ICLR 2026 Workshop Proposal
MemAgents: Memory for LLM-Based Agentic Systems

Zhengguang G. Cai[*], **Wenyue Hua**[†], **Keshuang Li**[‡], **Yunpu Ma**[§,♯], **Ercong Nie**[§,♯], **Hinrich Schütze**[§,♯], **Karolina Stańczak**[∂], **Matthew E. Taylor**[♭]

[*]The Chinese University of Hong Kong [†]Microsoft [‡]Huawei
[§]LMU Munich [♯] Munich Center for Machine Learning
[∂]ETH Zurich [♭]University of Alberta

1 Workshop Summary

Agentic systems are already being deployed in high-stakes settings such as robotics, autonomous web interaction, and software maintenance, and their capabilities ultimately hinge on memory (Mu et al., 2025; Wang et al., 2024b; Lei et al., 2025; Pink et al., 2025). While LLM memorization typically refers to static, in-weights retention of training data or recent context, agent memory is online, interaction-driven, and under the agent's control. This distinction brings new requirements, e.g., write policies during interaction, temporal credit assignment across episodes, and provenance-aware retrieval, that do not arise for standalone LLMs. Memory is therefore crucial for agentic systems. This is underscored by evidence that LLMs struggle to use long contexts robustly and often need external or structured memory. Agentic systems must operate over extended horizons, learn from interaction, and adapt as goals and contexts shift. The limiting factor is increasingly not raw model capability but memory (Du et al., 2025; Pink et al., 2025): how agents encode, retain, retrieve, and consolidate experience into useful knowledge for future decisions. Consistent with this view, recent commentary has argued that reinforcement learning can finally generalize when supplied with strong priors and explicit reasoning (Yao, 2025); however, current evaluations often underplay sequential accumulation of experience, where memory becomes decisive.

We propose a workshop devoted to memory for agentic systems, with a primary emphasis on the **memory layer for LLM-based agentic systems**. Our premise is that long-lived, safe, and useful agents require a principled memory substrate that supports single-shot learning of instances, context-aware retrieval, and consolidation into generalizable knowledge. The workshop is intentionally not a generic "LLM or agent" venue; it is a focused forum on *the memory layer* that underwrites agent behavior across domains, e.g., software tools, embodied and robotic tasks, and multi-agent settings. We will bridge three perspectives that center the memory layer: (1) **memory architectures and representations**, covering episodic, semantic, working, and parametric memory and their interfaces with external stores (Wang and Chen, 2025; Chhikara et al., 2025; Zhou et al., 2025; Li et al., 2025); (2) **systems and evaluation**, including data structures, retrieval and consolidation pipelines, and benchmarks that test non-i.i.d., long-horizon competence (Maharana et al., 2024; Kuratov et al., 2024; Wu et al., 2024a; Luo et al., 2025); and (3) **neuroscience-inspired memory**, including complementary learning systems and hippocampal–cortical consolidation as design inspiration (Squire et al., 2015; O'Reilly et al., 2014; Gutiérrez et al., 2025; Jimenez Gutierrez et al., 2024).

This workshop provides an ideal venue to advance the design of the memory layer for agentic systems and to convene researchers across reinforcement learning, memory research, large language models, agentic systems, and neuroscience, with an organizing team that spans these communities.

1.1 Workshop Goals

Why a dedicated memory layer for agents.

Before proceeding, we briefly distinguish agent memory from LLM memorization: (i) **When** learning happens, online during interaction versus mainly at pretraining or fine-tuning (Kiyomaru et al., 2024; Fang et al., 2025); (ii) **Where** it lives, hybrid external or structured stores plus in-weights versus primarily parametric (Yang

et al., 2024; Jimenez Gutierrez et al., 2024; Gutiérrez et al., 2025); (iii) **How** it is governed, explicit write and forget policies, temporal credit assignment, and provenance versus opaque parametric retention (Yang et al., 2025; Morris et al., 2025; Sakarvadia et al., 2024).

Given these differences, we next connect advances in reinforcement learning with neuroscience findings on memory. We focus on how an agent perceives, stores, and uses memories during interaction, and how the factual knowledge already present in LLMs interacts with agent-level memory. The workshop will motivate and define a dedicated memory module for agents, and will surface principled designs and benchmarks to support it. We will discuss interactions among memory types, including relationships between episodic and semantic memory and between working and long-term memory (Tresp et al., 2023; Modarressi et al., 2024; Tresp et al., 2015; Spens and Burgess, 2024). We will describe conversion pathways such as consolidation from episodic to semantic memory and the transition from explicit memory to in-weights implicit knowledge (Kim et al., 2025; Tian et al., 2025; Zhao et al., 2025), treating memory as part of the agent's cognitive loop rather than a passive log. We will also identify when specific memory mechanisms improve generalization, reduce forgetting, and keep agents safe and auditable over long horizons.

Integrate multiple disciplines.

Bring together researchers across generative AI, reinforcement learning, cognitive psychology, and neuroscience to share insights and approaches for designing **the memory layer of agents**, ranging from neuroscience-inspired mechanisms to practical RL and systems implementations, under the unifying theme of *agent memory design and operation*. Translate findings on episodic and semantic systems, replay, consolidation, and context binding into implementable mechanisms and metrics for agents, for example episodic control for reinforcement learning inspired by human episodic memory (Pritzel et al., 2017; Yang et al., 2025) and RL-managed external memory systems (Yan et al., 2025). This translation faces challenges such as efficient retrieval of relevant memories and balancing stability and plasticity in long-term learning (Rudroff et al., 2024).

Identify challenges and frameworks.

Discuss open challenges in making memory a core component of AI agents, including catastrophic forgetting, retrieval efficiency, and memory structure choices such as structured versus unstructured, symbolic versus neural, and graph-based versus vector-based, as well as interfaces between external and in-weights stores. Work toward common frameworks or benchmarks for evaluating memory capabilities in RL and agent settings, outlining clear research directions for memory systems that allow agents to learn cumulatively over time, adapt to new tasks, and approach human-like flexibility in learning.

1.2 Workshop Scope and Topics

The workshop will cover a broad range of topics related to the memory layer for agentic systems. Key themes and questions include:

Memory Architectures and Mechanisms.

What kinds of agent memory architectures best enable LLM-based agents to remember and use past information? We focus on mechanisms of the memory layer: long-term stores, retrieval and scheduling pipelines, context management such as chunking and summarization, key-value caches, and consolidation interfaces to in-weights memory. We will discuss how agents can store and retrieve important details from their histories to support planning, collaboration, and communication, safety monitoring, personalization, and decision-making, etc. Example papers from the community and organizers: Chhikara et al. (2025); Zhou et al. (2025); Li et al. (2025); Mei et al. (2024); Xu et al. (2025); Tresp et al. (2023, 2015); Ma et al. (2019); Yan et al. (2025); Modarressi et al. (2023).

Types of Memory in Agents.

In humans, long-term memory comprises explicit, or declarative, memories such as episodic events and facts, and implicit, or procedural, memories such as skills and habits. We discuss the analogy of these in agentic systems. For instance, an agent's episodic memory might store experiences, including trajectories of states, actions, and outcomes, while its procedural memory could manifest as policies or skills learned through repetition. How can we design agents that form episodic memories of significant events and *semantic memories* of factual knowledge, and then use them to plan or adapt? How can transient experiences be consolidated into

lasting knowledge? What algorithms or structures allow information to persist over an agent's lifetime? Example papers from the community and organizers: Shi et al. (2025); Pink et al. (2025); Du et al. (2025); Sumers et al. (2023); Wang and Chen (2025); Shinn et al. (2023); Park et al. (2023).

Explicit vs. In-Weights Memory in LLM Agents.
How should we represent and manage explicit memories in symbolic, textual, or graph form, such as gathered experiences or semantic facts consolidated from episodes, relative to in-weights memory? What mechanisms allow conversion, editing, or distillation from explicit stores into parametric representations, while preserving safety, privacy, and provenance? How should we balance an agent's explicit memory with its parametric memory? Example papers from the community and organizers: Anokhin et al. (2024); Jin et al. (2024); Jimenez Gutierrez et al. (2024); Gutiérrez et al. (2025); Borgeaud et al. (2022); Wang et al. (2024a); Meng et al. (2022a,b); İlhan et al. (2022); Hua et al. (2024a).

Comparisons to Human Memory.
What are the similarities and differences between an agentic system's memory and human memory? This topic invites insights from cognitive science and neuroscience on memory limitations and strengths. For example, both humans and AI systems exhibit biased recall. Human memories are influenced by attention and emotion, while AI recall might prioritize information aligned with training data, recent context, or task requirements. Humans also forget and abstract. Should agents similarly forget irrelevant details to avoid overload or use summarization techniques akin to human reflection? Example papers from the community and organizers: Maharana et al. (2024); Laban et al. (2025); Huff and Ulakçı (2024); Janik (2023); Huang et al. (2025); Hua et al. (2024c); Tresp and Li (2024); Cai et al. (2023).

Neuroscience-Inspired Memory Systems.
How can neuroscience inform the design of agent memory? We consider memory systems in the brain, for example, the hippocampus and cortex, consolidation, and replay, as inspiration and discuss how to translate these ideas into agent mechanisms and evaluation protocols such as complementary learning systems, dual-memory architectures, and targeted replay. We emphasize design principles and measurable outcomes rather than biological detail. Example papers from the community and organizers: Lin et al. (2025); Miao et al. (2024); Tresp et al. (2023, 2015); Jimenez Gutierrez et al. (2024); Gutiérrez et al. (2025).

Memory Dynamics and Lifelong Learning.
Memory is not static; it evolves over time. Key questions include how an agent should consolidate experiences into long-term memory and how dynamic retrieval can selectively recall relevant past experiences depending on context. Research on continual and lifelong learning in RL highlights the need for memory systems that support self-improving agents and enable knowledge accumulation across sequential tasks. Example papers from the community and organizers: Xiong et al. (2025); Wang et al. (2023); Ge et al. (2023); Zheng et al. (2025); Taylor and Stone (2009); Wang and Taylor (2018); İlhan et al. (2022).

Human and Multi-Agent Interaction Memory Over Time. How should agents write, align, and govern memories that arise from repeated interactions with other agents and with humans, across days or weeks, under constraints of identity, consent, and privacy? Key issues include persistence of roles and personas across sessions, memory synchronization and conflict resolution between agents, provenance and ownership for human-contributed memories, and conversion of dialogue episodes into shared plans and norms, etc. Example papers from the community and organizers: Wang and Chen (2025); Zhang et al. (2025); Ndousse et al. (2021); Li et al. (2024); Rezazadeh et al. (2025); Wu et al. (2024b); Hua et al. (2024b); Mei et al. (2024).

Evaluation and Benchmarks for Agent Memory.
How should we evaluate and compare memory capabilities in agents? We need principled benchmarks and tasks that test memory under controlled, interactive conditions. Discussion points include interactive environments that require long-term memory, multi-episode challenges that demand remembering past outcomes, standardized metrics for memory usage and forgetting, and robust methods for memorization and retrieval that can be audited. How do we detect when an agent is truly using memory rather than exploiting shortcuts? Example papers from the community and organizers: (Laban et al., 2025; Maharana et al., 2024; Hu et al., 2025; Wu et al., 2024a; Hua et al., 2024a; Modarressi et al., 2024).

Each of these topics touches on the central question of how to endow agents with **human-like capabilities in the memory layer** that improve their learning and performance. Ultimately, **advancing the memory layer for agentic systems** will bring us closer to AI agents that learn cumulatively, adapt readily to new challenges, and behave with the coherence and continuity we associate with memory in biological intelligence.

1.3 Differences from Previous Related Workshops

There are dedicated workshops on LLM-based agents and on memorization in LLMs. What is missing is a forum centered on the memory module for agents, because agent memory is dynamic, functioning during interaction and over time, coupling explicit stores with in-weight knowledge, retrieval, consolidation, and credit assignment across episodes, which differs materially from static memorization.

New Frontiers in Associative Memories @ ICLR 2025 (Kempe et al., 2025) foregrounded associative and memory-in-weights theory. L2M2 @ ACL 2025 (Jia et al., 2025) focuses on in-weights memorization, privacy risks, leakage, and editing or removal of memorized content in LLMs. ICLR 2024 Workshop on Large Language Models for Agents (Chen et al., 2024) takes a broad view of LLM agents, including tool use, reasoning, and multimodality, with one track on memory mechanisms. Our workshop stands apart. First, we focus specifically on the memory layer for agentic systems, examining how agents interact with and learn from their environments. Second, we adopt an agent-centric, neuroscience-informed lens that emphasizes transformations across memory types and their interrelations, using comparisons between agentic and human memory to derive design principles. We target the memory layer for agents: the theories, mechanisms, system designs, and evaluations that enable agents to learn from their memories and act better tomorrow.

2 Tentative Schedule

Tentative Important Dates We propose a one-day workshop featuring invited talks, poster sessions, a panel discussion, and short oral presentations for selected outstanding contributed papers.

- Workshop Proposal Submission Deadline: 10 October 2025
- Notification of Acceptance: 1 December 2025
- Workshop Paper Submission Deadline: 30 January 2026
- Workshop Paper Notification Date: 1 March 2026
- Camera-Ready Deadline: 11 March 2026
- Workshop Day: 26/27 April 2026 (co-located with ICLR 2026)

All deadlines are 11:59pm AoE (Anywhere on Earth), in line with ICLR policy.

Tentative Workshop Schedule We propose a single-day workshop, from 8:45 AM to 6:00 PM local time, in a hybrid format. The schedule is designed to maximize opportunities for knowledge sharing, discussion, and networking, and will be aligned with the official ICLR 2026 schedule. The tentative workshop schedule is detailed in Table 1. To maximize knowledge exchange and foster interdisciplinary collaboration, our workshop schedule is structured to provide substantial time for interactive group discussions, paper presentations, invited keynote talks, an expert panel, and recognition of outstanding contributions. Through a combination of oral presentations, poster sessions, and focused discussion periods, we aim to create an engaging environment where participants from interdisciplinary areas can meaningfully connect, share insights, and advance the state of memory-layer research for agentic systems. The final schedule will be closely coordinated with the official ICLR 2026 timetable to ensure smooth integration with the main conference and broad accessibility for all attendees.

Details on the Keynote Sessions. We will invite leading researchers and practitioners representing diverse backgrounds from LLMs, reinforcement learning, and neuroscience to share their perspectives on memory architectures, mechanisms, and evaluation for agentic AI systems. Each keynote will be allotted 25 minutes for presentation, followed by a 5-minute Q&A.

Time	Session
08:45 – 09:00	Welcome and Opening Remarks
09:00 – 09:30	Keynote 1: Aditi Raghunathan (CMU)
09:30 – 10:00	Keynote 2: Yuchen Eleanor Jiang (OPPO)
10:00 – 10:30	Keynote 3: Volker Tresp (LMU Munich)
10:30 – 11:30	Poster and Discussion Session (Coffee break)
11:30 – 12:30	Oral Presentation Session 1 (10 min × 6)
12:30 – 13:30	Lunch Break
13:30 – 14:00	Keynote 4: Jeff Z. Pan (University of Edinburgh)
14:00 – 14:30	Keynote 5: Zhuosheng Zhang (Shanghai Jiao Tong University)
14:30 – 15:00	Keynote 6: Timothy P. Lillicrap (Google DeepMind)
15:00 – 16:00	Poster and Discussion Session (Coffee break)
16:00 – 17:00	Panel discussion experts with diverse and well-balanced research backgrounds
17:00 – 17:50	Oral Presentation Session 2 (10 min × 5)
17:50 – 18:00	Paper Award Ceremony & Closing Remarks

Table 1: Tentative workshop schedule for the proposed one-day workshop.

Details on the Panel Session. Our expert panel will feature thought leaders from both academia and industry with balanced research backgrounds to foster an integrative discussion on the challenges and opportunities in building robust memory layers for long-lived agents. Panelists will address open questions such as balancing episodic and semantic memory, bridging explicit and in-weights memory, and evaluating real-world memory competence. To ensure broad engagement, we will solicit questions from submitting authors via the submission portal and gather community feedback and questions through our workshop's social media channels and website. The panel will begin by addressing pre-collected questions, then transition to live audience Q&A for an interactive and inclusive session.

Details on Oral and Poster Presentation Sessions. Accepted papers will be presented as either oral talks or posters, with the presentation format determined by the organizing committee based on quality, novelty, and relevance to the workshop themes. We will hold two oral presentation sessions, one in the morning and one in the afternoon, each featuring a series of 8-minute lightning talks followed by a 2-mintue Q&A. Two poster sessions will run concurrently with coffee breaks, enabling extended discussion and networking. To increase the accessibility and visibility of all contributions, presenters will be invited to optionally submit pre-recorded video summaries (5–7 minutes for spotlights, 1–3 minutes for posters), which will be featured on the workshop website. Outstanding contributions will be recognized with paper awards, announced prior to the closing remarks.

3 Invited Speaker/Panelists

Aditi Raghunathan (Assistant Professor, CMU) [Female, Confirmed]
Bio: Aditi Raghunathan is an Assistant Professor at Carnegie Mellon University. She is interested in building robust ML systems with guarantees for trustworthy real-world deployment. Previously, she was a postdoctoral researcher at Berkeley AI Research, and received her PhD from Stanford University in 2021. Her research has been recognized by Forbes 30 under 30, the Schmidt AI2050 Early Career Fellowship, the Arthur Samuel Best Thesis Award at Stanford, a Google PhD fellowship in machine learning, and an Open Philanthropy AI fellowship.
Research and Talk Topics: Memorization and LLM trainig

Yuchen Eleanor Jiang (Senior Researcher, OPPO) [Female, Confirmed]
Bio: Yuchen Eleanor Jiang is currently a senior researcher at OPPO. She was previously the CEO of AIWaves Inc. and earned her PhD and Master's degree in Computer Science from ETH Zurich. Her research focuses on large language models and agents, including language agents, web agents, self-evolving agents, etc.
Research and Talk Topics: Memory for self-evolving agents

Volker Tresp (Professor, LMU Munich) [Male, Confirmed]

Bio: Volker Tresp is a professor of informatics at Ludwig Maximilian University of Munich (LMU). He received his Ph.D. degree from Yale University in Yale's Image Processing and Analysis Group (IPAG). He was Siemens Inventor of the Year for his innovations in neural networks research, and the first Siemens Distinguished Research Scientist. He was co-editor of Advances in Neural Information Processing Systems 13. He is known for his work on Bayesian machine learning, in particular the Bayesian Committee Machine and his work on hierarchical learning with Gaussian processes. His team has been doing pioneering work on machine learning with knowledge graphs, temporal knowledge graphs, and scene graph analysis. The work on the Tensor Brain reflects his interest in mathematical models for cognition and neuroscience. In 2020, he became a Fellow of the European Laboratory for Learning and Intelligent Systems (ELLIS). As co-director (with Kristian Kersting and Paolo Frasconi), he leads the ELLIS program "Semantic, Symbolic and Interpretable Machine Learning".
Research and Talk Topics: Tensor Brain: Perception, Memory and Semantic Decoding

Jeff Z. Pan (Professor, The University of Edinburgh) [Male, Confirmed]
Bio: Jeff Pan is a professor of knowledge computing in the School of Informatics at the University of Edinburgh. He is the chair of the Knowledge Graphs group at the Alan Turing Institute. He received his Ph.D. in Computer Science from the University of Manchester. Prof. Pan's research focuses primarily on knowledge computing and artificial intelligence, in particular on knowledge-based learning and reasoning, and knowledge-based natural language understanding and generation, model fusion, model editing, and knowledge memorisation, as well as their applications, such as those in information retrieval, healthcare, software engineering, and open science. He is a key contributor of the W3C OWL (Web Ontology Language) standard. and led the development of the award-winning TrOWL reasoner, the only ontology reasoner that Oracle Spatial and Graph (from v12) uses via the OWL-DBC database connection. He is an internationally leading expert on Knowledge Graphs, the Chief Editor and main author of the first book on Knowledge Graph, and the Chief Scientist and Coordinator of the first EU Marie-Curie project on Knowledge Graph. He is an Editor of the Transactions on Graph Data and Knowledge (TGDK) and a Programme Chair of ISWC 2020, the premier international conference for the Semantic Web / Knowledge Graph.
Research and Talk Topics: Knowledge and memory representation, symbolic and parametric knowledge and memory

Zhuosheng Zhang (Assistant Professor, Shanghai Jiao Tong University) [Male, Confirmed]
Bio: Zhuosheng Zhang is a tenure-track assistant professor at Shanghai Jiao Tong University. His primary research interests include LLM reasoning, LLM agents, and LLM safety. He has published over 80 papers in top-tier conferences and journals, including TPAMI, ICML, ICLR, ACL, AAAI, EMNLP, TNNLS, TASLP, and COLING. Zhuosheng was selected as one of the Global Top 100 Chinese Rising Stars in Artificial Intelligence and won the Excellent Doctoral Thesis of Chinese Information Processing Society (CIPS), WAIC 2024 Youth Outstanding Paper Award, WAIC 2024 YunFan Award for Bright Star, and Baidu Scholarship. He has served as an action editor or (senior) area chair for NLP/ML venues such as ACL Rolling Review, ACL 2025, NeurIPS 2025, and EMNLP 2025. He actively contributes to NLP/ML community by giving tutorials at conferences such as CVPR 2024, LREC-COLING 2024, IJCNLP-AACL 2023, and IJCAI 2021.
Research and Talk Topics: Memory for Web Agents and Agent Safety

Timothy P. Lillicrap (Research Director, Google DepMind) [Male, Confirmed]
Bio: Timothy Lillicrap received an Hon. B.Sc. in Cognitive Science & Artificial Intelligence from the University of Toronto and a Ph.D. in Systems Neuroscience from Queen's University in Canada. He moved to the University of Oxford in 2012, where he worked as a Postdoctoral Research Fellow. In 2014, he joined Google DeepMind as a Research Scientist. His research focuses on machine learning for optimal control and decision-making, as well as using these mathematical frameworks to understand how the brain learns.
Research and Talk Topics: Deep reinforcement learning and memory

4 Organizers and Biographies

Wenyue Hua (Senior Researcher, Microsoft Research, Female)
Email: wenyuehua@microsoft.com
Bio: Wenyue Hua is currently a senior researcher at Microsoft Research, AI Frontiers. She was a CS postdoctoral researcher at UCSB working with Prof. William Wang. She received her Ph.D. from Rutgers University-

New Brunswick, under the supervision of Professor Yongfeng Zhang. Her research focuses on the safety and efficiency of LLM agents, multi-agent interaction, and LLM reasoning. She was selected as KAUST AI Rising Star in 2025, published over 40 papers at top natural language processing and machine learning conferences such as ACL, EMNLP, ICLR, NeurIPS, TACL.

Karolina Stańczak (Postdoctoral Researcher, ETH Zurich, Female)
Email: kstancza@ethz.ch
Bio: Karolina Stańczak is a postdoctoral researcher at the ETH AI Center. Previously, she was a postdoctoral researcher at Mila – Quebec AI Institute and McGill University School of Computer Science. She earned her PhD from the Department of Computer Science at the University of Copenhagen. Her thesis, titled "A Multilingual Perspective on Probing Gender Bias," was awarded the SCIENCE Faculty's PhD Award for advancing innovative techniques to detect gender bias both in natural language and language models. Her research interests encompass safety, alignment, and interpretability of large language models, with a focus on developing responsible AI systems in diverse, multicultural contexts.

Yunpu Ma (Lecturer, LMU Munich, Male)
Email: cognitive.yunpu@mail.com
Bio: Yunpu Ma is currently a lecturer at the Center for Information and Language Processing at LMU Munich. He is also affiliated with the Munich Center for Machine Learning and serves as a group leader at the TRESP Lab. His research focuses on designing AI agents with persistent memory, advanced reasoning capabilities, and the ability to operate autonomously in open-ended environments.

Zhengguang G. Cai (Professor, The Chinese University of Hong Kong, Male)
Email: zhenguangcai@cuhk.edu.hk
Bio: Zhengguang G. Cai is a Professor at the Department of Linguistics and Modern Languages at the Chinese University of Hong Kong. He received his PhD in psychology from the University of Edinburgh. Prior to joining CUHK, he was a lecturer in psychology at the University of East Anglia and an ESRC Future Research Leader fellow at University College London. Prof. Cai's research bridges the domains of language, memory, and cognition, employing behavioral, neuroscientific, and computational methods to investigate how the human brain encodes, stores, and retrieves linguistic information. His work has been supported by the Hong Kong General Research Fund and published in leading cognitive neuroscience journals. Prof. Cai also demonstrates strong academic leadership as the founder and organizer of the Virtual Psycholinguistics Forum.

Matthew E. Taylor (Professor, University of Alberta, Male)
Email: matthew.e.taylor@ualberta.ca
Bio: Matthew Taylor is a Fellow and Canada CIFAR AI Chair at Amii and a Professor of Computing Science at the University of Alberta. He is the Director of the Intelligent Robot Learning (IRL) Lab and a Principal Investigator at the Reinforcement Learning & Artificial Intelligence (RLAI) Lab, at the University of Alberta. His current fundamental and applied research interests are in reinforcement learning, human-in-the-loop AI, multi-agent systems, and robotics. His recent research focuses on developing intelligent agents, physical or virtual entities that interact with their environments. He has been a PI or co-PI on over $6M USD in competitively awarded research funding from federal, state, and industrial sources, including the National Science Foundation CAREER award. He has delivered invited talks for the Association for the Advancement of Artificial Intelligence (AAAI) and at the International Joint Conference on Artificial Intelligence (IJCAI).

Ercong Nie (Ph.D. Candidate, LMU Munich, Male)
Email: ecnie@foxmail.com
Bio: Ercong Nie is a final-year PhD candidate at the Center for Information and Language Processing (CIS), LMU Munich. His research centers on multilingual NLP, human-inspired interpretability of LLMs, and knowledge representation and editing in LLMs. He is an active member of the NLP and AI communities, with publications at leading conferences such as ACL, EMNLP, EACL, and COLING. He received a research grant from the German Federal Ministry of Education and Research for junior researchers. He has co-organized a Birds of a Feather (BoF) event at EMNLP and regularly serves as a program committee member for various NLP/ML venues.

Keshuang Li (Principle Strategetic Researcher, Huawei Technologies, Male)

Email: likeshuang@huawei.com
Bio: Keshuang Li is a Principal Strategic Researcher at Huawei. He received his Ph.D. from University College London. His research focuses on Embodied AI, intelligent Agent systems, and their applications in Multimodal scenarios.

Hinrich Schütze (Professor, LMU Munich, Male)
Email: hinrich@hotmail.com
Bio: Hinrich Schütze is Professor at the Center for Information and Language Processing at LMU Munich. His lab is engaged in research on representation learning, multilinguality and interpretability of large language models, knowledge expansion of NLP models, and the intersection of NLP and robotics. His research has been funded by NSF, the German National Science Foundation, and the European Research Council (ERC Advanced Grant), inter alia. Prof. Schütze is coauthor of two well-known textbooks (Foundations of Statistical Natural Language Processing and Introduction to Information Retrieval), a fellow of HessianAI, ELLIS (the European Laboratory for Learning and Intelligent Systems) and ACL (Association for Computational Linguistics), and (co-)awardee of several best paper awards and the ACL 2023 25-year test of time award. His extensive organizational experience within the NLP community is demonstrated by his service as a former ACL president and as General Chair of ACL 2013.

5 Diversity and Inclusion Commitment

We are deeply committed to advancing diversity and inclusion throughout every aspect of our workshop, from the selection of the organization team and invited speakers to our review process, program design, and outreach.

Organizers and Speakers. In formulating the organizational team and selecting speakers, we have prioritized diversity across gender, race, institutional affiliation, career stage, academic discipline, and industrial representation, in line with ICLR's diversity expectations. We bring together experts from LLMs/agents, RL, NLP, neuroscience, and cognitive psychology, ensuring a plurality of perspectives on the memory layer (architectures, algorithms, systems, and evaluations). Our roster includes individuals from various backgrounds and career stages. Among the 8 organizers, there are 6 from academia, including three full professors, 1 assistant professor, 1 postdoctoral scholar, and 1 Ph.D. candidate. Particularly, we have 2 industrial practitioners from frontier companies. Organizers are affiliated with different institutions covering North America, Europe, and Asia. Our commitment to fostering diversity is also evident in our choice of invited speakers. We invite speakers to introduce their perspectives on memory for LLM agents from diverse research fields, including agentic AI, RL, NLP, and neuroscience. Moreover, we are conscious of gender and geographic location representation, with both female and male speakers from North American, European, and Asian institutions.

Participant and Reviewer Diversity. To promote an inclusive environment for all participants, we will implement a double-blind review process on OpenReview to help minimize bias and ensure fairness. We aim to create a space where attendees can connect over shared perspectives and engage in meaningful discussions. The program will be curated to feature a wide representation of research topics while adhering to high standards of quality. Besides, we will actively recruit a diverse program committee and encourage submissions from underrepresented and under-resourced communities, including through affinity groups such as WiML and Black in AI. We will also offer a tiny/short-paper track designed to lower barriers for late-breaking and less-than-full-conference work of early-career researchers and contributors from outside the traditional ML/AI publication pipeline, in line with ICLR guidance.

Inclusion in program design. The program will emphasize interactive elements (posters, spotlights, panel, and Q&A) to broaden participation and enable mentoring and networking.

6 Workshop Modality, Virtual Access, and Reaching out

Workshop Modality. The workshop will be conducted in a **hybrid** format, with a primary emphasis on in-person participation to facilitate deep engagement and spontaneous collaboration. Recognizing the value of

broader accessibility, we will also provide comprehensive remote participation options. All sessions, including keynote talks, paper presentations, and panel discussions, will be live-streamed in accordance with ICLR 2026 guidelines. The conference center's standard equipment will be utilized to ensure high-quality streaming and technical reliability. Our organizing team brings extensive experience with platforms such as Zoom and GatherTown Live, ensuring smooth facilitation for both physical and virtual attendees. In-person interactions will be complemented by robust online tools to promote seamless communication and inclusive participation throughout the event.

Accessibility and Virtual Access. Accessibility is a central commitment of our workshop. All workshop materials, including the program schedule, organizer and speaker lists, and PDFs of accepted papers, will be made available in advance on the workshop website. Live-streamed sessions will be supplemented with asynchronous access: presentation recordings will be released on YouTube and archived on the website, and pre-recorded video summaries will be provided for contributed works. The public Slack channel will facilitate real-time and asynchronous interaction, enabling all participants to participate in Q&A and community dialogue regardless of time zone or location. Should any technical or accessibility issues arise, critical information and content will remain readily accessible via the website and Slack. Following the event, we will compile and publish a comprehensive report summarizing key outcomes and discoveries, ensuring the workshop's impact is sustained and broadly shared.

Funding. We will collaborate with industrial partners in sponsorship support, including Microsoft, Amazon, Huawei, and ByteDance. Besides, academic organizations such as the Munich Center for Machine Learning (MCML) have expressed interest in providing funding assistance. The specific level of support will be finalized once the workshop receives approval. These contributions will be allocated to prize awarding and travel grants for students and early-career researchers. Consistent with our commitment to diversity, equity, and inclusion, a portion of the funds will be reserved to broaden participation from underrepresented groups and to help reduce registration costs for attendees from low-income regions. This support will enable us to make the workshop more inclusive, engaging, and impactful for the entire community.

Advertising and Outreach. To maximize visibility and community engagement, we will employ a multi-channel approach to advertising and outreach. A dedicated workshop website will serve as the central hub for all information, including the call for papers, program schedule, organizer and speaker profiles, and accepted paper preprints. We will publicize the workshop through social media platforms such as X (Twitter) and LinkedIn, as well as through organizers' and speakers' personal and institutional websites. Announcements will be distributed via university and laboratory mailing lists, industry partner networks, and major research institutes and consortia, such as MCML and ELLIS. We will also connect with local and regional communities across North America, Europe, and Asia, and reach out to related interest groups in machine learning, NLP, reinforcement learning, neuroscience, and robotics. Engagement will be further encouraged via a public Slack channel, which will support ongoing discussion and Q&A before, during, and after the workshop.

7 Workshop Submission, Reviewing, and Anticipated Size

Workshop Submission. We welcome submissions in three tracks to accommodate a wide range of contributors and foster inclusivity. For **regular submissions**, we invite both short research papers (up to 4 pages) and full-length papers (up to 9 pages, excluding references and supplementary materials). All submissions must adhere to the ICLR 2026 template and be submitted in PDF format via the OpenReview platform. Accepted papers will be presented as posters, with 10–12 selected for oral spotlight presentations based on quality and relevance. Exceptional contributions will be recognized with two outstanding paper awards and one social impact award. All accepted papers will be made available through both the workshop homepage and OpenReview. In alignment with the ICLR Workshop Proposals 2026 guidance, we will establish a **Tiny Paper Track** to encourage participation from under-represented, under-resourced, and emerging researchers, including those who may not have the means to submit full papers or who are new to the field. Tiny papers (limited to 2 pages) are particularly suitable for late-breaking results, novel but concise theoretical insights, re-analyses of prior work, human-inspired perspectives, or interdisciplinary ideas—especially from neuroscience and cognitive science communities—seeking feedback and engagement at ICLR. This track is designed to broaden accessibility and foster interdisciplinary dialogue within the workshop community.

Verena Blaschke *PhD Student, LMU Munich*	**Yiwei Wang** *PhD Student, TU Darmstadt*	**Nadav Borenstein** *Postdoc Fellow, University of Copenhagen*	**Hongru Wang** *Postdoc Fellow, University of Edinburgh*
Di Wu *PhD Student, University of Amsterdam*	**Philipp Mondorf** *PhD Student, LMU Munich*	**Lei Li** *PhD Student, Hong Kong University*	**Tobias Schreieder** *PhD Student, TU Dresden*
Heming Xia *PhD Student, Hong Kong PolyU*	**Yongkang Liu** *Lecturer, Northeastern University*	**Ahmad Dawar Hakimi** *PhD Student, LMU Munich*	**Mengru Wang** *PhD Student Zhejiang University*
Tim Schopf *Postdoc Fellow, TU Dresden*	**Linyang He** *PhD Student, Columbia University*	**Qianli Wang** *PhD Student, TU Berlin*	**Lizhou Fan** *AP, CUHK*
Siddhesh Pawar *PhD Student, University of Copenhagen*	**Wenhao Zhu** *Research Scientist, ByteDance*	**Ali Modarressi** *PhD Student, LMU Munich*	**Mingyang Wang** *Applied Scientist, Amazon*
Abdullatif Köksal *Research Scientist, Google DeepMind*	**Nicholas Popovic** *PhD Student, TU Dresden*	**Rui He** *Postdoc Fellow, UPF Barcelona*	**Bo Shao** *PhD Student, CISPA Helmholtz*
Zheyu Zhang *PhD Student, TU Munich*	**Shaoxiong Ji** *AP, University of Turku*	**Xueli An** *Senior Researcher, Huawei*	**Robert Litschko** *Postdoc Fellow, LMU Munich*
Sikuan Yan *PhD Student, LMU Munich*	**Jinhe Bi** *Doctoral Researcher, Huawei*	**Greta Warren** *Postdoc Fellow, University of Copenhagen*	**Valentin Hofmann** *Postdoc Fellow, University of Washington*
Florian Eichin *PhD Student, LMU Munich*	**Qiyuan Zhang** *PhD Student, Hong Kong CityU*	**Zhaochen Su** *PhD Student, HKUST*	**Jiaming Ji** *PhD Student, Peking University*
Lei Lu *Research Director, Huawei*	**Amir Hossein Kargaran** *PhD Student, LMU Munich*	**Shaomu Tan** *PhD Student, University of Amsterdam*	**Zifeng Ding** *Postdoc Fellow, University of Cambridge*
Jiayuan Rao *PhD Student, SJTU*	**Xufeng Duan** *Research AP, CUHK*		

Table 2: Pool of Candidate PC Members.

Anticipated Size. Given the topical breadth and multidisciplinary appeal of this workshop, we anticipate an audience of approximately 300–500 participants. Based on related workshops at previous ICLR conferences and the broad interest in agentic AI and memory systems, we expect to receive between 50 and 100 paper submissions across all tracks.

Program Committee and Reviewing. Table 2 presents our proposed program committee, comprising 42 researchers from 28 institutions worldwide. We have intentionally assembled a diverse committee that spans all career stages—from PhD students and postdoctoral researchers to senior faculty—ensuring a broad range of perspectives and expertise. The committee members collectively bring strong backgrounds in natural language processing, machine learning, and neuroscience, reflecting the interdisciplinary nature of the workshop. All candidates are established contributors to their fields, with multiple publications in top-tier conferences and journals. This breadth and depth of experience will help ensure a rigorous and inclusive review process, as well as vibrant discussions throughout the workshop.

Each submission will undergo a double-blind review process via OpenReview, with every paper assigned to three to four reviewers. We will ensure that no reviewer is assigned more than four papers to maintain quality and provide detailed, thoughtful feedback. Our program committee comprises a diverse group of experts spanning NLP, reinforcement learning, AI/ML, LLM/agents, cognitive science, and neuroscience. To further enrich the review process, we will invite submitting authors to serve as additional reviewers. We will also recruit emergency reviewers as needed to ensure that all submissions receive timely evaluations and that the review process remains on schedule. Conflicts of interest will be rigorously managed in accordance with ICLR workshop policies: organizers and program committee members will not participate in the review of submissions from their own institutions or with whom they have a conflict.

References

Petr Anokhin, Nikita Semenov, Artyom Sorokin, Dmitry Evseev, Andrey Kravchenko, Mikhail Burtsev, and Evgeny Burnaev. 2024. Arigraph: Learning knowledge graph world models with episodic memory for llm agents. *arXiv preprint arXiv:2407.04363*.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, and 1 others. 2022.

Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*, pages 2206–2240. PMLR.

Zhenguang G Cai, Xufeng Duan, David A Haslett, Shuqi Wang, and Martin J Pickering. 2023. Do large language models resemble humans in language use? *arXiv preprint arXiv:2303.08014*.

Xinyun Chen, Xiangru Tang, Di Jin, Devamanyu Hazarika, Daniel Fried, Dawn Song, Shafiq Joty, and Meredith Ringel Morris. 2024. Workshop on large language models for agents. In *ICLR 2024 Workshops*.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*.

Yiming Du, Wenyu Huang, Danna Zheng, Zhaowei Wang, Sebastien Montella, Mirella Lapata, Kam-Fai Wong, and Jeff Z Pan. 2025. Rethinking memory in ai: Taxonomy, operations, topics, and future directions. *arXiv preprint arXiv:2505.00675*.

Runnan Fang, Yuan Liang, Xiaobin Wang, Jialong Wu, Shuofei Qiao, Pengjun Xie, Fei Huang, Huajun Chen, and Ningyu Zhang. 2025. Memp: Exploring agent procedural memory. *arXiv preprint arXiv:2508.06433*.

Yingqiang Ge, Wenyue Hua, Kai Mei, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang, and 1 others. 2023. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems*, 36:5539–5568.

Bernal Jiménez Gutiérrez, Yiheng Shu, Weijian Qi, Sizhe Zhou, and Yu Su. 2025. From rag to memory: Non-parametric continual learning for large language models. *arXiv preprint arXiv:2502.14802*.

Yuanzhe Hu, Yu Wang, and Julian McAuley. 2025. Evaluating memory in llm agents via incremental multi-turn interactions. *arXiv preprint arXiv:2507.05257*.

Wenyue Hua, Jiang Guo, Mingwen Dong, Henghui Zhu, Patrick Ng, and Zhiguo Wang. 2024a. Propagation and pitfalls: Reasoning-based assessment of knowledge editing through counterfactual tasks. *arXiv preprint arXiv:2401.17585*.

Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, and 1 others. 2024b. Game-theoretic llm: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*.

Wenyue Hua, Xianjun Yang, Mingyu Jin, Zelong Li, Wei Cheng, Ruixiang Tang, and Yongfeng Zhang. 2024c. Trustagent: Towards safe and trustworthy llm-based agents through agent constitution. In *Trustworthy Multimodal Foundation Models and AI Agents (TiFA)*.

Jen-tse Huang, Kaiser Sun, Wenxuan Wang, and Mark Dredze. 2025. Llms do not have human-like working memory. *arXiv preprint arXiv:2505.10571*.

Markus Huff and Elanur Ulakçı. 2024. Towards a psychology of machines: large language models predict human memory. *arXiv preprint arXiv:2403.05152*.

Ercüment İlhan, Srijita Das, Matthew E Taylor, and 1 others. 2022. Methodical advice collection and reuse in deep reinforcement learning. *arXiv preprint arXiv:2204.07254*.

Romuald A Janik. 2023. Aspects of human memory and large language models. *arXiv preprint arXiv:2311.03839*.

Robin Jia, Eric Wallace, Yangsibo Huang, Tiago Pimentel, Pratyush Maini, Verna Dankers, Johnny Wei, and Pietro Lesci. 2025. The first workshop on large language model memorization. ACL 2025 workshop page.

Bernal Jimenez Gutierrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. 2024. Hipporag: Neurobiologically inspired long-term memory for large language models. *Advances in Neural Information Processing Systems*, 37:59532–59569.

Mingyu Jin, Weidi Luo, Sitao Cheng, Xinyi Wang, Wenyue Hua, Ruixiang Tang, William Yang Wang, and Yongfeng Zhang. 2024. Disentangling memory and reasoning ability in large language models. *arXiv preprint arXiv:2411.13504*.

Julia Kempe, Dmitry Krotov, Hilde Kuehne, Daniel Lee, and Sara Solla. 2025. New frontiers in associative memories. Workshop page.

Joongwon Kim, Anirudh Goyal, Liang Tan, Hannaneh Hajishirzi, Srinivasan Iyer, and Tianlu Wang. 2025. Astro: Teaching language models to reason by reflecting and backtracking in-context. *arXiv preprint arXiv:2507.00417*.

Hirokazu Kiyomaru, Issa Sugiura, Daisuke Kawahara, and Sadao Kurohashi. 2024. A comprehensive analysis of memorization in large language models. In *Proceedings of the 17th International Natural Language Generation Conference*, pages 584–596.

Yury Kuratov, Aydar Bulatov, Petr Anokhin, Ivan Rodkin, Dmitry Sorokin, Artyom Sorokin, and Mikhail Burtsev. 2024. Babilong: Testing the limits of llms with long context reasoning-in-a-haystack. *Advances in Neural Information Processing Systems*, 37:106519–106554.

Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. 2025. Llms get lost in multi-turn conversation. *arXiv preprint arXiv:2505.06120*.

Mingcong Lei, Honghao Cai, Binbin Que, Zezhou Cui, Liangchen Tan, Junkun Hong, Gehan Hu, Shuangyu Zhu, Yimou Wu, Shaohan Jiang, and 1 others. 2025. Robomemory: A brain-inspired multi-memory agentic framework for lifelong learning in physical embodied systems. *arXiv preprint arXiv:2508.01415*.

Xinyi Li, Sai Wang, Siqi Zeng, Yu Wu, and Yi Yang. 2024. A survey on llm-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, 1(1):9.

Zhiyu Li, Shichao Song, Chenyang Xi, Hanyu Wang, Chen Tang, Simin Niu, Ding Chen, Jiawei Yang, Chunyu Li, Qingchen Yu, and 1 others. 2025. Memos: A memory os for ai system. *arXiv preprint arXiv:2507.03724*.

Yueqian Lin, Qinsi Wang, Hancheng Ye, Yuzhe Fu, Hai Li, Yiran Chen, and 1 others. 2025. Hippomm: Hippocampal-inspired multimodal memory for long audiovisual event understanding. *arXiv preprint arXiv:2504.10739*.

Haotian Luo, Huaisong Zhang, Xuelin Zhang, Haoyu Wang, Zeyu Qin, Wenjie Lu, Guozheng Ma, Haiying He, Yingsha Xie, Qiyang Zhou, and 1 others. 2025. Ultrahorizon: Benchmarking agent capabilities in ultra long-horizon scenarios. *arXiv preprint arXiv:2509.21766*.

Yunpu Ma, Volker Tresp, and Erik A Daxberger. 2019. Embedding models for episodic knowledge graphs. *Journal of Web Semantics*, 59:100490.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. Evaluating very long-term conversational memory of llm agents. *arXiv preprint arXiv:2402.17753*.

Kai Mei, Xi Zhu, Wujiang Xu, Wenyue Hua, Mingyu Jin, Zelong Li, Shuyuan Xu, Ruosong Ye, Yingqiang Ge, and Yongfeng Zhang. 2024. Aios: Llm agent operating system. *arXiv preprint arXiv:2403.16971*.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. 2022a. Locating and editing factual associations in gpt. *Advances in neural information processing systems*, 35:17359–17372.

Kevin Meng, Arnab Sen Sharma, Alex Andonian, Yonatan Belinkov, and David Bau. 2022b. Mass-editing memory in a transformer. *arXiv preprint arXiv:2210.07229*.

Xin Miao, Yongqi Li, Shen Zhou, and Tieyun Qian. 2024. Episodic memory retrieval from llms: A neuromorphic mechanism to generate commonsense counterfactuals for relation extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 2489–2511.

Ali Modarressi, Ayyoob Imani, Mohsen Fayyaz, and Hinrich Schütze. 2023. Ret-llm: Towards a general read-write memory for large language models. *arXiv preprint arXiv:2305.14322*.

Ali Modarressi, Abdullatif Köksal, Ayyoob Imani, Mohsen Fayyaz, and Hinrich Schütze. 2024. Memllm: Finetuning llms to use an explicit read-write memory. *arXiv preprint arXiv:2404.11672*.

John X Morris, Chawin Sitawarin, Chuan Guo, Narine Kokhlikyan, G Edward Suh, Alexander M Rush, Kamalika Chaudhuri, and Saeed Mahloujifar. 2025. How much do language models memorize? *arXiv preprint arXiv:2505.24832*.

Fangwen Mu, Junjie Wang, Lin Shi, Song Wang, Shoubin Li, and Qing Wang. 2025. Experepair: Dual-memory enhanced llm-based repository-level program repair. *arXiv preprint arXiv:2506.10484*.

Kamal K Ndousse, Douglas Eck, Sergey Levine, and Natasha Jaques. 2021. Emergent social learning via multi-agent reinforcement learning. In *International conference on machine learning*, pages 7991–8004. PMLR.

Randall C O'Reilly, Rajan Bhattacharyya, Michael D Howard, and Nicholas Ketz. 2014. Complementary learning systems. *Cognitive science*, 38(6):1229–1248.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Mathis Pink, Qinyuan Wu, Vy Ai Vo, Javier Turek, Jianing Mu, Alexander Huth, and Mariya Toneva. 2025. Position: Episodic memory is the missing piece for long-term llm agents. *arXiv preprint arXiv:2502.06975*.

Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adria Puigdomenech Badia, Oriol Vinyals, Demis Hassabis, Daan Wierstra, and Charles Blundell. 2017. Neural episodic control. In *International conference on machine learning*, pages 2827–2836. PMLR.

Alireza Rezazadeh, Zichao Li, Ange Lou, Yuying Zhao, Wei Wei, and Yujia Bao. 2025. Collaborative memory: Multi-user memory sharing in llm agents with dynamic access control. *arXiv preprint arXiv:2505.18279*.

Thorsten Rudroff, Oona Rainio, and Riku Klen. 2024. Neuroplasticity meets artificial intelligence: A hippocampus-inspired approach to the stability–plasticity dilemma. *Brain Sciences*, 14(11):1111.

Mansi Sakarvadia, Aswathy Ajith, Arham Khan, Nathaniel Hudson, Caleb Geniesse, Kyle Chard, Yaoqing Yang, Ian Foster, and Michael W Mahoney. 2024. Mitigating memorization in language models. *arXiv preprint arXiv:2410.02159*.

Yaorui Shi, Yuxin Chen, Siyuan Wang, Sihang Li, Hengxing Cai, Qi Gu, Xiang Wang, and An Zhang. 2025. Look back to reason forward: Revisitable memory for long-context llm agents. *arXiv preprint arXiv:2509.23040*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36:8634–8652.

Eleanor Spens and Neil Burgess. 2024. A generative model of memory construction and consolidation. *Nature human behaviour*, 8(3):526–543.

Larry R Squire, Lisa Genzel, John T Wixted, and Richard G Morris. 2015. Memory consolidation. *Cold Spring Harbor perspectives in biology*, 7(8):a021766.

Theodore Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas Griffiths. 2023. Cognitive architectures for language agents. *Transactions on Machine Learning Research*.

Matthew E Taylor and Peter Stone. 2009. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(7).

Wanxin Tian, Shijie Zhang, Kevin Zhang, Xiaowei Chi, Yulin Luo, Junyu Lu, Chunkai Fan, Qiang Zhou, Yiming Zhao, Ning Liu Siyu Lin, and 1 others. 2025. Seea-r1: Tree-structured reinforcement fine-tuning for self-evolving embodied agents. *arXiv preprint arXiv:2506.21669*.

Volker Tresp, Cristóbal Esteban, Yinchong Yang, Stephan Baier, and Denis Krompaß. 2015. Learning with memory embeddings. *arXiv preprint arXiv:1511.07972*.

Volker Tresp and Hang Li. 2024. How the (tensor-) brain uses embeddings and embodiment to encode senses and symbols. *arXiv preprint arXiv:2409.12846*.

Volker Tresp, Sahand Sharifzadeh, Hang Li, Dario Konopatzki, and Yunpu Ma. 2023. The tensor brain: A unified theory of perception, memory, and semantic decoding. *Neural Computation*, 35(2):156–227.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Peng Wang, Zexi Li, Ningyu Zhang, Ziwen Xu, Yunzhi Yao, Yong Jiang, Pengjun Xie, Fei Huang, and Huajun Chen. 2024a. Wise: Rethinking the knowledge memory for lifelong model editing of large language models. *Advances in Neural Information Processing Systems*, 37:53764–53797.

Yu Wang and Xi Chen. 2025. Mirix: Multi-agent memory system for llm-based agents. *arXiv preprint arXiv:2507.07957*.

Zhaodong Wang and Matthew E Taylor. 2018. Interactive reinforcement learning with dynamic reuse of prior knowledge from human/agent's demonstration. *arXiv preprint arXiv:1805.04493*.

Zora Zhiruo Wang, Jiayuan Mao, Daniel Fried, and Graham Neubig. 2024b. Agent workflow memory. *arXiv preprint arXiv:2409.07429*.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. 2024a. Longmemeval: Benchmarking chat assistants on long-term interactive memory. *arXiv preprint arXiv:2410.10813*.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, and 1 others. 2024b. Autogen: Enabling next-gen llm applications via multi-agent conversations. In *First Conference on Language Modeling*.

Zidi Xiong, Yuping Lin, Wenya Xie, Pengfei He, Jiliang Tang, Himabindu Lakkaraju, and Zhen Xiang. 2025. How memory management impacts llm agents: An empirical study of experience-following behavior. *arXiv preprint arXiv:2505.16067*.

Wujiang Xu, Kai Mei, Hang Gao, Juntao Tan, Zujie Liang, and Yongfeng Zhang. 2025. A-mem: Agentic memory for llm agents. *arXiv preprint arXiv:2502.12110*.

Sikuan Yan, Xiufeng Yang, Zuchao Huang, Ercong Nie, Zifeng Ding, Zonggen Li, Xiaowen Ma, Hinrich Schütze, Volker Tresp, and Yunpu Ma. 2025. Memory-r1: Enhancing large language model agents to manage and utilize memories via reinforcement learning. *arXiv preprint arXiv:2508.19828*.

Hongkang Yang, Zehao Lin, Wenjin Wang, Hao Wu, Zhiyu Li, Bo Tang, Wenqiang Wei, Jinbo Wang, Zeyun Tang, Shichao Song, and 1 others. 2024. Memory3: Language modeling with explicit memory. *arXiv preprint arXiv:2407.01178*.

Xidong Yang, Wenhao Li, Junjie Sheng, Chuyun Shen, Yun Hua, and Xiangfeng Wang. 2025. Agentic episodic control. *arXiv preprint arXiv:2506.01442*.

Shunyu Yao. 2025. The second half. Blog post.

Guibin Zhang, Muxin Fu, Guancheng Wan, Miao Yu, Kun Wang, and Shuicheng Yan. 2025. G-memory: Tracing hierarchical memory for multi-agent systems. *arXiv preprint arXiv:2506.07398*.

Linxi Zhao, Sofian Zalouk, Christian K Belardi, Justin Lovelace, Jin Peng Zhou, Kilian Q Weinberger, Yoav Artzi, and Jennifer J Sun. 2025. Pre-training large memory language models with internal and external knowledge. *arXiv preprint arXiv:2505.15962*.

Junhao Zheng, Chengming Shi, Xidi Cai, Qiuke Li, Duzhen Zhang, Chenxing Li, Dong Yu, and Qianli Ma. 2025. Lifelong learning of large language model based agents: A roadmap. *arXiv preprint arXiv:2501.07278*.

Zijian Zhou, Ao Qu, Zhaoxuan Wu, Sunghwan Kim, Alok Prakash, Daniela Rus, Jinhua Zhao, Bryan Kian Hsiang Low, and Paul Pu Liang. 2025. Mem1: Learning to synergize memory and reasoning for efficient long-horizon agents. *arXiv preprint arXiv:2506.15841*.