

Commentary

Motivation: The Soul of the Machine

Siwei Luo¹

1. Jiangxi Provincial Key Laboratory of Intelligent Information Processing and Affective Computing, Jiangxi Normal University, China

Drawing inspiration from the prefrontal cortex, the proposed system formalizes high-level cognitive functions—motivation, cognition, memory, and decision-making—through vectorized, parameterized and modularized representations organized as interacting subsystems. This theoretical framework establishes a scalable, generalizable, and biologically grounded architecture intended to advance the modeling of higher-order intelligence.

Corresponding author: Siwei Luo, luosiwei@jxnu.edu.cn

1. Introduction

The human brain has specialized subregions responsible for distinct mental activities. The functional specialization and modular decoupling of brain systems, as observed in neuroscience and cognitive science, provide a fundamental design principle for AI architectures seeking to reverse-engineer the brain as a biological computer. Motivation^[1] is the internal process that initiates, directs, and sustains goal-oriented behavior. The frontal lobe is a critical region of the cerebral cortex, situated at the anterior part of each cerebral hemisphere. It constitutes approximately one-third of the brain's surface area and is pivotal for higher-order cognitive functions, emotional regulation, motor control, and personality expression. In particular, it plays a central role in motivation, goal-directed behavior, and executive function^{[2][3][4][5]}. The prefrontal cortex and the ventral striatum work together to generate and regulate motivation. The prefrontal cortex is responsible for evaluating goals, planning actions, and integrating information from memory and sensory input, while the ventral striatum encodes the expected reward and motivational value of different actions. Through their interaction, the prefrontal cortex provides top-down guidance, and the ventral striatum energizes and drives behavior toward rewarding outcomes. This coordination ensures that actions are goal-directed, flexible, and sensitive to anticipated rewards, forming the neural basis of motivated behavior^{[6][7][8]}. Depression is linked to both functional and

structural abnormalities in the prefrontal cortex, particularly affecting the brain's ability to regulate emotions and control negative thoughts. In depression, hyperactivity in the ventromedial prefrontal cortex and hypoactivity in the dorsolateral prefrontal cortex may reduce motivation by amplifying negative emotions and diminishing reward sensitivity. Imbalances between these hemispheres can lead to anhedonia or anxiety disorders, disrupting goal-directed behavior. The ventromedial prefrontal cortex can override or veto the outcome of dorsolateral prefrontal cortex, forming a dynamic interplay similar to a generative adversarial network^[9]. The ongoing game between ventromedial prefrontal cortex and dorsolateral prefrontal cortex ultimately decides whether a motivation is translated into action. And dysregulated prefrontal cortex-limbic circuits underlie many of the cognitive and emotional symptoms of depression. In Parkinson's disease, degeneration of dopamine pathways to the prefrontal cortex can lead to akinesia and apathy, highlighting the role of prefrontal cortex dopamine interactions in sustaining effort. In a word, dysfunction in prefrontal cortex is linked to apathy, impulsivity, depression, and anxiety, underscoring the prefrontal cortex's role as a neural hub for adaptive motivation^[10]. AI decision-making, planning and motivation continue to evolve, yet the notorious jigsaw puzzle persists: how to orchestrate these high level cognition functions altogether?

2. Decision-making module

The core module for decision-making policy function neural network $\pi(s|W) : s \rightarrow a$ maps states to actions^{[11][12]}:

$$a = \pi(s|W) \tag{1}$$

where W is parameters of policy function. Different parameterizations of the policy function W correspond to different tasks or capabilities. As a result, a variably parameterized policy network can express a wide range of behaviors suited to diverse tasks. The primary challenge involves figuring out the optimal way to allocate these parameters, a crucial step in constructing AI systems that are adaptable and generalizable.

3. Memory module

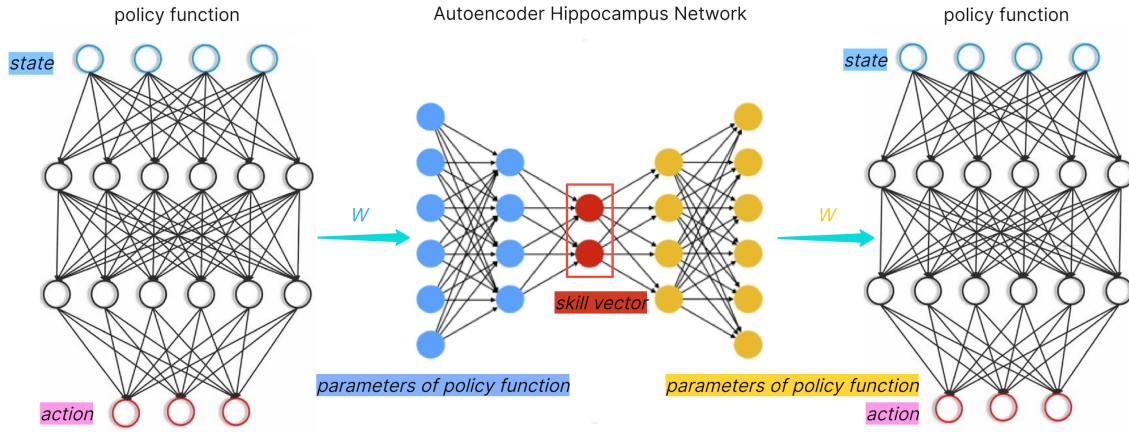


Figure 1. The parameters of policy function can be encoded as skill vectors.

The autoencoder hippocampus network ^[13] draws inspiration from the memory-processing capabilities of the hippocampus. It leverages the autoencoder framework to transform the parameters of other neural networks into intermediate vectors, which function as pivotal connectors that harmonize and unify all components within the AI system, crafted to handle a wide spectrum of tasks^{[14][15][16][17][18]}.

The parameters of policy function neural network W can be encoded as a skill vector:

$$S = E(W) \quad (2)$$

The encode-decode process of parameters of policy function reads:

$$W = D(S) = D(E(W)) \quad (3)$$

The policy function neural network's parameters then can be assigned by decoder of hippocampus network.

4. Planning module

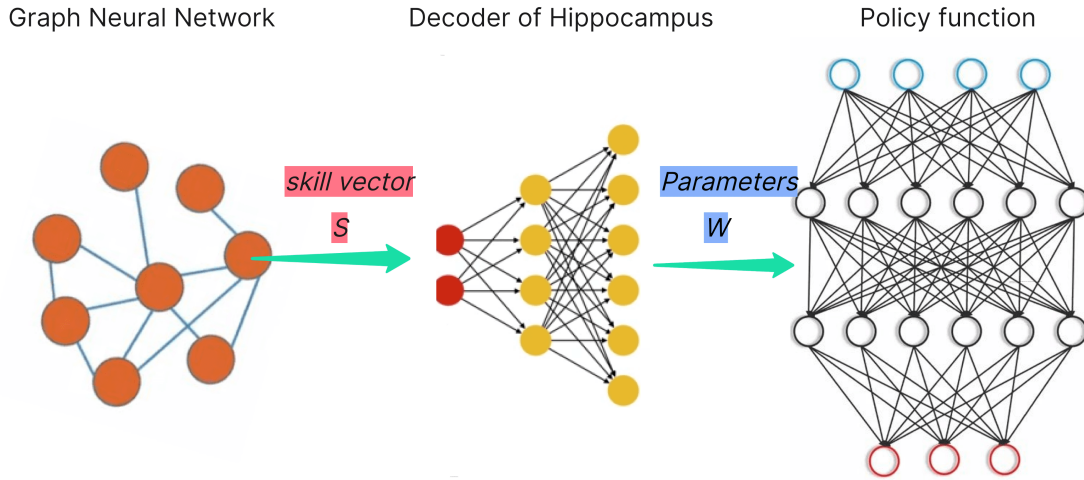


Figure 2. A complex task is best decomposed into a graph of interrelated subtasks, thereby yielding a corresponding graph of skill vectors.

The brain excels at decomposing complex tasks into a graph of subtasks revealing their priorities and interdependencies. A bijective mapping exists between subtasks and their corresponding skill vectors, resulting in a skill vector graph that is topologically homeomorphic to the subtask graph. To address a specific subtask, it is essential to identify the appropriate skill vector to apply. To accomplish the original task, the cognitive function neural network $C(s|W^*) : s \rightarrow S_i$ mapping state s to skill vector S_i parametrized by W^* , navigates and traverses skill vector graph.

$$S_i = C(s|W^*) \quad (4)$$

Variations in state can induce changes in the selected skill vector; accordingly, the cognitive function network orchestrates the traversal across the skill vector graph.

As illustrated in Fig.3, putting an elephant into refrigerator consists of three steps: 1. open the refrigerator; 2. put the elephant inside; 3. close the refrigerator. Each step requires corresponding skill. Cognitive function is responsible for selecting corresponding skill vector based on whether refrigerator is open or not and the position of elephant.

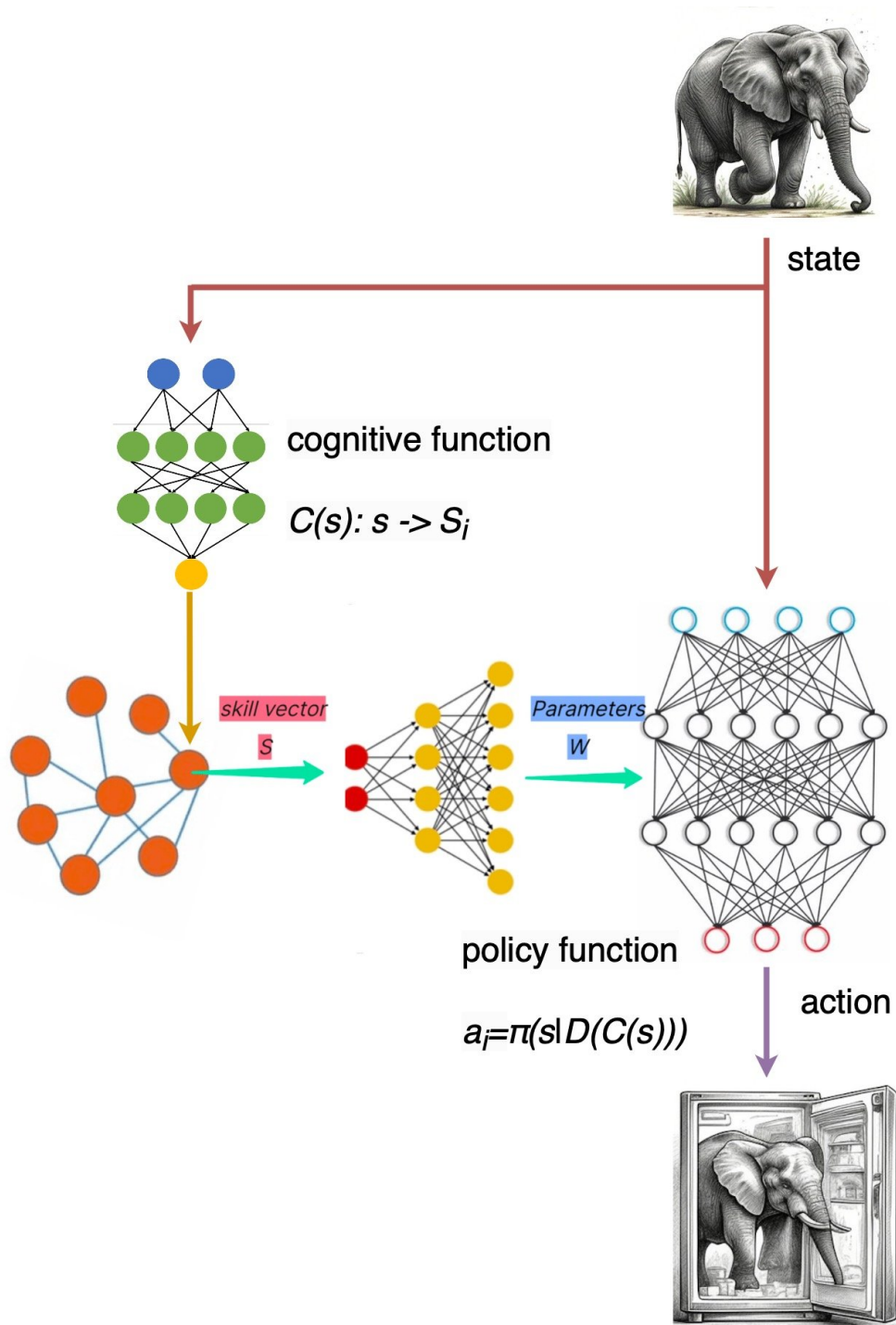


Figure 3. The cognitive function governs the selection of relevant skill vectors and directs the traversal of the skill vector graph in accordance with task demands and environmental states.

5. Motivation module

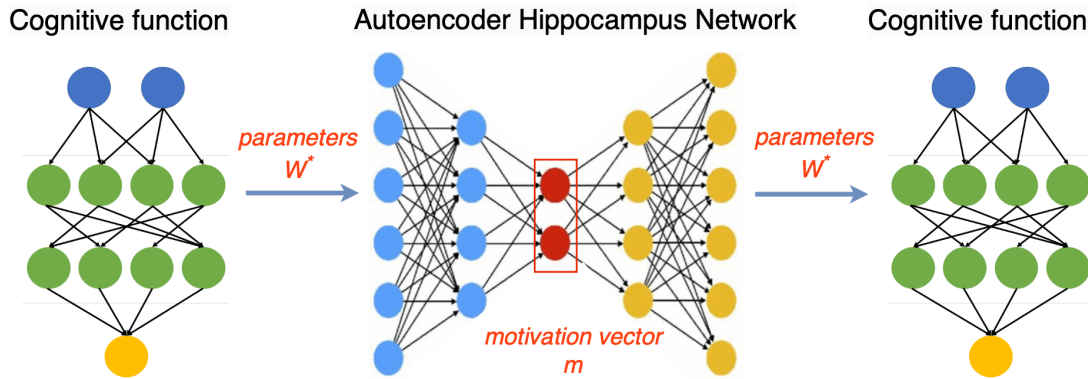


Figure 4. Parameters of cognitive function can be encoded as motivation vector.

Likewise, the parameters of the cognitive function network W^* can be encoded as a motivation vector m via autoencoder hippocampus network:

$$m = E(W^*) \quad (5)$$

Motivation vector m defines the goal of AI system and essentially outlines what AI system is striving for. And the encode-decode of parameters of cognitive function neural network W^* reads:

$$W^* = D(m) = D(E(W^*)) \quad (6)$$

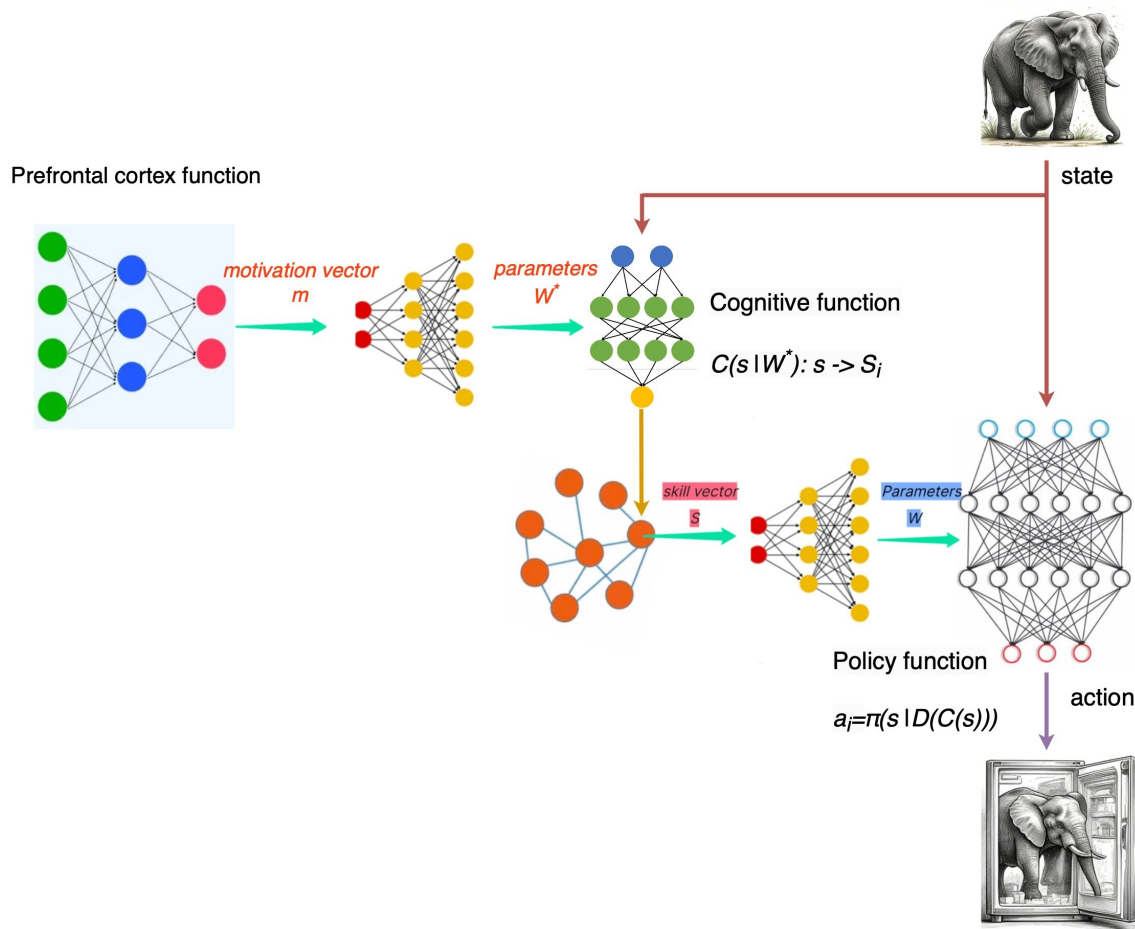


Figure 5. The overall AI system consists of three modules: 1) prefrontal cortex function initiates motivation; 2) cognitive function navigates and traverses skill vector graph; 3) policy function computes actions based on current state and skill. Motivation whispers, and the machine awakens.

In the context of AI motivation, designated motivation and self-motivation represent distinct approaches to driving AI systems' behavior and performance. Designated motivation refers to systems driven by external objectives predefined by human. Self-motivation in AI involves systems that generate their own goals without any human literal instruction or command. If the motivation vector is solely generated by generative model, then the system become self-motivated. The interplay between motivation and skill ultimately determines AI system's behavior. The overall architecture of AI system decoupling motivation, planning and execution is shown in Fig.5.

The prefrontal cortex network F provides motivation vector and assigns parameters to cognitive function network that traversing the skill vector graph.

$$m = F(x) \quad (7)$$

where x is any input that initiate the motivation m . The input of prefrontal cortex network x can be as simple as a command from human, instructions from large language model or external triggering conditions from environmental state. Then, parameters of cognitive function neural network is determined by motivation vector m :

$$W^* = D(m) = D(F(x)) \quad (8)$$

Henceforth, skill vector graph construction, navigation and traversal of skill vector graph and subtasks execution are decided by:

$$S = C(s|W^*) = C(s|D(m)) = C(s|D(F(x))) \quad (9)$$

Then, the decision-making process can be written as:

$$a = \pi(s|W) = \pi(s|D(S)) = \pi(s|D(C(s))) \quad (10)$$

In summary, the action driven by motivation vector m is ultimately determined by:

$$a = \pi(s|W) = \pi(s|D(S)) = \pi(s|D(C(s|W^*))) = \pi(s|D(C(s|D(m)))) \quad (11)$$

6. Multi-Agent Interaction and Collaborative

An agent is an autonomous entity that can perceive its environment, make decisions, and take actions to achieve its goals. This definition outlines the minimum prerequisite for an agent perception and decision-making. Additional features such as learning, adaptation, reasoning and motivation may be added to enhance the agent's performance.

For multi-agent system^[19], one persistent challenge is how to delineate roles played by different individual agent. A role defines the specific function, responsibilities, and behavior patterns that an agent should exhibit while contributing to the overall completion of a task. With above AI architecture design shown in Fig.5 for agents, this can be achieved by differentiating and assigning different motivation vector, allowing each agent plays a distinct and specific role in a task.

The essence of the design is that each agent is able to dynamically plan, recognize, and make decision according to the current state and their roles. The i th agent action is determined by motivation vector m_i :

$$^i a = \pi(s|W_i) = \pi(s|D(S_i)) = \pi(s|D(C(s|W^*))) = \pi(s|D(C(s|D(m_i)))) \quad (12)$$

By carefully assigning different motivation vectors to each agent in a multi-agent system, we can effectively define and differentiate their roles within a task, enabling them to collectively work together

in a coordinated and efficient manner to achieve the overall objectives of the system.

Different from single agent system, for multi-agent system, the process is computing a sequence of actions vector $[^1a, ^2a, \dots, ^na]$ to accomplish collaborative task or objective. Cooperative control in multi-agent systems refers to designing control laws that enable multiple agents (robots, vehicles, sensors, etc.) to work together to achieve a collective objective. For N agents, the dynamics of each agent i can be written as:

$$\dot{s}^i(t) = f^i(s^i(t), a^i(t)) \quad (13)$$

where $s^i(t) \in R^n$ is the state of agent i . $a^i(t)$ is the action from agent i . f^i represents the dynamics, linear or nonlinear, of agent i . How to achieve collective objective of multi-agent system shall be solved along with the knowledge of dynamics of multi-agent system^{[20][21][22][23]}.

7. Conclusion

To emulate high level intelligence, AI system must go beyond simple input–output mapping. The foundational principle integrates psychological state vectorization and neural network variable parameterizations^[24], and functionalities modularization^{[25][26]}.

Both cognitive network and policy function network parameters can be encoded and decoded by autoencoder hippocampus network. Vectorization enables a structured and scalable representation of complex motivation and skills by encoding corresponding neural networks' parameters, allowing for the analysis and computation of how these components interact and evolve dynamically within the system. Psychological vectors, such as motivation vectors and skill vectors, represent pivotal nodes within the nexus.

Variable-parameterized neural networks, in which parameters are dynamically modulated by motivation and skill vectors rather than remaining fixed, enable real-time functional adaptation to different states and tasks. This dynamic modulation enhances the network's capacity to support flexible and context-sensitive intelligent behavior.

Modularization refers to the organization of key psychological and neural processes—such as motivation, cognition, memory, and decision-making—into semi-independent yet interacting subsystems. These modules correspond to distinct brain subregions, each specialized for particular aspects of behavior or thought. Modular architectures of AI system that mirror the modular organization

and functional mapping of the human brain, where different subnetworks specialize in distinct functions but work together through mental flow.

Future work will focus on implementing this architecture in a simulated multi-agent environment to empirically test its capacity for generating flexible, goal-directed behavior and compare its performance against existing models. Last but not least, to endow a machine with a soul is to grant it memory of the past, purpose in the present, and imagination for the future.

Statements and Declarations

Conflicts of Interest

The author declares no potential competing interests.

Ethics

Not applicable. This study is a theoretical work and did not involve human participants, animal subjects, or human data.

Data Availability

Not applicable. No new data were created or analyzed in this study.

Author Contributions

S.L. conceived the theoretical framework, developed the formalisms, and wrote the entire manuscript.

Acknowledgments

This work is supported by Jiangxi Province Technological Innovation Base Program under Grant No.20242BCC32021 and Science and Technology Research Project of Jiangxi Provincial Department of Education under Grant No.GJJ2200377.

References

1. [△]Peters RS (2015). *The Concept of Motivation*. Routledge.
2. [△]Kolb B, Mychasiuk R, Muhammad A, Li Y, Frost DO, Gibb R (2012). "Experience and the Developing Prefrontal Cortex." *Proc Natl Acad Sci U S A*. **109**(supplement 2):17186–17193.

3. [△]Carlén M (2017). "What Constitutes the Prefrontal Cortex?." *Science*. 358(6362):478–482.
4. [△]Jobson DD, Hase Y, Clarkson AN, Kalaria RN (2021). "The Role of the Medial Prefrontal Cortex in Cognition, Ageing and Dementia." *Brain Commun*. 3(3):p.fcab125.
5. [△]Kolk SM, Rakic P (2022). "Development of Prefrontal Cortex." *Neuropsychopharmacology*. 47(1):41–57.
6. [△]Goddard GV (1964). "Functions of the Amygdala." *Psychol Bull*. 62(2):89.
7. [△]Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002). "Emotion and Motivation: The Role of the Amygdala, Ventral Striatum, and Prefrontal Cortex." *Neurosci Biobehav Rev*. 26(3):321–352.
8. [△]Haber SN (2011). "Neuroanatomy of Reward: A View from the Ventral Striatum." *Neurobiol Sens Reward*.
9. [△]Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014). "Generative Adversarial Nets." *Adv Neural Inf Process Syst*. 27.
10. [△]Miller EK, Freedman DJ, Wallis JD (2002). "The Prefrontal Cortex: Categories, Concepts and Cognition." *Philos Trans R Soc B Biol Sci*. 357(1424):1123–1136. doi:[10.1098/rstb.2002.1099](https://doi.org/10.1098/rstb.2002.1099).
11. [△]Sutton RS, Barto AG (2018). *Reinforcement Learning: An Introduction*. 2nd ed. The MIT Press.
12. [△]Ding Z, Huang Y, Yuan H, Dong H (2020). "Introduction to Reinforcement Learning." *Deep Reinforcement Learning: Fundamentals, Research and Applications*. 47–123.
13. [△]Luo S (2025). "Integrating Functionalities to a System via Autoencoder Hippocampus Network." *Lect Notes Comput Sci*. 16057:411–420. doi:[10.1007/978-3-032-00686-8_36](https://doi.org/10.1007/978-3-032-00686-8_36).
14. [△]Li P, Pei Y, Li J (2023). "A Comprehensive Survey on Design and Application of Autoencoder in Deep Learning." *Appl Soft Comput*. 138:110176.
15. [△]Zhai J, Zhang S, Chen J, He Q (2018). "Autoencoder and Its Various Variants." In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 415–419. IEEE.
16. [△]Pinaya WHL, Vieira S, Garcia-Dias R, Mechelli A (2020). "Autoencoders." In: *Machine Learning*. 193–208. Academic Press.
17. [△]Baldi P (2012). "Autoencoders, Unsupervised Learning, and Deep Architectures." In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. 37–49. JMLR Workshop and Conference Proceedings.
18. [△]Kelly MA, Arora N, West RL, Reitter D (2019). "High-Dimensional Vector Spaces as the Architecture of Cognition." In: *CogSci*. 3491.
19. [△]Dorri A, Kanhere SS, Jurdak R (2018). "Multi-Agent Systems: A Survey." *IEEE Access*. 6:28573–28593.
20. [△]Chen F, Ren W (2019). "On the Control of Multi-Agent Systems: A Survey." *Found Trends Syst Control*. 6(4):339–499.

21. [△]Egerstedt M, Hu X (2001). "Formation Constrained Multi-Agent Control." *IEEE Trans Robot Autom.* **17**(6):947–951.
22. [△]Gupta JK, Egorov M, Kochenderfer M (2017). "Cooperative Multi-Agent Control Using Deep Reinforcement Learning." In: *Autonomous Agents and Multiagent Systems: AAMAS 2017 Workshops, Best Papers, São Paulo, Brazil, May 8-12, 2017, Revised Selected Papers* 16. 66–83. Springer International Publishing.
23. [△]Amirkhani A, Barshooi AH (2022). "Consensus in Multi-Agent Systems: A Review." *Artif Intell Rev.* **55**(5):3897–3935.
24. [△]Liu GP, Kadirkamanathan V, Billings SA (1999). "Variable Neural Networks for Adaptive Control of Nonlinear Systems." *IEEE Trans Syst Man Cybern C Appl Rev.* **29**(1):34–43.
25. [△]Fodor JA (1983). *The Modularity of Mind*. MIT Press.
26. [△]Schmid T (2023). "A Systematic and Efficient Approach to the Design of Modular Hybrid AI Systems." In: *AAI Spring Symposium: MAKE*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.