A Survey of Memory Models for Virtual Agents and Humans: From Psychological Foundations to Computational Architectures

Alessandro Emmanuel Pecora[0009−0007−6826−5514], Francesco Strada[0000−0001−9197−9100], and Andrea Bottino[0000−0002−8894−5089]

Department of Control and Computer Engineering
Politecnico di Torino, Turin, ITALY
alessandro.pecora@polito.it, francesco.strada@polito.it, andrea.bottino@polito.it

Abstract. Virtual Humans are an advanced class of Virtual Agents characterized by human-like embodiment and cognitive capabilities. Central to their adaptivity is the integration of computational cognitive architectures, in which memory models play a key role in learning, continuity, and contextual reasoning. This review bridges psychological theories of memory and their computational implementations by comparing symbolic and connectionist approaches and exploring new paradigms such as Memory-Augmented Neural Networks and Large Language Models. We propose a unified framework for analyzing the components of artificial memory —Working, Semantic, Episodic, Procedural, Spatial, and Autobiographical Memory — and review their applications in domains such as education, games, and social simulation. Finally, we discuss open challenges and future works.

Keywords: Virtual Agents · Virtual Humans · Memory Models · Computational Cognitive Architectures · Artificial Intelligence · Memory-Augmented Neural Networks

1 Introduction

Virtual Humans (VHs) are a specialized class of Virtual Agents (VAs) [16], characterized by anthropomorphic embodiment and advanced cognitive capabilities. They replicate human-like presence, intelligence, and complex behavioral mechanisms, enabling naturalistic interactions through verbal and non-verbal communication [67,84]. VHs and VAs are increasingly adopted in immersive, interactive, and XR environments. They serve practical roles in various domains, such as education as adaptive tutors [40], social simulations for modeling human behavior [57], and video games for enhancing NPC realism [89].

A key requirement for cognitive and adaptive VHs is a robust memory system that enables them to store, retrieve and apply knowledge over an extended period. This capability supports continuity across interactions, contextual coherence and adaptation by tracking interactions with the environment, maintaining context, and consolidating knowledge.

To conceptualize and structure these complex capabilities, cognitive psychology has developed Memory Models (MMs) – theoretical frameworks that describe how information is encoded, stored, consolidated, forgotten, and retrieved in the human mind. These models include distinctions between Short-Term Memory and Long-Term Memory [1], Working Memory [2], Episodic and Semantic Memory, Procedural learning [73], Autobiographical structuring of life events [11] and Spatial Memory [69]. MMs provide a basis for the understanding of human cognition and for the development of artificial memory systems in VAs and VHs.

In artificial systems, MMs are implemented within Computational Cognitive Architectures (CCAs), which structure cognitive functions like perception, reasoning, memory, and learning [44]. These architectures typically follow *symbolic* paradigms, based on structured rules and logic-driven reasoning [52], *connectionist* paradigms, relying on emergent cognition from distributed numerical computations [64], or *hybrid* paradigms, integrating the advantages of both structured and adaptive approaches [82].

Given the relevance of VAs and VHs in immersive and interactive environments, this work provides a comprehensive analysis of MMs tailored to these agents, as understanding how MMs support their adaptive behavior, continuity, and interaction is crucial to advance research and improve their capabilities.

Although previous work has investigated the memory of VAs, existing surveys mostly focus on specific aspects. Lim et al. [45] analyze MMs in social companions, emphasizing personalized interaction and continuity. Zhang et al. [85] examine memory in LLM-based agents and categorize their storage and retrieval mechanisms. He et al. [22] provide a broad survey of MMs in AI and establish important theoretical foundations. However, while these studies focus on memory in artificial agents at a conceptual or system-wide level, our work narrows the scope to the specific needs and characteristics of VAs and VHs.

To address this gap, we categorize existing artificial MMs based on their alignment with cognitive psychology and their integration with CCAs. We then (i) conduct a comparative analysis of symbolic and connectionist approaches; (ii) introduce Spatial Memory as a key element of VHs cognition that has not been explored in depth in previous surveys, and discuss the relevance of Autobiographical Memory, which extends Episodic Memory to support self-awareness and continuity of identity — an essential feature of ideal VHs; and (iii) link memory implementations to real-world applications in domains such as education, gaming, and social simulation to illustrate how different MMs affect different scenarios involving VHs.

The paper is organized as follows. Section 2 defines VAs and VHs, highlighting their differences and characteristics. Section 3 reviews psychological memory theories as a foundation for artificial implementations. Section 4 examines CCAs paradigms. Section 5 categorizes computational MMs and their domain-specific applications. Section 6 discusses key challenges and future directions. Finally, Section 7 presents a summary and conclusions.

2 Virtual Agent and Virtual Human

Understanding Virtual Agents (VAs) is fundamental to the study and development of Virtual Humans (VHs), as the latter builds on the fundamental capabilities of the former to achieve more complex, human-like behaviors. VAs and VHs are on a continuum and differ in aspects such as intelligence, cognitive abilities, embodiment, autonomy and interactivity. These agents are used in a variety of domains, including assistive technologies, training environments, research and entertainment. However, their classification is often fluid, with terms such as chatbots, conversational agents and embodied conversational agents often used interchangeably [14]. Other taxonomies emphasize relational capabilities, such as socially intelligent agents that maintain long-term engagement or companions designed for ongoing social interaction [13]. In the absence of rigid boundaries and universally accepted definitions, we adopt an operational perspective to distinguish between VAs and VHs, focusing on the features most relevant to this study rather than treating them as strictly separate categories.

VAs are artificial systems that are able to perceive their environment and interact autonomously with users or other agents to achieve specific goals [16], exhibiting varying degrees of decision-making and adaptability [8]. Their interactions can be through text, voice or graphical interfaces, and although some of them include visual representation, embodiment is not a defining characteristic.

VHs, on the other hand, are a specialized subset of VAs that mimic human presence and behavior. In addition to anthropomorphic embodiment, they have advanced cognitive and social capabilities [47,67]. In contrast to traditional VAs, VHs combine both verbal and nonverbal modalities, allowing for richer and more contextualized interactions.

In this review, the term VAs is used generically to refer to both VAs and VHs unless a distinction is explicitly required.

3 Psychological Foundations of Memory Models

Psychological theories of memory provide a foundational blueprint for the development of memory systems in VAs. By modeling how humans encode, store, and retrieve information, these theories inform the development of Computational Cognitive Architectures (CCAs) that support long-term memory, contextual thinking, and adaptive interaction. This section provides an overview of the main psychological models that underlie current computational approaches to Memory Models (MMs).

The Multi-Store Model: One of the earliest and most influential theories, the Multi-Store Model [1], conceptualizes memory as three components: Sensory Register, which briefly stores stimuli; Short-Term Memory (STM), a temporary processing workspace; and Long-Term Memory (LTM), a stable repository for long-term retention. In addition to these structural components, the model introduces Control Processes, i.e. strategies that regulate information flow. These include rehearsal (the repetition of information for retention), transfer (the movement of information from STM to LTM), and retrieval (the access of stored knowledge when needed).

The Working Memory Model: Despite its foundational impact, the Multi-Store Model has limitations, treating STM as a passive buffer rather than a dynamic workspace, overemphasizing rehearsal while neglecting semantic encoding and emotional arousal. It also fails to account for implicit memory, which underlies skill learning and habit formation in motor tasks and social behavior.

To address these issues, Baddeley and Hitch [3] proposed the Working Memory (WM) model, which refines the concept of STM by introducing a more dynamic framework in which information is actively processed rather than just stored. At the core of WM, the Central Executive manages attention, task coordination, and cognitive resources for multitasking. It regulates three subsystems: the phonological loop for verbal information, the visuospatial sketchpad for processing spatial and visual data, and the episodic buffer for integrating memories and linking WM to LTM, though its transfer mechanism remains unclear.

Hierarchical and Functional Models of Long-Term Memory: Tulving [70,71,73] refined LTM structure by distinguishing between Semantic Memory and Episodic Memory. Semantic Memory stores general knowledge and facts, independent of specific personal experiences; for example, knowing that a "bed" is a piece of furniture used for sleeping, or that Paris is the capital of France. In contrast, Episodic Memory encodes personal experiences anchored in time and space. An example of Episodic Memory would be remembering that this morning you got out of bed at 6 a.m., focusing on the specific event and moment rather than on the conceptual knowledge of what a bed is. Both forms of memory require conscious recall—Episodic Memory reconstructs past episodes, while Semantic Memory supports conceptual and categorical knowledge. In a later work [72], Tulving introduced Procedural Memory, which emphasizes unconscious mechanisms involved in skill learning and priming, such as riding a bicycle without needing to consciously recall specific training episodes.

Conway [11] extended Tulving's model by introducing Autobiographical Memory, which integrates episodic and semantic elements into a coherent, self-relevant narrative. Autobiographical Memory is hierarchically organized: Lifetime Periods define broad phases (e.g., childhood, university years); General Event Representations group related episodes (e.g., a series of family holidays); and Events-Specific Knowledge capture detailed, event-specific memories (e.g., the comment a professor made about your thesis). Autobiographical Memory plays a central role in the construction of personal identity by enabling both narrative coherence and detailed episodic recall.

Finally, in his early work [69], Tolman introduced the concept of Spatial Memory, which supports navigation and spatial reasoning through internal cognitive maps; for instance, mentally visualizing the layout of your home to find your keys.

4 Computational Cognitive Architecture

The psychological Memory Models (MMs) discussed earlier provide a conceptual lens on human cognition, but their application in VAs requires a structured computational framework. In this context, memory is not an isolated module, but a core component of Computational Cognitive Architectures (CCAs) — systems that integrate perception, reasoning, learning, and adaptation to enable coherent, intelligent behavior [44]. This section examines how MMs are embedded in CCAs and compares the main architectural paradigms that shape cognitive processes in VAs.

4.1 Symbolic AI

Symbolic AI structures cognition through explicit rules and human-readable knowledge representations [19,52]. Inspired by Newell's Unified Theories of Cognition [55], symbolic architectures model memory as a structured repository of facts and procedural rules that support logic-based reasoning. Prominent examples are SOAR [37] and ACT-R [60].

SOAR is based on the Problem Space Hypothesis, which conceptualizes logical reasoning as a search process through goals, states and operators. Its memory system comprises a Working Memory, which stores the current state, and a production memory, i.e. a set of if-then rules that guide decision-making. Learning takes place through chunking, in which experiences are combined to form new production rules, gradually improving the efficiency of thinking.

Although ACT-R is formally categorized as a hybrid architecture, it is predominantly symbolic, as both declarative and procedural knowledge is represented in structured, rule-based formats. The connectionist layer plays an auxiliary role, as it does not serve broader cognitive adaptation, but only supports activation-based retrieval to prioritize knowledge and influence memory access.

4.2 Connectionist (or sub-symbolic) AI

While symbolic AI encodes knowledge via explicit rules and structured representations, connectionist AI models cognition as an emergent property of distributed computation [64]. Among its various implementations, Artificial Neural Networks are the most prominent. In these networks, intelligence emerges through pattern learning across artificial neurons, where knowledge is embedded in the network's parameters.

According to [32], memory in these systems falls into two categories: Internal Parametric Memory, stored in the network parameters, and external memory structures, known as Memory-Augmented Neural Networks (MANNs), which provide explicit mechanisms for storage and retrieval beyond the internal capacity of the model.

Internal Parametric Memory encodes knowledge directly in the parameters of Artificial Neural Networks, which also determine the model's ability to process sequences and recognize patterns. Several architectures have explored this approach, each offering advantages in specific contexts. For example, Recurrent Neural Networks and Long Short-Term Memory Networks, can process sequential information, but often struggle with long-term dependencies due to vanishing gradients [25,62]. Hopfield networks act as autoassociative memory systems that retrieve stored patterns even from noisy inputs [26], while correlation matrix memories rely on Hebbian learning to enable associative recall

[23,35]. Although these models provide valuable insights into biologically inspired memory mechanisms, their use in VAs remains limited as they are better suited for specialized tasks than for core cognitive functions such as reasoning or adaptation.

More flexible and scalable architectures are needed to support higher level cognitive functions such as contextual reasoning, multi-modal integration and cross-task generalization. Transformers [74] have played a central role in this process. Originally developed for sequence modeling, they now form the basis of most VAs thanks to their mechanism of self-attention, which dynamically weighs the contextual relevance of input elements — crucial for tasks such as language understanding and generation. Large Language Models (LLMs) [86], which build on Transformers, use extensive pre-training to model language. LLMs treat prompts as a form of temporary memory. These prompts form the contextual basis for response generation and are conceptually similar to Working Memory in Baddeley's model [2]. However, LLMs are constrained by fixed prompt sizes, limiting the amount of accessible information during inference.

Despite these advances, learning in all connectionist models usually occurs by updating the memory of the network (i.e. its internal parameters) through retraining or fine-tuning to new data. However, this approach lacks mechanisms to retain previously learned knowledge in a controlled way. Therefore, when the model is retrained on new tasks or experiences, it often overwrites previous information — a phenomenon known as catastrophic forgetting [17]. This limitation can lead to difficulties in implementing a stable and persistent LTMs, especially for agents that need to accumulate and retain knowledge over an extended period of time.

MANNs extend neural architectures with external memory mechanisms that enable more persistent and flexible information processing [32]. They can be roughly divided into parametric and non-parametric approaches, depending on how the memory is stored and accessed.

Parametric MANNs include trainable memory modules that are separate from the main weights of the model. These modules evolve dynamically through experience and offer potential solutions to problems such as catastrophic forgetting and limited context windows in LLMs. However, they remain largely experimental. Notable examples include Titans, which adapts memory via novelty detection[4], Metalearned Neural Memory [54] and Heterogeneous MANNs with trainable memory tokens [59].

Non-parametric MANNs, on the other hand, are based on explicit external storage and retrieval processes. By decoupling knowledge retention from trainable modules, this approach allows for persistent memory independent of gradient updates, effectively mitigating catastrophic forgetting. Models such as Neural Turing Machines and Differentiable Neural Computers [20,21], Memory Networks [79] and Memformer [80] exemplify this approach by using structured external memories for persistent retention.

A particularly practical example is Retrieval-Augmented Generation (RAG) [41], a scalable approach that integrates external memory into LLMs by retrieving relevant content into the prompt. In this sense, it serves as a dynamic form of WM that controls the generation of responses. These mechanisms underlies cognitive architectures based on LLM [56], such as CoALA [65] (Cognitive Architecture for Language Agents), which explicitly models episodic, semantic and procedural memory through iterative retrieval, and LARP [83] (Language Agent Role Play), which improves coherence over time by combining memory decay with environmental adaptation to filter out outdated or irrelevant content over time.

Despite their promising properties, LLMs produce stochastic results that can lead to incoherent or unwanted behavior. To address this problem, emerging hybrid architectures aim to integrate symbolic reasoning with LLM-based systems [61,81] using symbolic constraints to "guardrail" LLMs flexibility, i.e. to guide generation, enforce consistency and improve interpretability without sacrificing flexibility or adaptability.

5 Memory Models in Virtual Agents and Virtual Humans

Based on the key concepts introduced in Section 3, we identify three core components—Sensory Register (SR), Working Memory (WM), and Long-Term Memory (LTM)—as the basis of a conceptual framework for analyzing the implementation of computational Memory Models in Virtual Agents (VAs) (Figure 1). This framework refines the one proposed by He et al. [22] by including Spatial Memory and extending Episodic Memory with Autobiographical Memory. These additions introduce spatial awareness and a broader form of self-representation, both of which are essential for designing more human-like VAs.

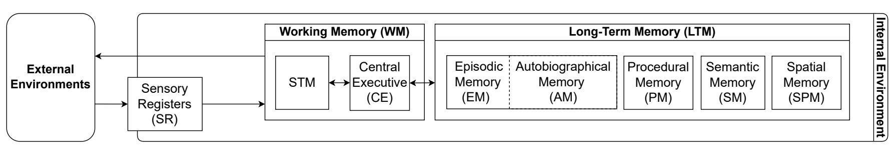

Fig. 1. Memory framework used in this review, illustrating key psychological memory components commonly adopted in computational implementations.

In this framework, the **external environments** from which agents receive stimuli include virtual spaces, physical settings, and human interactions, while the **internal environment** consists of memory systems, emotional states, and other conditions influencing behavior. The **SR** processes multimodal raw signals (text, audio, images, or sensor data), acting as a buffer between external and internal environments. It temporarily stores stimuli before transferring them to WM, where in practice, SR is often embedded.

WM, based on Baddeley's model [2], comprises specialized memory buffers and the **Central Executive (CE)** that manages them. In our framework, these buffers are unified into a single temporary store referred to as STM. The CE coordinates interactions between memory systems and supports reasoning and decision-making through processes such as inference, summarization, consolidation, and other executive functions. Some implementations explicitly model WM as a distinct module, while others distribute its roles across decentralized components.

LTM is a persistent repository composed of modules inspired by cognitive science: **Episodic Memory (EM)** stores self-relevant experiences and can be extended into **Autobiographical Memory (AM)**, which structures these experiences into coherent life narratives to support self-awareness and identity continuity; **Procedural Memory (PM)** encodes learned skills, **Semantic Memory (SM)** retains general knowledge, and **Spatial Memory (SPM)** manages cognitive maps, object locations, and spatial reasoning.

The following sections explore how LTM is implemented across its main components, examining the role of WM within each one rather than as a standalone system.

5.1 Episodic Memory and Autobiographical Memory

Episodic Memory (EM) in VAs supports the storage and retrieval of past experiences and thus enables long-term interactions and behavioral continuity. A key feature of EM is its temporal structure, as episodes represent events anchored to specific points in time. However, the identification of episode boundaries remains an open challenge [31,51,66]. While human memory uses temporal schemas and salient events to segment experiences [38,39,68], artificial systems typically rely on timestamps that lack the nuanced flexibility of human cognition. Moreover, since storing all experience is computationally impractical, artificial EMs must incorporate explicit forgetting mechanisms to selectively discard data of low relevance and maintain memory efficiency.

Implementation strategies vary across paradigms: connectionist approaches often use ad hoc memory structures tailored to specific domains, while symbolic systems tend to follow formalized cognitive models based on psychological theories.

Connectionist approaches: In [14], an Embodied Conversational Agent (ECA) for General Interaction domain—which involves open-ended, unconstrained dialogues—stores conversational episodes, containing dialogues and extracted topics, in a relational database. Similarly, [42] presents a Conversational Agent that timestamps dialogue history in WM, using a fine-tuned Large Language Model (LLM) to summarize past interactions and discard less relevant data. A similar memory approach is applied in [89] for video game non-playable characters, ensuring behavioral consistency and improving player immersion.

A notable contribution in Social Simulation is in [57]. This domain models human behavior via multi-agent systems, where agents act as proxies for humans to simulate social interactions. Another application is in education, where [40] simulates a classroom with virtual and human participants. Unlike previous implementations that organize LTM into its distinct components, [57] employs a unified Memory Stream that continuously records agent perceptions in natural language. Though undifferentiated, it primarily serves an episodic function by capturing temporally structured experiences.

Symbolic approaches: MIACE (Model of Student Knowledge Acquisition, [48]) is a cognitive architecture that simulates a student for teacher training purposes. It uses EM to store learning episodes, capturing problem-solving steps, cognitive strategies, and decision-making processes. These episodes are later retrieved to address similar problems, enabling the system to adapt and refine its instructional approach based on past experiences.

Other symbolic models extend EM into Autobiographical Memory (AM), integrating semantic knowledge into a structured self-narrative. Conway's hierarchy [11] defines AM across three levels: Lifetime Periods (LPs), General Event Representations (GERs), and Events-Specific Knowledge (ESKs). While this framework models human LTM, direct applications in VAs are complex.

A key challenge is that VAs operate within narrow timeframes, while humans accumulate memories continuously throughout life. LPs require long-term structures, which conflict with session-based AI interactions. As a result, current AM implementations focus solely on GERs and ESKs. For instance, [33,34] present AM models for General Interaction that combine GERs and ESKs to preserve conversational context. Here, GERs are enriched with emotion metrics, expressed as categorical labels [33] or via the Pleasure-Arousal-Dominance model [49], and with sentence polarity computed using VADER sentiment analysis [30]. ESKs resources specify the data type—such as text, audio, or image—along with an activation level that influences forgetting or retention. In contrast, [24] employs GERs alone as goal templates in LTM, dynamically activating them in WM based on internal or environmental triggers. The goal's status (active, completed, or violated) guides AM updates. When a goal is violated, the system retrieves emotionally intense episodes, allowing the agent to refine its strategies.

Retrieval and Forgetting Mechanisms: In artificial EM, retrieval and forgetting are closely interdependent: retrieval mechanisms rely on prioritization, while forgetting regulates memory load and relevance over time. Most systems use weighted schemes where memories are ranked based on factors such as recency, frequency of access and semantic similarity. For example, [42] combines semantic relevance, topic overlap and time decay to filter out low-priority memories to ensure that frequently accessed or contextually important episodes are preserved. Similarly, [27] introduces an exponential decay function to compute consolidation weights that influence retrieval ranking over time. In [57], the episode importance is estimated at storage time using a prompted LLM so that less relevant memories can be deprioritized at future retrieval. Instead of storing entire dialog transcripts, [89] implements forgetting by summarizing conversations and indexing them with a Dual-Tower Deep Semantic Similarity Model, which enables efficient retrieval through semantically encoded memory units.

Challenges and Future Directions: While AM plays a crucial role in building continuity and self-awareness in human cognition, its computational implementation in VAs remains limited. In particular, there is no connectionist approach that incorporates AM, and symbolic models typically omit LPs — the highest level in Conway's hierarchy. This omission reflects a deeper challenge: modeling long-term temporal structures in artificial agents that often engage in short or session-based interactions. Two key problems arise: (1) defining episode boundaries and (2) structuring memory across sessions. In domains such as social simulation [57], where agents interact continuously, the primary challenge is to meaningfully divide experiences into time segments in order to improve agents' self-awareness. In contrast, in session-based domains such as education [40,48] both core challenges are present. First, clear episode boundaries need to be defined to track learning activities across tasks and sessions. Second, effective session management is essential to maintain continuity across temporally fragmented interactions. In this context, LPs could serve as an overarching framework to represent learning phases, while session structures naturally support episodic segmentation.

Another critical limitation lies in the current retrieval and forgetting mechanisms, which often depend on manually set thresholds for factors such as recency, frequency or relevance that lack adaptability and generalization. Further development towards self-adaptive, learning-based mechanisms would allow agents to dynamically regulate the relevance of memory based on evolving context and task requirements, improving long-term coherence and scalability.

5.2 Procedural Memory

Procedural Memory (PM) allows VAs to acquire, store and refine actions and skills through execution and feedback, enabling lifelong learning. Unlike EM, which stores specific events, PM encodes how actions are performed, enabling their automation and refinement over time. In computational implementations, procedural knowledge is represented in formats such as symbolic structures, text-based instructions, programming functions, and hierarchical action plans.

Connectionist approaches: PM plays a central role in domains such as Open-Ended World Exploration and Artificial General Intelligence, where VAs continuously acquire and refine skills to adapt to novel, diverse tasks and dynamic environments. Several works have proposed LLM-based approaches to develop Generally Capable Agents — VAs specifically designed to handle these requirements — exploiting Minecraft as a testbed [15,75,77,88,90].

PM for Generally Capable Agents is structured around two components: action plans, which define high-level sequences decomposing tasks into sub-goals, and skills, which represent executable actions or routines needed for task execution.

For action plan retrieval, procedural knowledge is often stored as structured action plans retrievable based on task requirements and used as examples to guide new plan generation. A multimodal Retrieval-Augmented Generation approach is proposed in [77], where PM is structured as a key-value memory: input instructions and visual observations serve as keys, while corresponding action plans are stored as values. Differently, [90], adopts a simpler text-based PM. Both methods decompose tasks into subgoals but differ in retrieval: [77] uses multimodal queries enhanced by LLM-driven reasoning, while [90] relies only on text similarity.

Besides action plans, the approach in [87] maintains a skill database composed of executable code snippets and textual descriptions, retrieving relevant methods via cosine similarity. Likewise, [75] organizes PM as a skill library of executable programs. When encountering novel tasks, the agent generates new skills by modifying or combining existing ones. In all these works, skills are refined through feedback processed by a LLM, enabling continuous adaptation.

A different paradigm is proposed in [15], where PM is acquired directly through LLM fine-tuning instead of relying on external memories modules. Unlike methods based on complex prompts (e.g., [75,90]), the LLM is trained to generate next-step actions based on observations and task descriptions. This process unfolds in two stages: (1) an exploration phase, where the agent collects a dataset of observations and feedback through trial-and-error interactions; and (2) a learning phase, where this dataset is used to fine-tune the LLM.

Another key mechanism in this context concerns automatic skill acquisition, whereby agents autonomously generate tasks, to accumulate experience and improve their procedural competence. One such approach is self-instruct, introduced in [77], which enables agents to autonomously generate and refine tasks by writing instructions for themselves. This mechanism is further enhanced by the use of parallel agents that share a distributed PM, allowing for faster skill acquisition. Building on this concept, [88] integrates self-instruct within a hierarchical multi-agent system that combines centralized planning with decentralized execution, supporting complex task decomposition, agent coordination, and knowledge sharing. The same mechanism is also employed by individual agents in [75,87], where it is referred to as curriculum learning.

Symbolic approaches: Symbolic AI structures PM via goal-oriented planning frameworks. In [48], a student simulation system defines procedural knowledge as goal-driven procedures that track success rates and episodic occurrences. These are categorized into: complex procedures that define subgoals and whose execution leaves traces in episodic memory; primitive procedures executed automatically without such traces; and associative procedures that retrieve stored solutions (e.g., basic arithmetic). Similarly, [36] models PM for a VH in a domain-specific environment (a museum) using the Belief-Desire-Intention framework. In these systems, symbolic PM explicitly encodes structured procedures and conversational sequences—such as dialog plans and rule-based interactions—to ensure predictable, context-aware behavior.

Working memory and PM: WM plays a key role in managing and retrieving procedural knowledge, though its function and structure differ across symbolic and LLM-based systems. In symbolic approaches [36,48], it is an explicit module that regulates retrieval and activation of procedures based on contextual demands. It also interacts with EM to integrate past experiences, enabling dynamic adjustment of strategies. In contrast, LLM-based systems rely on the limited context window as an implicit WM [75,77,88,90]. The key difference lies in procedural generation: symbolic systems follow rule-based mechanisms, while

LLMs use generative capabilities to flexibly adapt, refine, and generate new procedures, supporting more open-ended reasoning.

Challenges and future directions: PM implementations are highly domain-dependent: in Generally Capable Agents, they support adaptive skill learning and action planning, while in ECAs, they are typically more constrained to ensure behavioral consistency. This divergence makes balancing adaptability and reliability a key challenge. LLM-based approaches offer flexibility and dynamic skill generation but may yield inconsistent behaviors due to their stochastic nature. Symbolic methods, in contrast, ensure predictable, structured procedures but require greater manual configuration and lack adaptability.

As previously noted in the Computational Cognitive Architecture section, hybrid architectures represent a promising avenue to address this trade-off. Specifically, combining symbolic constraints with LLM-driven procedural learning enables both structure and flexibility. In the context of PM, recent proposals integrating Belief-Desire-Intention framework with LLM-based agents [18] illustrate this synergy. Such models can use symbolic plans to guide procedural generation and mitigate issues like goal misinterpretation, unintended behavior, and value misalignment (i.e., conflicts between LLMs outputs and human norms) [29].

5.3 Semantic Memory

Semantic Memory (SM) in VAs serves as a knowledge base, providing general [24,34] or domain-specific [36,48,87,90] information – such as facts, concepts, and entities – independent of specific experiences. It provides context and underpins comprehension, reasoning, decision-making, and response generation. Unlike EM and PM, which vary significantly between connectionist and symbolic approaches, SM is generally structured at a higher level in a similar way across both paradigms.

Connectionist approaches: In LLM-based Generally Capable Agent, SM supports procedural knowledge by providing factual and conceptual grounding for translating high-level plans into executable actions. For instance, [90] leverages an external wiki for real-world knowledge, while [87] relies on parametric knowledge encoded in the LLM to retrieve task-relevant information.

In healthcare applications, [84] introduces a conversational agent that adapts to patient histories and communication styles by extracting semantic knowledge from dialogues via Dual-Process enhanced Memory Mechanisms. Recent interactions are stored in WM, while an external buffer (STM) acts as a cache for LTM. A Rehearsal Process extracts notes from dialogues, and an Executive Process uses frequency to decide whether to store them in STM or directly in LTM. Extracted information is labeled as commonsense or user-specific and used as semantic cues during retrieval. To balance precision and generalization, STM entries are retrieved via Levenshtein distance for near-exact string matches, whereas LTM entries use cosine similarity to capture broader semantic relations.

In social simulation (previously discussed in the EM Section), [57] employs a unified Memory Stream to log episodic experiences. Beyond storing raw observations, the system uses a reflection process in which an LLM periodically analyzes past interactions to infer semantic patterns – such as social relationships, preferences, and behavioral tendencies – and consolidate them into long-term knowledge that guides future behavior.

Symbolic approaches: Several symbolic approaches design semantic knowledge for ECAs in the General Interaction domain. For instance, [34] introduces a self-memory module containing the agent's name, age, and preferences to support personal queries, alongside a Common Sense ontology derived from WordNet [58]. Similarly, [24] employs a World Knowledge ontology. Both provide broad real-world knowledge and support the interpretation of experiences stored in EM.

In [34], semantic data from the Common Sense ontology is stored as metadata in EM, enhancing episode retrieval. Similarly, [14] attaches semantic cues to episodes using Named Entity Recognition and Pointwise Mutual Information [10]. This system also includes a User Model that stores user-specific details, such as name and behavior, to personalize interactions and avoid repetition. In contrast, [24] keeps semantic and episodic data separate, using World Knowledge to expand episodic queries. The museum guide in [36] follows a similar approach, applying domain-specific semantic knowledge to tailor responses based on user intent, selectively retrieving from SM, augmenting EM queries, or combining both to provide contextually appropriate answers.

SM also supports PM by providing foundational domain knowledge for procedural learning. In the student simulation presented in [48], this knowledge includes abstract representations of concepts, problem-solving strategies, and instructional examples, organized into two classes: the Goal class, which stores procedures learned from the teacher that can be instantiated with specific parameters, and the Domain Knowledge class, which contains core concepts, examples, and answers to contextualize procedural execution.

Challenges and Future Directions: The reviewed works show that SM often supports both PM and EM, with WM serving as a bridge. Some systems adopt dedicated SM [24,48,90], while others embed semantic information within EM [14,34], or adopt a unified memory system [57]. These approaches highlight the need to balance modularity and integration to improve adaptability, reasoning, learning and recall [76].

Another key direction is the development of personal, subjective SM, which encodes self-related knowledge such as beliefs, preferences, and perspectives [53]. These can be derived from episodic experiences. To this end, the reflection process used in [57] could be adapted to extract and consolidate such content, enriching SM with a subjective layer.

5.4 Spatial Memory

Spatial Memory (SPM) enables VAs to encode, store, and retrieve information about environmental layout and object positions [46]. It supports navigation, spatial reasoning, and object localization, mirroring the way human memory often relies on spatial context to organize and recall experiences [9]. For VAs, the integration of dedicated SPM modules increases realism by supporting natural movements and spatial references in the interaction. Early research focused primarily on navigation and environmental learning [5]. However, explicitly modeling SPM as a cognitive map enables more advanced capabilities, such as remembering what is where, i.e. associating objects with their location even when they are no longer in sight. In games and simulations, this allows non-playable characters to navigate purposefully rather than randomly, improving believability and immersion.

Within the broader LTM, SPM functions as an independent yet complementary module. While EM encodes "when" events occur, SPM encodes "where" they happen—for instance, EM may recall finding a key, whereas SPM retains the environment layout and the key's exact location. Similarly, SM could generalize that "keys are usually found on tables". SPM bridges these systems by linking events to physical locations and anchoring abstract knowledge in the environmental structure [9]. This functionality is particularly important for LLMs-based VAs, which lack inative spatial awareness and therefore require explicit SPM modules to support situated reasoning and spatial continuity across tasks and environments [43].

Drawing from cognitive psychology, in which SPM combines egocentric and allocentric representations[7], computational implementations adopt a similar dual structure. Egocentric memory encodes space relative to the agent's position and updates dynamically to support real-time interactions (e.g., obstacle avoidance, pointing). Allocentric memory builds a stable, position-independent map for long-term planning (e.g., navigating a building). Humans use both in parallel, and VAs can similarly benefit from this dual system [6].

Symbolic approaches: The museum guide in [36], is the only and notable example of symbolic approach. It uses an egocentric SPM to track visitors and objects in real time. By continuously updating its SPM in relation to its position, it enables spatially grounded references (e.g. "behind you" or "to my right") and integrates verbal and non-verbal interactions (e.g., pointing) for natural engagement.

Connectionist approaches: For navigation tasks, agents typically rely on allocentric spatial memory. In home navigation and robotic simulation, [78] introduces an Instructable Embodied Agent—a system operating in simulated physical environments that performs tasks based on user instructions—which implements SPM via an allocentric 3D scene graph. In this graph, nodes represent objects, rooms, and edges encode spatial relations (e.g., containment, adjacency). The graph updates dynamically as the agent explores, allowing it to track changes and optimize multi-step planning (e.g., get an apple and put it on a plate). The work in [63] adopts a similar structure but extends it with semantic labeling and an occupancy map that tracks not only spatial availability and objects, but also area types. It further integrates a vision-language model to enhance recall, spatial reasoning, and task planning.

Challenges and future directions: Most applications have focused on navigation [5,63,78], with limited use of SPM in conversational contexts [36].

Yet, domains like Open-Ended World Exploration could benefit from SPM to structure and retain spatial knowledge, opening new research opportunities.

A major challenge lies in efficiently representing large, dynamic environments. As the world expands, allocentric maps grow complex. Hierarchical 3D scene graphs help manage this by reducing redundancy [78], but maintaining scalability and real-time updates remains difficult—especially in settings with frequent object movement and continuous discovery of new areas.

Another key issue is the integration of SPM and EM. Ideally, EM would index spatial locations stored in SPM. However, most current systems lack mechanisms for such links, limiting the ability to retrieve memories by location or update spatial knowledge over time. While such integration is more common between SM and other memory modules – thanks to their predominantly textual and abstract nature – SPM poses additional challenges due to its complex and dynamic spatial structure.

Table 1 summarizes the main types of LTM, highlighting the computational approaches adopted and representative works for each.

LTM Type	**Connectionist Approaches**	**Symbolic Approaches**
Episodic Memory (EM)	[14,42,89,57,40,27]	[48]
Autobiographical Memory (AM)	-	[33,34,24]
Procedural Memory (PM)	[77,90,87,75,15,88]	[48,36]
Semantic Memory (SM)	[90,87,84,57]	[34,24,14,36,48]
Spatial Memory (SPM)	[78,63]	[36]

Table 1. Memory types, corresponding computational approaches, and representative works.

6 Discussion

Memory modeling in VA is an evolving field, shaped by advances in both cognitive science and AI. In the previous sections, we have addressed several challenges. In this discussion, we consider some of these issues from an overarching perspective and highlight additional under-researched challenges that deserve further investigation.

Toward Self-Awareness and Subjectivity: Integrating Episodic and Semantic Memory. An important step towards more believable and self-aware agents is the combined development of EM and SM. On the episodic side, improving event segmentation and implementing agent-specific temporal structures support more coherent recall and structuring of life phases (LPs), features that are crucial for domains such as social simulations [57] and beneficial in educational contexts [40,48]. On the semantic side, the enrichment of memory with self-related knowledge, such as beliefs, preferences and opinions, strengthens the personality and continuity of the agent. This information may originate from episodic experiences or other structured inputs [53,57]. By considering these dimensions together, VHs can develop deeper forms of self-awareness and subjectivity that enable more human-like behavior and interaction.

Resource Demands and Real-Time Execution. VAs require real-time outputs while integrating multiple intensive processes such as speech-to-text, emotion recognition, and image understanding. In addition, the embodied nature if VHs introduces a demanding graphical component. For LLM-based systems, GPU and latency constraints further exacerbate the challenge. These limitations are even more pronounced in XR contexts, where untethered devices have stricter resource constraints. Despite progress, current hardware struggles to support fully responsive, real-time VAs, making the development of efficient, parallelizable architectures essential until future platforms can natively sustain such AI workloads.

The Need for Standardized Frameworks and Evaluation Methods. Cognitive theories remain largely conceptual and fragmented, resulting in non-standardized implementations and a lack of best practices [12]. Evaluation methods typically rely on qualitative studies and task-based tests. Among these, ablation studies—currently the most effective strategy—enable the isolation and assessment of each memory module [22], though they often rely on human judgment and thus remain inherently subjective. More generalizable protocols inspired by cognitive science [50] and automated assessment tools [28] are sorely needed to improve reproducibility and enable consistent benchmarking across artificial MMs.

Forgetting Challenges. Forgetting is essential to ensure memory efficiency across all memory types. In Episodic Memory, the continuous accumulation of episodic traces—particularly in open-ended interactions—demands selective forgetting to retain only relevant experiences. Recent work [22] highlights its importance also in Semantic Memory, where overlapping or outdated knowledge requires active forgetting and compression to reduce semantic interference. In Procedural Memory, eliminating obsolete or unused skills supports flexibility and adaptability. Similarly, in Spatial Memory, removing outdated spatial information prevents overload and preserves navigation accuracy [6]. Emerging methods—such as weight-based forgetting, semantic relevance filtering, and dual egocentric/allocentric encoding—offer promising directions. Future research should refine these strategies to balance precision, adaptability, and computational efficiency across memory systems.

7 Conclusion

In this survey, we have provided a comprehensive overview of Memory Models for Virtual Humans, linking psychological foundations with computational architectures. By proposing a unified framework, we explored how symbolic and connectionist approaches implement key memory components — such as Episodic, Semantic, Procedural, Autobiographical, and Spatial Memory — in different domains. Our analysis has highlighted both current opportunities and limitations, including the need for efficient forgetting mechanisms and challenges related to self-awareness, scalability and evaluation. Addressing these open challenges is crucial to advance the development of coherent, adaptive and human-like Virtual Humans.

References

1. Atkinson, R.C., Shiffrin, R.M.: Human memory: A proposed system and its control processes! Human Memory (1968)
2. Baddeley, A.: Working memory and language: an overview. Journal of Communication Disorders **36**(3), 189–208 (2003)
3. Baddeley, A., Hitch, G.: Working memory. In: Bower, G. (ed.) The Psychology of Learning and Motivation, pp. 47–89. Academic Press (1974)
4. Behrouz, A., Zhong, P., Mirrokni, V.: Titans: Learning to memorize at test time. arXiv preprint arXiv:2501.00663 (2024)
5. Brom, C., Korenko, T., Lukavský, J.: How Do Place and Objects Combine? “What-Where” Memory for Human-Like Agents. Springer Berlin Heidelberg (2009)
6. Brom, C., Vyhnánek, J., Lukavskỳ, J., Waller, D., Kadlec, R.: A computational model of the allocentric and egocentric spatial memory by means of virtual agents, or how simple virtual agents can help to build complex computational models. Cognitive Systems Research **17**, 1–24 (2012)
7. Burgess, N.: Spatial memory: how egocentric and allocentric combine. Trends in Cognitive Sciences **10**(12), 551–557 (2006)
8. Castelfranchi, C.: To be or not to be an “agent”. In: International Workshop on Agent Theories, Architectures, and Languages. pp. 37–39. Springer (1996)
9. Cho, J., Yoon, J., Ahn, S.: Spatially-aware transformer for embodied agents (2024)
10. Church, K., Hanks, P.: Word association norms, mutual information, and lexicography. Computational linguistics **16**(1), 22–29 (1990)
11. Conway, M.A., Pleydell-Pearce, C.W.: The construction of autobiographical memories in the self-memory system. Psychological review **107**(2), 261 (2000)
12. Criss, A.H., Howard, M.W.: Models of episodic memory. The Oxford handbook of computational and mathematical psychology **399**, 165–183 (2015)
13. Dautenhahn, K.: The art of designing socially intelligent agents: Science, fiction, and the human in the loop. Applied artificial intelligence **12**(7-8), 573–617 (1998)
14. Elvir, M., Gonzalez, A.J., Walls, C., Wilder, B.: Remembering a conversation–a conversational memory architecture for embodied conversational agents. Journal of Intelligent Systems **26**(1), 1–21 (2017)
15. Feng, Y., Wang, Y., Liu, J., Zheng, S., Lu, Z.: Llama rider: Spurring large language models to explore the open world. arXiv preprint arXiv:2310.08922 (2023)
16. Franklin, S., Graesser, A.: Is it an agent, or just a program?: A taxonomy for autonomous agents. In: International workshop on agent theories, architectures, and languages. pp. 21–35. Springer (1996)
17. French, R.M.: Catastrophic forgetting in connectionist networks. Trends in cognitive sciences **3**(4), 128–135 (1999)
18. Frering, L., Steinbauer-Wagner, G., Holzinger, A.: Integrating belief-desire-intention agents with large language models for reliable human–robot interaction and explainable artificial intelligence. Engineering Applications of Artificial Intelligence **141**, 109771 (2025)
19. Garnelo, M., Shanahan, M.: Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. Current Opinion in Behavioral Sciences **29**, 17–23 (2019)
20. Graves, A.: Neural turing machines. arXiv preprint arXiv:1410.5401 (2014)
21. Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S.G., Grefenstette, E., Ramalho, T., Agapiou, J., et al.: Hybrid computing using a neural network with dynamic external memory. Nature **538**(7626), 471–476 (2016)

22. He, Z., Lin, W., Zheng, et al.: Human-inspired perspectives: A survey on ai long-term memory. arXiv preprint arXiv:2411.00489 (2024)
23. Hebb, D.O.: The organization of behavior: A neuropsychological theory. Psychology press (2005)
24. Ho, W.C., Dautenhahn, K., Lim, M.Y., Vargas, P.A., Aylett, R., Enz, S.: An initial memory model for virtual and robot companions supporting migration and long-term interaction. pp. 277–284. IEEE (2009)
25. Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)
26. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences **79** (1982)
27. Hou, Y., Tamoto, H., Miyashita, H.: " my agent understands me better": Integrating dynamic human-like memory recall and consolidation in llm-based agents. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. pp. 1–7 (2024)
28. Howard, C.W., Johnson, et al.: The autonomous cognitive examination: Machine-learning based cognitive examination. medRxiv pp. 2024–09 (2024)
29. Hristova, T., Magee, L., Soldatic, K.: The problem of alignment. AI & SOCIETY pp. 1–15 (2024)
30. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. vol. 8, pp. 216–225 (2014)
31. Khemlani, S.S., Harrison, A.M., Trafton, J.G.: Episodes, events, and models. Frontiers in human neuroscience **9**, 590 (2015)
32. Khosla, S., Zhu, Z., He, Y.: Survey on memory-augmented neural networks: Cognitive insights to ai applications. arXiv preprint arXiv:2312.06141 (2023)
33. Knob, P., Dias, W.S., Kuniechick, N., Moraes, J., Musse, S.R.: Arthur: a new eca that uses memory to improve communication. In: 2021 IEEE 15th International Conference on Semantic Computing (ICSC). pp. 163–170. IEEE (2021)
34. Knob, P.R., Pizzol, N.D., Musse, S.R., Pelachaud, C.: Arthur and bella: multi-purpose empathetic ai assistants for daily conversations. The Visual Computer **40**(4), 2933–2948 (2024)
35. Kohonen, T.: Correlation matrix memories. IEEE transactions on computers **100**(4), 353–359 (1972)
36. Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A conversational agent as museum guide–design and evaluation of a real-world application. In: Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5. pp. 329–343. Springer (2005)
37. Laird, J.E.: The Soar cognitive architecture. MIT press (2019)
38. Larsen, S.F., Conway, M.A.: Reconstructing dates of true and false autobiographical memories. European Journal of Cognitive Psychology **9**(3), 259–272 (1997)
39. Larsen, S.F., Thompson, C.P.: Reconstructive memory in the dating of personal and public news events. Memory & Cognition **23**(6), 780–790 (1995)
40. Lee, U., Lee, S., Koh, et al.: Generative agent for teacher training: Designing educational problem-solving simulations with large language model-based agents for pre-service teachers. In: NeurIPS'23 Workshop on Generative AI for Education (GAIED) (2023)
41. Lewis, P., Perez, E., Piktus, A., Petroni, et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems **33**, 9459–9474 (2020)
42. Li, H., Yang, C., Zhang, et al.: Hello again! llm-powered personalized agent for long-term dialogue. arXiv preprint arXiv:2406.05925 (2024)

43. Li, Z., Lu, Y., Mu, Y., Qiao, H.: Cog-ga: A large language models-based generative agent for vision-language navigation in continuous environments (2024)
44. Lieto, A.: Cognitive design for artificial minds. Routledge (2021)
45. Lim, M.Y.: Memory models for intelligent social companions. In: Human-computer interaction: The agency perspective, pp. 241–262. Springer (2012)
46. Madl, T., Chen, K., Montaldi, D., Trappl, R.: Computational cognitive models of spatial memory in navigation space: A review. Neural Networks **65**, 18–43 (2015)
47. Magnenat-Thalmann, N., Thalmann, D.: Handbook of virtual humans. John Wiley & Sons (2005)
48. Mayers, A., Lefebvre, B., Frasson, C.: Miace, a human cognitive architecture. ACM SIGCUE Outlook **27**(2), 61–77 (2001)
49. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology (1996)
50. Mejía-Arango, S., Wong, R., Michaels-Obregón, A.: Normative and standardized data for cognitive measures in the mexican health and aging study. Salud publica de Mexico **57**, s90–s96 (2015)
51. Michelmann, S., Hasson, U., Norman, K.A.: Evidence that event boundaries are access points for memory retrieval. Psychological Science **34**(3), 326–344 (2023)
52. Minker, J.: Logic-based artificial intelligence, vol. 597. Springer Science & Business Media (2012)
53. Munir, A.e.h., Qazi, W.M.: Artificial subjectivity: Personal semantic memory model for cognitive agents. Applied Sciences **12**(4), 1903 (2022)
54. Munkhdalai, T., Sordoni, A., Wang, T., Trischler, A.: Metalearned neural memory. Advances in Neural Information Processing Systems **32** (2019)
55. Newell, A.: Unified theories of cognition. Harvard University Press (1994)
56. Niu, Q., Liu, J., Bi, Z., Feng, P., Peng, et al.: Large language models and cognitive science: A comprehensive review of similarities, differences, and challenges. arXiv preprint arXiv:2409.02387 (2024)
57. Park, J.S., O'Brien, J., Cai, et al.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th annual acm symposium on user interface software and technology. pp. 1–22 (2023)
58. Princeton University: WordNet: A Lexical Database for English (2024)
59. Qiu, Z., Liu, Z., Yan, S., Zhang, S., Fu, J.: Heterogenous memory augmented neural networks. arXiv preprint arXiv:2310.10909 (2023)
60. Ritter, F.E., Tehranchi, F., Oury, J.D.: Act-r: A cognitive architecture for modeling cognition. Wiley Interdisciplinary Reviews: Cognitive Science **10**(3), e1488 (2019)
61. Romero, O.J., Zimmerman, J., Steinfeld, A., Tomasic, A.: Synergistic integration of large language models and cognitive architectures for robust ai: An exploratory analysis. In: Proceedings of the AAAI Symposium Series. vol. 2, pp. 396–405 (2023)
62. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. nature **323**(6088), 533–536 (1986)
63. Sarch, G., Wu, Y., Tarr, M.J., Fragkiadaki, K.: Open-ended instructable embodied agents with memory-augmented large language models. arXiv preprint arXiv:2310.15127 (2023)
64. Smolensky, P.: Connectionist ai, symbolic ai, and the brain. Artificial Intelligence Review **1**(2), 95–109 (1987)
65. Sumers, T.R., Yao, S., Narasimhan, K., Griffiths, T.L.: Cognitive architectures for language agents. arXiv preprint arXiv:2309.02427 (2023)
66. Swallow, K.M., Zacks, J.M., Abrams, R.A.: Event boundaries in perception affect memory encoding and updating. Journal of Experimental Psychology: General **138**(2), 236 (2009)

67. Swartout, W.R., Gratch, J., Jr, H., et al.: Toward virtual humans. AI Magazine **27**(2), 96–96 (2006)
68. Thompson, C.P., Skowronski, J.J., Larsen, S.F., Betz, A.L.: Autobiographical memory: Remembering what and remembering when. Psychology Press (2013)
69. Tolman, E.C.: Cognitive maps in rats and men. Psychological review **55** (1948)
70. Tulving, E.: Episodic and semantic memory. In: Tulving, E., Donaldson, W. (eds.) Organization of Memory, pp. 381–402. Academic Press (1972)
71. Tulving, E.: Elements of Episodic Memory. Oxford University Press (1983)
72. Tulving, E., Schacter, D.: Priming and human memory systems. Science **247**(4940), 301–306 (1990)
73. Tulving, E.: How many memory systems are there? American Psychologist **40**(4), 385–398 (1985)
74. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
75. Wang, G., Xie, Y., Jiang, Y., Mandlekar, et al.: Voyager: An open-ended embodied agent with large language models. arXiv:2305.16291 (2023)
76. Wang, W., Tan, A.H., Teow, L.N.: Semantic memory modeling and memory interaction in learning agents. IEEE Transactions on Systems, Man, and Cybernetics: Systems **47**(11), 2882–2895 (2016)
77. Wang, Z., Cai, S., Liu, A., Jin, Y., Hou, et al.: Jarvis-1: Open-world multi-task agents with memory-augmented multimodal language models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
78. Wang, Z., Yu, B., Zhao, J., Sun, et al.: Karma: Augmenting embodied ai agents with long-and-short term memory systems. arXiv:2409.14908 (2024)
79. Weston, J., Chopra, S., Bordes, A.: Memory networks. arXiv preprint arXiv:1410.3916 (2014)
80. Wu, Q., Lan, Z., Qian, K., Gu, J., Geramifard, A., Yu, Z.: Memformer: A memory-augmented transformer for sequence modeling. arXiv:2010.06891 (2020)
81. Wu, S., Oltramari, A., Francis, J., Giles, C.L., Ritter, F.E.: Cognitive llms: Towards integrating cognitive architectures and large language models for manufacturing decision-making. arXiv preprint arXiv:2408.09176 (2024)
82. Xiong, H., Wang, Z., Li, et al.: Converging paradigms: The synergy of symbolic and connectionist ai in llm-empowered autonomous agents. arXiv preprint arXiv:2407.08516 (2024)
83. Yan, M., Li, R., Zhang, H., Wang, H., Yang, Z., Yan, J.: Larp: Language-agent role play for open-world games. arXiv preprint arXiv:2312.17653 (2023)
84. Zhang, K., Kang, Y., Zhao, F., Liu, X.: Llm-based medical assistant personalization with short-and long-term memory coordination. arXiv:2309.11696 (2023)
85. Zhang, Z., Bo, X., Ma, et al.: A survey on the memory mechanism of large language model based agents (2024)
86. Zhao, W.X., Zhou, K., Li, J., Tang, et al.: A survey of large language models. arXiv:2303.18223 **1**(2) (2023)
87. Zhao, Z., Chai, W., Wang, et al.: See and think: Embodied agent in virtual environment. In: European Conference on Computer Vision. Springer (2024)
88. Zhao, Z., Chai, W., Wang, X., Ma, K., Chen, et al.: Steve series: Step-by-step construction of agent systems in minecraft. arXiv preprint arXiv:2406.11247 (2024)
89. Zheng, S., He, K., Yang, et al.: Memoryrepository for ai npc. IEEE Access (2024)
90. Zhu, X., Chen, Y., Tian, H., Tao, et al.: Ghost in the minecraft: Generally capable agents for open-world environments via large language models with text-based knowledge and memory. arXiv:2305.17144 (2023)