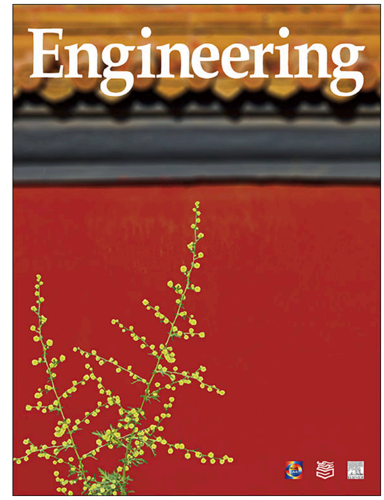
Journal Pre-proofs



Perspective article

Machine Memory Intelligence: Inspired by Human Memory Mechanisms

Qinghua Zheng, Huan Liu, Xiaoqing Zhang, Caixia Yan, Xiangyong Cao, Tieliang Gong, Yong-Jin Liu, Bin Shi, Zhen Peng, Xiaocen Fan, Ying Cai, Jun Liu

PII: S2095-8099(25)00029-3
DOI: https://doi.org/10.1016/j.eng.2025.01.012
Reference: ENG 1804

To appear in: *Engineering*

Received Date: 22 August 2024
Revised Date: 24 December 2024
Accepted Date: 24 January 2025

Please cite this article as: Q. Zheng, H. Liu, X. Zhang, C. Yan, X. Cao, T. Gong, Y-J. Liu, B. Shi, Z. Peng, X. Fan, Y. Cai, J. Liu, Machine Memory Intelligence: Inspired by Human Memory Mechanisms, *Engineering* (2025), doi: https://doi.org/10.1016/j.eng.2025.01.012

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company

Journal Pre-proofs

Artificial Intelligence—Perspective

Machine Memory Intelligence: Inspired by Human Memory Mechanisms

Qinghua Zheng [a,b], Huan Liu [a,b,*], Xiaoqing Zhang [c,d], Caixia Yan [a,b], Xiangyong Cao [a,b], Tieliang Gong [a,b], Yong-Jin Liu [e], Bin Shi [a,b], Zhen Peng [a,b], Xiaocen Fan [c,d], Ying Cai [f], and Jun Liu [a,b]

[a] *School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China*

[b] *Shaanxi Provincial Key Laboratory of Big Data Knowledge Engineering, Xi'an Jiaotong University, Xi'an 710049, China*

[c] *Translational Research Institute of Brain and Brain-Like Intelligence, Shanghai Fourth People's Hospital, Shanghai 200434, China*

[d] *School of Medicine, Tongji University, Shanghai 200092, China*

[e] *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

[f] *Department of Computer Science, Iowa State University, Ames, IA 50011, USA*

* Corresponding author.

E-mail address: huanliu@xjtu.edu.cn (H. Liu).

Journal Pre-proofs

ARTICLE INFO

Article history:
Received xxx
Revised xxx
Accepted xxx
Available online

Keywords:

Machine memory intelligence
Neural mechanism
Associative representation
Continual learning
Collaborative reasoning

ABSTRACT

Large models, exemplified by ChatGPT, have reached the pinnacle of contemporary artificial intelligence (AI). However, they are plagued by three inherent drawbacks: excessive training data and computing power consumption, susceptibility to catastrophic forgetting, and a deficiency in logical reasoning capabilities within black-box models. To address these challenges, we draw insights from human memory mechanisms to introduce "machine memory," which we define as a storage structure formed by encoding external information into a machine-representable and computable format. Centered on machine memory, we propose the brand-new machine memory intelligence (M^2I) framework, which encompasses representation, learning, and reasoning modules and loops. We explore the key issues and recent advances in the four core aspects of M^2I, including neural mechanisms, associative representation, continual learning, and collaborative reasoning within machine memory. M^2I aims to liberate machine intelligence from the confines of data-centric neural networks and fundamentally break through the limitations of existing large models, driving a qualitative leap from weak to strong AI.

Journal Pre-proofs

1. Introduction

Artificial intelligence (AI) is experiencing a rapid development phase from weak to strong. Breakthroughs in deep learning algorithms, along with the expansion of data and computing power, have enabled large models to achieve or even surpass human-level performance in various fields. However, these models also exhibit notable drawbacks, such as excessive data and computing power consumption, susceptibility to catastrophic forgetting, and a deficiency in logical reasoning capabilities within black-box models. To truly realize machine intelligence, it is essential to explore new and sustainable approaches that diverge from the current deep neural network paradigms.

1.1. Large models and their inherent drawbacks

Large models, exemplified by ChatGPT, have adopted the technical route map of big data, massive computing power, and strong algorithms. This has significantly advanced fields such as language understanding, intent recognition, and content generation, reaching the pinnacle of contemporary AI. However, large models are plagued by three inherent drawbacks. First, they consume data and computing power excessively, resulting in significant energy consumption. For example, merely training GPT-3 requires 1.287×10^6 kW·h of electricity [1], and ChatGPT consumes 5.6×10^5 kW·h of electricity daily during the problem-solving phase [2]. Second, they are susceptible to catastrophic forgetting. When large models learn new tasks, they often quickly forget previously learned knowledge, leading to a significant decline in performance on earlier tasks [3]. Finally, they are black-box models with a deficiency in logical reasoning capabilities. For instance, Turing Award winner Yann LeCun believes that the existing autoregressive large models can hardly perform reasoning and planning tasks.

The root causes of the aforementioned limitations can be attributed to the architecture of artificial neural networks, the training mechanisms used for large models, and the models' data-driven reasoning mechanisms. First, the architecture of artificial neural networks lacks alignment with the underlying mechanisms of problems, resulting in poor adaptability and interpretability. Second, during the training phase, the backpropagation algorithm iteratively updates all parameters based on global errors. This process consumes a large amount of training data and computing power; it can also lead to the forgetting of earlier knowledge. Third, during the reasoning phase, the forward propagation mechanism involves all parameters in the computation, which also consumes a substantial amount of computing power.

1.2. Inspiration from the human brain's memory mechanisms

Research in brain science and cognitive science has been a consistent driving force throughout the 70-year development of AI. Studies have shown that memory plays a fundamental role in human brain intelligence [4–6]. More specifically, the impact of memory on the brain's intellectual activities (e.g., learning, abstraction, association, and reasoning) runs through the three main stages of encoding, storage, and retrieval.

The unique memory mechanisms of the human brain provide important insights for overcoming the inherent flaws of traditional artificial neural networks. First, the brain's memory-activation mechanism can retrieve a small amount of knowledge from the long-term memory (LTM) into the working memory for reasoning. Although the human brain contains approximately 10^{11} neurons and 10^{15} synapses, its energy consumption is only 20–23 W, while the energy consumption of equivalently scaled large models is five orders of magnitude higher [7]. Second, the synaptic plasticity [8] of the neurons in the human brain is the foundation of learning and memory formation. This plasticity is characterized by local, autonomous, and unsupervised features, which prevent the catastrophic forgetting problem observed in large models, in which all parameters are updated, resulting in new things being learned at the cost of old ones being forgotten. Lastly, the dual-system cooperation mechanism [9] of brain memory underpins complex reasoning in humans. This mechanism holds the potential to enhance the weak logical reasoning capabilities of black-box models in large-scale AI systems.

1.3. Machine memory intelligence

Inspired by the memory mechanisms of the human brain, we propose the fundamental concept of "machine memory," which we define as a multilayered, distributed network storage structure formed by encoding external information into a machine-representable and computable format. This structure supports dynamic updates and reorganization, spatiotemporal associations, abstract concrete associations, and fuzzy hash access. Centered on machine memory, we also propose a machine memory intelligence (M^2I) framework (Fig. 1), which utilizes representation, learning, and reasoning modules, along with brain-inspired mechanisms, to form two interactive loops that enable the machine memory's associative representation, continual learning, and collaborative reasoning. Using this architecture, M^2I can fundamentally break through the difficulties of brute-force computation, catastrophic forgetting, and deficient logical reasoning inherent in existing large models. Table 1 provides a summarized comparison of M^2I and large models in the three capabilities of representation, learning, and reasoning.

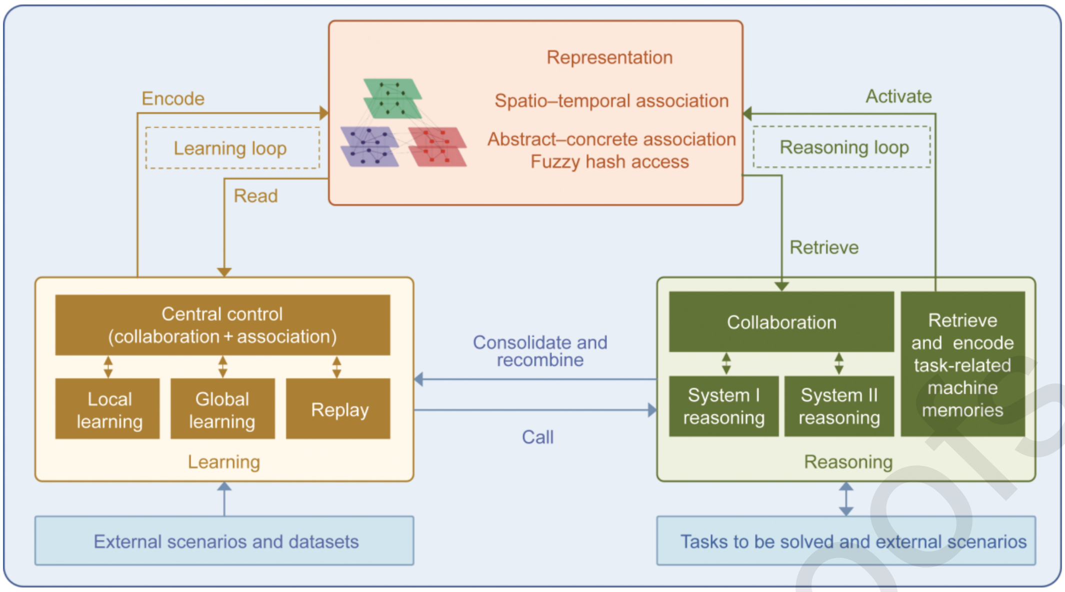


Fig. 1. The framework of the proposed M²I.

M²I contains three main modules:

- **A representation module.** This module is responsible for the structured representation of various types of knowledge or information encoded in the learning module; it then realizes hierarchical classified organization and storage. Inspired by the associative mechanisms of the human brain and the distinct representations of various memory structures, the representation module should incorporate the key features of associative memory. These include the ability to support spatiotemporal associations, abstract-concrete associations, and fuzzy hash access, among others. Concurrently, it offers fundamental operations for heterogeneous knowledge or information post-encoding, such as abstraction, association, updating, and access.

- **A learning module.** This module is designed to extract new knowledge from external scenarios and datasets, encode this knowledge, and form the basic units that constitute machine memory. Emulating the dual nature of the human brain's learning process and its memory-replay mechanisms, the learning module integrates both local and global learning strategies. This integration enables continual learning and equips the system with the capacity for autonomous reconstruction and consolidation of machine memory.

- **A reasoning module.** This module activates and retrieves relevant machine memories iteratively, according to the task to be solved and the current scene, and synergizes different reasoning modes to obtain the solution result. Informed by the dual-process theory and the reward mechanisms of the human brain, the reasoning module is designed to integrate both intuitive and logical reasoning modes, fostering collaborative problem-solving. Moreover, the module is capable of migrating redundant system II processes to system I, thereby facilitating intuitive reasoning and increasing the speed and energy efficiency of the reasoning process.

The interaction between the learning module and the representation module creates a learning loop. Here, the learning module collaborates with both local and global learning mechanisms to encode external scenes and datasets into the basic units of machine memory, which are then represented by the representation module as spatio–temporal associations and abstract–concrete associations of machine memory. Meanwhile, the learning module also needs to read (or retrieve) existing machine memories, collaborating in local learning, global learning, and the playback process, thereby forming a loop.

The interaction between the reasoning module and the representation module creates a reasoning loop. The reasoning module extracts key information from the tasks to be solved and external scenes for encoding, activating the machine memory within the representation module based on this information. It then utilizes the retrieved memories to facilitate intuitive reasoning and logical reasoning. The above process is executed iteratively, forming a loop.

Table 1

Comparison of the capabilities of M^{2}I versus large models.

Item	Capability		
	Representation	Learning	Reasoning
Large models	Parameterized knowledge; hierarchical network structure with weak plasticity; unable to extract content-related knowledge	Backpropagation (global learning); excessive training data and computing power consumption; catastrophic forgetting	System I reasoning; all parameters involved; catastrophic forgetting; learning and reasoning are relatively independent
M^{2}I	Dynamic decomposition and reconstruction of knowledge; hierarchical network structure with strong plasticity; content-based fuzzy hash memory access	Fusion of Hebbian learning and backpropagation (global + local); low training data and computing power consumption; memory-based continual learning	Collaborative reasoning between System I and System II; partial parameters involved; memory can be consolidated, updated, and forgotten; learning, reasoning, and machine memory interact and iterate

M^{2}I focuses on four directions: ①What neural mechanism supports machine memory? ② How can associative representation of machine memory be achieved? ③ How can continual learning be carried out under low-power conditions? ④How can the dual-system cooperation of intuition and logic in reasoning be realized? Sections 2–5 systematically address the key issues and significant advancements in each of these four directions. In summary, M^{2}I adopts an entirely new technical methodology, which holds the potential to thoroughly overcome the inherent bottlenecks encountered by current large models.

2. Related work on brain-memory-inspired machine models

In this section, we review the latest advancements in machine models that draw inspiration from the memory functions of the human brain. First of all, spiking neural networks (SNNs) represent the forefront of artificial neural networks, embracing a brain-inspired approach to information processing [10]. Within the biological brain, neurons transmit information across synapses through electrical pulses, operating in a sparse and asynchronous fashion. SNNs emulate this functionality, with their neurons exchanging information via temporal events known as spikes and adhering to integrate-and-fire (IF) dynamics. Spike-timing-dependent plasticity (STDP) is a locally derived learning rule and has been proposed to train SNNs. As a synaptic plasticity mechanism, STDP fine-tunes synaptic weights in accordance with the precise timing of spikes between neurons. This temporal modulation allows neural networks to learn and adapt based on the sequence of input spikes, mirroring the learning and memory processes observed in biological nervous systems.

Secondly, in brain-memory-inspired models, replay is a key point of connection between AI and neuroscience. Replay in the brain has been viewed as rehearsal or as sampling from a transition model and is often used to address the continual learning issues [11] of deep learning models. A straightforward idea is to approximate and recover old data distributions by storing a few old training samples or training a generative model. Typical sub-directions [12] include experience replay, which saves a few old training samples in a memory buffer; generative replay, which trains a generative model to provide generated samples; and feature replay, which recovers the distribution of old features through saving prototypes, saving statistical information or training a generative model.

In addition to the replay mechanism, active forgetting of the biological brain is employed to overcome catastrophic forgetting with continual learning. For example, by modeling a robust *Drosophila* learning system that actively regulates forgetting with multiple learning modules, Wang et al. [13] proposed a generic approach that appropriately attenuates old memories in parameter distributions to improve learning plasticity and accordingly coordinates a multi-learner architecture to ensure solution compatibility. Furthermore, Kurth-Nelson et al. [14] proposed that replay in the brain has a more general function: deriving new knowledge through compositional computation. Therefore, they suggested that replay—as a compositional computation embedded in the larger brain system including the cortex (which behaves like a deep neural network in many ways)—should be mapped to machine learning techniques that hybridize compositional computation with deep learning. Additionally, the replay mechanism can be combined with

Journal Pre-proofs

Journal Pre-proofs

combinations during the replay process, thereby achieving structured reasoning.

Thirdly, the hippocampus is crucial for the formation of human memory, and recent studies have explored the relationships between the hippocampus and promising novel neural network architectures. For instance, Whittington et al. [15] mathematically associated and compared representation methods of transformer architecture with those of the human hippocampus, providing a fresh perspective on explaining the biological plausibility of transformers. Inspired by such findings, Kim et al. [16] proposed a new non-linear activation function that mimics *N*-methyl-D-aspartate receptor (NMDAR) dynamics. NMDAR-like nonlinearity shifts short-term working memory into long-term reference memory in transformers, thus enhancing a process that is similar to memory consolidation in the mammalian brain.

Recently, Jiang et al. [17] suggested that attention should be paid to the self-evolution of AI models. Compared with the use of large-scale data to train models, self-evolution may only use limited data or interactions. Drawing inspiration from the columnar organization of the human cerebral cortex, the researchers hypothesized that AI models could potentially develop emergent cognitive capabilities and construct internal representational models through iterative interactions with their environment. To achieve this, Jiang et al.[17] proposed that models must be equipped with LTM, which stores and manages processed real-world interaction data. LTM not only enables the representation of long-tail individual data in statistical models but also facilitates self-evolution by supporting diverse experiences across various environments and agents.

3. The neural mechanisms of machine memory

Unveiling and translating human learning and memory principles contribute to the development of M^{2}I. In this section, we elaborate on the key issues and progress in neuronal and neural network pre-configuration, the neuronal and circuit plasticity of learning and memory, and the generative grammar of brain intelligence.

3.1. Key issues regarding the neural mechanisms of machine memory

- **The neuronal and neural network pre-configuration of brain intelligence.** It is now well-acknowledged that human intelligence originates from the reception of external stimuli and subsequent processing of these input signals for learning and memory. During human brain development, neuroectoderm cells are patterned to variable regional neural progenitors, which then differentiate into specific neuronal subtypes, constructing the basic framework for sensing and processing external signals. Fully elucidating how neuronal and neural networks are pre-configured and how the human brain is constructed step-by-step into neurobiologically and neuropsychologically responsive neuronal ensembles during embryonic development is essential in order to build functional modules for M^{2}I.

- **The neuronal and neural circuit plasticity of brain learning and memory.** The competence and performance of the human brain's neurobiological or neuropsychological function are attributed to the temporal and spatial activation of neural networks formed of neurons and synapses. Plasticity at the neuronal or neural circuit level accounts for adaptive processes regarding experiences, resulting in the generation or evolution of learning and memory. Tracing the participating neurons or neural circuits and unveiling their cellular and molecular adaptations in a specific learning and memory paradigm are the key to elucidating brain intelligence.

- **The generative grammar of brain intelligence.** The newly developed brain is just like a pre-configured blank slate. Immediate experience exposure sequentially activates or inhibits a cluster of neurons, which internally generate representations of the external stimuli. Repeated, similar, or related experience exposures might lead to a replay of these neuronal ensembles and therefore strengthen or amend these internally generated representations. The internally generated representation repertoires are selected upon new stimuli and used to form and express novel representations of the external world. We hypothesize that, with pre-configured neurons and neural networks, which are capable of generating intelligence, there is a generative grammar that writes, erases, and reads external experiences at either the cellular or circuit level.

3.2. Major progress in the neural mechanisms of machine memory

During embryonic development, pluripotent epiblast cells are defaulted into neuroectoderm cells, which are arranged into a single-layer neural plate. The neural plate then curves up, and its left and right edges melt at the dorsal midline to form a neural tube [18]. It is during this neural tube formation period that, by patterning morphogens, neuroectoderm cells gradually take regional identities, such as the cortical and hippocampal neural progenitors at the dorsal region of the forebrain; and the lateral, medial and caudal ganglionic eminence and pre-optic area neural progenitors at the ventral forebrain [19]. Neural progenitors at the dorsal forebrain largely differentiate into g [illegible] id (GABAergic) inhibitory neurons or cholinergic neurons [20,21]. Neurons can either migrate or project to local or distal brain regions, where they form pre-configured neural networks through synapses. The newly developed clusters of neurons and neural networks are pre-configured to sense and process specific external stimuli just like a pre-configured blank slate and purposely receive, record, and rework information from the external world (Fig. 2).

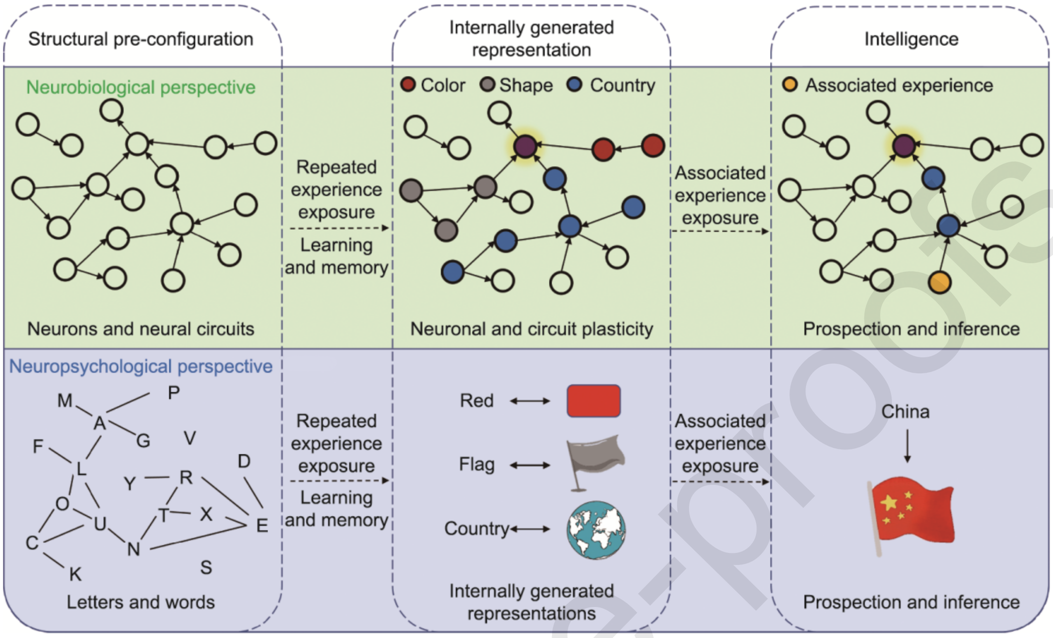


Fig. 2. Working model of the generative grammar of brain intelligence.

Owing to the rapid development of single-cell transcriptomic analysis technologies, neuronal identities have been carefully depicted, along with their molecular features and developmental trajectories, in most brain regions of model organisms and of humans during development [22]. Human pluripotent stem cells, including human embryonic stem cells and human induced pluripotent stem cells, also hold the potential to develop into different brain regional neural progenitors and then neurons in culture [23,24]. The *in vitro* generated neurons and neural networks might serve as a pre-configured blank slate for stimuli sensing and processing and therefore provide an unprecedented model for studying or mimicking intelligence generation.

During the learning process, stimuli are sensed by primary neurons, which transmit signals to downstream neurons through action-potential-mediated synaptic activation or inhibition [25]. In another type of learning, multiple signals can converge and integrate on a downstream neuron through several input neurons [26]. During the memory process, signals with attention are processed to internally generate representations at either the neuronal or neural circuit level [27].

Learning and memory behavioral studies are used to characterize learning and learned states in animal models. Calcium imaging [28], immediate early gene reporter [29], and *in vivo* electrophysiological recording [30] are recent technologies that are capable of mapping the neurons and neural circuits involved in a behavior event. High-resolution imaging technologies are now used to investigate synaptic adaptations in learning, such as newly generated or decayed synapses and dynamic changes in the number of receptors for neural transmitters at the post-synaptic membrane [31]. Our recent study developed genetic nuclear membrane labeling followed by a single nuclear sequencing strategy, which can be applied to capture functionally related neurons and determine the underlying molecular mechanisms of learning and memory [32]. Projection-specific chemogenetic and optogenetic manipulations are now widely applied to reveal functional projections in the acquisition, consolidation, and retrieval of behavior, including learning and memory [33].

The formation of brain intelligence is somewhat similar to the generation of linguistic performance, where the brain's neurons can be viewed as letters and its sequentially activated neuronal ensembles as a vocabulary list. The most challenging aspects of brain intelligence are how the brain summarizes and extracts a specific meaning from these vocabulary repertoires, how a sentence is organized with a putative linguistic grammar, and how a strong

Journal Pre-proofs

Journal Pre-proofs

efforts from neurobiologists, neuropsychologists, and scientists in machine intelligence might pave a way to decode the brain's general generative grammar, which could then be applied to unveiling brain intelligence and constructing the next generation of machine intelligence.

4. Associative representations in machine memory

The representation of machine memory is the foundational construct that allows AI systems to encode, store, and retrieve information effectively. The key challenges and main progress in this field can be summarized as follows.

4.1. Key issues related to associative representations in machine memory

- **The abstract–concrete association issue.** Memory reconstruction theory posits that memory is not a mere copy of past sensory experiences but rather a reconstruction based on semantic content. Indeed, neuroscience research has discovered the existence of an "abstract–concrete" joint representation mechanism in both rodents and humans. Notably, the fan cells in the lateral entorhinal cortex of the human brain can create associations between entities, resulting in associative memories. Consequently, leveraging this representational mechanism to associatively encode various forms of sensory information, such as visual and linguistic data, along with semantic abstractions in machine perception, presents a significant challenge. This endeavor aims to improve the plasticity and flexibility of activation methods within machine memory's associative representations.

- **The spatiotemporal dynamic association issue.** Neuroscience research has revealed that human memory storage is characterized by its fragmented and dynamic nature. Spatially, memories are spatially decomposed into distinct components and allocated to different brain regions. Temporally, the complexity of interactions between various brain systems leads to the dynamic updating of memory fragments over time. While dynamic associations across space and time are fundamental to the construction and consolidation of memory in the human brain, current machine learning models represent memory information as specific parametric knowledge, which lacks explicit spatiotemporal encoding and the capacity for associative mechanisms related to memory content. As a result, it is extremely challenging to replicate the human brain's memory-reconstruction mechanisms to develop a topological structure that evolves and associates along spatial and temporal dimensions in machine memory.

- **The efficient hashing access issue.** The brain employs the concept of distributed storage to maintain an extensive repository of memories; through an efficient fuzzy hash access mechanism, it can swiftly pinpoint the necessary memories using specific cues or critical information. However, current machine learning models encounter challenges with low access efficiency when confronted with a plethora of memory information, struggling to rapidly retrieve and identify the memories associated with given prompts. Therefore, it is challenging to emulate the brain's high-performance hashing access mechanism to construct a distributed hash storage structure and a mapping mechanism that is sensitive to localized machine memory.

4.2. Major progress related to associative representations in machine memory

Here, we review the progress that has been achieved in associative representations in machine memory within three areas: the abstract–concrete association of memory, the spatiotemporal association of memory, and efficient memory access.

In the realm of abstract-concrete memory association, researchers from the Institute of Cognitive Neuroscience, University College London, have combined the use of reinforcement learning algorithms and brain imaging techniques, demonstrating a mechanism of abstraction built upon the valuation of sensory features [34]. Professor David W. Tank at the Princeton Neuroscience Institute, Princeton University, investigated the neural coding mechanisms within the hippocampus that underlie the formation of abstract knowledge [35], revealing how spatial and non-spatial information is integrated to create a geometric framework for storing and retrieving complex concepts. A cooperative team from Fudan University and the University of York discovered the common neural circuits that assist in the retrieval of semantic memory and episodic memory and thereby enable the brain to flexibly extract long-term memories with different functions [36]. Professor Neil Burgess from University College London proposed a generative model that captures the abstract-concrete association in memory (Fig. 3) [37], illustrating how memories are dynamically formed and reified from the interplay of abstract concepts and concrete instances.

Regarding spatiotemporal memory association, researchers at the University of Oxford have constructed the Tolman–Eichenbaum machine (TEM) as a theoretical model that unifies spatial and relational memory through the process of generalization within the hippocampal formation [38]. Collaborative research between Peking University and the Chinese Academy of Sciences Brain Science and Intelligence Technology Innovation Center found that the neural activity state space during the maintenance of sequential memory is composed of mutually orthogonal

al. [40] developed a bio-inspired spiking content-addressable memory model based on the CA3 region of the hippocampus with the ability to learn, forget, and recall memories.

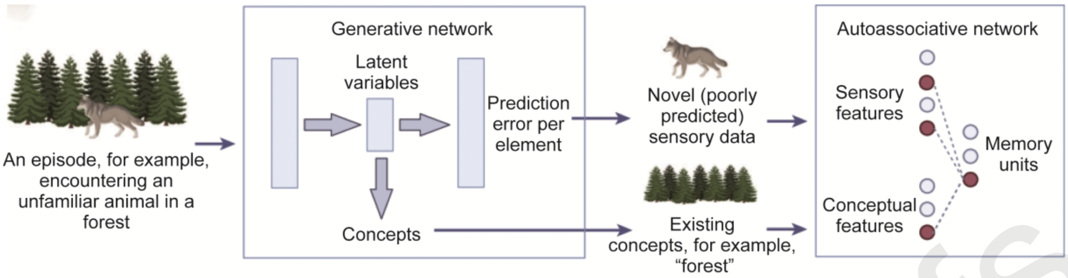


Fig. 3. Illustration of the associative generative model for memory construction and consolidation proposed in Ref. [37].

As for the mechanisms of memory access, Dasgupta et al. [41] proposed an efficient hash-based similarity search algorithm inspired by the working mechanism of the olfactory cells of fruit flies, whose input is an odor and whose output is a tag (called a "hash") for that odor. This work provides a new perspective on efficient access to memory in machines. Professor Tajana Simunic at the University of California, San Diego, proposed a fully binary brain-inspired classifier based on hyperdimensional computing for energy-efficient and high-accuracy classification [42]. Luo et al. [43] revealed the interactive mechanisms between the hippocampus and the cerebral cortex during the learning process, discovering an efficient coding method that integrates distributed storage with multi-sensory perception. Inspired by the working principles of brain neurons, researchers have introduced a nanowire network capable of instant learning and memory, marking a crucial step toward the development of highly efficient, energy-saving machine intelligence [44].

These works not only deepen our understanding of the mechanism of memory encoding but also provide theoretical support for future research on machine memory representation and learning.

4.3. Discussion

4.3.1.. What human brain memory research inspired associative representation in machine memory?

From the perspective of information theory, human memory is a process of encoding, storing, and retrieving external input information. Endel Tulving, a Canadian cognitive psychologist and a member of the National Academy of Sciences, proposed the Systems of Processing Information (SPI) theory [45,46] of memory functions. This theory posits that the memory system consists of multiple memory modules that perform specific functions, such as the perceptual input system, working memory system, language processing system, conceptual system, and LTM system. The relationships between these modules are characterized by serial information encoding, parallel information storage, and independent information retrieval (e.g., when recalling an event, we can independently retrieve relevant visual, auditory, emotional, and other types of information as needed).

Inspired by this theory, we aim to explore the associative collaborative representation mechanisms of machine memory, focusing on distributed machine memory representations that combine abstraction and generalization, dynamic spatial-temporal association, and efficient fuzzy hashing retrieval. This research endeavors to achieve distributed machine memory representations that support multi-level abstraction, spatial-temporal association, and efficient retrieval.

4.3.2 How does the M^2I model achieve associative representations in machine memory?

The whole-brain mapping theory [47,48] suggests that the brain utilizes the principle of distributed storage to preserve vast amounts of memories. That is, memories stored in the brain are not concentrated in a specific area but are stored in a dispersed manner throughout the neural circuits of the entire brain. Therefore, we intend to construct a distributed hash storage structure composed of multilevel storage nodes for machine memory and to design a consistent hash mapping function for distributed storage nodes, providing a foundation for the efficient and stable retrieval of memories.

Journal Pre-proofs

an efficient fuzzy hashing mechanism for machine memory, enabling the rapid association and retrieval of similar but distinct memories and ensuring the efficiency, fault tolerance, and robustness of memory retrieval. This approach offers insights into addressing the shortcomings of large models that excessively consume data and computing resources.

Inspired by the reconstructive mechanisms [50] of human memory, we construct a distributed multi-channel machine memory semantic network and study the generative self-associative connections of abstract-concrete machine memory. On this basis, we aim to establish a machine memory storage mechanism that decouples cues from content and to construct an explicit coding and association mechanism for machine memory across time and space. This mechanism will overturn the traditional structure of machine learning models based on implicit parameterized knowledge, enabling the extraction of specific content or related knowledge. It will also help to address the weakness of traditional black-box models in terms of logical reasoning capabilities.

5. Continual learning in machine memory

Continual learning in machine memory focuses on addressing the catastrophic forgetting phenomenon. It transcends traditional continual-learning frameworks by incorporating biomimetic principles that emulate the human brain's remarkable capacity for lifelong learning and adaptation. The architectural foundation of this methodology is predicated on achieving a tradeoff between synaptic plasticity and memory stability, thereby facilitating robust knowledge acquisition while preserving previously encoded information. The key challenges and fundamental processes in this domain can be systematically categorized and analyzed through the lens of neuroscientific principles and computational requirements, as delineated in the following.

5.1. Key issues related to continual learning in machine memory

- **Adaptive regulation in machine memory control.** The adaptive regulation mechanism of the human brain's memory-learning patterns forms the foundation of memory learning. Neuroscientific research has revealed that different learning modes in human brain memory may excel at various types of scenario requirements and memory objectives. Simultaneously, the memory-learning process usually involves the collaboration of multiple learning modes. Therefore, a significant challenge in the design of machine memory control lies in how to draw inspiration from and simulate the adaptive regulation mechanism of the human brain's memory-learning patterns. The goal is to automatically select appropriate single or combined learning modes for different scenarios, thereby enhancing the efficiency of machine memory learning.

- **Collaborative associative continual learning in machine memory.** Humans' capacity for continual learning forms the basis for both reinforcing existing knowledge and assimilating new information. However, current continual learning models often suffer from drawbacks such as high computational costs and difficulties in knowledge transfer, making them inefficient in handling complex and dynamic big data scenarios. Consequently, the question of how to draw inspiration from the Hebbian learning characteristics of the human brain to establish a collaborative mechanism integrating local and global learning, as well as an associative replay learning mechanism, is a major challenge.

- **Updating and reorganization of machine memory.** Memory updating and reorganization enable the brain to rapidly adapt to dynamically changing scenarios. However, existing machine learning models lack effective mechanisms for updating and reorganization, making it difficult for them to swiftly adapt to complex and dynamic scenarios. Therefore, a notable challenge that needs to be addressed is how to draw inspiration from Bayesian brain theory and sleep-related neuroscience research to organically incorporate Bayesian principles and memory-replay mechanisms into methods for machine memory updating and reorganization.

5.2. Major progress related to continual learning in machine memory

Replay-based approaches represent a fundamental paradigm in neurologically inspired approaches to continual learning, serving as a critical mechanism for mitigating catastrophic forgetting and facilitating knowledge retention in neural networks. This section focuses on reviewing recent developments in neurologically inspired continual learning from the perspectives of experience replay and generative replay.

Experience replay typically involves storing a subset of old training samples in a memory buffer and strategically utilizing them during the learning process. A great deal of effort has been put toward experience-replay methodologies, with a focus on two principal dimensions: the construction of efficient memory buffers and the effective exploitation of stored information. For example, Rolnick et al. [51] explored a straightforward solution by

Journal Pre-proofs

using experience-replay buffers for all past events and adopted a mixture of on- and off-policy learning by leveraging l … nt task to previously learned tasks through a discriminative embedding space inspired by complementary learning systems theory. Wang et al. [53] proposed memory replay with data compression (MRDC) to reduce the storage cost of old training samples and thus increase the number of stored samples in the memory buffer. Wang et al. [54] suggested a brain-inspired deep retrieval and imagination (DRI) framework, in which a generative model is designed to produce imaginary data and leverages knowledge distillation to retrieve past experiences. Boschini et al. [55] described a simple and effective approach that combines rehearsal and knowledge distillation; this method allows a deep-learning-based model to revise its replay memory to welcome novel information regarding past data, paving the way for learning as-yet-unseen classes. Wang et al. [56] developed a generic approach that appropriately attenuates old memories in parameter distributions to improve learning plasticity by modeling a robust *Drosophila* learning system that actively regulates forgetting with multiple learning modules, as shown in Fig. 4.

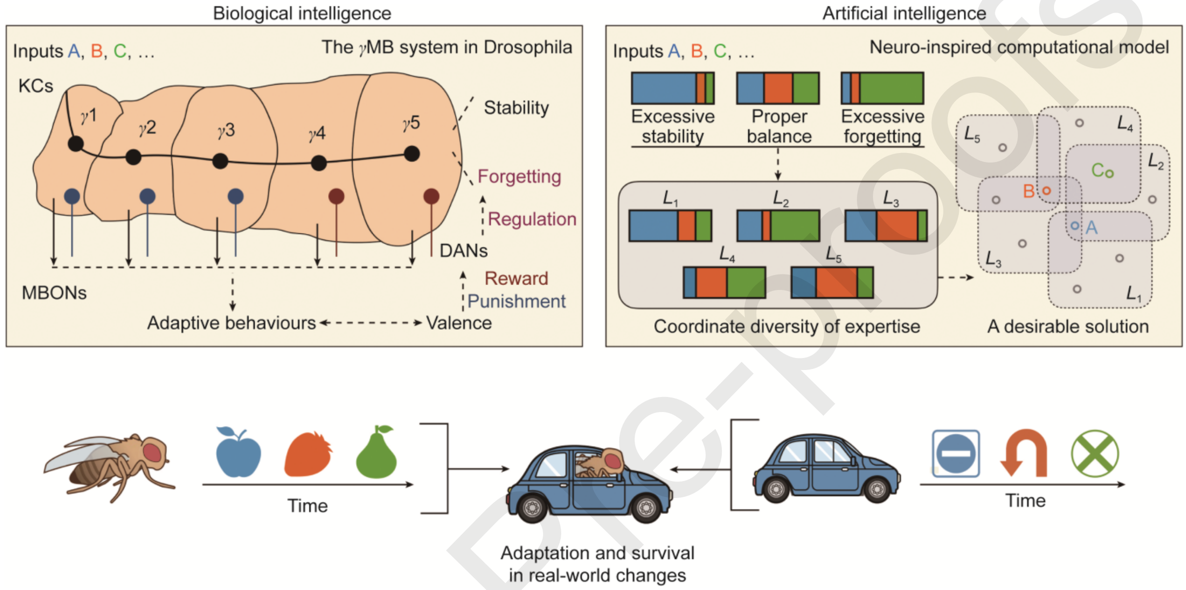


Fig. 4. Continual learning with reference to a biological learning system. KCS: Kenyon cells; MBONs: mushroom body output neurons; γMB: the γ subset of the *Drosophila* mushroom body; L_1–L_5: five continual learners corresponding to the five compartments. Reproduced from Ref. [56] with permission.

Generative replay usually requires training an additional generative model to replay generated data—that is, recover the old data distribution. Many methods have recently been proposed to implement the generative replay methodology. For example, Shin et al. [57] suggested a preliminary framework that incorporates the replaying of generated data from the old generative model into the process of learning each generation task; this approach enables the inheritance of previously acquired knowledge. Wu et al. [58] extended this approach by utilizing replay alignment to ensure the consistency of generative data sampled with the same random noise across both the old and new generative models, in a process that resembles function regularization in enforcing coherence. Rostami et al. [59] developed a computational model that can efficiently expand its previously learned concepts to new domains using a few labeled samples.Van de Ven et al. [60] proposed a new, brain-inspired variant of replay in which internal or hidden representations are replayed that are generated by the network's own context-modulated feedback connections without storing data. Inspired by the organization principles of the brain's memory system, Wang et al. [61] applied a triple network generative adversarial network (GAN) architecture to model the interplay of the hippocampus, prefrontal cortex, and sensory cortex.

5.3. Discussion

5.3.1. What human brain memory research inspired the continual learning of machine memory?

For the continual learning of machine memory, we drew inspiration from the adaptive nature of biological neural networks, which exhibit the ability to dynamically adjust synaptic plasticity in response to changing inputs. This adaptability includes: ① preserving learned synaptic changes to combat interference, which underpins the concept of weight regularization to selectively control parameter adjustments [62,63]; ② expanding or pruning functional

Journal Pre-proofs

space [64,65]; ③ regulating synaptic plasticity based on activity and persistence, akin to meta-plasticity or the adaptability of adaptability, in line with the principles of meta-learning [66,67]; and ④ utilizing inhibitory synapses to temper the activity of stimulated neurons, which resembles the role of binary masks in parameter allocation [68,69]. We were also inspired by the Hebbian learning theory [70] to design the parameter update mechanism of our machine memory model; that is, only the local parameter is updated, allowing our model to alleviate the issue of catastrophic forgetting.

5.3.2. How does the M^2I model achieve the continual learning of machine memory?

Machine memory resolves catastrophic forgetting through three interconnected, brain-inspired mechanisms that work synergistically to maintain knowledge stability while enabling continual learning.

At its core, the adaptive regulation mechanism [71] mirrors the human brain’s memory-learning patterns by dynamically selecting and coordinating multiple learning modes, automatically optimizing the learning process for different scenarios while efficiently allocating computational resources. This adaptive control is seamlessly integrated with a collaborative associative learning system that implements Hebbian learning principles [70], combining local and global learning approaches to maintain both specific task performance and overall knowledge coherence. The local learning component focuses on task-specific updates while the global learning preserves broader knowledge structures, with an associative replay learning mechanism reinforcing critical connections and facilitating efficient knowledge transfer across different domains.

This collaborative associative approach overcomes the limitations of current continual learning models, which include high computational costs and difficulties in knowledge transfer, making it efficient for handling complex and dynamic big data scenarios.

The collaborative associative continual learning of M^2I is further enhanced by a memory updating and reorganization mechanism inspired by Bayesian brain theory [72] and sleep-related neuroscience research [73], which enables rapid adaptation to dynamic scenarios through probabilistic knowledge updates and selective memory replay. By incorporating Bayesian principles and memory-replay mechanisms, the system can efficiently update and reorganize stored information while preserving essential knowledge structures, significantly reducing computational overhead compared with traditional continual learning approaches. This mechanism, along with the implementation of biologically inspired learning processes and advanced resource-optimization techniques, ensures the robust preservation of existing knowledge while facilitating the incorporation of new information.

6. Collaborative reasoning by machine memory

Based on the efficient associative representations and continual learning of machine memory, the performance of interpretable reasoning based on a task or a problem to be solved and the consolidation and updating of memories are at the core of M^2I. The key challenges and main progress in this field can be summarized as follows.

6.1. Key issues in collaborative reasoning by machine memory

- **Multi-type machine memory activation for complex scenes and tasks.** During the process of memory retrieval in the human brain, the working memory selectively attends to important information during specific scenes and tasks while actively suppressing irrelevant stimuli. Subsequently, various types of memories are activated in brain regions such as the hippocampus and prefrontal cortex based on content relevance. Therefore, the key to designing an efficient mechanism for memory activation in machines has two aspects: first, addressing the problem of information purification and filtering for scenes and tasks in order to enhance the representation of underlying relational elements across different scenes and tasks, thereby achieving the highly scalable encoding of scene and task information; and, second, tackling the problem of efficient and comprehensive activation for multiple types of memory to avoid the vast search space that results from the ambiguity and complexity of scene semantics, thereby improving retrieval efficiency while reducing resource consumption.

- **Cooperative reasoning between different types of memory.** Although existing large language models (LLMs) are capable of reasoning with various modalities, such as text, images, and three dimensions (3D), they still lack interpretability in the reasoning process. The dual-process collaborative reasoning mechanism of intuition and reasoning in the human brain provides promising insight for addressing the low reasoning efficiency and poor interpretability of existing LLMs. To incorporate this mechanism into machine memory models, the major challenges lie in constructing a generative intuition reasoning system, denoted as system I, and an iterative logical

Journal Pre-proofs

reasoning system, denoted as system II; designing pathway-selection mechanisms for complex reasoning tasks;

Journal Pre-proofs

- **The design of consolidation and transfer mechanisms for different types of memory.** In order to address complex and dynamic tasks in real-world scenes, existing LLMs require task-specific fine-tuning and inference adaptation, which results in high energy consumption. Inspired by the reward mechanisms of learning and memory in the human brain, constructing new associations between memories and consolidating them to achieve rapid intuitive reasoning in system I can greatly enhance reasoning speed and reduce machine inference energy consumption. The challenge lies in how to implement sparse reward modeling for complex reasoning tasks, stimulate memory consolidation, establish mechanisms for generating and transferring associations between memories, and migrate high-energy logical reasoning results to system I to achieve fast intuitive reasoning.

6.2. Major progress in collaborative reasoning by machine memory

The AI community is currently striving to develop various approaches in order to acquire the complex reasoning capability inherent in the notion of intelligence, which currently represents a major bottleneck. Although several influential experts believe that the integration route is the most promising [74] and propose a so-called neuro-symbolic AI approach [75], theoretical research on human cognition and decision reasoning, which aims to understand how humans evolved to acquire advanced capabilities and contributes to raising key fundamental research questions for the future development of AI, is still lacking. At present, Kahneman's theory of thinking fast and slow [76] (also known as dual process theory) has become a widely recognized principle and is rapidly being absorbed into the AI environment. This section briefly reviews the progress that has been achieved by related work along this line.

Before going further, we introduce a well-known, successful case to illustrate how powerful an intelligent agent with complex human-like reasoning capabilities can be. In 2016, AlphaGo shocked the world by defeating humans in the strategy game Go [77]. Its huge success is widely believed to come from the adoption of dual process cognitive systems [78]. To be specific, AlphaGo and its many variants employ a collaborative framework that combines fast deep neural networks with slower tree search methods [77,79,80]. The tree search model deduces all possible action outcomes based on the current state and selects the optimal one, while a deep neural network assigns a policy value to specific board configurations based on feedback from the actions selected and played [81]. As system I, the deep neural network acts as a pre-attention stage, building fast associations to select relevant representations for system II powered by the tree search. Although both systems can operate independently, the best performance beyond human levels is achieved only when they interact in synergy.

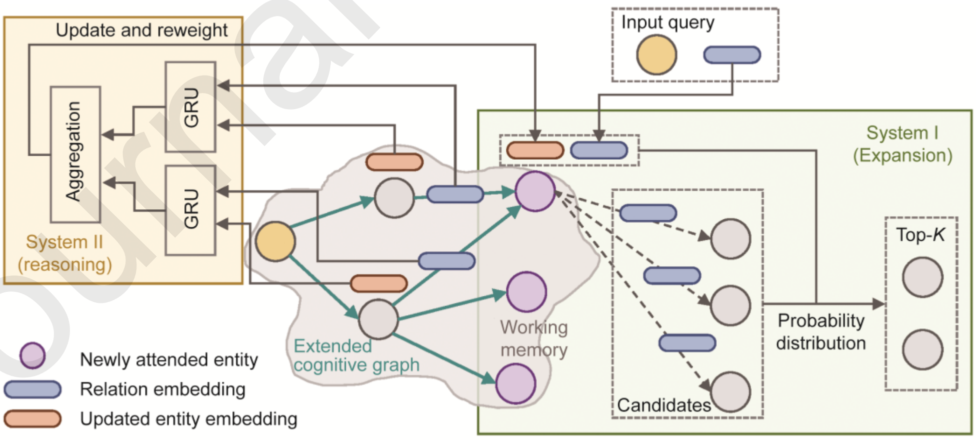


Fig. 5. Overview of the dual system architecture for explainable recommendation in CogKR [82]. GRU: gated recurrent unit. *K* is conventionally a positive integer and is the number of items in the returned list.

Recently, dual process theory has been increasingly integrated with AI, especially in the field of natural language processing. For example, CogKR [82] adopts a representative dual system architecture, as shown in Fig. 5, which iteratively coordinates an extension module and a reasoning module to perform multi-hop knowledge graph reasoning in the form of subgraphs instead of individual paths. Hua et al. [83] built a two-system architecture that utilizes neural logic reasoning (system II) on top of representation learning models (system I). CogER [84] is a cognition–aware knowledge graph reasoning model for explainable recommendation, in which system I generates an intuitive estimation of the next entity, and system II conducts explicit reasoning to select the most promising entity. Motivated

Journal Pre-proofs

in complex reasoning [85]. The Statler framework [86] equips LLMs with a persistent, memory-like representation of the world state. By maintaining the world state, LLMs can access context-related information over extended periods, greatly enhancing their reasoning capabilities. Memory-of-thought (MoT) [87] is a memory-of-thought framework in which LLMs save high-confidence thoughts as external memory through pre-thinking and then recall relevant memory to help their reasoning. Despite some empirical success, research on collaborative reasoning in advanced AI is still in its initial stage.

6.3. Discussion

6.3.1. What human brain memory research inspired the collaborative reasoning of machine memory?

As discussed earlier, the data-driven inference approach of deep learning involves all parameters in the computation, which is highly resource intensive. Moreover, this approach primarily learns correlations from data, presenting significant limitations in content neutrality and interpretability. However, a revisitation of earlier symbolic reasoning methods reveals that they excel in these areas. Nevertheless, symbolic reasoning has limitations: ①It often struggles with handling uncertainty and ambiguity, making it difficult to adapt to new data or unseen scenarios and leading to weaker generalization abilities; and ② it is challenging to use to represent knowledge involving high-dimensional data such as text, images, and audio-video content. In fact, this is the motivation behind our approach.

According to the dual process theory [81] in cognitive neuroscience, the human brain possesses two markedly different information-processing mechanisms. System I relies on experience and intuition and quickly engages in association and pattern matching, akin to deep learning. System II, on the other hand, operates slowly and consciously, requiring more cognitive load and attention; it is suitable for analysis, reasoning, and solving complex problems, similar to symbolic reasoning. Drawing inspiration from the dual system mechanism of the human brain and integrating the parametric knowledge (memory) of deep learning with the symbolic knowledge (memory) from early knowledge engineering holds great promise for addressing the aforementioned issues.

6.3.2 How does the M^{2}I model achieve collaborative reasoning by machine memory?

Admittedly, a substantial body of work has focused on the design of individual reasoning modules, with typical representatives including various types of knowledge graph reasoning models [88]. However, the great majority of existing techniques—whether embedding-based [89–91], path-based [92–94], or rule-based [95–97]—only involve the learning and reasoning process of a single model for relatively simple tasks such as head reasoning and relation inferring. In contrast, the concept of collaborative reasoning we have elaborated includes multiple machine-memory-related mechanisms and is oriented to more complicated reasoning scenarios such as advanced algebra and complex task planning. To this end, we plan to draw on dual process theory from cognitive science and the neurobiological bases of brain memory to build a reasoning module that contains a system for fast intuitive reasoning (–system I) and a system for rational logical reasoning (system II).

The reasoning module can retrieve different types of task-related machine memories, such as abstract semantics and spatiotemporal associations, which are various forms of knowledge organized and stored in a hierarchical and categorized form by the representation module, according to the external scenarios and tasks to be solved. Based on the retrieved memory, the reasoning module integrates the two reasoning patterns of the generative-based system I and iterative-based system II to work collaboratively, during which new relevant memories are extracted as supplements to promote long-chain reasoning. In the process of collaborative work, the dynamic selection of reasoning pathways ensures that the most appropriate system is selected at each reasoning stage, while the conflict-resolution mechanism aims to resolve conflicts between the multiple types of retrieved memories to maintain the rationality and correctness of the reasoning. Furthermore, the module is intended to realize the formation of associations between memories, the consolidation of memories (i.e., the enhancement of correct memories and rectification of incorrect memories), and their migration from system II to system I, enabling fast intuitive reasoning (when faced with the same problem) that significantly improves the reasoning speed and energy efficiency.

Although some efforts have been made to explore collaborative reasoning models [82,98,99] for different tasks such as knowledge graph reasoning, multi-hop question answering, and classification under continual learning, such efforts predominantly focus on the iterative interaction between two functionally independent systems, neglecting to consider memory consolidation and its transfer mechanism between systems, which is the key to efficient reasoning in the human brain. Unlike existing studies, the collaborative reasoning we advocate is expected to fundamentally overcome the inherent defect of weak logical reasoning in black box models and achieve advanced reasoning with

Journal Pre-proofs

This kind of human-brain-like collaborative reasoning will greatly enhance the development of advanced AI in various fields. For example, in the medical field, given a series of medical records such as a patient's temperature sheets, laboratory test reports, medical imaging examination results, and so on, the machine-memory-based reasoning module could automatically derive professional diagnostic opinions for the patient, including a traceable and interpretable diagnostic basis and prescriptions. The experience from each diagnosis will form specific machine memories to promote the accuracy and efficiency of the subsequent reasoning. At the same time, these memories will make various associations with existing related memories, including supplementation, enhancement, and weakening. The organic integration of multiple memories is likely to generate new insights that will help humans identify unknown diseases.

7. Conclusions and future work

LLMs and multimodal models, such as ChatGPT and SORA, have attracted widespread attention from academia and industry due to their robust capabilities for understanding, generating, and interacting, as well as their potential for cross-domain applications. These models represent a pivotal direction in current AI research. However, their reliance on data-driven neural network frameworks inevitably brings inherent limitations, including excessive consumption of training data and computational power, catastrophic forgetting, weak logical reasoning abilities, and a lack of interpretability.

The memory mechanisms of the human brain play a crucial role in human intelligence. Specifically, the brain's synaptic plasticity, multi-level integration, and associative activation mechanisms endow it with high neural plasticity, preventing catastrophic forgetting and enabling the execution of complex reasoning and planning tasks with low power consumption. Inspired by the memory mechanisms of the human brain, this perspective article introduces the interdisciplinary research direction of machine memory, which is positioned at the intersection of AI, neuroscience, and cognitive science. We identify a series of challenges in this domain, including the neurological mechanisms of machine memory, associative representation, continual learning, and collaborative reasoning. Research in the direction of machine memory aims to address the path dependency on traditional artificial neural networks for achieving machine intelligence, thereby overcoming the issues of brute-force computation, catastrophic forgetting, and weak logical reasoning capabilities inherent in current LLMs and multimodal models. Furthermore, leveraging M^2I can help validate theories and hypotheses in brain science, particularly those related to human brain memory, thus promoting the deep integration, cross-validation, and collaborative development of AI and brain science.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62137002, 62250009, 62202367, 82025020, and 82230072).

Compliance with ethics guidelines

Qinghua Zheng, Huan Liu, Xiaoqing Zhang, Caixia Yan, Xiangyong Cao, Tieliang Gong, Yong-Jin Liu, Bin Shi, Zhen Peng, Xiaocen Fan, Ying Cai, and Jun Liu declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Luccioni AS, Viguier S, Ligozat AL. Estimating the carbon footprint of BLOOM, a 176B parameter language model. J Mach Learn Res 2023;24(253):1–15.

[2] de Vries A. The growing energy footprint of artificial intelligence. Joule 2023;7(10):2191–4.

[3] Kotha S, Springer JM, Raghunathan A. Understanding catastrophic forgetting in language models via implicit inference. 2023. arXiv:2309.10105.

[4] Nęcka E. Cognitive analysis of intelligence: the significance of working memory processes. Pers Individ Dif 1992;13(9):1031–46.

[5] Yoo J, Wood F. BayespCN: a continually learnable predictive coding associative memory. In: Proceedings of the 36th International Conference on Neural Information Processing Systems; 2022 Nov 28–Dec 9; New Orleans, LA, USA. Red Hook: Curran Associates Inc.; 2022. p. 29903–14.

[7] Mehonic A, Kenyon AJ. Brain-inspired computing needs a master plan. Nature 2022;604(7905):255–60.

[8] Goto A, Bota A, Miya K, Wang J, Tsukamoto S, Jiang X, et al. Stepwise synaptic plasticity events drive the early phase of memory consolidation. Science 2021;374(6569):857–63.

[9] Evans JSB. Dual-process theories of reasoning: contemporary issues and developmental applications. Dev Rev 2011;31(2–3):86–102.

[10] Dampfhoffer M, Mesquida T, Valentian A . Backpropagation-based learning techniques for deep spiking neural networks: a survey. IEEE Trans Neural Netw Learn Syst 2024;35(9):11906–21.

[11] Ostapenko O, Lesort T, Rodriguez P, Rifat Arefin MD, Douillard A, Rish I, et al. Continual learning with foundation models: an empirical study of latent replay. In: Proceedings of the Conference on Lifelong Learning Agents; 2022 Aug 22–24; Montréal, QC, Canada. Online: PMLR; 2022, p. 60–91.

[12] Wang L, Zhang X, Su H.A comprehensive survey of continual learning: theory, method and application. IEEE Trans Pattern Anal Mach Intell 2024;46:5362–83.

[13] Wang L, Zhang X, Li Q . Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nat Mach Intell 2023;5(12):1356–68.

[14] Kurth-Nelson Z, Behrens T, Wayne G . Replay and compositional computation. Neuron 2023;111(4):454–69.

[15] Whittington JCR, Warren J, Behrens TEJ. Relating transformers to models and neural representations of the hippocampal formation. 2021. arXiv:2112.04035.

[16] Kim DK, Kwon J, Cha M. Transformer as a hippocampal memory consolidation model based on NMDAR-inspired nonlinearity. In: Proceedings of the 37th International Conference on Neural Information Processing Systems; 2023 Dec 10–16; New Orleans, LA, USA. Red Hook: Curran Associates Inc.; 2024. p. 14637–64.

[17] Jiang X, Li F, Zhao H, Wang J, Shao J, Xu SH, et al. Long term memory: the foundation of AI self-evolution. 2024. arXiv:2410.15665.

[18] Sadler T, Greenberg D, Coughlin P, Lessard J. Actin distribution patterns in the mouse neural tube during neurulation. Science 1982;215(4529):172–4.

[19] Chi L, Fan B, Feng D, Chen Z, Liu Z, Hui Y, et al. The dorsoventral patterning of human forebrain follows an activation/transformation model. Cereb Cortex 2017;27(5):2941–54.

[20] Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 2016;352(6293):1586–90.

[21] Mi D, Li Z, Lim L, Li M, Moissidis M, Yang Y, et al. Early emergence of cortical interneuron diversity in the mouse embryo. Science 2018;360(6384):81–5.

[22] Ma L, Du Y, Hui Y, Li N, Fan B, Zhang X, et al. Developmental programming and lineage branching of early human telencephalon. EMBO J 2021;40(21):e107277.

[23] Tao Y, Zhang SC. Neural subtype specification from human pluripotent stem cells. Cell Stem Cell 2016;19(5):573–86.

[24] Zhang X, Huang CT, Chen J, Pankratz MT, Xi J, Li J, et al. Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell 2010;7(1):90–100.

[25] Yamada T, Yang Y, Valnegri P, Juric I, Abnousi A, Markwalter KH, et al. Sensory experience remodels

Journal Pre-proofs

genome architecture in neural circuit to drive motor learning. Nature 2019;569(7758):708–13.
Journal Pre-proofs
[26] Yadav N, Noble C, Niemeyer JE, Terceros A, Victor J, Liston C, et al. Prefrontal feature representations drive memory recall. Nature 2022;608(7921):153–60.

[27] Bradley C, Nydam AS, Dux PE, Mattingley JB. State-dependent effects of neural stimulation on brain function and cognition. Nat Rev Neurosci 2022;23(8):459–75.

[28] Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nat Methods 2019;16(7):649–57.

[29] Kitamura T, Ogawa SK, Roy DS, Okuyama T, Morrissey MD, Smith LM, et al. Engrams and circuits crucial for systems consolidation of a memory. Science 2017;356(6333):73–8.

[30] Zhou Y, Zhu H, Liu Z, Chen X, Su X, Ma C, et al. A ventral CA1 to nucleus accumbens core engram circuit mediates conditioned place preference for cocaine. Nat Neurosci 2019;22(12):1986–99.

[31] Michalska JM, Lyudchik J, Velicky P, Štefaničková H, Watson JF, Cenameri A, et al. Imaging brain tissue architecture across millimeter to nanometer scales. Nat Biotechnol 2024;42(7):1051–64.

[32] Chen X, Du Y, Broussard GJ, Kislin M, Yuede CM, Zhang S, et al. Transcriptomic mapping uncovers Purkinje neuron plasticity driving learning. Nature 2022;605(7911):722–7.

[33] Fan X, Song J, Ma C, Lv Y, Wang F, Ma L, et al. Noradrenergic signaling mediates cortical early tagging and storage of remote memory. Nat Commun 2022;13(1):7623.

[34] Cortese A, Yamamoto A, Hashemzadeh M, Sepulveda P, Kawato M, De Martino B, et al. Value signals guide abstraction during learning. eLife 2021;10:e68943.

[35] Nieh EH, Schottdorf M, Freeman NW, Low RJ, Lewallen S, Koay SA, et al. Geometry of abstract learned knowledge in the hippocampus. Nature 2021;595(7865):80–4.

[36] Vatansever D, Smallwood J, Jefferies E. Varying demands for cognitive control reveals shared neural processes supporting semantic and episodic memory retrieval. Nat Commun 2021;12(1):2134.

[37] Spens E, Burgess N. A generative model of memory construction and consolidation. Nat Hum Behav 2024;8(3):526–43.

[38] Whittington JC, Muller TH, Mark S, Chen G, Barry C, Burgess N, et al. The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. Cell 2020;183(5):1249–63.

[39] Xie Y, Hu P, Li J, Chen J, Song W, Wang XJ, et al. Geometry of sequence working memory in macaque prefrontal cortex. Science 2022;375(6581):632–9.

[40] Casanueva-Morato D, Ayuso-Martinez A, Dominguez-Morales JP, Jimenez-Fernandez A, Jimenez-Moreno G. Bio-inspired computational memory model of the hippocampus: an approach to a neuromorphic spike-based content-addressable memory. Neural Netw 2024;178:106474.

[41] Dasgupta S, Stevens CF, Navlakha S. A neural algorithm for a fundamental computing problem. Science 2017;358(6364):793–6.

[42] Imani M, Morris J, Messerly J, Shu H, Deng Y, Rosing T, et al. BRIC: locality-based encoding for energy-efficient brain-inspired hyperdimensional computing. In: Proceedings of the 56th Annual Design Automation Conference; 2019 Jun 2–6; Las Vegas, NV, USA. New York City: Association for Computing Machinery (ACM); 2019. p. 1–6.

[43] Luo W, Yun D, Hu Y, Tian M, Yang J, Xu Y, et al. Acquiring new memories in neocortex of hippocampal-lesioned mice. Nat Commun 2022;13(1):1601.
Journal Pre-proofs

memory with neuromorphic nanowire networks. Nat Commun 2023;14(1):6697.

[45] Tulving E. Organization of memory. In: Gazzaniga MS, editor. The cognitive neurosciences. Cambridge: The MIT Press; 1995. p. 839–53.

[46] Wu C, Huang S, Yuan Q. Seven important theories in information system empirical research: a systematic review and future directions. Data Inf Manag 2022;6(1):100006.

[47] Li Z, Shang Z, Liu J,Zhen H, Zhu E, Zhong S, et al. D-LMBmap: a fully automated deep-learning pipeline for whole-brain profiling of neural circuitry. Nat Methods 2023;20(10):1593–604.

[48] Patow G, Martin I, Sanz Perl Y. Whole-brain modelling: an essential tool for understanding brain dynamics. Nat Rev Methods Primers 2024;4(1):53.

[49] Frankland PW, Josselyn SA, Köhler S. The neurobiological foundation of memory retrieval. Nat Neurosci 2019;22(10):1576–85.

[50] Xue G. From remembering to reconstruction: the transformative neural representation of episodic memory. Prog Neurobiol 2022;219:102351.

[51] Rolnick D, Ahuja A, Schwarz J, Lillicrap T, Wayne G. Experience replay for continual learning. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada. Red Hook: Curran Associates Inc.; 2019. p. 350–60.

[52] Rostami M, Kolouri S, Pilly PK. Complementary learning for overcoming catastrophic forgetting using experience replay. 2019. arXiv:1903.04566.

[53] Wang L, Zhang X, Yang K, Yu L, Li C, Hong L, et al. Memory replay with data compression for continual learning. 2022. arXiv:2202.06592.

[54] Wang Z, Liu L, Duan Y, Tao D. Continual learning through retrieval and imagination. In: Proceedings of the 36th AAAI/ACM Conference on Artificial Intelligence; 2022 Feb 22–Mar 1; held Virtually. Washington: AAAI Press; 2022. p. 8594–602.

[55] Boschini M, Bonicelli L, Buzzega P, Porrello A, Calderara S, et al. Class-incremental continual learning into the extended der-verse. IEEE Trans Pattern Anal Mach Intell 2022;45(5):5497–512.

[56] Wang L, Zhang X, Li Q, Zhang M, Su H, Zhu J, et al. Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nat Mach Intell 2023;5(12):1356–68.

[57] Shin H, Lee JK, Kim J, Kim J. Continual learning with deep generative replay. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA. Red Hook: Curran Associates Inc.; 2017. p. 2994–3003.

[58] Wu C, Herranz L, Liu X, van de Weijer J, Raducanu B. Memory replay GANs: learning to generate new categories without forgetting. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems; 2018 Dec 3–8; Montréal, Canada. Red Hook: Curran Associates Inc.; 2018. p. 5966–76.

[59] Rostami M, Kolouri S, Pilly P, McClelland J. Generative continual concept learning. In: Proceedings of the AAAI Conference on Artificial Intelligence; 2020 Feb 7–12; New York City, NY, USA. Washington: AAAI Press; 2020. p. 5545–52.

[60] Van de Ven GM, Siegelmann HT, Tolias AS. Brain-inspired replay for continual learning with artificial neural networks. Nat Commun 2020;11(1):4069.

[61] Wang L, Lei B, Li Q, Su H, Zhu J, Zhong Y, et al. Triple-memory networks: a brain-inspired method for continual learning. IEEE Trans Neural Netw Learn Syst 2021;33(5):1925–34.

Journal Pre-proofs

[62] Hayashi-Takagi A, Yagishita S, Nakamura M. Labelling and optical erasure of synaptic memory traces in

[63] Aljundi R, Babiloni F, Elhoseiny M. Memory aware synapses: learning what (not) to forget. In: Proceedings of the 15th European Conference on Computer Vision (ECCV); 2018 Sep 8–14; Munich, Germany. Berlin: Springer; 2018. p. 139–54.

[64] Bailey CH, Kandel ER, Harris KM. Structural components of synaptic plasticity and memory consolidation. Cold Spring Harb Perspect Biol 2015;7(7):a021758.

[65] Schwarz J, Czarnecki W, Luketina J. Progress & compress: a scalable framework for continual learning. In: Proceedings of the 35th International Conference on Machine Learning (PMLR 2018); 2018 Jul 10–15; Stockholm, Sweden. Online: PMLR; 2018. p. 4528–37.

[66] Abraham WC. Metaplasticity: tuning synapses and networks for plasticity. Nat Rev Neurosci 2008;9(5):387.

[67] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, Australia. New York City: JMLR.org; 2017. p. 1126–35.

[68] Arevian AC, Kapoor V, Urban NN. Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. Nat Neurosci 2008;11(1):80–7.

[69] Serra J, Suris D, Miron M. Overcoming catastrophic forgetting with hard attention to the task. In: Proceedings of the 35th International Conference on Machine Learning (ICML-2018); 2018 Jul 10–15; Stockholm, Sweden. Online: PMLR; 2018. p. 4548–57.

[70] Caporale N, Dan Y. Spike timing-dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci 2008;31(1):25–46.

[71] Bluestone JA, Abbas AK. Natural versus adaptive regulatory T cells. Nat Rev Immunol 2003;3(3):253–7.

[72] Lange RD, Shivkumar S, Chattoraj A. Bayesian encoding and decoding as distinct perspectives on neural coding. Nat Neurosci 2023;26(12):2063–72.

[73] Grandner MA, Fernandez FX. The translational neuroscience of sleep: a contextual framework. Science 2021;374(6567):568–73.

[74] Booch G, Fabiano F, Horesh L, Kate K, Lenchner J, Linck N, et al. Thinking fast and slow in AI. In: Proceedings of the AAAI Conference on Artificial Intelligence; 2021 Feb 2–9; Held virtually. Washington: AAAI Press; 2021. p. 15042–6.

[75] Besold TR, d'Avila Garcez A, Bader S, Bowman H, Domingos P, Hitzler P, et al. Neural-symbolic learning and reasoning: a survey and interpretation. In: Neuro-symbolic artificial intelligence: the state of the art. Amsterdam: IOS Press; 2021. p. 1–51.

[76] Kahneman D. Thinking fast and slow. New York City: Farrar, Strauss and Giroux; 2011.

[77] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature 2016;529(7587):484–9.

[78] Anthony T, Tian Z, Barber D. Thinking fast and slow with deep learning and tree search. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4– 9; Long Beach, CA, USA. Red Hook: Curran Associates Inc.; 2017.

[79] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science 2018;362(6419):1140–4.

[80] Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, et al. Mastering Atari, Go, chess

Journal Pre-proofs

[81] Kelly M, Barron AB. The best of both worlds: dual systems of reasoning in animals and ai. Cognition 2022;225:105118.

[82] Du Z, Zhou C, Yao J, Tu T, Cheng L, Yang H, et al. CogKR: cognitive graph for multi-hop knowledge reasoning. IEEE Trans Knowl Data Eng 2021;35(2):1283–95.

[83] Hua W, Zhang Y. System 1+ system 2 = better world: neural-symbolic chain of logic reasoning. In: Proceedings of the Association for Computational Linguistics: EMNLP; 2022 Dec 7–11; Abu, Dhabi. Stroudsburg: Association for Computational Linguistics (ACL); 2022. p. 601–12.

[84] Bing Q, Zhu Q, Dou Z. Cognition-aware knowledge graph reasoning for explainable recommendation. In: Proceedings of the 16th ACM International Conference on Web Search and Data Mining; 2023 Feb 27–Mar 3; New York City, NY, USA. New York City: Association for Computing Machinery (ACM); 2023. p. 402–10.

[85] Sun J, Zheng C, Xie E, Liu Z, Chu R, Qiu J, et al. A survey of reasoning with foundation models. 2023. arXiv:2312.11562.

[86] Yoneda T, Fang J, Li P, Zhang H, Jiang T, Lin S, et al. Statler: state-maintaining language models for embodied reasoning. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA); 2024 May 13–17, Yokohama, Japan. New Yor City: IEEE; 2024. p. 15083–91.

[87] Li X, Qiu X. Mot: memory-of-thought enables ChatGPT to self-improve. 2023. arXiv:2305.05181.

[88] Liang K, Meng L, Liu M. A survey of knowledge graph reasoning on graph types: static, dynamic, and multi-modal. IEEE Trans Pattern Anal Mach Intell 2024;46(12):9456–78.

[89] Yang B, Yih SW, He X. Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015); 2015 May 7–9; San Diego, CA, USA. Trier: dblp; 2015.

[90] Balaževic I, Allen C, Hospedales T. Multi-relational Poincaré graph embeddings. In: Proceedings of the 33rd Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada. Red Hook: Curran Associates Inc.; 2019. p. 4465–75.

[91] Mai S, Zheng S, Yang Y. Communicative message passing for inductive relation reasoning. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI-21); 2021 Feb 2–9; Held virtually. Washington: AAAI Press; 2021. p. 4294–302.

[92] Das R, Neelakantan A, Belanger D. Chains of reasoning over entities, relations, and text using recurrent neural networks. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics; 2017 Apr 3–7; Valencia, Spain. Stroudsburg: Association for Computational Linguistics (ACL); 2017. p.132–41.

[93] Xiong W, Hoang T, Wang WY. DeepPath: a reinforcement learning method for knowledge graph reasoning. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; 2017 Sep 9–11; Copenhagen, Denmark. Stroudsburg: Association for Computational Linguistics (ACL); 2017. p. 564–73.

[94] Lin XV, Socher R, Xiong C. Multi-hop knowledge graph reasoning with reward shaping. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; 2018 Oct 31–Nov 4; Brussels, Belgium. Stroudsburg: Association for Computational Linguistics (ACL); 2018. p. 3243–53.

[95] Yang F, Yang Z, Cohen WW. Differentiable learning of logical rules for knowledge base reasoning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA. Red Hook: Curran Associates Inc.; 2017. p. 2316–25.

[96] Sadeghian A, Armandpour M, Ding P. DRUM: end-to-end differentiable rule mining on knowledge graphs.

Journal Pre-proofs

In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–
Journal Pre-proofs

[97] Qu M, Tang J. Probabilistic logic neural networks for reasoning. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada. Red Hook: Curran Associates Inc.; 2019. p. 7712–22.

[98] Ding M, Zhou C, Chen Q. Cognitive graph for multi-hop reading comprehension at scale. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019 Jul 28–Aug 2; Florence, Italy. Stroudsburg: Association for Computational Linguistics (ACL); 2019. p. 2694–703.

[99] Qi B, Chen X, Gao J. Interactive continual learning: fast and slow thinking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 Jun 16–22; Seattle, WA, USA. New Yor City: IEEE; 2024. p. 12882–92.

Journal Pre-proofs

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: