SCIENCE CHINA
Technological Sciences

Special Topic: Brain-Inspired AI Research
•Review•

August 2024 Vol.67 No.8: 2282–2296
https://doi.org/10.1007/s11431-024-2732-9

Brain-inspired artificial intelligence research: A review

WANG GuoYin[1,2,3*], BAO HuaNan[1,2], LIU Qun[1,2], ZHOU TianGang[4,10], WU Si[5], HUANG TieJun[6], YU ZhaoFei[6], LU CeWu[7], GONG YiHong[8], ZHANG ZhaoXiang[9] & HE Sheng[10]

[1] *Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;*
[2] *Key Laboratory of Cyberspace Big Data Intelligent Security, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;*
[3] *College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China;*
[4] *State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;*
[5] *School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China;*
[6] *State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871, China;*
[7] *Department of Computer Science, School of Electronics, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;*
[8] *Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;*
[9] *The Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;*
[10] *Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China*

Received March 6, 2024; accepted June 26, 2024; published online July 30, 2024

Artificial intelligence (AI) systems surpass certain human intelligence abilities in a statistical sense as a whole, but are not yet the true realization of these human intelligence abilities and behaviors. There are differences, and even contradictions, between the cognition and behavior of AI systems and humans. With the goal of achieving general AI, this study contains a review of the role of cognitive science in inspiring the development of the three mainstream academic branches of AI based on the three-layer framework proposed by David Marr, and the limitations of the current development of AI are explored and analyzed. The differences and inconsistencies between the cognition mechanisms of the human brain and the computation mechanisms of AI systems are analyzed. They are found to be the cause of the differences and contradictions between the cognition and behavior of AI systems and humans. Additionally, eight important research directions and their scientific issues that need to focus on brain-inspired AI research are proposed: highly imitated bionic information processing, a large-scale deep learning model that balances structure and function, multi-granularity joint problem solving bidirectionally driven by data and knowledge, AI models that simulate specific brain structures, a collaborative processing mechanism with the physical separation of perceptual processing and interpretive analysis, embodied intelligence that integrates the brain cognitive mechanism and AI computation mechanisms, intelligence simulation from individual intelligence to group intelligence (social intelligence), and AI-assisted brain cognitive intelligence.

artificial intelligence, cognitive science, brain science, intelligence science, large language model

Citation: Wang G Y, Bao H N, Liu Q, et al. Brain-inspired artificial intelligence research: A review. Sci China Tech Sci, 2024, 67: 2282–2296, https://doi.org/10.1007/s11431-024-2732-9

*Corresponding author (email: wanggy@cqupt.edu.cn)

1 Introduction

Building machines that can think and infer like humans has been a human endeavor for centuries. The birth of artificial intelligence (AI) is regarded to have occurred at the 1956 Dartmouth Conference [1]. Since then, AI has undergone numerous fluctuations and advancements to reach its current state [2,3], surpassing human capabilities in numerous specific aspects [4–6]. In this process, cognitive science has repeatedly inspired the progress of AI. Cognitive science is a cutting-edge, interdisciplinary field that emerged in the mid-1970s [7]. Its primary objective is to investigate the nature and laws of human cognition and intelligence, and it focuses on studying how information is formed, represented, and recognized by the brain. The prevailing viewpoint maintains that current AI algorithms are different from human cognitive functions because they are still constructed by statistical learning that depends on massive data [8]. Leveraging insights from cognitive science research to enhance AI models and algorithms may be a promising avenue for the future development of AI.

Although some breakthroughs, such as ResNet [9], AlphaGo [10], ChatGPT [11], and Sora [12], represent the statistical surpassing of certain single human intelligence capabilities, AI systems do not truly achieve human intelligence. Most of them pay more attention to "fitting" rather than "cognition" by designing loss functions. Contemporary AI is beleaguered by a plethora of issues, including its black-box nature [13] and vulnerability [14]. It is very important and urgent to understand what the intelligence of the human brain is. After conducting a full study and understanding the cognitive intelligence of the brain, we may know what intelligence should be provided to machines. Currently, there is a scarcity of papers in which authors systematically analyze the intersection of cognitive science and AI, particularly a lack of analysis from the perspective of cognitive science on the limitations of AI. In this study, we analyze the challenges of existing AI systems using the three-layer framework proposed by Marr [15].

The consensus within the academic community suggests that inspiration derived from cognitive science may serve as a crucial opportunity for the next leap forward in the development of AI. Prof. Chen [16] posited that the core fundamental science problem of the new generation of AI is building the relationship between cognition and computation, specifically, the relationship between "global-first" and "local-first". Recently, Ohki et al. [17] reviewed the advantages that the human brain possessed in terms of learning efficiency, continuity, and generalization over AI. The pursuit of a novel machine learning methodology inspired by brain cognition is a promising approach to addressing the challenges of AI. It is also an inevitable development trend of future AI technology. In response to this, we begin by analyzing the limitations of brain-inspired AI at three levels and then propose our perspectives on future AI from two aspects. First, given that the historical development of AI has generally been inspired by cognitive science, based on the limitations analysis mentioned previously, we propose seven future research directions based on Marr's three-layer framework. Second, just as AI is increasingly being combined with other areas of research, AI increasingly inspires cognitive science. In this study, we provide a preliminary summary of AI-assisted brain intelligence (AI4BI).

This study is organized as follows: In Section 2, we primarily introduce the developmental history of AI. In Section 3, we primarily analyze the limitations of AI systems from the three layers framework. In Section 4, we address the limitations of existing brain-inspired AI and propose eight research directions from the perspectives of both brain intelligence-assisted AI (BI4AI) and AI4BI. Finally, we draw conclusions in Section 5. Table 1 summarizes the abbreviations frequently used in this study.

2 Inspiration for AI from brain cognition

Generally, AI is categorized into three primary schools: symbolism AI, behaviorism AI, and connectionism AI. The representative achievements include expert systems, reinforcement learning (RL), and deep neural networks (DNNs), which have played a pivotal role in the advancement of AI. Next, we review the histories of these three schools inspired by brain cognition in the context of cognitive science.

2.1 Inspiration for symbolism AI

Symbolism AI was one of the most active schools in the early stages of the development of AI. Its fundamental concept is that cognition is a type of symbolic processing and the

Table 1 Abbreviation definitions

Abbreviation	Definition
AI	Artificial intelligence
ANNs	Artificial neural networks
SNNs	Spiking neural networks
LLMs	Large language models
AI4BI	Artificial intelligence assisted brain intelligence
DNNs	Deep neural networks
RL	Reinforcement learning
CNNs	Convolutional neural networks
GAI	Generative artificial intelligence
RT-2	Robotics transformer 2

processes of human thinking can be computed by symbols. Hence, computers can simulate human intelligence through the operation of various symbolic rules.

Much of the early researches on symbolism AI came from cognitive psychology. In the 1950s, researchers proposed computational theory [18,19]. In 1976, Newell and Simon [20] proposed the physical symbol system hypothesis, which posited that the fundamental building blocks of knowledge are symbols and intelligence is contingent on knowledge. They believed that computer software and cognitive psychology methods could be used to imitate the functioning of the human brain at a macro level [21,22].

Logic theorist is a significant milestone in the history of AI. It is primarily a computer program based on symbolism that specializes in imitating human cognition processes [23,24]. In 1958, Prof. Hao [25] proved over 350 theorems in the first-order logic section of Principia Mathematica in only 9 min on an IBM 704 computer. Wu Wenjun proposed the Wu method in 1977. This method is a groundbreaking approach to geometric theorem proving using computers that inherited and developed the algorithm-based tradition of ancient Chinese mathematics, and revolutionized the field of automatic reasoning. In 2024, a new state-of-the-art approach for automated theorem proving was achieved by combining the Wu method and neuro-symbolic models[1)].

Machine learning algorithms based on symbolism AI have the advantage of strong interpretability, traceable inference processes, and flexible knowledge representation. However, as the problems to be solved have become increasingly complex, symbolism AI algorithms have been not up to the task. Although some researchers have found that human and AI cognitive behaviors have a symbolic nature[2)], the complexity and abstraction of cognitive intelligence expressed by the brain are much higher than that of machine intelligence based on symbolism AI.

2.2 Inspiration for behaviorism AI

Behaviorism AI, also known as evolutionism AI, is a "perception-action"-based behavioral intelligence imitation methodology derived from evolutionary and cybernetic theory. Behaviorism AI researchers believe that intelligence depends on perception and behavior. Intelligence may not require knowledge, knowledge representation, and knowledge reasoning. Hence, perception and control are the core issues of behaviorism AI.

In 1950, Turing [26] first proposed embodied intelligence in his paper "Computing machinery and intelligence," that is, robots or simulators that can perceive and interact with the environment, autonomously plan, make decisions, act, and have the ability to execute tasks like a human. This is the ultimate form of AI. Embodied intelligence has a physical body, collects environmental information through sensors, uses mechanical actuators to perform physical operations, or interacts with humans and the environment in real time through specific entities, such as robots [27]. In the 1980s, Brooks' [28] hexapod walking robot (Genghis) was a control system based on a perception-action model that simulated insect behavior. In July 2023, the team of DeepMind released the vision-language-action model called Robotics Transformer 2 (RT-2). RT-2 is regarded as an initial exploration of representing embodied multimodal large models with three major capabilities: symbolic understanding, inference, and human recognition[3),4)].

In recent years, RL has gained increasing attention because of the development of general AI. Many core ideas in RL were inspired by imitating animal behavior, cognitive psychology, and cognitive neuroscience [29].

RL is the core technology of AlphaGo that sparked global attention and discussion. It uses Q-learning algorithms to estimate the expected payoff of each state-action pair and select the optimal action. Additionally, RL plays a crucial role in large language models (LMMs) [10]. In 2020, the pretraining process of GPT-3 incorporated RL learning techniques that enabled the model to optimize its parameters based on the specific requirements of the task. This innovation aims to make natural language processing technology more powerful and accurate [11].

Although the behaviorism AI breaks the traditional cognitive psychology sandwich model of "perception-thinking-acting" in traditional cognitive science, it still has significant shortcomings in mimicking human cognition and learning. We analyze this in detail in the subsequent sections.

2.3 Inspiration for connectionism AI

Some researchers believe that AI should be derived from bionics, particularly emphasizing the imitation of human brain models.

In 1943, psychologist McCulloch and mathematician Pitts [30] first proposed the M-P model. It assumed that the membrane potential of a neuron had different states at different times. When the membrane potential exceeds a certain threshold, the neuron generates a pulse signal. In 1958, Rosenblatt [31] proposed a neural network consisting of two layers of neurons called Perception. In 1962, Hubel and

1) https://arxiv.org/abs/2404.06405
2) https://arxiv.org/abs/2305.01939
3) https://arxiv.org/abs/2212.06817
4) https://arxiv.org/abs/2307.15818

Wiesel [32] proposed the concept of receptive field in their research on the visual cortical cells of cats and monkeys. They were the first to suggest that neural networks in the visual cortex have a hierarchical structure. Inspired by this research, the Neocognitron was developed by Fukushima [33] in 1980. It is a hierarchical multilayer neural network. The model consists of several layers, each containing a set of convolution filters and pooling operations. In 1986, Rumelhart et al. [34] successfully solved the parameter optimization of multilayer perceptual neural networks using the back propagation algorithm. Another important development during this period was the LeNet-5 model developed by LeCun et al. [35] in 1998. LeNet-5 was the first convolutional neural network (CNN) to achieve significant success in image classification tasks. Since the advent of the 21st century, with the advancement of hardware and massive data [36], the performance of artificial neural networks (ANNs) has improved rapidly. In 2014, Inception Net won the ILSVRC-2014 image classification competition [37]. This model was inspired by Hebbian theory derived from cognitive science. A squeeze-and-excitation network was the winner of the last ILSVRC competition. Its performance improved greatly as a result of the introduction of the attention mechanism. The attention mechanism was mainly inspired by research on visual attention in cognitive neuroscience [38,39]. The visual attention mechanism allows the brain to focus on specific visual information while ignoring irrelevant stimuli. Visual cognition consists of two pathways: “bottom-up” and “top-down”. Figure 1 illustrates the relationship between the two pathways in detail. Both of them interact and work together to facilitate effective visual processing and attention allocation. By mimicking the cognitive attention allocation mechanism, many types of attention mechanisms have been extended gradually in DNNs, such as self-attention mechanisms, channel attention mechanisms, and spatial attention mechanisms. These mechanisms aim to emphasize important parts of information and minimize irrelevant parts.

As AI application domains have become more complex, deep learning based on connectionism has achieved great success in areas such as vision and language because of its superb function-fitting capabilities. However, the debate that started 50 years ago about whether AI should mimic the structure of the nervous system is ongoing [40]. Current AI, represented by deep learning, is becoming increasingly less like the nervous system. Many AI experts no longer insist that “machines need to think like humans to be intelligent”. However, as research progresses, deep learning models cannot be adapted to critical and sensitive applications, and cannot replicate human cognition and integrative inference. Current AI models only exceed some of the capabilities of human intelligence in a statistical sense rather than achieving the true realization of individual intelligence.

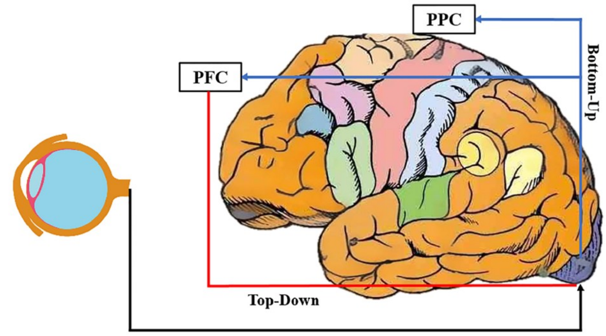


Figure 1 (Color online) Attention mechanism inspired by brain cognition [38].

In conclusion, cognitive science provides important theoretical support and practical guidance for the development of AI models. These guidelines have strongly promoted the continuous innovation of AI technology and the expansion of AI applications.

3 Limitations of current brain cognition-inspired AI

Marr [15] believed that “most of the phenomena that are central to us as human beings—the mysteries of life and evolution, of perception and feeling and thought—are primarily phenomena of information processing. One of the fascinating features of information-processing machines is that to understand them completely, one has to be satisfied with one’s explanations at many different levels”. He described an information processing system as a loosely connected three-layer structure, that is, the hardware implementation layer, the representation and algorithm layer, and the computational theory layer.

Marr mentioned that the study of the mechanism of human intelligence by cognitive scientists can be summarized in the following three stages. Take the study of visual mechanisms as an example. First, researchers have studied human visual mechanisms primarily from the perspective of the cellular function. This field is often referred to as neuroanatomy. Although these researchers identified mechanisms behind some low-level visual cognitive behaviors, they failed to explain the role of visual centers in higher-level visual cognition [41]. Second, in the 1970s, cognitive scientists studied some higher-level visual cognitive behaviors in terms of representing image features [42,43]. These studies were usually categorized under psychophysics. Despite numerous research results, these studies still did not clarify the connection between vision and cognitive processes. Finally, researchers have concluded that the description of cognitive behavior requires a level of abstraction that captures behaviors and motivations at a higher level. For example, the visual system of spiders is very sensitive to “V”-shaped textures. To fully explain this mechanism, it is necessary to incorporate the spider’s courtship behavior as a motivating factor.

Collectively, these three stages illustrate the scholars’ evolving comprehension of cognitive mechanisms. Marr also summarized the content and research logic of these three stages into a three-layer model. The hardware implementation layer centers on the physical computation infrastructure. The representation and algorithm layer focus on information depiction and articulation. The computational theory layer offers a philosophical abstraction of cognitive processes that is crucial for understanding information processing systems.

Figure 2 illustrates three corresponding layers between brain intelligence and AI. In the following sections, we analyze the shortcomings of brain-inspired AI in the three layers.

3.1 Limitations in the hardware implementation layer

The hardware implementation layer primarily involves the physical realization of an intelligence system. In the brain, this corresponds to neural structures and biochemical processes. In AI systems, it corresponds to the infrastructure of the model.

The fundamental building block of the brain is the biological neuron, which comprises a cell body, dendrites, and an axon. The cell body contains the nucleus and cytoplasm, and provides structure and support to the neuron. Dendrites are extensions of the cell body that receive signals from other neurons. The axon is a long extension that carries outgoing signals to other neurons or effector organs. Neuroscientists have mapped the hemispheric subdivisions of the brain and defined four regions. The parietal lobe is responsible for the senses of touch, pain, and temperature. It is located just after the central sulcus. The occipital lobe, which processes visual information, is located at the back of the head. The temporal lobe is responsible for auditory processing and is located in the lower part of the lateral fissure. The frontal lobe takes part in motor control and cognitive activities, such as planning, decision-making, and goal-setting, and is located above the lateral sulcus and before the central sulcus. Complex cross-regional connections exist between these areas, which are necessary for advanced brain function.

The ANN is a digital analog of a biological neural network and the basic unit is an artificial neuron. The artificial neuron of first and second-generation ANNs is a mathematical function, which typically involves a weighted sum and nonlinear function. Another common type of ANN is the spiking neural network (SNN), which is considered to be the third generation of neural network models and is a detailed imitation of the brain. Unlike traditional ANNs, SNNs produce a series of discrete impulse signals rather than continuous values. They mimic the way that biological neurons emit pulses at specific moments [44,45].

The first two generations of neural networks and SNNs represent two different ideas for the imitation of the brain structure. The former focuses more on abstracting the mechanisms of the brain rather than replicating its physical mechanisms in detail. Such neural networks do not fully mimic the details of neurons, but rather learn the structure and function of the brain at a macroscopic layer. This structure is more in line with the way computers process signals and is more suitable for stacking. Hence, these neural networks are very large and achieve strong cognitive performance. The latter focuses more on a detailed simulation of the physical mechanisms of the brain. The physical structure of the neurons and neuronal organization are very similar to

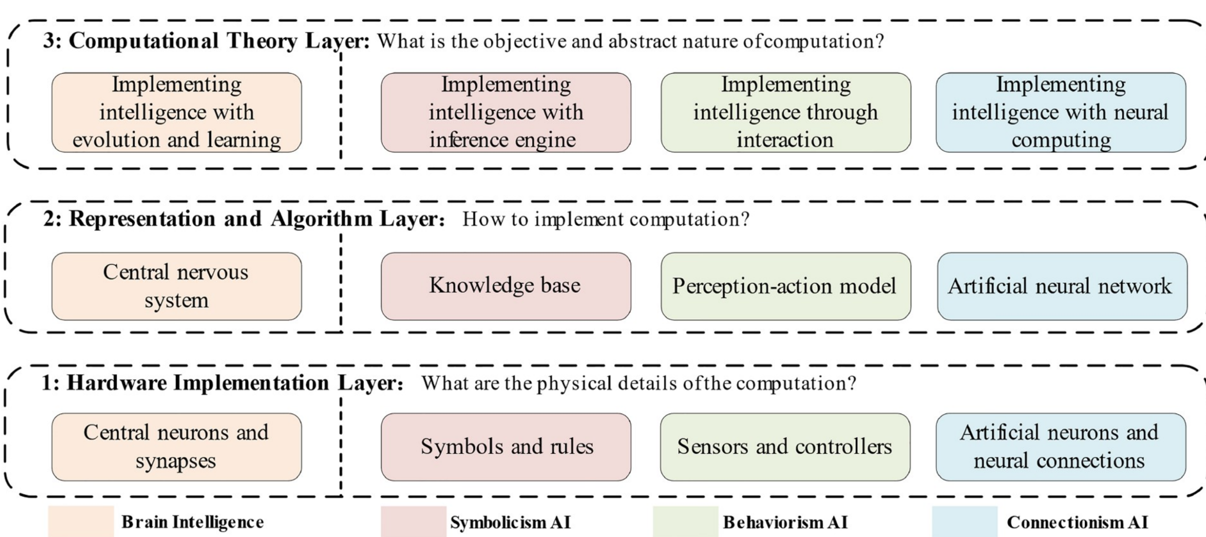


Figure 2 (Color online) Relationship between Marr’s three-layer theory and the brain intelligence, artificial intelligence.

the biological neural structure. These neural networks replicate the way brain neurons interact with each other, and reproduce the working principle of the brain by imitating the activation and inhibition states of brain neurons, in addition to the connection and communication between them. Compared with second-generation ANNs, SNNs are characterized by low power consumption, robustness, and strong temporal processing capability.

However, they all have limitations. First, SNNs imitate the central nervous system at a fine granularity; however, they struggle to do this at a coarser granularity because of the difficulty in scaling up. This results in a significant performance gap between SNNs and traditional ANNs. Undertaking further in-depth research on the intricate workings of the nervous system and gaining a profound comprehension of its elaborate mechanisms that underlie the execution of advanced cognitive and physical functions has the immense potential to help us to solve this complex problem. When the brain is tasked with processing a significant volume of information, inhibitory neurons effectively attenuate or suppress unwanted or interfering connections, thereby enabling the brain to process pertinent information with greater precision and evade distraction from irrelevant data. This intricate mechanism may significantly help the SNN to process increasingly complex cognitive tasks. Second, traditional ANNs only mimic brain mechanisms without considering the physical processes of biological neurons. They place greater emphasis on fitting not cognition, and only pay attention to how to set the loss function to achieve the best fitting performance. Traditional artificial neurons are more simple than biological neurons, and less flexible. With the same performance, ANNs not only have far more neurons than biological neural systems [46] but also consume far more power [47]. Spiking neurons have an advantage in terms of their ability to accrue and retain more historical state information. They have a capability that is immensely valuable for investigating the network's short-term memory, working memory, and related functions. Conversely, traditional artificial neurons excel in the simplicity of their structure and ease of stacking. The fusion of these two types of neurons has the potential to propel ANNs toward a broader transition from mere function fitting to comprehensive cognitive computation. This integration is anticipated to empower AI systems with enhanced robustness and adaptability, thereby making them more resilient to noise and adversarial attacks.

3.2 Limitations in the representation and algorithm layer

This layer concerns the process of solving a problem from input to output. It bridges the gap between abstract goals and specific implementations. Although both human brains and AI systems can perform similar functions, their information-processing mechanisms are fundamentally distinct. A significant number of studies in the field of AI have been dedicated to addressing this discrepancy.

Along with breakthroughs in AI systems, many mind-boggling phenomena have emerged. In the recognition task, the adversarial attack is a typical phenomenon that results from this discrepancy [48–50][5]. In the field of visual cognition, researchers discovered that ResNet-50, which was trained on ImageNet, primarily relied on texture to recognize images [13], as shown in Figure 3(b). Adversarial attacks, which introduce small perturbations into input data, cause deep learning models to make incorrect decisions, as shown in Figure 3(c). This differs significantly from human cognitive habits, which prioritize shape when recognizing objects. In the generation task, OpenAI's Sora, a highly advanced AI system, generates content that does not adhere to the laws of physics. Sora created a four-legged ant, whereas all known ants on Earth have six legs.

The fundamental reason for these problems is the difference in the visual cognition mechanisms between human brains and ANNs. Because of the structural differences mentioned regarding the previous layer, brain cognition exhibits a multimodal and multilevel characteristic. From a macroscopic perspective, human brains prioritize global-first properties in visual perception, also known as the global-first topological nature theory of perception [51,52]. This theory posits that the visual process begins with a broad range of topological properties, which can be described in terms of global topological invariants. Cognitive scientists have found that the perception of topological properties is rooted in the intrinsically photosensitive retinal ganglion cells. In terms of processing, the brain receives raw image data from photoreceptor cells in the retina and further deals with the information to extract higher-layer features. These features are abstracted and integrated in various layers, ultimately resulting in our understanding and perception of the image. However, ANNs are more akin to powerful fitting functions. Although they possess a multi-layered structure, their information processing mechanisms differ significantly from those of the brain. The adversarial attack phenomenon mentioned previously is believed by some researchers to stem from the complexity of decision boundaries [49]. The texture bias phenomenon is attributed to the inability of CNNs to effectively abstract low-level features into high-level representations [53].

Taking large models and generative AI (GAI) as examples, we believe that the above limitations can be specifically manifested in the following points: (1) Because of disparities

5) https://arxiv.org/abs/1312.6199

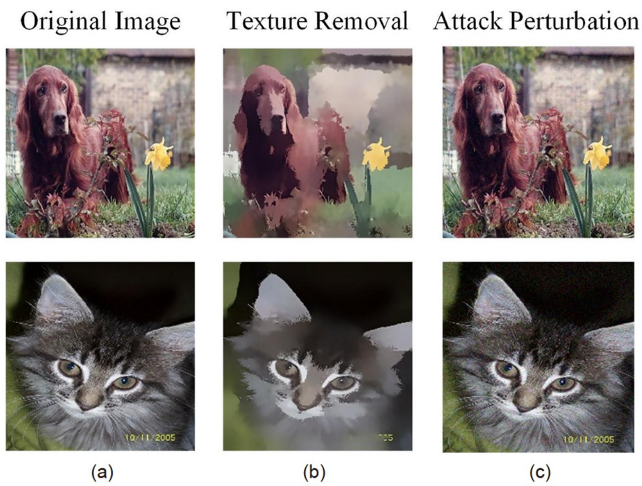
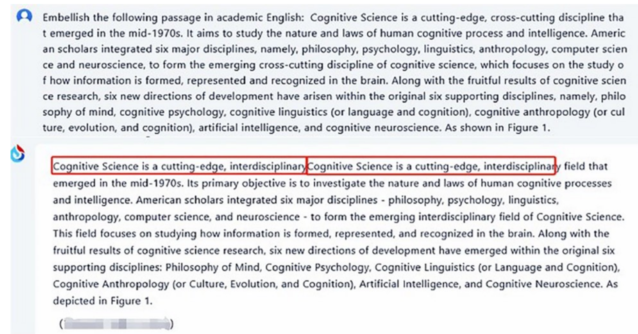


Figure 3 (Color online) Phenomena inconsistent with human cognitive machinery [13,48].

in cognitive mechanisms, they differ from human training and learning methods. LLMs with extensive parameter scales are exposed to much more text during pre-training than people would encounter in a lifetime of reading. Prof. Yonatan Bisk described these models as "primarily memory engines" [54]. (2) They fail to learn the highly abstract nature of knowledge. LLMs can mimic language domains, but they cannot learn human mental language. Prof. Jacob Browning and Prof. Yann Lecun presented:"A system trained on language alone will never approximate human intelligence, even if trained from now until the heat death of the universe", and "it is clear that these systems are doomed to a shallow understanding that will never approximate the full-bodied thinking we see in humans" [55]. (3) Unlike humans, they lack the ability to judge the plausibility of generated content, for instance, the repetitive sentences in Figure 4. The model produced such obvious mistakes, but could not assess them.

The processing of mental language in the brain involves the coordinated work of multiple brain regions. In particular, Wernicke's area and Broca's complex play pivotal roles in language production, comprehension, memory, and expression, thereby enabling individuals to engage in complex mental activities using mental language [56]. Notably, the brain regions involved in language learning do not exist in isolation; instead, they are intricately interconnected and interact with other cognitive functions. By contrast, LLMs primarily rely on probability to predict the next word. LLMs do not share the common mental language that human brains use.

Researchers posited that contemporary AI research compartmentalizes cognitive functions into distinct domains, such as speech, vision, and natural language, which results in isolated information and the sparse integration of data. Conversely, the human brain is an integrated cognitive system. Deep learning-based image recognition models ne-


Embellish the following passage in academic English: Cognitive Science is a cutting-edge, cross-cutting discipline tha t emerged in the mid-1970s. It aims to study the nature and laws of human cognitive process and intelligence. Americ an scholars integrated six major disciplines, namely, philosophy, psychology, linguistics, anthropology, computer scien ce and neuroscience, to form the emerging cross-cutting discipline of cognitive science, which focuses on the study o f how information is formed, represented and recognized in the brain. Along with the fruitful results of cognitive scien ce research, six new directions of development have arisen within the original six supporting disciplines, namely, philo sophy of mind, cognitive psychology, cognitive linguistics (or language and cognition), cognitive anthropology (or cul ture, evolution, and cognition), artificial intelligence, and cognitive neuroscience. As shown in Figure 1.

Cognitive Science is a cutting-edge, interdisciplinaryCognitive Science is a cutting-edge, interdisciplinary field that emerged in the mid-1970s. Its primary objective is to investigate the nature and laws of human cognitive processes and intelligence. American scholars integrated six major disciplines - philosophy, psychology, linguistics, anthropology, computer science, and neuroscience - to form the emerging interdisciplinary field of Cognitive Science. This field focuses on studying how information is formed, represented, and recognized in the brain. Along with the fruitful results of cognitive science research, six new directions of development have emerged within the original six supporting disciplines: Philosophy of Mind, Cognitive Psychology, Cognitive Linguistics (or Language and Cognition), Cognitive Anthropology (or Culture, Evolution, and Cognition), Artificial Intelligence, and Cognitive Neuroscience. As depicted in Figure 1.

Figure 4 (Color online) Sentence repetition problems of GPT3.5. The two red boxes contain two repeated sentences.

cessitate the collection and categorization of millions of photos to enable machines to "recognize" cats, whereas the human brain excels at capturing and recognizing the flow of information without requiring the backing of big data. We believe that the fundamental approach to address these issues lies in imitation, which allows AI systems to simulate the cognitive mechanisms of the nervous system from various perspectives, potentially resolving numerous challenges encountered in the representation and algorithm layer.

3.3 Limitations in the computational theory layer

This layer concerns the theory of abstract computation, and focuses on the goals of computation and the strategies for performing computation.

The learning process of the brain is complex neural activity that involves collaboration and interaction among multiple brain regions. Researchers have indicated that the brain's learning primarily involves four processes: (1) information input and perception, where the brain receives information from the external world through various sensory organs, such as vision, hearing, and touch, which is converted into neural signals and transmitted to different regions of the brain for processing; (2) information processing and storage, where different regions of the brain analyze, integrate, classify, and store the received information; (3) memory formation, where the brain forms memories by consolidating and reinforcing the connections between neurons that are strengthened through continuous learning and repeated practice, which enables the brain to better extract and recall stored information; (4) outputting knowledge, where the brain extracts and processes stored information as needed and transmits commands to effectors through motor neurons. Ohki et al. [17] discovered that the human brain maximizes its learning efficiency through a self-organizing mechanism. The brain has cognitive neural mechanisms for lifelong continuous learning, which is accomplished through memory playback during sleep [33]. Therefore, it is inaccurate to describe the brain's learning behavior using a single fixed loss function. Additionally, in studies on human brain cognitive intelligence [57], researchers have shown that the generation of human brain cognition can be based not only on individual units but also the mutual influence between groups, and the survival of the fittest in an open environment makes the objective function of the human brain cognitive mechanism constantly change and evolve.

By contrast, taking DNN training as an example, the learning process of AI can be viewed as an optimization of a complex function, typically involving the following steps: (1) data collection; (2) forward, the neural network uses the current parameters and inputs data to calculate a predicted output; (3) loss calculation: the difference between the predicted output and ground truth is measured; (4) backward: with the aim to optimize the parameters, the gradient of each parameter is calculated using a backward algorithm; (5) parameter updating: based on the calculated gradient, optimization algorithms (e.g., gradient descent optimization method) are used to adjust the network parameters and iterated until convergence.

Computational logic is also different for AI systems and the human brain. For instance, they are completely different in the way they process uncertain information. AI systems express uncertain information as data and then use mathematical tools, such as rough sets [58] and fuzzy sets [59]. The brain processes uncertain information through inhibitory filtering mechanisms. It has specific brain regions to process this information. In 2021, Mukherjee et al. [60] indicated that two types of midback projection exist in the prefrontal cortex.

The exponential increase in data has led to the development of deep learning models to achieve higher performance. However, AI systems still fall short in terms of continuous learning and new concept discovery in unfamiliar environments compared with the human brain. The primary limitation of AI systems' learning ability is that they can only imitate human cognition methods to a certain extent. We believe that the investigation of factors that contribute to the emergence of intelligence and the processes involved in its implementation can be used to overcome the limitations associated with this layer. Further research into biological behavior and survival goals will be beneficial for the development of more realistic AI models. Specifically, adaptive loss functions and continuous learning capabilities are potential future solutions.

4 Future trends and insights into the integration of AI and cognitive science

The achievements of cognitive science have inspired much of the development of AI and provided important theoretical foundations and insights. Based on the analysis presented in Section 3, it is evident that AI merely mimics the manner in which the human brain processes information. These simulations are not entirely accurate, resulting in the numerous limitations discussed in Section 3. To achieve this, we propose mutual inspiration between cognitive science and AI. Inspiration from cognitive science to AI can be categorized into three levels and seven points based on Marr's three-layer model, whereas inspiration from AI to cognitive science comprises a single, standalone point.

4.1 Brain cognition-inspired hardware implementation layer

ANN models can be regarded as imitations of the nervous system. As discussed in Section 3.1, the two most notable limitations at this level are the difficulty of achieving high performance through the large-scale implementation of SNNs and the challenge of mimicking the cognitive mechanisms of biological neural systems using traditional artificial neurons. We believe that the following two paths may inspire future research in AI in the hardware implementation layer.

4.1.1 Highly imitated bionic information processing model

We believe that the physical mechanisms of neurons should be imitated more realistically. Some biological neuron structures have not been well used in ANNs. For example, brain neurons have different functions and include excitatory (E) neurons and inhibitory (I) neurons [61]. Inhibitory mechanisms play a very important role in central nervous systems and are not easily mimicked by traditional artificial neurons.

Existing brain-inspired computing is an initial attempt in this direction. Brain-inspired computing, also known as neuromorphic computing, is a comprehensive term that encompasses computing theories, architectures, chip designs, application models, and algorithms that draw inspiration from the information-processing modes and structures of biological neural systems. Neuromorphic computing uses neurons and synapses as the basic units, and simulates the central nervous system in terms of its structure and function [45]. Huang's group [62] proposed a novel spike sampling method that mimics the retina's signal processing to support high-speed photography. By adopting this method, a high-speed camera was developed, which reconstructs visual images at 40000 frames per second in both normal and high-speed scenes. Additionally, SNNs have gained widespread attention because of their low power consumption and fast inference capabilities when implemented on neuromorphic hardware [47]. These achievements demonstrate the superiority of SNNs over traditional neural networks in specific tasks.

We believe two issues exist for SNNs that are worthy of further study. First, they still lack widely used machine learning algorithms because of the intricate discontinuities and implicit nonlinear mechanisms inherent in SNNs. Huang's group [63–66] proposed a series of ANN-SNN conversing algorithms with the aim of developing effective machine learning algorithms for SNNs. Additionally, optimizing the network structure of SNNs with multi-scale plasticity is a promising research direction. The traditional spike timing dependent plasticity learning rule requires measuring and storing the entire population activity state of synapses when calculating individual synaptic changes, which is computationally intensive and challenging to implement, thus limiting the construction of large-scale deep network models [44].

4.1.2 Large-scale deep learning model balancing structure and function

The central nervous system is a carbon-based intelligence system, whereas AI is a silicon-based intelligence system. Their physical and chemical properties are fundamentally different. Thus, detailed imitation is no substitute for macro imitation. Traditional ANNs are better suited to imitate some brain cognitive mechanisms. For example, neurons and neural connections can develop and expand under favorable conditions, thereby increasing the complexity of the neural network by forming new synaptic connections [44]. Researchers have made initial attempts to mimic the structure of central nervous systems.

Zhang et al. [67] used task-specific neural unit search and architectural growth to continuously learn new scenes, thereby addressing issues that exist in online learning. He proposed "Reusable Architecture Growth" and "Scene Router" to divide the neural network architecture search into cellular and layer-layer means, thereby forming a continuously changing network structure. This strategy has been applied favorably in the field of autonomous driving. Gong introduced a CNN design called the cognitive-inspired network (CogNet), which mimics the structure of visual cognitive mechanisms. This innovative architecture incorporates three distinct components: global paths, local paths, and top-down modulators. The local features extracted by the local paths are modulated by the global features from the global paths, which simulates the brain's hierarchical modulation mechanism. This network combines global and local information and avoids texture bias by mimicking the brain structure [53]. Furthermore, some researchers have directly emulated the neural systems of lower organisms to achieve balanced intelligence. For example, the nematode, which measures approximately 1 mm in length, possesses a total of 302 neurons in its body, which enables it to exhibit complex intelligent behaviors, including sensing, foraging, escaping, and mating. It has been demonstrated that the computational complexity exhibited by a single neuron in the nematode life model can be analogous to that of five to eight layers of DNNs [68]. Researchers designed a network for autopilot functionality which composed of 19 fine neurons by simulating the neurons of the nematode [46]. Therefore, the imitation of biological mechanisms can significantly affect the realization of general intelligence.

These research findings demonstrate the advantages derived from a balanced structure and function. Recently, in new neuroscience studies, researchers have shown that hierarchical cortical processing is integrated with a massively parallel process to which subcortical areas substantially contribute and proposed the shallow brain hypothesis [69,70]. Inspired by these achievements, it is helpful to construct new high-efficiency structures of DNNs. In particular, the architecture discovered in [69,70] could inspire new frameworks of AI models, such as modular deep learning architectures and width learning [71].

4.2 Brain cognition-inspired representation and algorithm layer

Based on the analysis in Section 3.2, we propose that simulating the information processing mechanisms of the human brain could serve as the primary approach to address limitations in the representation and algorithm layer. This simulation can be categorized into three distinct types. First, simulating the hierarchical structure of the biological neural system [72], which can be summarized as Section 4.2.1. Second, mimicking specific brain information processing mechanisms, encompassed by Section 4.2.2. Finally, achieving a balanced simulation that combines performance and interpretability, can be described as a Section 4.3.3.

4.2.1 Multi-granularity joint problem solving bidirectionally driven by data and knowledge

Granular computing constitutes a theoretical framework that investigates cognitive methodologies, problem-solving techniques, and information-processing paradigms depending on multi-granularity structures. As is commonly understood, humans can integrate information at multiple granularities to facilitate more exhaustive and precise cognitive computations. Conversely, most AI models undergo a unidirectional transformation from fine-grained to coarse-grained in their progression from data to knowledge, which violates the "global-first" principle.

Wang [73,74] proposed an innovative cognitive computing framework called data-driven granular cognitive computing (DGCC). For DGCC, data represent the finest granularity layer and knowledge is the abstraction of data in different granularity layers. The DGCC model provides a new idea for research on cognitive computing and lays the foundation for solving the problem of the "separation of knowledge and data". Based on this framework, Prof. Xia et al. [75–77] proposed a granular ball framework that can be applicable to diverse machine learning models. This framework segregates the dataset into distinct subsets with different granularities and improves predictive efficacy. It also exhibits robust generalization ability and interpretability, which can effectively process large-scale datasets. This idea has a wide range of applications. Prof. Dai et al. [78] proposed a method called "sketch less face image retrieval", which endeavors to retrieve the desired facial photograph in a minimal number of strokes. This method was built based on multi-granularity concepts and divided into two stages, from coarse-grained to fine-grained. The joint embedding space of the complete sketch and photo is learned in the first stage and the embeddings of the partial sketch are optimized in the second stage.

The success of the DGCC model and other multi-granularity models [79] demonstrated that simulating the multi-granularity cognitive mechanisms of the human brain can significantly enhance the performance of AI models. However, the current integration of these mechanisms is limited to certain tasks and the exploration of broader applications remains a promising research direction.

4.2.2 *AI models that simulate specific brain structures*

As mentioned previously, the cognitive mechanisms of AI and the human brain are very different. Directly using ANNs to simulate brain cognitive mechanisms represents a promising research direction to address these limitations.

In recognition tasks, several researchers have already emulated the cognitive mechanisms discovered by cognitive science, thereby addressing the limitations of existing AI systems. Prof. Chen [51] posited that the fundamental scientific issue of the next generation of AI is "the interplay between cognition and computation", specifically the computational theories of "global-first" versus "local-first". He first discovered that human visual cognition followed the "global-first" theory. Borrowing the "global-first" mechanism, Dong et al. [53] proposed a novel CNN architecture known as the CogNet, which comprises global paths, local paths, and top-down modulators. This architecture solves the phenomenon of texture bias well. Additionally, visual perception mechanisms in the retina have received extensive attention. There are many types of cells in the retina, such as optic rod cells and optic cone cells; hence, the retina has a strong ability to detect targets at different scales. To address the problem of the poor detection of multi-scale targets by traditional neural networks, Zhang's [80] proposed a method called TridentNet for target detection. It constructs a parallel multi-branch architecture in which each branch shares the same parameters, but has different receptive fields. The different branches are trained by sampling object instances at appropriate scales. TridentNet can achieve fast convergence without any additional parameters and computational costs. The concept of "global-first" effectively characterizes how the brain manages the relationship between local and global processing. Similar attempts have been made in other fields, such as federated machine learning [81].

In language generation tasks, cognitive science findings are potentially helpful. For example, Elizabeth Spelke, a psychologist at Harvard University, proposed the theory of core knowledge. She believed that humans are born with a small, mutually independent system of core knowledge. Pinker [82] proposed the mental language, which refers to the process by which humans use and understand language in a mental layer. Because humans essentially have the same neural structure, everyone has inborn grammar knowledge. For GAI, adding such an interpretable structure is likely to address many of the aforementioned limitations [83]. Attempts have been made to integrate physical rules into AI systems for some cognitive tasks, and we anticipate that this area will become a focal point of research endeavors.

Currently, a number of cognitive science research findings are challenging the rationality of existing deep learning architectures [69,70]. We believe that these findings may inspire new AI model structures that better simulate human intelligence.

4.2.3 *Collaborative processing mechanism with the physical separation of perceptual processing and interpretive analysis*

The black-box nature of deep learning has been a challenge that has plagued connectionism AI. Although numerous attempts have been made to enhance the interpretability of deep learning, few have imitated the interpretation mechanisms of the human brain. In the human brain, perception and its interpretation occur in different regions. The brain's perception and interpretation processes are separate, yet synergistic. Investigations have been conducted to address this challenge.

Local interpretable model-agnostic explanations is representative work based on this idea. It uses an understandable linear model to simulate the original model in a local classification hyperplane and analyzes the contribution of each feature using the feature weights of the linear model [84]. Based on the bi-directional cognitive ability of the cloud model (data→concept and concept→data), Liu et al. [85] proposed a Cloud-VAE model that embeds understandable concepts. First, the cloud model-based clustering algorithm transforms the initial constraint of latent space into a prior distribution of the concept. Second, the reparameterization trick based on the forward cloud transformation algorithm is designed to estimate the latent space concept by increasing the randomness of latent variables. This model decouples the feature space for deep learning and enhances the interpretability of the model using interpretable concepts obtained from decoupling. Zhang et al. [86] proposed an explainer of graph neural networks. Two pathways are designed: one is a predictor to predict the category of the graph, and the other is an explainer to provide the critical subgraph and nodes for this prediction. Each of these previously mentioned methods uses an additional structure that explains the unexplained prediction mechanism, which effectively reduces the black-box nature of the models.

This separation mechanism satisfies the interpretation needs of most model users, and achieves a balance between model performance and interpretability. Applying this interpretation mechanism to more domains is a goal for further research.

4.3 Brain cognition-inspired computational theory layer

As described in Section 3.3, both the environment and behavior of other individuals affect the decisions of intelligent individuals. From the perspectives of factors that influence intelligence and the specific implementation process of intelligence, we propose two research directions for brain-inspired AI.

4.3.1 Embodied intelligence that integrates the brain cognitive mechanism and AI computation mechanisms

According to Prof. Yao Qizhi, the next challenge in the field of AI is to realize "embodied general AI"[6]. With the development of deep learning, particularly LLMs and visual models, the development of embodied intelligence has accelerated greatly. However, there is no consensus on the main scientific issues and technical routes of embodied intelligence. Prof. Lu Cewu proposed the following directions that may help the development of embodied intelligence. (1) Computer vision models can detect and recognize objects effectively, whereas the objects that agents encounter or operate in the real environment are still not recognized accurately. How to obtain this type of data on a large scale with the corresponding large model is still being explored. (2) Embodied intelligence needs to integrate the senses of sight, audition, and even smell, and combine with LLMs to form a new multimodal large model. Although RT-2 represents the initial exploration of embodied multimodal micromodels, it still has great difficulties in building a theoretically complete and practically feasible technological framework. (3) Imitation learning and augmentation learning are regarded as two major tools for embodied intelligence. Imitation learning records real human operations, but it cannot traverse all feasible human operations and has poor scalability. Augmented learning has strong self-exploration capability, but still relies on the imitation engine. Further research is needed on how to integrate both of them to make embodied intelligence learning realistic and scalable. (4) The imitation engine is a fundamental tool of embodied intelligence. How to build an efficient and physically realistic imitation engine involves the promising intersection research direction. This is a complex and difficult challenge.

Additionally, the learning objectives of existing AI systems are mainly determined by loss functions or reward functions. These functions are usually designed manually and remain fixed during the training process. The computational objectives required by embodied intelligence are more complex. Hence, the design of loss functions inspired by biological evolution or biological perception can enhance the demand for AI models for complex tasks. For example, the gradual maturation of the human brain is the result of the interaction between genes and the environment that obeys the law of use and disuse. Researchers have made initial explorations in this area [87–89].

4.3.2 Intelligence imitation from individual intelligence to group intelligence (social intelligence)

Human intelligent behavior is related not only to individuals but also groups. Swarm intelligence is an intelligence system that consists of many simple individuals, which realizes intelligent behavior through interaction and collaboration between individuals. In an ecosystem, the relationships between populations or individuals are more complex. Maslow's [57] hierarchy of needs suggests that human purposes can be divided into five levels, including basic survival and social culture. In most studies, researchers pay more attention to how to improve the performance of individuals. However, the relationships of interaction and collaboration between different AI models have been ignored.

Studies have been conducted that are inspired by natural phenomena, such as evolution and biological behavior [90–92]. However, with the aim of next-generation AI, the following two research directions still deserve further study. (1) The emergence mechanism of swarm intelligence remains an important topic [93]. How do groups generate intelligence? (2) It is worth studying how to model the tasks and decompose them into small tasks when solving large-scale and complex problems with human-machine hybrid swarm intelligence [93]. This also relates to the relationship between global and local, as mentioned previously.

4.4 AI4BI

The research directions discussed previously are all inspirations from cognitive science to AI. Furthermore, as AI advances, it also has numerous beneficial reciprocal influences on cognitive science.

The application of AI methods in the field of cognitive science is mainly divided into two levels. The first is the use

6) https://www.reemanrobot.com/info/towards-embodied-general-artificial-intelligen-87258676.html

of AI models to replace traditional methods to aid cognitive science research. Prof. Dai Qionghai et al. [78] proposed an ultra-wide, ultra-resolution, and ultra-fast microscope imaging instrument called RUSH-I. RUSH-I is a multi-dimensional and multi-scale high-resolution computational camera that can be used to observe the cellular activity of the brain. It provides a new tool to study the structure and function of subshells, cells, tissues, and organs from an *in vivo* study, and solve the conflict between the field of view and resolution [94,95]. Second, AI as a tool to advance brain science. In 2016, Huth et al. [96] investigated the language processing mechanisms in the brain. They visualized active areas of the brain using a data-driven approach. Specifically, they used word embedding to build relationships between words and brain regions, and then used a regression algorithm to obtain semantic maps. Developed in 2022 to explore the relationship between language understanding and deep language representation, Caucheteux et al. [97] uses a linear model to predict brain activity based on GPT-2 activation. Because LLMs have strong interactive abilities and their responses are very similar to those of humans, these large models are replacing the human to become the objects of study for cognitive psychologists. This approach avoids issues such as environmental variables and ethics, and is beneficial for the development of psychology. Wu investigated the biological neural information processing mechanism using a continuous attractor neural network, and proposed a series of original models and mathematical tools. This network is used to explain various brain functions [98–103]. Moreover, BrainPy, which is an international leading programming platform for neural modeling and brain-like computation, has been further developed [104].

Additionally, certain ideas from AI are shaping the development of cognitive science. The prevailing concept in AI is the data-driven strategy, which relies on large datasets to acquire knowledge. However, there is a lack of such datasets in cognitive science research. The construction of datasets consistent with brain cognition is a very important future task. Prof. Wu Si believes that the future construction of cognition-compliant datasets encompasses the following aspects: (1) memory research-related datasets: to study the mechanisms of short-term and long-term memory, in addition to the process of information encoding, storage, and retrieval, datasets should include the results of memory experiments, such as recall tasks and recognition tasks; (2) decision-making and problem-solving related datasets: to study how people make decisions, in addition to problem-solving strategies and processes, these datasets should include information about choices, inference processes, and results in different decision-making scenarios; (3) social cognition-related datasets: to study how humans understand and process social information, including understanding the intentions, emotions, and behaviors of others, the datasets should include the experimental results of social interactions, such as imitation and emotional contagion; (4) cognitive science-related datasets: to focus on the cognitive development of children and adolescents, the datasets should cover, for example, language acquisition, conceptual development, and moral judgment; (5) mental health and cognitive disorder-related datasets: to study the changes of cognitive function in mental illness and neurodegenerative diseases, the datasets should cover the results of cognitive assessments and clinical diagnostic information.

5 Conclusions

In this study, we presented a comprehensive analysis of the current state, limitations, and research directions of brain-inspired AI. From the day it was born, the goal of AI researchers has been to develop machines that can emulate human thinking. However, despite significant progress in various areas of AI, current data-driven AI differs fundamentally from the intricate cognitive processes of humans. Cognitive science, specifically cognitive psychology and cognitive neuroscience, has provided valuable inspiration for the development of AI. Currently, both generalized AI and GAI have made significant progress and yielded unforeseen outcomes. Some researchers even believe that they can develop independently from cognitive science. However, we believe that to imbue machines with intelligence, it is necessary to understand brain intelligence first. Only with a deep understanding of this can we effectively implement intelligence using machines.

Based on the three-layer structure proposed by Marr, in this review, we discussed and analyzed the limitations of the current development of brain-inspired AI regarding possessing the cognitive ability and cognitive behavior of a human. We proposed eight research directions, exploring both brain-inspired AI and AI-assisted brain intelligence, with a focus on future AI research.

In conclusion, interdisciplinary research between AI and brain cognitive science is an important scientific research direction. More intelligent, efficient, and humanized AI systems are expected to be developed by understanding the working mechanism of the human brain and integrating it with advanced AI algorithms. For instance, could novel cognitive insights such as the shallow brain hypothesis inspire the creation of innovative artificial neural architectures that address limitations, such as a lack of interpretability and inadequate robustness? Furthermore, could cutting-edge large-scale models, for example, assist cognitive scientists in advancing our understanding of the interaction patterns between different brain regions across multiple modalities? The integration of these two fields has immense potential for future research.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62221005, 61936001, and 62376045), the Natural Science Foundation of Chongqing, China (Grant Nos. cstc2021ycjh-bgzxm0013), and the Project of Chongqing Municipal Education Commission, China (Grant No. HZ2021008).

1 Moor J. The dartmouth college artificial intelligence conference: The next fifty years. Ai Mag, 2006, 27: 87
2 Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA, 1982, 79: 2554–2558
3 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput, 1997, 9: 1735–1780
4 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Commun ACM, 2017, 60: 84–90
5 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518: 529–533
6 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, 2017. 30
7 Miller G A. The cognitive revolution: A historical perspective. Trends Cogn Sci, 2003, 7: 141–144
8 Nie J, Jiang J, Li Y, et al. Data and domain knowledge dual-driven artificial intelligence: Survey, applications, and challenges. Expert Syst, 2023, e13425
9 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778
10 Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529: 484–489
11 Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. In: Proceedings of the 34th International Conference on Neural Information Processing Syste. Vancouver, 2020. 33: 1877–1901
12 Peebles W, Xie S. Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 4195–4205
13 Geirhos R, Michaelis C, Wichmann F A, et al. Imagenet-trained CNNs are biased towards texture increasing shape bias improves accuracy and robustness. In: Proceedings of the 7th International Conference on Learning Representations (ICLR). New Orleans, 2019
14 Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. In: Proceedings of the 2nd International Conference on Learning Representations (ICLR). Banff, 2014
15 Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge: MIT Press, 2010
16 Chen L. Core fundamental scientific issues in next-generation artificial intelligence: The relationship between cognition and computation. Bull Chin Acad Sci, 2018, 33: 1104–1106
17 Ohki T, Kunii N, Chao Z C. Efficient, continual, and generalized learning in the brain-neural mechanism of Mental Schema 2.0. Rev Neurosci, 2023, 34: 839–868
18 Yu Z, Zhang Y, You J, et al. Adaptive semi-supervised classifier ensemble for high dimensional data classification. IEEE Trans Cybern, 2019, 49: 366–379
19 von Neumann J, Kurzweil R. The Computer and the Brain. New Haven: Yale University Press, 2012
20 Newell A, Simon H A. Computer science as empirical inquiry. Commun ACM, 1976, 19: 113–126
21 Newell A, Simon H A. The Simulation of Human Thought. Santa Monica: Rand Corporation, 1959
22 Newell A, Simon H A. Computer simulation of human thinking. Science, 1961, 134: 2011–2017
23 Newell A, Shaw J C, Simon H A. Empirical explorations of the logic theory machine: A case study in heuristic. In: Proceedings of the Western Joint Computer Conference: Techniques for Reliability. Los Angeles, 1957, 218–230
24 Newell A, Simon H. The logic theory machine—A complex information processing system. IEEE Trans Inform Theor, 1956, 2: 61–79
25 Hao W. A Survey of Mathematical Logic. Beijing: Science Press, 1962
26 Turing A M. I.—Computing machinery and intelligence. Mind, 1950, LIX: 433–460
27 Duan J, Yu S, Tan H L, et al. A survey of embodied AI: From simulators to research tasks. IEEE Trans Emerg Top Comput Intell, 2022, 6: 230–244
28 Brooks R A. A robot that walks; emergent behaviors from a carefully evolved network. Neural Comput, 1989, 1: 253–262
29 Subramanian A, Chitlangia S, Baths V. Reinforcement learning and its connections with neuroscience and psychology. Neural Networks, 2022, 145: 271–287
30 McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys, 1943, 5: 115–133
31 Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev, 1958, 65: 386–408
32 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol, 1962, 160: 106–154
33 Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern, 1980, 36: 193–202
34 Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature, 1986, 323: 533–536
35 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE, 1998, 86: 2278–2324
36 Jia D, Wei D, Socher R, et al. ImageNet: A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255
37 Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 1–9
38 Katsuki F, Constantinidis C. Bottom-up and top-down attention. Neuroscientist, 2014, 20: 509–521
39 Hubel D H, Wiesel T N. Brain mechanisms of vision. Sci Am, 1979, 241: 150–162
40 Minsky M, Papert S. Perceptrons: An Introduction to Computational Geometry. Cambridge: MIT Press, 1969
41 Barlow H B. Single units and sensation: A neuron doctrine for perceptual psychology? Perception, 1972, 1: 371–394
42 Rosenfeld A, Thurston M. Edge and curve detection for visual scene analysis. IEEE Trans Comput, 1971, C-20: 562–569
43 Horn B K P. Determining lightness from an image. Comput Graph Image Process, 1974, 3: 277–299
44 Zhang T, Xu B. Research advances and perspectives on spiking neural networks (in Chinese). Chin J Comput, 2020: 1767–1785
45 Huo B, Gao Y, Qi X. Research on spiking neural networks for brain-inspired computing. J Image Graph, 2023, 28: 401–417
46 Lechner M, Hasani R, Amini A, et al. Neural circuit policies enabling auditable autonomy. Nat Mach Intell, 2020, 2: 642–652
47 Pei J, Deng L, Song S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature, 2019, 572: 106–111
48 Kurakin A, Goodfellow I J, Bengio S. Adversarial examples in the physical world. In: Artificial Intelligence Safety and Security. Carabas: Chapman and Hall/CRC, 2018. 99–112
49 Carlini N, Wagner D. Towards evaluating the robustness of neural networks. In: Proceedings of the 2017 IEEE Symposium on Security

and Privacy (SP). San Jose: IEEE, 2017. 39–57
50 Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, 2015
51 Chen L. Topological structure in visual perception. Science, 1982, 218: 699–700
52 Tsao T, Tsao D Y. A topological solution to object segmentation and tracking. Proc Natl Acad Sci USA, 2022, 119: e2204248119
53 Dong S, Gong Y, Shi J, et al. Brain cognition-inspired dual-pathway CNN architecture for image classification. IEEE Trans Neural Netw Learn Syst, 2024, 35: 9900–9914
54 Hutson M. Robo-writers: The rise and risks of language-generating AI. Nature, 2021, 591: 22–25
55 Mitchell M, Krakauer D C. The debate over understanding in AI's large language models. Proc Natl Acad Sci USA, 2023, 120: e2215907120
56 Blank S C. Speech production: Wernicke, Broca and beyond. Brain, 2002, 125: 1829–1838
57 Maslow A H. A theory of human motivation. Psychol Rev, 1943, 50: 370–396
58 Pawlak Z. Rough set theory and its applications to data analysis. Cybern Syst, 1998, 29: 661–688
59 Zadeh L A. Fuzzy sets. Inf Control, 1965, 8: 338–353
60 Mukherjee A, Lam N H, Wimmer R D, et al. Thalamic circuits for independent control of prefrontal signal and noise. Nature, 2021, 600: 100–104
61 Shabel S J, Proulx C D, Piriz J, et al. GABA/glutamate co-release controls habenula output and is modified by antidepressant treatment. Science, 2014, 345: 1494–1498
62 Zhu L, Dong S, Li J, et al. Ultra-high temporal resolution visual reconstruction from a fovea-like spike camera via spiking neuron model. IEEE Trans Pattern Anal Mach Intell, 2022, 45: 1233–1249
63 Zheng Y, Zheng L, Yu Z, et al. High-speed image reconstruction through short-term plasticity for spiking cameras. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 6358–6367
64 Zhao J, Xiong R, Liu H, et al. Spk2imgnet: Learning to reconstruct dynamic scene from continuous spike stream. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 11996–12005
65 Ding J, Yu Z, Tian Y, et al. Optimal ANN-SNN conversion for fast and accurate inference in deep spiking neural networks. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. Montreal, 2021. 2328–2336
66 Zhu Y, Fang W, Xie X, et al. Exploring loss functions for time-based training strategy in spiking neural networks. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2023
67 Zhang C, Tian K, Fan B, et al. Continual stereo matching of continuous driving scenes with growing architecture. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. 18879–18888
68 Beniaguev D, Segev I, London M. Single cortical neurons as deep artificial neural networks. Neuron, 2021, 109: 2727–2739.e3
69 Suzuki M, Pennartz C M A, Aru J. How deep is the brain? The shallow brain hypothesis. Nat Rev Neurosci, 2023, 24: 778–791
70 Hawkins J, Lewis M, Klukas M, et al. A framework for intelligence and cortical function based on grid cells in the neocortex. Front Neural Circuits, 2019, 12: 121
71 Chen C L P, Liu Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE Trans Neural Netw Learn Syst, 2017, 29: 10–24
72 Lin X, Li L, Shi B, et al. Slow and weak attractor computation embedded in fast and strong EI balanced neural dynamics. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2024
73 Wang G. DGCC: Data-driven granular cognitive computing. Granul Comput, 2017, 2: 343–355
74 Wang G. Data-driven granular cognitive computing. In: Rough Sets. IJCRS 2017. Cham: Springer, 2017
75 Xia S, Zheng S, Wang G, et al. Granular ball sampling for noisy label classification or imbalanced classification. IEEE Trans Neural Netw Learn Syst, 2023, 34: 2144–2155
76 Xia S, Liu Y, Ding X, et al. Granular ball computing classifiers for efficient, scalable and robust learning. Inf Sci, 2019, 483: 136–152
77 Xia S, Peng D, Meng D, et al. A fast adaptive k-means with no bounds. IEEE Trans Pattern Anal Mach Intell, 2020, 1
78 Dai D, Li Y, Wang L, et al. Sketch less face image retrieval: A new challenge. In: Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023. 1–5
79 Zhang L, Zhang B. Theory of fuzzy quotient space (methods of fuzzy granular computing). J Softw, 2003, 14: 770–776
80 Li Y, Chen Y, Wang N, et al. Scale-aware trident networks for object detection. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 6054–6063
81 Chen L, Zhang W, Dong C, et al. FedTKD: A trustworthy heterogeneous federated learning based on adaptive knowledge distillation. Entropy, 2024, 26: 96
82 Pinker S. The Language Instinct: How the Mind Creates Language. London: Penguin UK, 2003
83 Piloto L S, Weinstein A, Battaglia P, et al. Intuitive physics learning in a deep-learning model inspired by developmental psychology. Nat Hum Behav, 2022, 6: 1257–1267
84 Ribeiro M T, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. San Diego, 2016. 97–101
85 Liu Y, Liu Z, Li S, et al. Cloud-VAE: Variational autoencoder with concepts embedded. Pattern Recogn, 2023, 140: 109530
86 Zhang Y, Cheung W K, Liu Q, et al. Towards explaining graph neural networks via preserving prediction ranking and structural dependency. Inf Process Manage, 2024, 61: 103571
87 Zeng Y, Zhang T, Xu B. Improving multi-layer spiking neural networks by incorporating brain-inspired rules. Sci China Inf Sci, 2017, 60: 052201
88 Gütig R, Sompolinsky H. The tempotron: A neuron that learns spike timing-based decisions. Nat Neurosci, 2006, 9: 420–428
89 Wu J, Huang Z, Huang W, et al. Prioritized experience-based reinforcement learning with human guidance for autonomous driving. IEEE Trans Neural Netw Learn Syst, 2024, 35: 855–869
90 Holland J H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Cambridge: MIT Press, 1992
91 Storn R, Price K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim, 1997, 11: 341–359
92 Beni G, Wang J. Swarm intelligence in cellular robotic systems. In: Robots and Biological Systems: Towards a New Bionics? Berlin, Heidelberg: Springer, 1993. 703–712
93 Wang G Y, Cheng D D, Xia D Y, et al. Swarm intelligence research: From bio-inspired single-population swarm intelligence to human-machine hybrid swarm intelligence. Mach Intell Res, 2023, 20: 121–144
94 Wu J, Lu Z, Jiang D, et al. Iterative tomography with digital adaptive optics permits hour-long intravital observation of 3D subcellular dynamics at millisecond scale. Cell, 2021, 184: 3318–3332
95 Zhou T, Lin X, Wu J, et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. Nat Photon, 2021, 15: 367–373
96 Huth A G, de Heer W A, Griffiths T L, et al. Natural speech reveals

the semantic maps that tile human cerebral cortex. Nature, 2016, 532: 453–458

97 Caucheteux C, Gramfort A, King J R. Evidence of a predictive coding hierarchy in the human brain listening to speech. Nat Hum Behav, 2023, 7: 430–441

98 Dong X, Ji Z, Chu T, et al. Adaptation accelerating sampling-based bayesian inference in attractor neural networks. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, 2022. 21534–21547

99 Chu T, Ji Z, Zuo J, et al. Firing rate adaptation in continuous attractor neural networks accounts for theta phase shift of hippocampal place cells. Neuron, 2022, 3: 5

100 Zhang W, Wu Y N, Wu S. Translation-equivariant representation in recurrent networks with a continuous manifold of attractors. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, 2022. 15770–15783

101 Dong X, Chu T, Huang T, et al. Noisy adaptation generates Lévy flights in attractor neural networks. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. Sydney, 2021. 16791–16804

102 Lin X, Li L, Shi B, et al. Slow and weak attractor computation embedded in fast and strong EI balanced neural dynamics. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2023

103 Zhang W H, Wu S, Josić K, et al. Sampling-based Bayesian inference in recurrent circuits of stochastic spiking neurons. Nat Commun, 2023, 14: 7074

104 Wang C, Zhang T, Chen X, et al. BrainPy, a flexible, integrative, efficient, and extensible framework for general-purpose brain dynamics programming. eLife, 2023, 12: e86365