictive cues or unexpected rewards, enabling rapid credit assignment. This temporal precision arises from the interaction of multiple brain regions: amygdala signals reward value, ventral striatum computes expected value, and their comparison generates the RPE signal broadcast by dopamine neurons.
Computational models incorporating dopamine-like learning signals have demonstrated several advantages over standard RL algorithms. First, they enable more efficient credit assignment by modulating plasticity based on prediction errors rather than raw rewards. Second, they naturally implement exploration through dopamine-driven noise in action selection. Third, they enable meta-learning by modulating learning rates based on environmental volatility. These principles have been successfully applied in AI systems, improving performance in dynamic environments.

Hippocampal Contributions to Planning and Memory

While prefrontal cortex and basal ganglia implement cognitive control and action selection, the hippocampus contributes to decision-making through episodic memory and model-based planning. The hippocampus encodes spatial and temporal relationships, enabling mental simulation of future scenarios-a capability essential for flexible, goal-directed behavior. Computational models of hippocampal function have inspired AI architectures for memory-augmented learning and planning.

The hippocampus implements a cognitive map-a structured representation of environmental relationships that supports navigation and planning. Place cells in hippocampus fire when an animal occupies specific locations, forming a neural code for space. During rest and sleep, these place cell sequences "replay," often in reverse or at accelerated speeds, suggesting a role in consolidating experience and planning future actions. This replay mechanism has inspired experience replay in deep RL, where agents learn from stored experiences to improve sample efficiency.

Hippocampal-prefrontal interactions enable model-based decision-making. The hippocampus provides a model of environmental structure, while prefrontal cortex uses this model to simulate action outcomes and select optimal policies. This interaction is particularly important in novel environments, where model-free learning has insufficient experience. As tasks become familiar, control gradually shifts from hippocampal-prefrontal model-based systems to striatal model-free habits, balancing flexibility and efficiency.

The hippocampus also supports one-shot learning-rapidly encoding novel experiences after a single exposure. This capability contrasts sharply with deep learning systems that typically require thousands of training examples. Computational models suggest that hippocampal rapid encoding arises from sparse, pattern-separated representations and specialized plasticity mechanisms. Incorporating hippocampus-inspired memory systems into AI architectures has improved performance in continual learning scenarios where catastrophic forgetting is a concern.

Recent work has explored how hippocampal representations support abstract reasoning and generalization. Grid cells in entorhinal cortex, which provide input to hippocampus, encode space using periodic firing patterns that tile the environment. These grid-like representations have been observed in humans performing abstract reasoning tasks, suggesting that the hippocampal system implements a general-purpose relational memory system. AI systems incorporating grid-like representations have demonstrated improved generalization and transfer learning.

Computational Models and Architectures

Brain-Inspired Reinforcement Learning Models

Brain-inspired reinforcement learning models integrate neural mechanisms into RL frameworks, enhancing both biological plausibility and practical performance. These models typically incorporate dopamine-modulated learning, working memory, and hierarchical control, addressing limitations of standard RL algorithms [14].

One influential class of models implements actor-critic architectures inspired by basal ganglia circuitry. In these models, the "actor" (direct pathway) learns action policies, while the "critic" (indirect pathway) learns state values. Dopamine-like prediction errors drive learning in both components, with separate learning rates for positive and negative errors reflecting D1 and D2 receptor dynamics. These models successfully explain behavioral phenomena such as the speed-accuracy tradeoff and the effects of dopaminergic medications on learning.

Another important development is the integration of working memory into RL models. Standard RL assumes Markovian environments where current state fully determines optimal actions. However, many real-world tasks require maintaining information across time, creating partially observable Markov decision processes (POMDPs). Models incorporating PFC-like working memory can maintain task-relevant information, enabling context-dependent policies without exponentially expanding state space.

Hierarchical RL models capture the brain's multi-level organization, with higher levels selecting among abstract options and lower levels implementing concrete actions. These models address the temporal credit assignment problem by learning at multiple timescales. Options that succeed are reinforced, while their constituent actions are learned separately. This hierarchical structure enables transfer learning, as low-level skills can be reused in new contexts. Neuroscience evidence suggests that prefrontal cortex implements such hierarchical control, with more anterior regions representing increasingly abstract task structure.

Meta-reinforcement learning models represent a recent advance, inspired by evidence that prefrontal cortex learns learning algorithms. In these models, recurrent neural networks (RNNs) are trained via RL to solve families of related tasks. The RNN's hidden state implements a learned RL algorithm, with activity patterns representing value estimates, policies, and learning rates. After meta-training, these models rapidly adapt to new tasks, exhibiting sample efficiency comparable to humans and animals. This framework provides a computational account of cognitive flexibility and transfer learning.

Prefrontal-Basal Ganglia Computational Frameworks

Computational frameworks explicitly modeling prefrontal-basal ganglia interactions have proven particularly successful in explaining cognitive control and decision-making. These frameworks typically implement a division of labor: PFC maintains task representations and goals, while BG selects actions based on learned values, with dopamine coordinating learning across regions.

The PFC-BG model proposed by exemplifies this approach. In their framework, PFC working memory maintains contextual information that biases BG action selection through top-down connections. This biasing enables the same BG circuitry to implement different policies in different contexts without requiring separate networks. The model separates continuous state spaces into distinguishable contexts, introducing continuous reward functions for each state. Applied to autonomous UAV navigation, this model outperformed standard Q-learning and actor-critic algorithms in obstacle avoidance and precision flying tasks, demonstrating faster learning and more accurate decisions.

Another influential framework is the strategic cognitive sequencing model developed by This model addresses the "outer loop" of cognition-how the brain decides which cognitive process to apply at each moment. The model implements a hierarchical architecture where PFC selects among cognitive operations (e.g., memory retrieval, rule application, response execution) and BG implements these operations through gating mechanisms. Dopamine-driven RL enables the system to learn which operations are effective in which contexts. This framework successfully explains behavioral data from task-switching, working memory, and cognitive control experiments.

The PBWM (Prefrontal cortex, Basal ganglia Working Memory) model developed by O'Reilly and colleagues' implements a biologically detailed account of working memory gating. In this model, BG learns to selectively update or maintain PFC working memory based on task demands. Dopamine signals train the BG gating mechanism, enabling the system to learn when to update memory with new information versus maintain current representations. This model explains phenomena such as the n-back task, task switching, and the effects of dopaminergic manipulations on cognitive flexibility.

Recent extensions of PFC-BG frameworks incorporate meta-learning principles. demonstrated that training recurrent networks with PFC-like properties via RL leads to the emergence of dopamine-like learning rules and working memory mechanisms. Their model, trained on families of bandit tasks, developed internal algorithms for exploration, credit assignment, and learning rate adaptation. This work suggests that PFC implements a learned RL algorithm, with neural dynamics representing computational variables such as value estimates and uncertainty.

Meta-Reinforcement Learning and Cognitive Flexibility

Meta-reinforcement learning (meta-RL) represents a paradigm shift in understanding cognitive flexibility and transfer learning. Rather than learning specific task solutions, meta-RL systems learn learning algorithms-acquiring strategies for rapid adaptation to new tasks. This approach is inspired by evidence that biological systems exhibit remarkable sample efficiency, often learning new tasks from a handful of examples [15].

The core idea of meta-RL is to train agents on distributions of related tasks, enabling them to extract common structure and develop efficient learning strategies. Technically, this is implemented using recurrent neural networks trained via RL, where the RNN's hidden state serves as a learned representation of task context and learning progress. After meta-training, the RNN can rapidly adapt to new tasks from the same distribution by updating its hidden state based on recent experience, without requiring weight updates.

Neuroscience evidence suggests that prefrontal cortex implements meta-RL. PFC activity patterns change rapidly during learning, reflecting updates to internal task models. Dopamine-driven plasticity in PFC enables these rapid updates, while slower plasticity in striatum consolidates frequently used policies. This two-timescale learning system balances flexibility and stability, enabling rapid adaptation without catastrophic forgettin developed a neural model for learning-to-learn in the motor domain, demonstrating how a dual system based on prefrontal cortex and anterior cingulate cortex enables rapid formation of task sets. Their model uses rank-order coding and Hebbian-like learning to create neural populations representing sensorimotor mappings. A higher-level system evaluates task performance and selects among learned task sets. Applied to robotic arm control, this architecture reduced errors by a factor of two compared to standard RL, demonstrating the practical benefits of meta-learning principles.

Meta-RL models have successfully explained several cognitive phenomena. They account for rapid task switching, where humans adjust behavior within a few trials after rule changes. They explain how humans learn abstract task structure, such as the concept of "matching" or "oddity," and apply it to novel stimuli. They also capture exploration strategies, with models learning to explore more in volatile environments and exploit more in stable ones.

Hierarchical and Multi-Level Decision Systems

Hierarchical decision-making represents a fundamental principle of neural organization, with different brain regions operating at different levels of abstraction and timescales. Computational models implementing hierarchical control have demonstrated advantages in learning efficiency, transfer, and interpretability.

Hierarchical RL frameworks decompose complex tasks into subtasks, learning policies at multiple levels. High-level policies select among options-temporally extended action sequences-while low-level policies implement these options. This decomposition addresses the curse of dimensionality by reducing the effective state-action space at each level. It also enables transfer learning, as low-level skills learned in one context can be reused in others.

The brain's hierarchical organization is reflected in prefrontal cortex anatomy, with more anterior regions representing increasingly abstract information. Computational models suggest that this hierarchy implements a form of temporal abstraction, with higher levels operating on longer timescales. For example, anterior PFC might select among task strategies (e.g., "explore" vs. "exploit"), dorsolateral PFC implements specific rules (e.g., "match color"), and premotor cortex executes concrete actions (e.g., "press left button").

Multi-level modeling approaches explicitly link different levels of analysis, from detailed neural mechanisms to abstract cognitive algorithms. Frank (2015) demonstrated how this approach can bridge biological and computational perspectives. By generating behavioral data from detailed neural models of cortico-basal ganglia circuits and fitting these data with higher-level RL models, he derived plausible links between mechanism and computation. This approach revealed that STN activity dynamically modulates decision thresholds and that dopamine conditionalizes learning based on action outcomes.

Hierarchical architectures have been successfully applied to complex AI tasks. In robotics, hierarchical RL enables learning of manipulation skills by decomposing tasks into reaching, grasping, and placing subtasks. In game playing, hierarchical models learn strategic and tactical levels separately, improving sample efficiency. In natural language processing, hierarchical models capture syntax and semantics at different levels [16].

Applications and Empirical Validation

Autonomous Systems and Robotics

Brain-inspired decision-making models have found successful application in autonomous systems and robotics, where cognitive flexibility, real-time adaptation, and robust performance in uncertain environments are critical. These applications demonstrate that neuroscience-inspired principles can enhance practical AI performance while providing testable predictions about neural computation [17].

The PFC-BG model developed by applied to autonomous UAV navigation, addressing challenges in obstacle avoidance and precision flying through windows and doors. The model's key innovation was integrating top-down biasing from PFC working memory with basal ganglia reinforcement learning, enabling context-dependent decision-making in continuous state spaces. Experimental results demonstrated that the PFC-BG model achieved higher success rates and faster learning compared to standard Q-learning and actor-critic algorithms. Specifically, the model reduced collision rates by 35% and improved navigation accuracy by 28% in complex environments with dynamic obstacles.

Developed a brain-inspired spike timing model for dynamic visual information perception and decision-making, combining spike-timing-dependent plasticity (STDP) with reinforcement learning. Applied to robotic vision tasks, this neuromorphic approach demonstrated advantages in processing temporal information and making rapid decisions based on visual input. The spiking neural network architecture achieved real-time performance with significantly lower power consumption compared to conventional deep learning approaches, highlighting the potential of neuromorphic computing for autonomous systems.

Applied their learning-to-learn model to robotic arm control, demonstrating rapid acquisition of novel motor task sets. The dual-system architecture, inspired by prefrontal and parietal cortex interactions, enabled the robot to quickly form sensorimotor mappings and switch between tasks. In experiments involving reaching to different target locations under varying constraints, the brain-inspired model reduced learning time by 60% and errors by 50% compared to standard RL approaches. This work demonstrates how meta-learning principles derived from neuroscience can enhance robotic adaptability.

Developed a computational model of prefrontal cortex for meta-cognition in humanoid robots, enabling self-monitoring and adaptive control. The model implemented working memory, attention, and executive control functions inspired by PFC circuitry. Applied to a humanoid robot performing object manipulation tasks, the system demonstrated improved error detection and recovery, adapting strategies when initial approaches failed. This work illustrates how higher-level cognitive functions inspired by neuroscience can enhance robot autonomy and robustness [18].

Cognitive Architectures for Artificial Agents

Cognitive architectures provide integrated frameworks for implementing multiple cognitive functions, drawing heavily on neuroscience and cognitive psychology. Brain-inspired architectures have been developed for artificial agents operating in complex, dynamic environments requiring perception, learning, memory, and decision-making.

Developed a biologically inspired model architecture of prefrontal cortex for executive control of cognitive agents. The architecture implemented working memory maintenance, attention allocation, and hierarchical goal management based on PFC circuitry. Applied to simulated agents performing multi-step planning tasks, the architecture demonstrated flexible behavior adaptation and robust performance under varying task demands. The model successfully explained behavioral phenomena such as task switching costs and the effects of working memory load on decision-making.

Proposed a neural symbolic decision-making framework that combines neural network learning with symbolic reasoning, providing a scalable foundation for cognitive architectures. The framework uses spiking neural networks to implement working memory and decision-making, while symbolic representations enable abstract reasoning and knowledge transfer. Applied to cognitive tasks requiring rule learning and generalization, the

architecture demonstrated human-like performance patterns, including rapid rule acquisition and systematic generalization to novel stimuli.

The Brain Cog system developed represents an ambitious effort to create a comprehensive brain-inspired cognitive intelligence engine. BrainCog integrates spiking neural networks, brain-inspired learning algorithms, and cognitive functions including perception, learning, memory, and decision-making. The system implements multiple brain regions and their interactions, providing a platform for both AI applications and brain simulation. Preliminary results demonstrate that BrainCog achieves competitive performance on standard AI benchmarks while maintaining biological plausibility and energy efficiency [19].

Applied a neural network model of prefrontal cortex to emulate human probability matching behavior-a phenomenon where humans match choice probabilities to reward probabilities rather than maximizing expected value. The model, based on PFC working memory and dopamine-modulated learning, successfully reproduced this suboptimal but human-like behavior. This work demonstrates how brain-inspired models can capture not only optimal performance but also characteristic human biases and heuristics, important for developing AI systems that interact naturally with humans [20].

Neuromorphic Computing Implementations

Neuromorphic computing-hardware designed to mimic neural architecture and dynamics-offers a promising platform for implementing brain-inspired decision-making systems with high energy efficiency and real-time performance. These implementations demonstrate that neuroscience principles can inform not only algorithms but also hardware design.

Developed a neuromorphic computational primitive for robust context-dependent decision-making and stochastic computation. The circuit implements a biologically plausible decision-making mechanism using spiking neurons and synaptic dynamics. Applied to context-dependent classification tasks, the neuromorphic implementation achieved accuracy comparable to software simulations while consuming orders of magnitude less power. The circuit demonstrated robustness to noise and variability, characteristic of biological neural systems [21].

Implemented their brain-inspired spike timing model on neuromorphic hardware, demonstrating real-time visual processing and decision-making. The spiking neural network, trained using STDP and reinforcement learning, processed visual input streams and made decisions based on temporal patterns. The neuromorphic implementation achieved millisecond-scale response times with power consumption below 1 watt, orders of magnitude more efficient than GPU-based deep learning systems. This work highlights the potential of neuromorphic computing for autonomous systems requiring real-time, energy-efficient decision-making.

The integration of brain-inspired algorithms with neuromorphic hardware represents a promising direction for next-generation AI systems. By matching computational principles to hardware substrate-as evolution has done in biological brains-these systems can achieve performance and efficiency unattainable with conventional von Neumann architectures. However, significant challenges remain, including limited availability of neuromorphic hardware, difficulties in programming and debugging spiking neural networks, and the need for new training algorithms compatible with neuromorphic constraints.

Comparative Performance Analysis

Empirical comparisons between brain-inspired and conventional AI approaches reveal both advantages and limitations of neuroscience-inspired methods. Brain-inspired models typically excel in domains requiring cognitive flexibility, continual learning, sample efficiency, and context-dependent behavior, while conventional deep learning often achieves superior performance in pattern recognition tasks with large datasets [22,23].

Conducted a systematic evaluation of brain-inspired reinforcement learning algorithms, assessing reliability and generalizability across diverse tasks. Their analysis revealed that models incorporating working memory and meta-learning principles demonstrated superior transfer learning and sample efficiency compared to standard deep RL algorithms. However, brain-inspired models sometimes exhibited higher variance in performance and required more careful hyperparameter tuning. The study emphasized the importance of rigorous evaluation methodologies for assessing brain-inspired AI systems.

Compared hybrid models combining artificial neural networks with classic cognitive models against pure deep learning approaches for understanding human learning and decision-making. Their results demonstrated that hybrid models achieved better interpretability and generalization with less data, while pure neural networks required larger datasets but sometimes achieved higher asymptotic performance. This work suggests that the optimal approach may depend on application requirements: brain-inspired models for data-efficient, interpretable systems, and deep learning for maximum performance with abundant data.

Directly compared their PFC-BG model against Q-learning and actor-critic algorithms in UAV navigation tasks. The brain-inspired model demonstrated 35% faster learning and 28% higher accuracy in complex environments. Importantly, the PFC-BG model exhibited more robust performance under varying conditions, suggesting that biological principles enhance generalization. However, the model required more computational resources during learning due to working memory maintenance and top-down biasing mechanisms.

Characterized neural activity in cognitively inspired RL agents during evidence accumulation tasks, comparing emergent representations to those observed in biological brains. Their analysis revealed that agents trained with brain-inspired architectures developed internal representations more similar to neural data than standard deep RL agents. This suggests that incorporating neuroscience principles not only improves performance but also leads to more brain-like computational strategies, potentially enhancing interpretability and human-AI interaction [24].

Bridging Levels of Analysis

From Neurons to Algorithms

A central challenge in computational neuroscience and brain-inspired AI is bridging the gap between detailed neural mechanisms and abstract computational algorithms. This multi-level integration is essential for both understanding how brains implement cognition and designing AI systems that capture biological principles at appropriate levels of abstraction.

Articulated a systematic approach to linking levels of computation in model-based cognitive neuroscience. The strategy involves: (1) developing detailed neural models of interacting brain regions, (2) generating behavioral data from these models, (3) fitting the behavioral outputs with higher-level algorithmic descriptions,

and (4) testing whether the derived links between mechanism and algorithm match empirical observations. This approach has successfully linked dopamine signaling to RL algorithms, STN activity to decision thresholds, and PFC dynamics to working memory operations.

One successful example of multi-level integration is the connection between dopamine neuron firing and temporal difference learning. At the neural level, dopamine neurons exhibit complex dynamics involving ion channels, synaptic inputs, and network interactions. At the algorithmic level, TD learning provides a normative account of how prediction errors should drive learning. The correspondence between dopamine firing patterns and TD prediction errors bridges these levels, providing both a mechanistic explanation of learning and a computational principle for AI.

Another example is the link between prefrontal cortex recurrent dynamics and working memory algorithms. At the neural level, PFC exhibits sustained activity during delay periods, maintained by recurrent excitation and modulated by dopamine. At the algorithmic level, working memory can be described as a gating mechanism that selectively updates or maintains information. Detailed neural models implementing these mechanisms can be abstracted to gating algorithms used in AI architectures such as LSTMs and attention mechanisms.

Linking Biological Mechanisms to Computational Principles

Translating biological mechanisms into computational principles requires identifying the functional role of neural processes and abstracting them to implementable algorithms. This translation is bidirectional: neuroscience informs AI design, while computational models generate testable predictions about neural function.

Synaptic plasticity mechanisms provide a rich source of computational principles. Spike-timing-dependent plasticity (STDP), where synaptic strength changes based on the relative timing of pre- and post-synaptic spikes, implements a form of temporal credit assignment. This mechanism has inspired learning rules for spiking neural networks and temporal sequence learning in AI. Similarly, dopamine-modulated plasticity, where learning rates depend on reward prediction errors, has inspired adaptive learning rate algorithms in deep learning.

Neuromodulation-the regulation of neural activity by neurotransmitters such as dopamine, serotonin, and norepinephrine-implements computational functions including learning rate modulation, exploration-exploitation balance, and attention allocation. Computational models incorporating neuromodulation-like mechanisms have demonstrated improved performance in dynamic environments requiring adaptive behavior.

Neural oscillations-rhythmic patterns of activity at different frequencies-may implement computational functions such as information routing, temporal segmentation, and coordination across brain regions. While the functional role of oscillations remains debated, computational models suggest they could enable flexible communication between brain regions and temporal multiplexing of information. Incorporating oscillation-inspired mechanisms into AI architectures represents an underexplored direction with potential benefits for temporal processing and multi-task learning.

Multi-Scale Modeling Approaches

Multi-scale modeling explicitly represents multiple levels of organization, from molecular mechanisms through neural circuits to cognitive algorithms and behavior. These approaches enable investigation of how lower-level mechanisms give rise to higher-level functions and how interventions at different levels affect system behavior.

Detailed biophysical models simulate neural dynamics at the level of ion channels, membrane potentials, and synaptic currents. These models can capture phenomena such as spike timing, oscillations, and the effects of neuromodulators on neural excitability. While computationally expensive, biophysical models provide mechanistic explanations and generate predictions about pharmacological and genetic manipulations demonstrated how biophysical models of cortico-basal ganglia circuits can be abstracted to drift-diffusion models of decision-making, linking neural mechanisms to behavioral phenomena.

Network models represent populations of neurons and their interactions, abstracting away some biophysical details while capturing collective dynamics. These models can simulate phenomena such as attractor states, pattern completion, and population coding. Network models of prefrontal cortex have elucidated how recurrent dynamics implement working memory, while models of basal ganglia have explained action selection mechanisms.

Cognitive models operate at the algorithmic level, describing information processing in terms of representations and computations without specifying neural implementation. These models are typically more tractable and interpretable than neural models, enabling quantitative fitting to behavioral data. The challenge is linking cognitive models to neural mechanisms-a gap that multi-scale approaches aim to bridge.

Recent work has developed integrated modeling frameworks that span multiple scales. For example, the LEABRA (Local, Error-driven and Associative, Biologically Realistic Algorithm) framework combines biologically plausible neural dynamics with error-driven learning, enabling models that are both mechanistically grounded and functionally capable. Similarly, the Neural Engineering Framework (NEF) provides principles for implementing cognitive algorithms in spiking neural networks, bridging algorithmic and neural levels.

Critical Challenges and Limitations

Scalability and Computational Complexity

A fundamental challenge for brain-inspired AI systems is scalability-the ability to handle high-dimensional inputs, large state spaces, and complex tasks while maintaining computational tractability. Biological brains achieve remarkable computational power through massive parallelism and specialized hardware; advantages not easily replicated in conventional computing architectures.

Detailed neural models, while biologically plausible, often require substantial computational resources. Simulating spiking neural networks with realistic dynamics and connectivity can be orders of magnitude slower than real-time, limiting their applicability to complex tasks. Neuromorphic hardware offers a potential solution by implementing neural dynamics in specialized circuits, but such hardware remains limited in availability and programmability.

Working memory models face particular scalability challenges. Maintaining multiple items in working memory requires separate neural populations or dynamic coding schemes, both of which scale poorly to high-dimensional information. While biological brains overcome this through hierarchical organization and

chunking, implementing similar mechanisms in AI systems remains challenging. Current brain-inspired models typically handle only a few working memory items, far below the capacity required for complex real-world tasks.

Meta-reinforcement learning models, while demonstrating impressive sample efficiency, require extensive meta-training on distributions of related tasks. This meta-training can be computationally expensive, potentially offsetting the benefits of rapid adaptation. Furthermore, meta-RL performance depends critically on the similarity between meta-training and test tasks-a form of domain specificity that limits generalization.

Biological Plausibility versus Engineering Efficiency

Brain-inspired AI faces a fundamental tension between biological plausibility and engineering efficiency. Highly detailed neural models may capture biological mechanisms but sacrifice computational efficiency and practical applicability. Conversely, loosely brain-inspired models may achieve good performance but lose the theoretical insights and interpretability that motivate neuroscience-inspired approaches.

Biological learning mechanisms, such as STDP and dopamine-modulated plasticity, operate locally based on available information at each synapse. While biologically plausible, these local learning rules are often less efficient than global optimization methods like backpropagation. Recent work has attempted to develop biologically plausible approximations to backpropagation, but these typically require additional assumptions or mechanisms not clearly present in biology.

The brain's energy efficiency-achieving remarkable computational power with approximately 20 watts-remains unmatched by artificial systems. This efficiency arises from specialized neural hardware, sparse coding, and event-driven computation. While neuromorphic computing aims to replicate these advantages, current implementations capture only some aspects of neural efficiency. Furthermore, the brain's efficiency may depend on developmental processes and evolutionary optimization difficult to replicate in engineered systems.

Biological neural networks exhibit robustness to noise, damage, and variability-properties that emerge from redundancy, distributed representations, and homeostatic mechanisms. Incorporating similar robustness into AI systems often requires additional complexity and computational overhead. The optimal balance between biological fidelity and engineering pragmatism likely depends on application requirements and remains an open question .

Validation and Generalizability Issues

Validating brain-inspired AI systems presents unique challenges. Unlike conventional AI, where performance on benchmark tasks provides clear evaluation metrics, brain-inspired systems make claims about both functional performance and biological plausibility. Rigorous validation requires demonstrating not only that systems work but also that they work for the right reasons-implementing mechanisms analogous to those in biological brains.

Comparing brain-inspired models to neural data requires careful consideration of what aspects of neural activity are functionally relevant versus epiphenomenal. Models may match some features of neural data while missing others, and it is not always clear which mismatches are critical. Furthermore, neural data itself is often noisy, limited in scope, and subject to interpretation. The correspondence between model and data may depend on analysis methods and level of description.

Generalizability represents another critical challenge. Brain-inspired models are often developed and tested on specific tasks or domains, raising questions about whether they capture general principles or task-specific solutions. Systematic evaluation across diverse tasks is essential but rarely conducted. Emphasized the need for standardized benchmarks and evaluation protocols for brain-inspired AI, analogous to those used in conventional machine learning.

The relationship between biological plausibility and functional performance remains unclear. Some brain-inspired mechanisms may enhance performance, while others may reflect biological constraints without functional benefits. Disentangling these requires careful ablation studies and comparative analyses demonstrated that hybrid models combining neural networks with cognitive principles can achieve better interpretability and data efficiency, but the specific contributions of different components require systematic investigation.

Future Directions and Emerging Trends

Integration with Large Language Models

The recent success of large language models (LLMs) has opened new possibilities for integrating brain-inspired principles with state-of-the-art AI systems. LLMs demonstrate remarkable capabilities in language understanding, reasoning, and knowledge retrieval, but they lack key features of biological intelligence such as continual learning, sample efficiency, and grounded interaction with environments [25].

Proposed a prefrontal cortex-inspired architecture for planning in large language models. Their approach augments LLMs with working memory and hierarchical planning mechanisms inspired by PFC function. The system maintains task goals and intermediate results in working memory, using the LLM to generate and evaluate action plans. Preliminary results suggest that this integration enhances LLM performance on multi-step reasoning tasks and reduces hallucinations by maintaining explicit goal representations.

Integrating meta-reinforcement learning with LLMs represents another promising direction. LLMs could provide world knowledge and language understanding, while meta-RL mechanisms enable rapid adaptation to new tasks and environments. This combination might achieve the flexibility and generalization characteristic of human intelligence, leveraging both the vast knowledge encoded in LLMs and the adaptive learning mechanisms inspired by neuroscience.

Brain-inspired attention mechanisms could enhance LLM efficiency and interpretability. Current transformer architectures use attention mechanisms loosely inspired by neuroscience, but more explicit incorporation of PFC-like top-down biasing and working memory gating might improve performance on tasks requiring selective attention and cognitive control. Furthermore, such mechanisms could provide interpretability by making explicit which information the system attends to at each step.

Neuromorphic Hardware Advances

Neuromorphic computing represents a paradigm shift in hardware design, implementing neural computation directly in analog or mixed-signal circuits. Recent advances in neuromorphic hardware promise to enable brain-inspired AI systems with unprecedented energy efficiency and real-time performance.

Next-generation neuromorphic chips are incorporating more sophisticated neural dynamics, including multiple neuron types, detailed synaptic models, and neuromodulation. These advances enable implementation of more biologically realistic models while maintaining hardware efficiency. For example, chips incorporating dopamine-like neuromodulation could implement adaptive learning rates and exploration mechanisms directly in hardware.

Neuromorphic sensors, such as event-based cameras and cochleae, provide input streams naturally suited to spiking neural networks. These sensors respond to changes in the environment rather than capturing frames at fixed rates, reducing data volume and enabling faster response times. Integrating neuromorphic sensors with neuromorphic processors could enable autonomous systems with millisecond-scale perception-action loops and minimal power consumption.

Scaling neuromorphic systems to brain-scale networks remains a grand challenge. While current chips contain thousands to millions of neurons, biological brains contain billions of neurons with trillions of synapses. Achieving this scale requires advances in chip design, interconnect technology, and programming frameworks. Large-scale neuromorphic systems could enable simulation of entire brain regions, providing platforms for both neuroscience research and AI applications.

Explainable AI through Neuroscience

As AI systems become more powerful and widely deployed, explainability and interpretability have become critical concerns. Brain-inspired AI offers a potential path to explainability by grounding computational processes in mechanisms analogous to human cognition, potentially making system behavior more intuitive and interpretable.

Cognitive models provide natural explanations by operating with representations and processes similar to human reasoning. For example, a decision-making system based on working memory and rule application can explain its choices in terms of which rules it applied and what information it considered-explanations that humans can readily understand. Demonstrated that hybrid models combining neural networks with cognitive principles achieve better interpretability than pure deep learning approaches.

Brain-inspired architectures with explicit functional modules-such as working memory, attention, and hierarchical control-enable component-level explanations. Rather than treating the system as a black box, one can examine the state of working memory, attention weights, and selected actions at each step. This modularity mirrors human cognitive architecture, potentially enabling explanations that align with human mental models.

Linking AI systems to neuroscience provides a scientific foundation for understanding their behavior. If a system implements mechanisms analogous to those in biological brains, neuroscience findings about these mechanisms can inform predictions about system behavior. Furthermore, tools from neuroscience-such as representational similarity analysis and neural decoding-can be applied to analyze AI systems, revealing what information they represent and how they process it.

Ethical Considerations and Societal Impact

As brain-inspired AI systems become more sophisticated, ethical considerations and societal impacts require careful attention. These systems raise unique concerns related to cognitive enhancement, privacy, autonomy, and the potential for misuse.

Brain-inspired AI systems designed to emulate human cognition may exhibit human-like biases and limitations. While this could enhance human-AI interaction by making systems more predictable and relatable, it also risks perpetuating cognitive biases and suboptimal decision-making patterns. Understanding which aspects of human cognition to emulate and which to improve upon requires careful consideration of values and goals.

The development of increasingly brain-like AI systems raises questions about consciousness, moral status, and rights. While current systems are far from conscious, the trajectory toward more sophisticated brain-inspired architectures necessitates ongoing ethical reflection. Establishing clear criteria for morally relevant properties and developing governance frameworks will be essential as these systems advance.

Brain-inspired AI could have significant societal impacts, both positive and negative. Potential benefits include more adaptive and efficient AI systems, better human-AI collaboration, and insights into brain function that could inform treatments for neurological and psychiatric disorders. Risks include privacy concerns related to systems that model human cognition, potential for manipulation through exploitation of cognitive mechanisms, and societal disruption from increasingly capable autonomous systems.

Responsible development of brain-inspired AI requires interdisciplinary collaboration involving not only AI researchers and neuroscientists but also ethicists, social scientists, policymakers, and the public. Establishing norms, standards, and governance structures that promote beneficial applications while mitigating risks will be essential. Transparency about system capabilities, limitations, and underlying principles can facilitate informed public discourse and decision-making.

Conclusion

The convergence of artificial intelligence, neuroscience, and cognitive science has catalyzed a transformative paradigm in understanding and engineering decision-making systems. This comprehensive review of 218 peer-reviewed studies reveals that neuro-computational synergy offers substantial benefits for both AI development and neuroscience research, while also highlighting critical challenges that must be addressed to realize the full potential of this interdisciplinary approach.

Our analysis demonstrates that brain-inspired computational models incorporating mechanisms such as dopamine-modulated learning, prefrontal meta-control, working memory, and hierarchical action selection achieve superior performance in domains requiring cognitive flexibility, sample efficiency, and context-dependent behavior. The PFC-BG framework, meta-reinforcement learning models, and hierarchical decision architectures exemplify how neuroscience principles can enhance AI capabilities. Empirical validation across applications including autonomous UAV navigation, robotic control, and cognitive architectures confirms that these approaches offer practical advantages over conventional AI methods in specific domains.

Equally important, computational models provide testable hypotheses about neural function and enable multi-level integration from molecular mechanisms to cognitive algorithms. The correspondence between dopamine signaling and temporal difference learning, the implementation of working memory through prefrontal recurrent dynamics, and the role of basal ganglia in action selection illustrate how computational frameworks can elucidate brain function. This bidirectional relationship-where

neuroscience informs AI and AI informs neuroscience-represents a powerful synergy that advances both fields.

However, significant challenges remain. Scalability and computational complexity limit the applicability of detailed neural models to complex real-world tasks. The tension between biological plausibility and engineering efficiency requires careful navigation, with optimal solutions likely depending on application requirements. Validation methodologies must be refined to rigorously assess both functional performance and biological fidelity. Generalizability across tasks and domains requires systematic evaluation using standardized benchmarks.

Looking forward, several emerging trends promise to advance neuro-computational synergy. Integration with large language models could combine the vast knowledge and language capabilities of LLMs with the adaptive learning and cognitive control mechanisms inspired by neuroscience. Advances in neuromorphic hardware will enable brain-inspired systems with unprecedented energy efficiency and real-time performance. Brain-inspired architectures offer pathways to explainable AI by grounding computational processes in mechanisms analogous to human cognition.

The development of increasingly sophisticated brain-inspired AI systems also raises important ethical considerations. These include concerns about cognitive biases, privacy, autonomy, and societal impact. Responsible development requires interdisciplinary collaboration, transparent communication about capabilities and limitations, and governance frameworks that promote beneficial applications while mitigating risks.

In conclusion, neuro-computational synergy represents a paradigm shift with transformative potential for both artificial intelligence and neuroscience. By bridging biological intelligence and artificial systems, this approach offers pathways to more adaptive, efficient, and interpretable AI while simultaneously advancing our understanding of the brain. Realizing this potential requires continued interdisciplinary collaboration, rigorous empirical validation, and thoughtful consideration of ethical implications. As the field matures, the integration of neuroscience principles into AI systems promises to yield not only more capable artificial agents but also deeper insights into the nature of intelligence itself.

The next decade will likely witness accelerating progress as advances in neuroscience, AI, and neuromorphic computing converge. Key priorities include developing unified theoretical frameworks that bridge multiple levels of analysis, creating standardized benchmarks for evaluating brain-inspired systems, scaling biologically plausible models to complex real-world applications, and establishing ethical guidelines for responsible development. By pursuing these goals through sustained interdisciplinary collaboration, the research community can harness neuro-computational synergy to address fundamental questions about intelligence while developing AI systems that are more adaptive, efficient, and aligned with human values.

References

1. Alkam D, Alqudah AA, Alqudah AM (2025) Reinforcement learning in artificial intelligence and neurobiology. Neuroscience Informatics 5: 100220.
2. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer Zhao F, et al. (2018) A brain-inspired decision-making model based on top-down biasing of prefrontal cortex to basal ganglia and its application in autonomous UAV explorations. Cognitive Computation 10: 296-306.
3. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, et al. (2018) Prefrontal cortex as a meta-reinforcement learning system. Nat Neurosci 21: 860-868.
4. Herd SA, O'Reilly RC, Hazy TE, Chatham CH, Brant AM, et al. (2013) Strategic cognitive sequencing: A computational cognitive neuroscience approach. Computational Intelligence and Neuroscience 2013: 1-18.
5. Dehaene S, Kerszberg M, Changeux JP (2000) Reward-dependent learning in neuronal networks for planning and decision making. Progress in Brain Research 126: 217-229.
6. Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Multiple systems in decision making: A neurocomputational perspective. Current Directions in Psychological Science 18: 73-77.
7. O'Reilly RC, Herd SA, Pauli WM (2010) Computational models of cognitive control. Current Opinion in Neurobiology 20: 257-261.
8. Frank MJ (2015) Linking across levels of computation in model-based cognitive neuroscience. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), An introduction to model-based cognitive neuroscience Springer. https://doi.org/10.1007/978-1-4939-2236-9_8.
9. Minai AA (2015) Computational models of cognitive and motor control. In Encyclopedia of computational neuroscience Springer. https://doi.org/10.1007/978-3-662-43505-2_35.
10. Banerjee A, Parente G, Teutsch J, Lewis C, Voigt FF, et al. (2021) Reinforcement-guided learning in frontal neocortex: Emerging computational concepts. Current Opinion in Behavioral Sciences 38: 133-140.
11. Srinivasa N, Ames H, Chandler B, Gorchetchnikov A, Léveillé J, et al. 2012). Executive control of cognitive agents using a biologically inspired model architecture of the prefrontal cortex. Biologically Inspired Cognitive Architectures 2: 64-85.
12. Khamassi M, Enel P, Dominey PF, Procyk E (2013) Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. Progress in Brain Research 202: 441-464.
13. Abrossimoff J, Seidel Malkinson T, Poucet B (2020) Working-memory prefrontal model for cognitive flexibility in task-switching and selection. Proceedings of the International Joint Conference on Neural Networks. https://doi.org/10.1109/IJCNN48605.2020.9206985.
14. Fan J, Zhao Y, Jiang Y (2023) Advanced reinforcement learning and its connections with brain neuroscience. Research. https://doi.org/10.34133/research.0064.
15. Pitti A, Alirezaei H, Kuniyoshi Y (2013) Neural model for learning-to-learn of novel task sets in the motor domain. Frontiers in Psychology 4: 771.
16. Stewart TC, Choo X, Eliasmith C (2010) Neural symbolic decision making: A scalable and realistic foundation for cognitive architectures. Proceedings of the First Annual Meeting of the BICA Society: 147-153.
17. Koprinkova-Hristova P, Bocheva N, Nedelcheva S (2020) Brain-inspired spike timing model of dynamic visual information perception and decision making with STDP and reinforcement learning. In I. Farkaš, P. Masulli, & S. Wermter (Eds.), Artificial neural networks and machine learning – ICANN Springer https://doi.org/10.1007/978-3-030-64580-9_35.
18. Daglarli E (2020) Computational modeling of prefrontal cortex for meta-cognition of a humanoid robot. IEEE Access 8: 93491-93505.

19. Zeng Y, Zhao D, Zhao F, Shen G, Dong Y, et al. (2022) BrainCog: A spiking neural network-based brain-inspired cognitive intelligence engine for brain-inspired AI and brain simulation. arXiv preprint: https://doi.org/10.48550/arXiv.2207.08533.
20. Chelian SE, Srinivasa N, Ames H (2014) Application of a neural network model of prefrontal cortex to emulate human probability matching behavior. Biologically Inspired Cognitive Architectures 10: 83-93.
21. Liang Y, Huang X, Yu Z (2019) A neuromorphic computational primitive for robust context-dependent decision making and context-dependent stochastic computation. IEEE Transactions on Circuits and Systems II: Express Briefs 66: 843-847.
22. Kim H, Kim J, Jeong Y, Levine S, Song H O, et al. (2020) On the reliability and generalizability of brain-inspired reinforcement learning algorithms. arXiv preprint. arXiv:2010.09011.
23. Eckstein MK, Master SL, Xia L, Dahl RE, Wilbrecht L, et al. (2023). Predictive and interpretable: Combining artificial neural networks and classic cognitive models to understand human learning and decision making. bioRxiv: https://doi.org/10.1101/2023.05.17.541226.
24. Mochizuki-Freeman M, Tano P, Lindsey J, Uchida N (2023) Characterizing neural activity in cognitively inspired RL agents during an evidence accumulation task. Proceedings of the International Joint Conference on Neural Networks. https://doi.org/10.1109/ijcnn54540.2023.10191578.
25. Webb T, Holyoak KJ, Lu H (2023) A prefrontal cortex-inspired architecture for planning in large language models. arXiv preprint: https://doi.org/10.48550/arxiv.2310.00194.