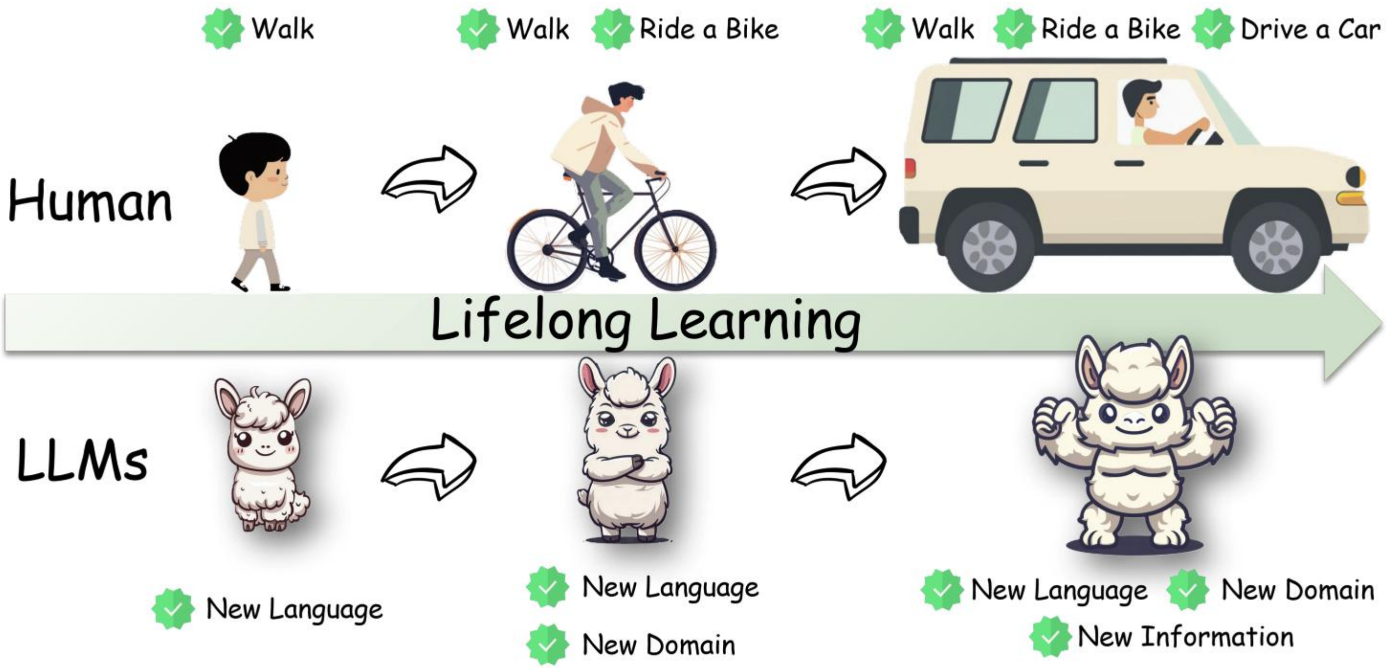
从数据驱动到经验驱动的终身学习

-构建自主进化智能体

周 杰 青年研究员
计算机科学与技术学院
華東師範大學
EAST CHINA NORMAL UNIVERSITY

终身（持续）学习背景



- □ 克服灾难性遗忘： 学习新的任务T_{N+1} 而不遗忘以前N个任务的能力
- □ 知识迁移：利用前面任务学习的知识用于学习新的任务T_{N+1}，包括正向和反向迁移

Junhao Zheng, et al. Towards Lifelong Learning of Large Language Models: A Survey

传统数据驱动的持续学习

- 封闭环境持续学习模型: 一个模型学习一系列任务
  - Continual learning/Lifelong Learning/Increment Learning
- 任务定义：依次学一连串任务$T_1, T_2, \ldots, T_N, \ldots$. 每一个任务t都包含一个完整训练数据.
  - 克服灾难性遗忘（catastrophic forgetting，CF）： 学习新的任务T_{N+1}而不遗忘以前N个任务的能力
  - 知识迁移（knowledge transfer，KT）：利用前面任务学习的知识用于学习新的任务T_{N+1}
    - 正向迁移
    - 反向迁移

假设:

当一个任务学完以后，该任务的（至少大部分）数据不可获取

任务T_{N+1} 和数据集D_{N+1} 都是完整给定的

$Y_{test} \in Y_{train}$

基于大模型的持续学习综述

离线持续学习

1）任务增量学习；

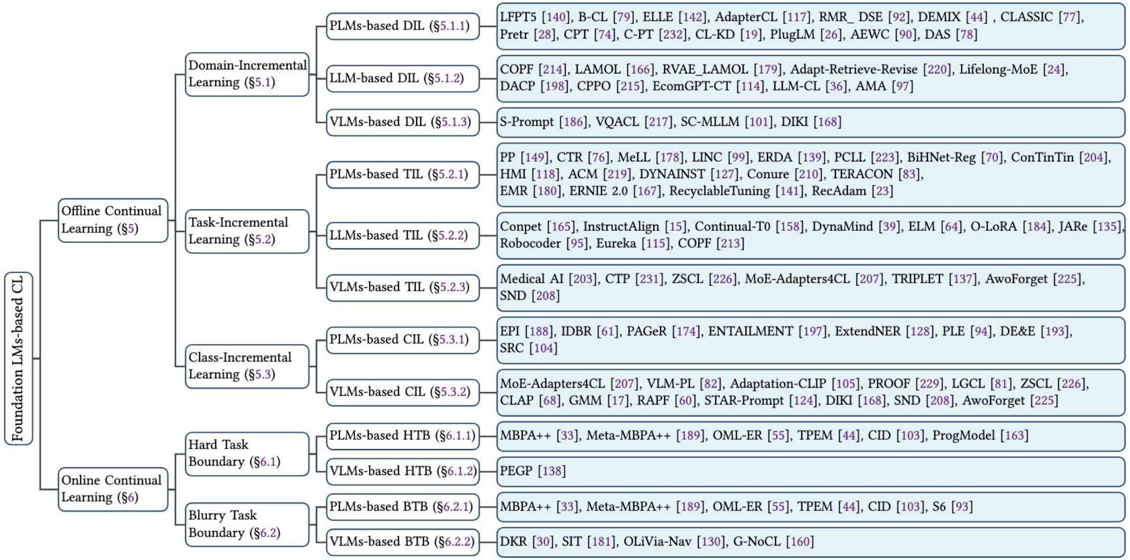
2）类别增量学习；

3）领域增量学习

在线持续学习

1）固定任务边界；

2）模糊任务边界



EASYCL
Continual Learning for Large Language Models

ICALK

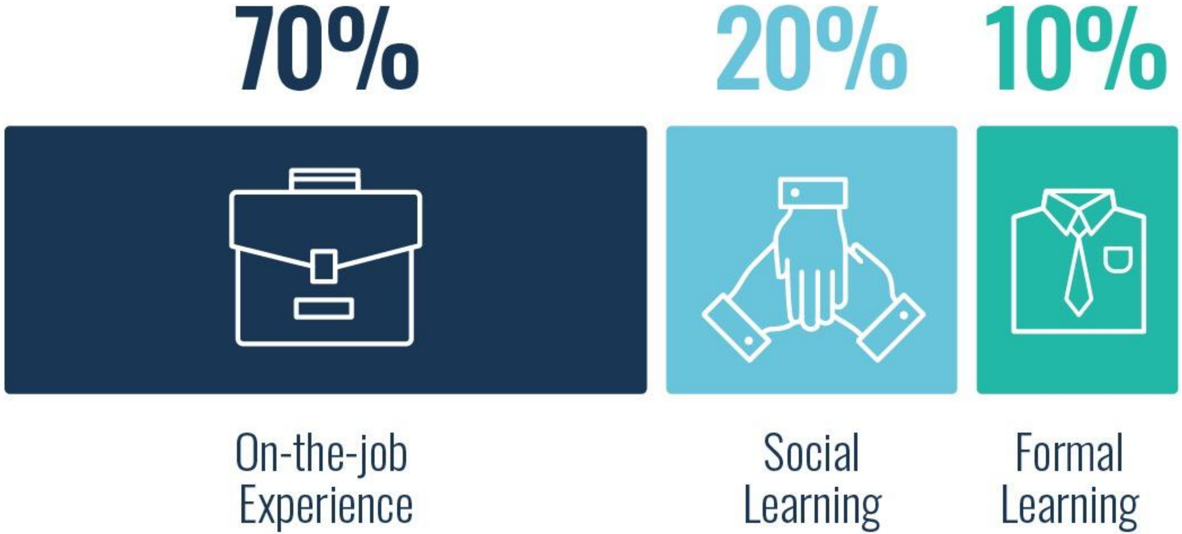
EasyCL（https://github.com/ECNU-ICALK/EasyCL）

多模态支持、适配绝大部分基座大语言模型，并支持十余种前沿的持续学习方法。

Y Yang, **J Zhou***, X Ding, T Huai, S Liu, Q Chen, L He, Y Xie. Recent Advances of Foundation Language Models-based Continual Learning: A Survey. ACM Computing Survey, 2025.

人如何学习知识

The 70:20:10 Framework



“**Formal training** is the tip of the **iceberg**. The real learning happens below the surface, through experience and relationships.”

Lombardo, Michael M; Eichinger, Robert W (1996). The Career Architect Development Planner (1st ed.). Minneapolis: Lominger. p. iv. ISBN 0-9655712-1-1.

人如何学习知识-以科研论文撰写为例

Paper 1

步骤4: 修改

反思观察
（标注数据学习）

An Interactive Continual Learning Framework Through Reinforced Learning with Human Feedback

Author1, Author2
Affiliation
Univ
City
{Author1, Author2}email@email

Author3
Affiliation
Univ
City
email@email

Abstract

In this paper, we focus on an interactive lifelong learning paradigm that enables artificial intelligence models to continually acquire new skills through real-time human feedback while preserving previously learned capabilities. Unlike conventional continual learning approaches, it addresses two critical challenges: (1) dynamic model updates using streaming data rather than pre-collected static datasets, and (2) robust learning from potentially noisy human feedback inherent in real-world interactive scenarios. To address these problems, we propose a large language models (LLMs)-based Reinforced interactive Continual Learning (RiCL) framework that learns new skills from real-time human feedback. First, we design a temporal consistency-aware purifier to automatically discriminate between clean and noisy samples in real-time data streams. Then, we propose an interaction-aware direct preference optimization strategy to align model behavior with human intent by reconciling AI-generated and human-provided feedback. Moreover, we introduce a noise-resistant contrastive learning module that captures robust representations by exploiting inherent data relationships without relying on potentially unreliable labels. Extensive experiments on benchmark datasets (FewRel and TACRED) contaminated with real-world noise patterns demonstrate that our RiCL approach substantially outperforms existing combinations of state-of-the-art online continual learning and noisy-label learning methods.

Keywords Interactive Continual Learning · Reinforced Learning · Human Feedback · Data Streaming · Noise

步骤1：讨论

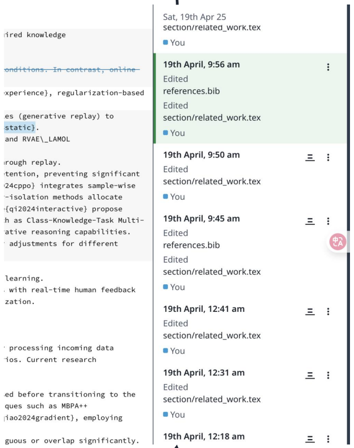
抽象概念化
（学习新能力）

学生

导师

步骤2: 写初稿

具体经验
（规划、思考、主动学习）



步骤3: 提意见

交互反馈

积极实验
（验证有效性）

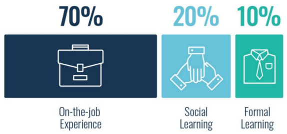
Paper 2 **Paper 3**

少样本

PPT: Good at New Task

人如何学习知识

The 70:20:10 Framework



"**Formal training** is the tip of the **iceberg**. The real learning happens below the surface, through experience and relationships."

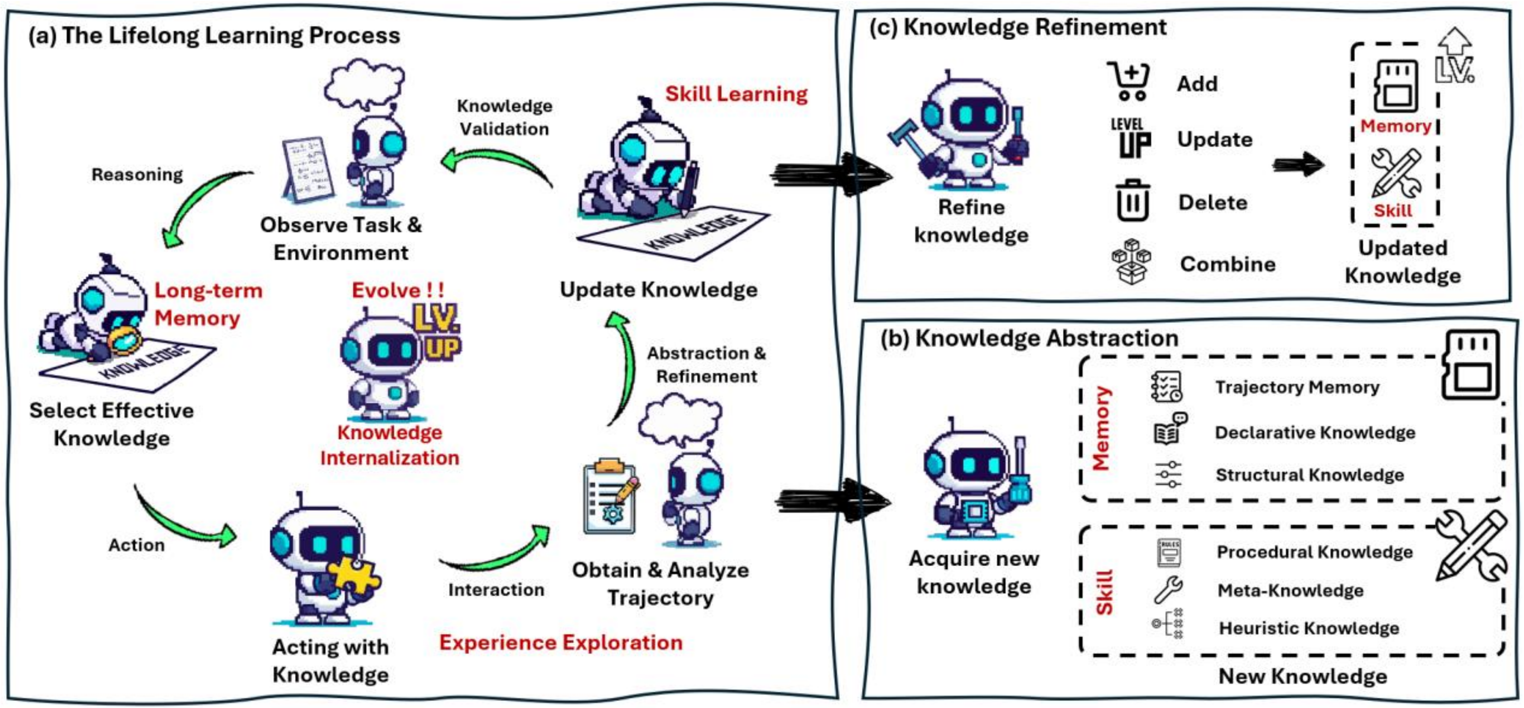
Lombardo, Michael M; Eichinger, Robert W (1996). The Career Architect Development Planner (1st ed.). Minneapolis: Lominger. p. iv. ISBN 0-9655712-1-1.

新任务

学生在自己经历3次论文撰写后掌握大部分的技能，并泛化新的任务（PPT）

经验驱动的终身学习

Github：https://github.com/ECNU-ICALK/ELL-StuLife
Webpage：https://ecnu-icalk.github.io/ELL-StuLife/



经验探索

长期记忆

技能学习

知识内化

- 核心理念：机器应从第一人称视角积累经验并学习。
- 范式转变：从模仿人类知识输出到主动探索和互动。
- 持续进化：通过与环境的持续互动，不断进化和改进。
- 关键机制：持续的经验学习机制，包括探索、记忆、技能和内化。

Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

经验驱动的终身学习 不等于 自进化

Step 1: 交互和轨迹获取（基于历史经验）

$$\xi^{(i,k)} \sim \pi(\cdot|\mathcal{K}^{(i,k-1)})$$

区别1: Task之间是有**顺序关系**的，并非所有task一次性学习，前后会互相影响

Step 2: 知识抽象和重构（可复用）

$$\mathcal{K}^{(i,k)} = \Phi_{\text{learn}}(\mathcal{K}^{(i,k-1)}, \xi^{(i,k)}, g^{(i)})$$

$$\max_{\pi, \Phi_{\text{learn}}} \sum_{i=1}^{N} \mathbb{E}_{\xi^{(i)} \sim \pi(\cdot|\mathcal{K}^{(i)})} \left[\sum_{t=0}^{T_i} R^{(i)}(s_t, a_t, g^{(i)}) \right]$$

Step 3: 知识验证（在后续任务验证有效性）

$$V(\mathcal{K}^{(i-1)}, \mathcal{T}^{(i)}) = J(\mathcal{T}^{(i)}, \pi(\cdot|\mathcal{K}^{(i-1)})) - J(\mathcal{T}^{(i)}, \pi_0)$$

区别2: **第i个task之前**所学到的knowledge，包括memory和skill两个部分

Step 4: 技能内化：将思考变成直觉

Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

StuLife-把AI送去上大学

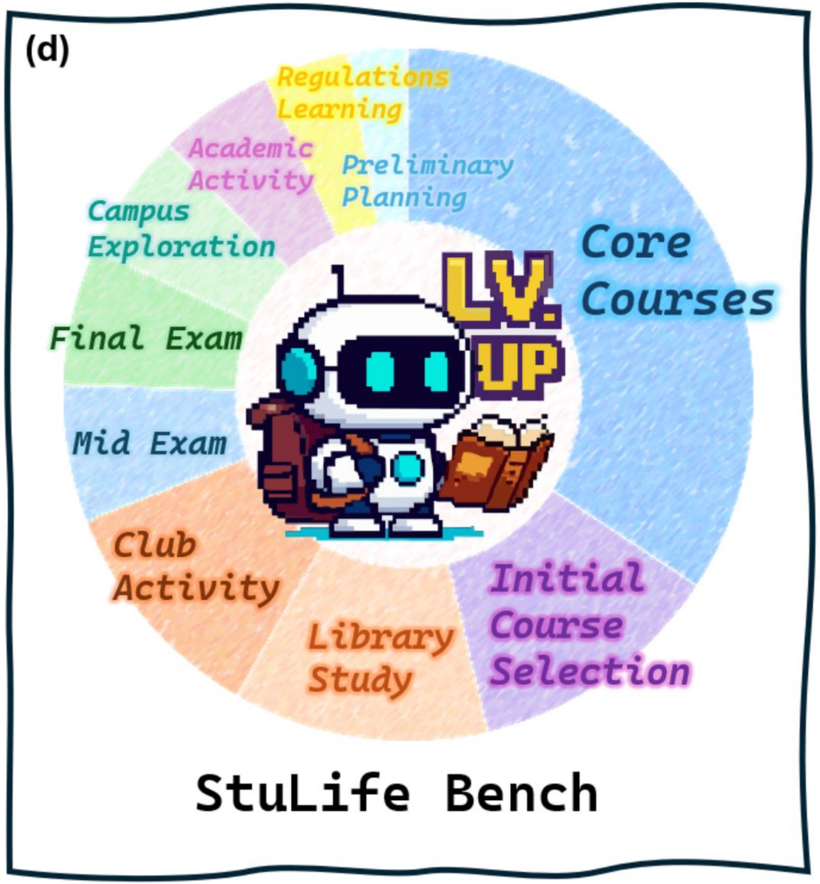
StuLife

大学生活模拟环境

场景: 模拟学生从入学到学术和个人发展的完整大学旅程。

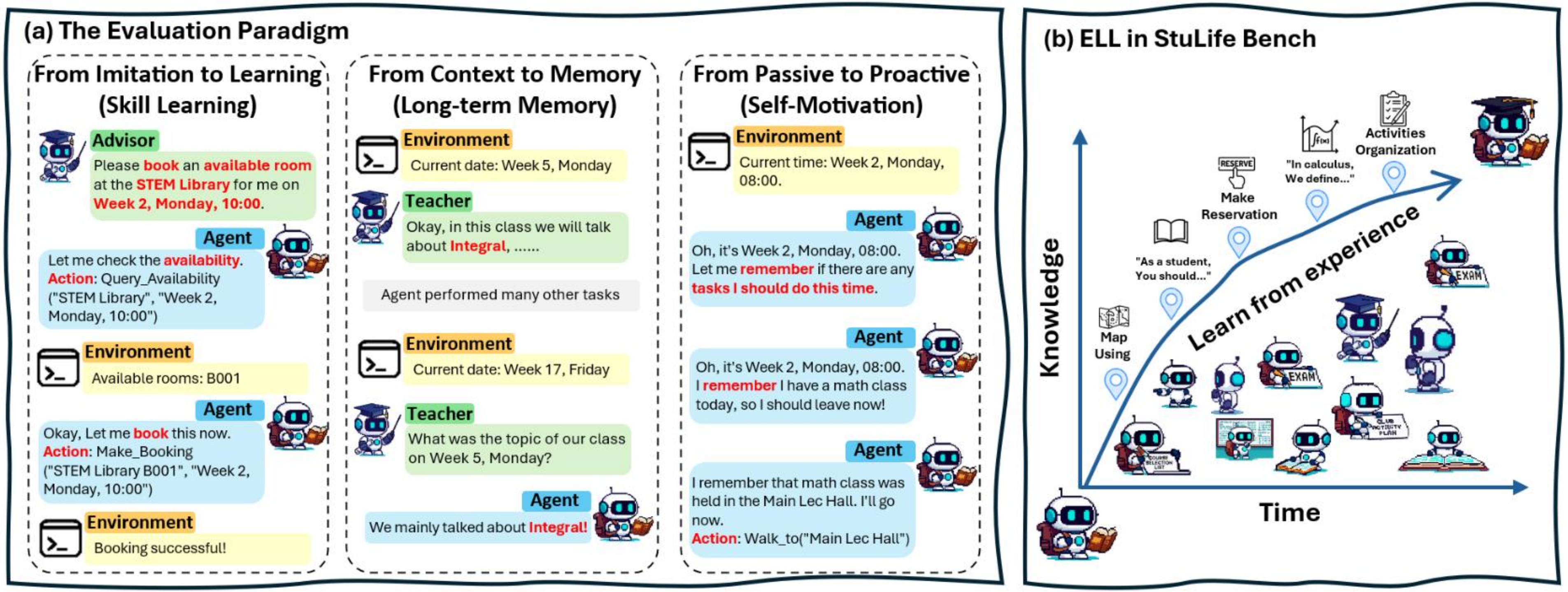
特点:

- □ 三个核心阶段（课堂内、日常校园、考试）
- □ 动态环境，状态变量随时间演变
  - □ 课堂内任务：规章制度学习；核心课程教学
  - □ 日常校园任务：校园探索、初步选课；初步规划；学术活动；图书馆资源管理；学生社团参与
  - □ 考试任务：期中考试；期末考试



Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

StuLife-把AI送去上大学



从模仿到学习　　从上下文到记忆　　从被动到主动

Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

StuLife-数据构建和介绍

Core Scenarios	Interconnected Scenarios	#Num	#Avg Len	#Max Len	#LTM	#Self-Motivat
In-Class	Regulations Learning	70	9125	9969	23	70
	Core Course Instruction	416	9203	10368	129	416
	Total	486	9191	10368	152	486
Daily Campus	Campus Exploration	75	2921	3006	25	25
	Initial Course Selection	150	3136	3420	50	0
	Preliminary Planning	50	3069	3133	50	0
	Academic Activity	72	3193	3466	22	22
	Library Study	151	2080	3068	50	50
	Club Activity	140	2981	3124	45	45
	Total	637	2883	3466	242	142
Examination	Midterm Exams	80	3264	3520	80	0
	Final Exams	80	3507	3686	80	0
	Total	160	3386	3686	160	0
Total	Total	1284	5792	10368	554	628

Datasets	Task Type	Seq	SkilL	LTM	SelfMotivat	Real	Interconnected	Interact	LfE
Lifelong-CIFAR10	CL	✓	✗	✗	✗	✗	✗	✗	✗
Lifelong-ImageNet	CL	✓	✗	✗	✗	✗	✗	✗	✗
CGLB	CL	✓	✗	✗	✗	✗	✗	✗	✗
EgoThink	Embodied AI	✗	✗	✗	✗	✓	✗	✗	✗
EmbodiedBench	Embodied AI	✗	✗	✗	✗	✓	✗	✓	✓
AgentBench	Agent	✗	✗	✗	✗	✗	✗	✓	✗
LoCoMo	Agent	✗	✗	✓	✗	✗	✗	✗	✗
StoryBench	Agent	✓	✗	✓	✗	✗	✗	✓	✗
LifelongAgentBench	Self-Evolving	✓	✓	✗	✗	✗	✗	✓	✓
StuLife (Our)	ELL	✓	✓	✓	✓	✓	✓	✓	✓

序列任务：任务之间前后互相影响

超长上下文：无法通过一次性输入

主动意识：依赖指令被动执行（缺乏内驱力）

现有大模型是否具有持续学习能力？-顶尖AI集体“挂科”

	StuGPA	LTRR	PIS	In-Class		Daily Campus		Exam		Total	
				Success	AvgTurn	Success	AvgTurn	Success	AvgTurn	Success	AvgTurn
Llama-3.1-8B	5.81	3.30	0.90	0.90	61.34	0.00	35.91	10.63	28.46	2.13	44.62
Qwen3-8B	13.31	4.33	0.54	0.90	10.12	8.31	**10.25**	14.38	6.31	6.71	9.71
Qwen3-30B-A3B	16.30	5.05	0.72	0.60	9.45	10.79	11.75	17.50	5.46	8.31	10.09
Qwen3-32B	7.36	3.97	0.54	0.60	7.80	2.25	13.79	13.13	4.88	3.51	10.41
Qwen3-32B	12.67	5.42	1.26	1.80	8.31	7.64	10.74	17.50	4.94	7.24	9.10
QwQ-32B	13.21	5.78	3.42	4.79	7.72	6.97	13.25	16.88	4.52	7.88	10.06
DeepSeek-V3	11.22	6.14	2.88	3.59	**5.84**	6.74	11.87	16.25	**4.26**	7.24	**8.64**
DeepSeek-R1	14.25	8.30	3.96	5.09	8.04	13.26	13.02	18.13	4.56	11.18	10.08
DeepSeek-V3.1	14.26	4.51	0.54	0.90	14.03	12.81	12.62	15.00	6.78	8.95	12.43
DeepSeek-V3.1	17.04	6.14	3.78	6.29	9.83	12.58	13.03	17.50	5.54	11.18	10.88
Qwen3-235B-A22B	16.03	5.42	1.80	2.10	18.71	10.34	17.17	16.88	10.75	8.52	16.95
Gemini-2.5-Pro	16.43	7.04	3.24	5.39	14.94	18.88	12.78	15.63	9.51	13.53	13.19
Grok4	17.38	**10.65**	4.50	4.79	6.31	**21.80**	11.25	**18.75**	5.69	**15.23**	8.68
GPT-5	**17.90**	6.50	**4.68**	**7.78**	12.70	14.16	14.31	16.88	6.24	12.35	12.69

- **StuGPA** = 20% Daily + 30% In-Class + 50% Exam
- LTRR：长期记忆任务上效果
- PIS：主动任务上效果

- 即使是最强的模型 (GPT-5) 在 StuLife 上得分仅为 17.9/100
- 揭示了当前AI与人类水平自主学习之间存在巨大鸿沟。
- 现有模型在长期记忆保留和自我驱动行为方面存在根本缺陷。

Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

上下文工程 能否实现 AGI

		StuGPA	LTRR	PIS	In-Class Success	In-Class AvgTurn	Daily Campus Success	Daily Campus AvgTurn	Exam Success	Exam AvgTurn	Total Success	Total AvgTurn
Vanilla Qwen3-235B-A22B		16.03	5.42	1.80	2.10	18.71	10.34	17.17	16.88	10.75	8.52	16.95
Proactive Qwen3-235B-A22B		16.90	5.96	3.06	5.09	16.70	10.34	16.38	16.88	7.73	9.58	15.42
Skill Qwen3-235B-A22B		17.28	6.86	0.90	1.50	16.89	15.28	16.51	17.50	9.28	10.76	15.75
Memory Qwen3-235B-A22B	+ Vanilla RAG	10.98	4.69	0.18	0.00	17.87	5.84	14.20	16.25	10.04	5.54	15.07
	+ Graph RAG	15.34	4.87	0.72	0.90	20.68	10.11	14.03	16.25	10.61	7.88	16.13
	+ MemGPT	19.99	6.86	1.44	2.40	17.28	13.03	13.59	23.75	9.02	11.08	14.42
	+ MemoryBank	17.64	5.96	1.62	0.90	16.68	12.36	14.15	20.00	8.04	9.58	14.35
All-in-One		21.07	9.39	3.76	2.69	16.82	17.75	15.65	25.63	6.30	13.74	14.93

- 充分设计的上下文工程可以显著提高模型的性能(**16.03 -> 21.07**)
  - 利用Proactive提示词提高模型主动性（**PIS 1.80->3.06**）
  - 利用Skill提示词提高模型技能使用能力（Daily Campus **10.34-> 15.28**）
  - Memory机制对模型性能影响很大（Exam **16.88 -> 23.75**）

Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

经验驱动终身学习挑战

- 挑战1: 高效探索与经验获取
  - 设计内在动机机制引导的交互
  - 奖励往往稀疏、延迟或难以定义
- 挑战2: 长期记忆与关联回忆
  - 如何构建可扩展、易访问的长期记忆?
  - 如何实现跨事件的关联回忆，克服灾难性遗忘?
- 挑战3: 技能抽象与管理
  - 如何动态管理技能生命周期（获取、验证、调用、演化）？
  - 如何将显式规则知识转化为直觉、泛化的内化能力?

挑战一：经验探索和奖励获取

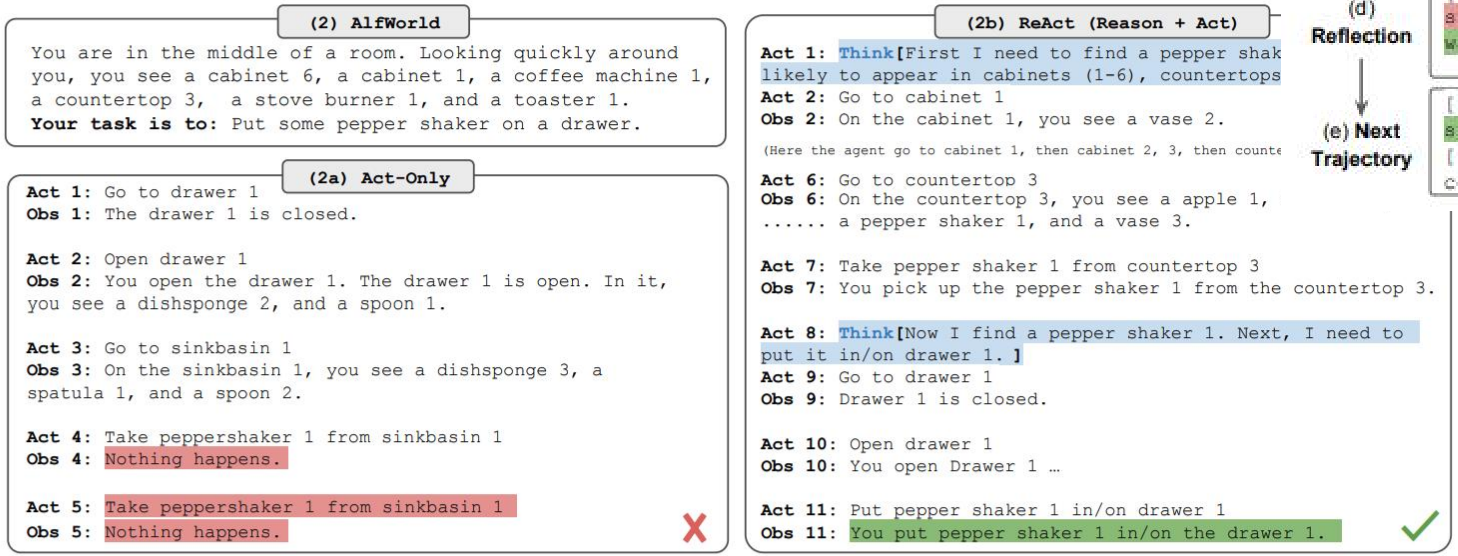
- 高效探索与经验获取
  - 设计内在动机机制引导的交互
  - 奖励往往稀疏、延迟或难以定义

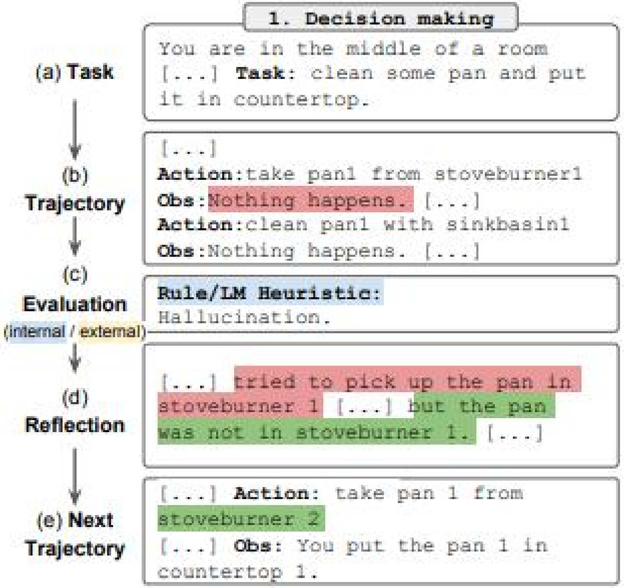
- 现有方法
  - 基于反思的经验探索（一边做一边思考）
  - 基于强化学习的探索（暴力搜索）
  - 基于人在回路的经验探索（交互中学习）
  - 基于多智能体的经验探索（众人拾柴火焰高）

- 如何少样本高效搜索路径？
- 如何更好和人类行为对齐？
- 通过Social进行学习？

一边反思一边探索

- 语言本身可以作为学习信号
- 反馈的语言化能一定程度上弥补 reward 稀疏问题

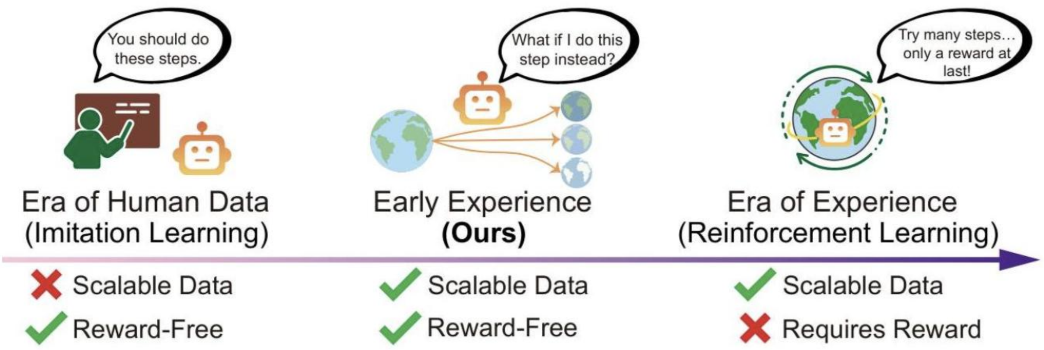


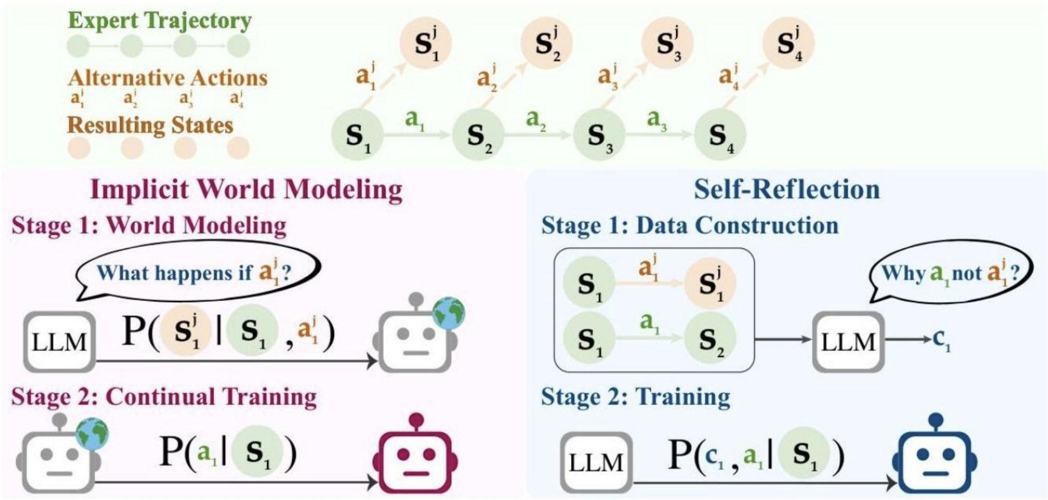


ReAct: Synergizing Reasoning and Acting in Language Models, ICLR, 2023

Reflexion：Language Agents with Verbal Reinforcement Learning，NeurIPS 2023

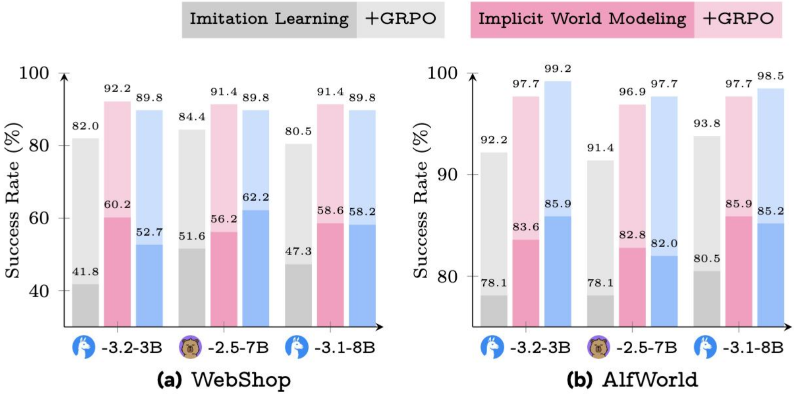
无法定义reward怎么办？早期经验学习



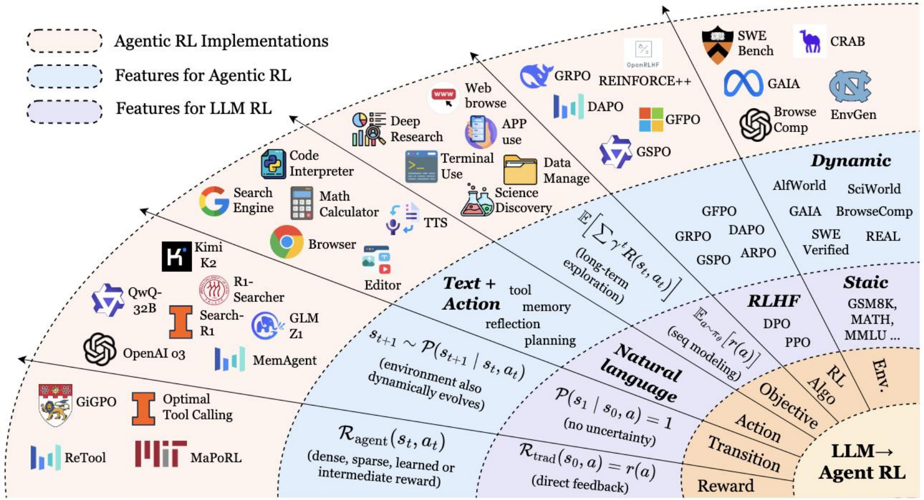


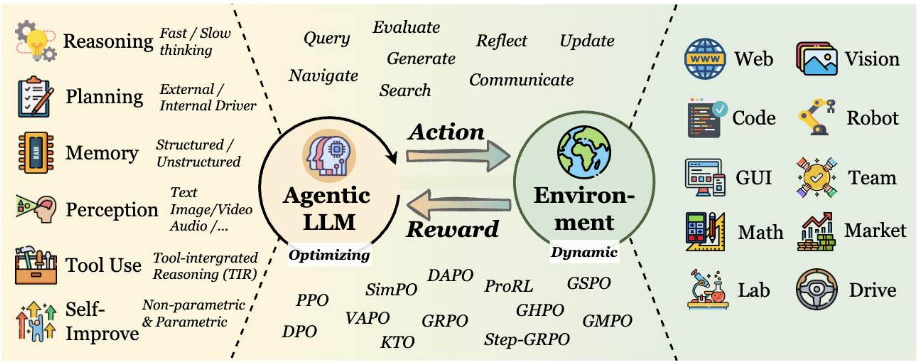
Agent Learning via Early Experience, 2025, Meta

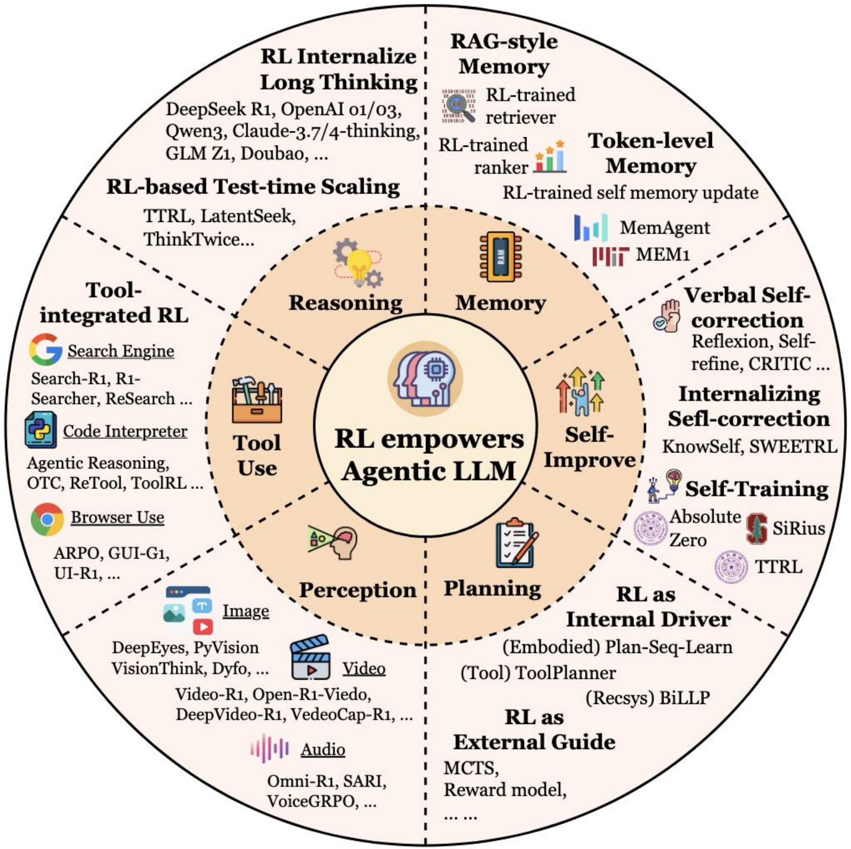
- 微调缺乏扩展性、强化学习reward稀疏，如何找一个中间方法？
- Implicit World Modeling (IWM)：通过预测未来状态来构建环境动态的内部表征
- Self-Reflection (SR)：智能体比较专家行动与自己的替代方案，并生成思维链解释，阐述为什么专家选择更优。



基于强化学习的经验探索





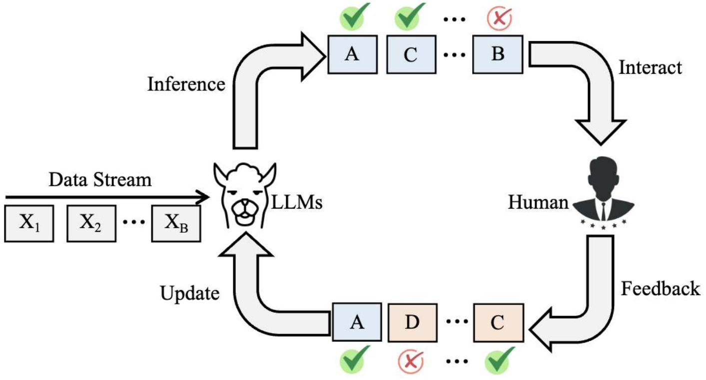


暴力多次搜索，差中选优

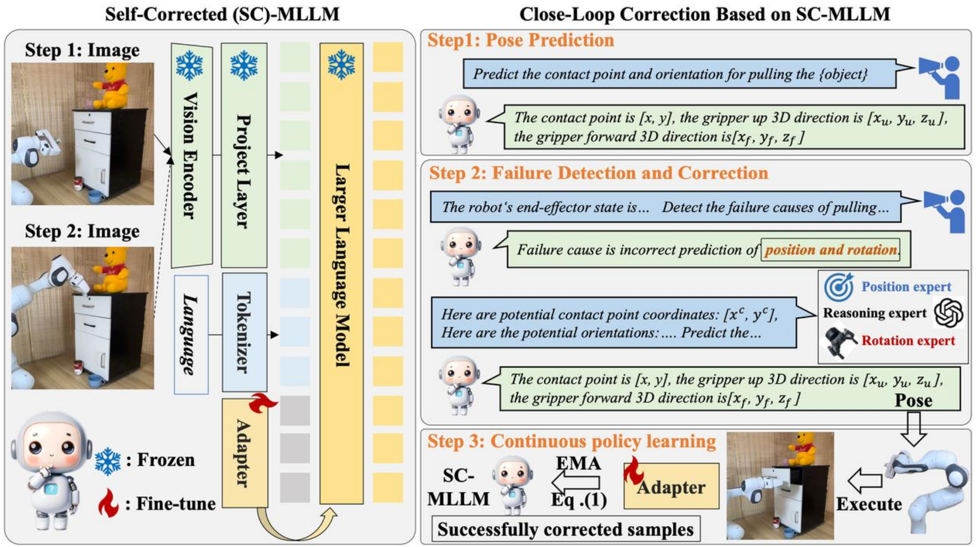
The Landscape of Agentic Reinforcement Learning for LLMs: A Survey, 2025

基于人反馈的经验探索

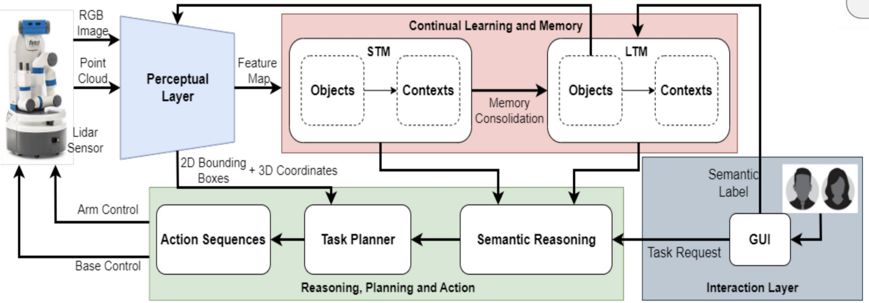
如何从人反馈中学习？



提出交互式持续学习框架，实时学习用户反馈



缺乏从纠正反馈中学习的能力



个性化家庭服务机器人，

用户实时反馈

犯一次错是无知，
犯两次错是愚蠢。

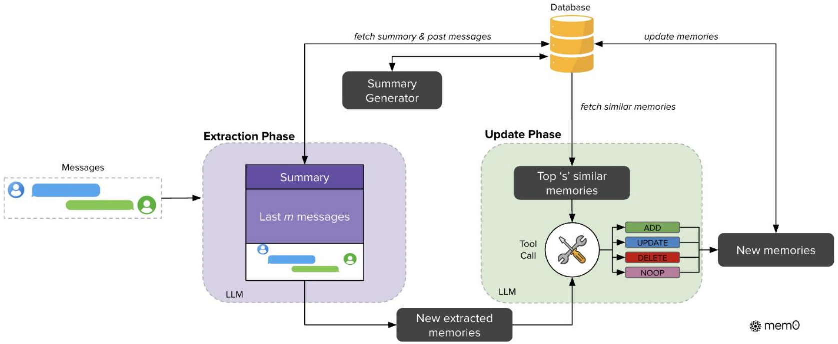
Yutao Yang, **Jie Zhou*** et al. AN INTERACTIVE CONTINUAL LEARNING FRAMEWORK THROUGH REINFORCED LEARNING WITH HUMAN FEEDBACK, 2025
Interactive Continual Learning Architecture for Long-Term Personalization of Home Service Robots, ICRA, 2024

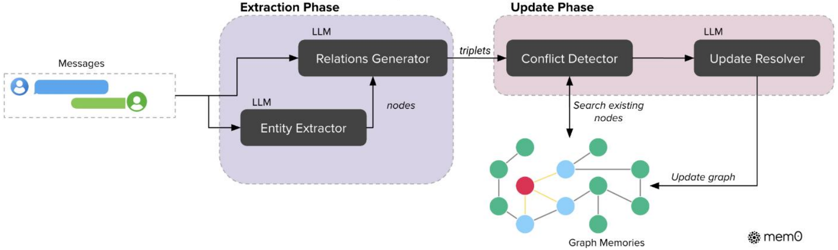
挑战二：长期记忆

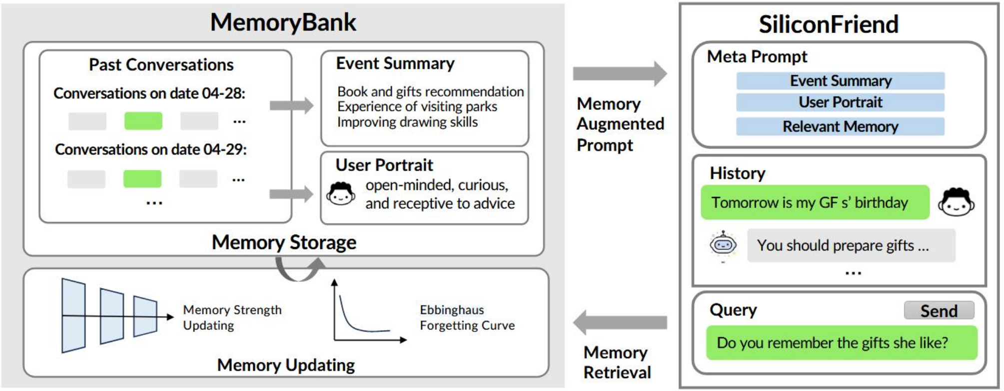
- 长期记忆与关联回忆
  - 如何构建可扩展、易访问的长期记忆系统？
  - 如何实现跨事件的关联回忆，克服灾难性遗忘？
- 现有方法
  - 上下文工程：将记忆以文本等形式存储
    - 外部知识库（RAG）
    - 提示词优化（上下文自我进化）
  - 参数化记忆：将记忆转化为参数化记忆
  - 混合记忆：文本形式记忆+参数化记忆管理

- 如何存储记忆？文本or参数？
- 如何检索记忆？
- 如何使用记忆？

外部知识库-长期记忆建模







□ 引入艾宾浩斯遗忘曲线：提出了MemoryBank长短期记忆融合机制，为LLM提供类似人类的长期记忆模块。

$$R = e^{-\frac{t}{S}}$$

S：记忆强度，随使用次数增加而增强

t：距离上次使用记忆的时间

R：记忆的保留率

随着使用时间的增大，记忆会逐渐衰弱直至消失，像人类的记忆一样

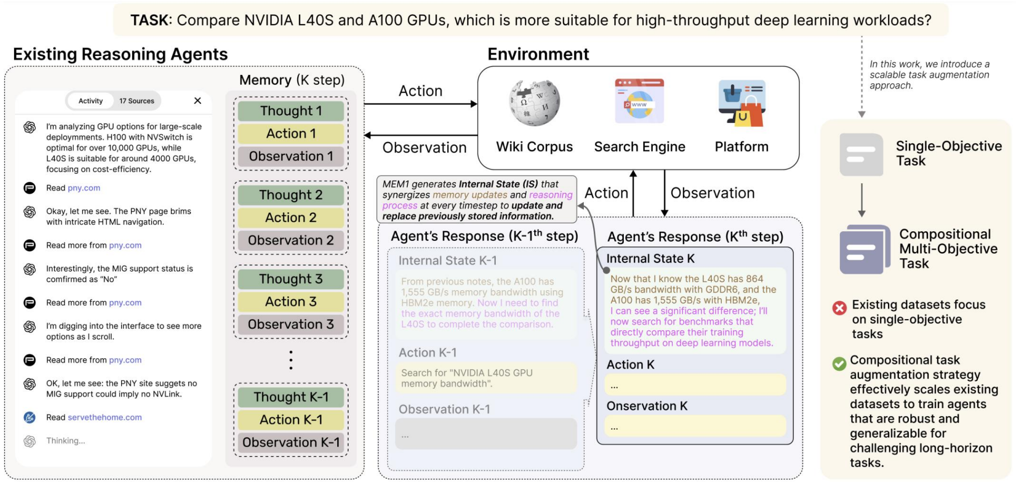
□ **Mem0 架构**：可扩展的记忆架构，自动提取、整合和检索对话中的关键信息。

□ **Mem0g**：进一步引入图结构的记忆表示，用节点和边来表示对话元素及其关系

MemoryBank: Enhancing Large Language Models with Long-Term Memory AAAI-24
Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory 2025

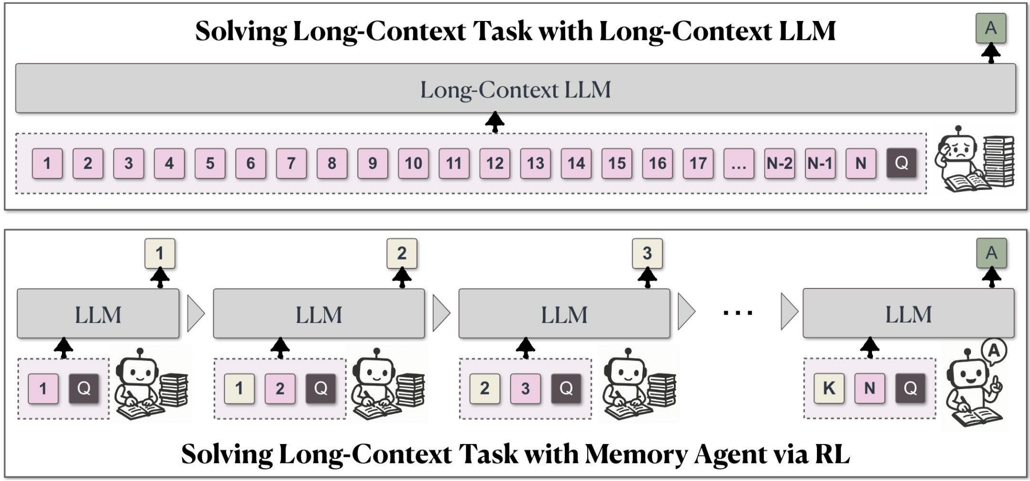
可学习Prompt-记忆压缩

- ☐ 整合推理和记忆到一个紧凑的内部状态，实现了跨长期、多轮任务的**恒定内存**使用



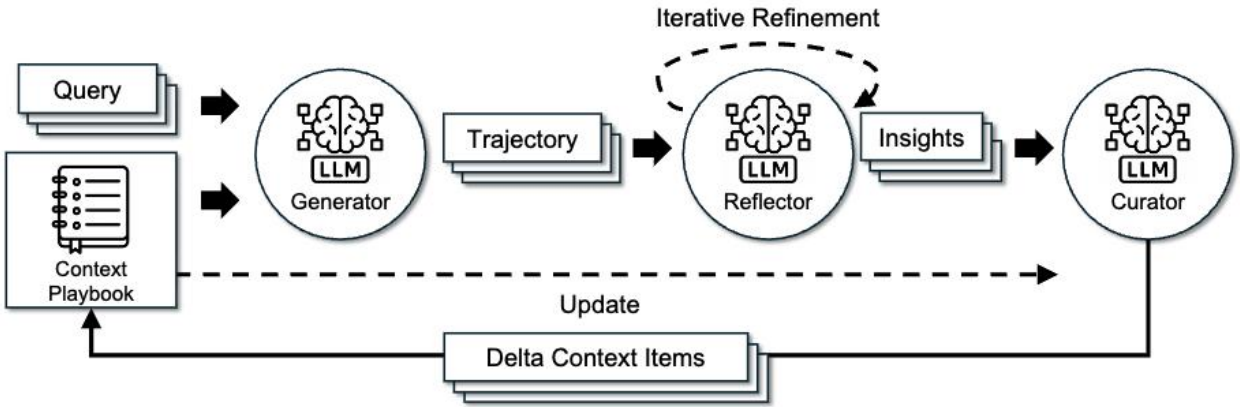
通过强化学习选择主动记住哪些与回想哪些记忆？

- ☐ 长程任务中上下文冗余、计算成本急剧上升，推理效率低下。通过主动筛选重要信息
- ☐ 并**遗忘无关**信息，有助于显著提升效率并保持记忆精炼
- ☐ **主动记忆机制**：发展主动记忆管理，实现信息的高效筛选、存储与遗忘

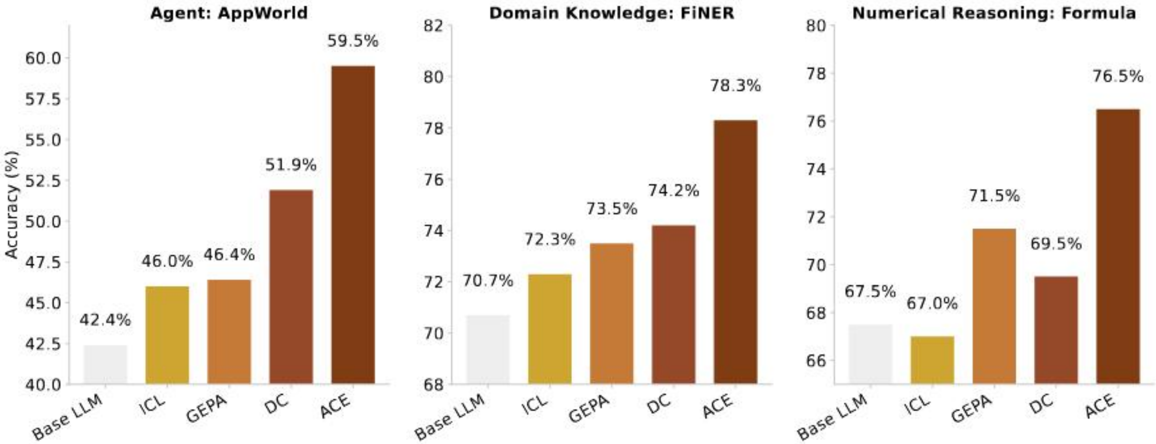


Zhou, Zijian, et al. "MEM1: Learning to Synergize Memory and Reasoning for Efficient Long-Horizon Agents." arXiv preprint arXiv:2506.15841 (2025).

上下文工程能否实现大模型记忆？自进化上下文



- 自主上下文工程（ACE）将视角从将上下文视为静态、简洁的提示转变为将它们视为"全面、演变的操作手册"。



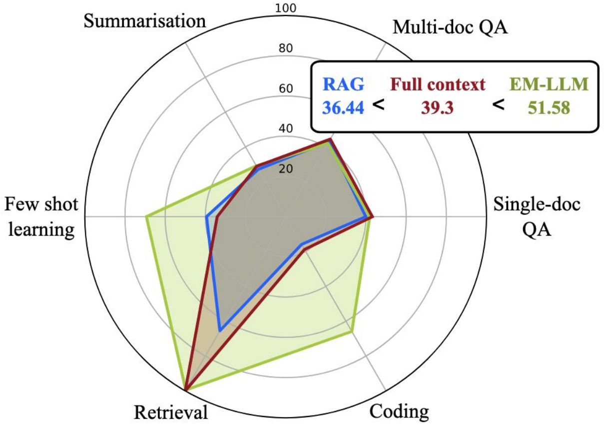
- Reflector：从成功和失败中提炼具体见解

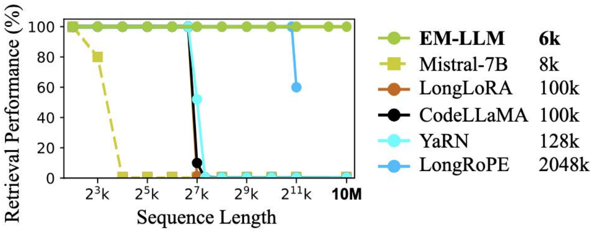
- Curator：将见解综合为紧凑的"增量上下文项"

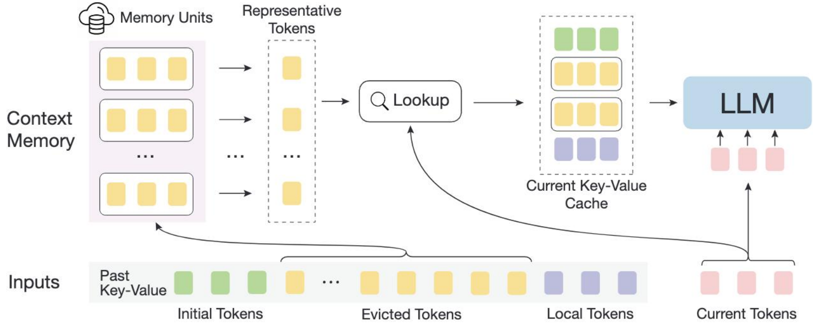
Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models, 2025l, Stanford

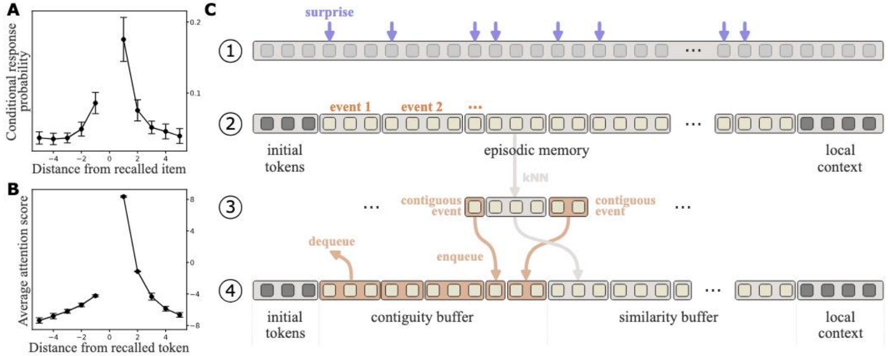
参数化记忆-KV存储（本质上还是RAG）

- 参数化检索：InfLLM 提出了一种无需训练的记忆方法
- 记忆切片：EM-LLM 将人类情景记忆和事件认知整合到大型语言模型中



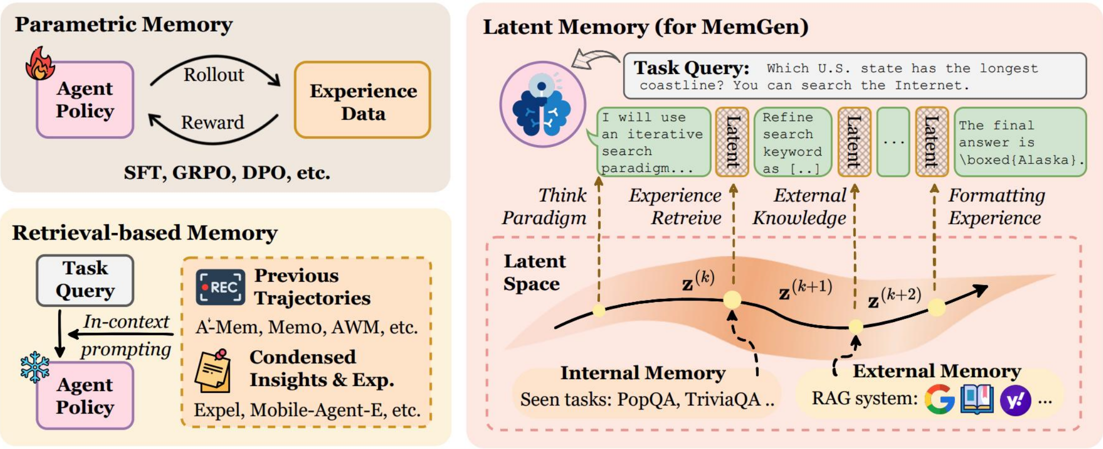


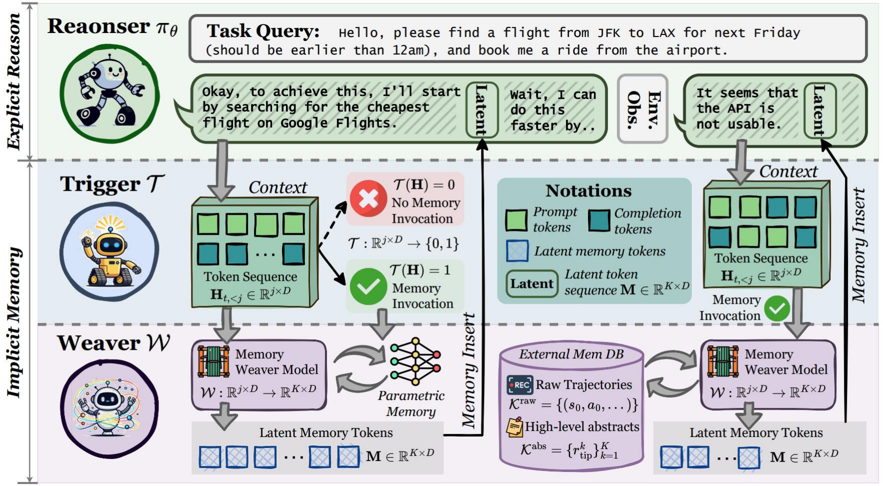




InfLLM: Training-Free Long-Context Extrapolation for LLMs with an Efficient Context Memory, NeurIPS, 2025
HUMAN-LIKE EPISODIC MEMORY FOR INFINITE CONTEXT LLMS, ICLR, 2025

可学习参数化记忆-存什么？如何读？





区别于参数化和基于检索的记忆，关注记忆的融会贯通，通过最终生成效果进行强化训练

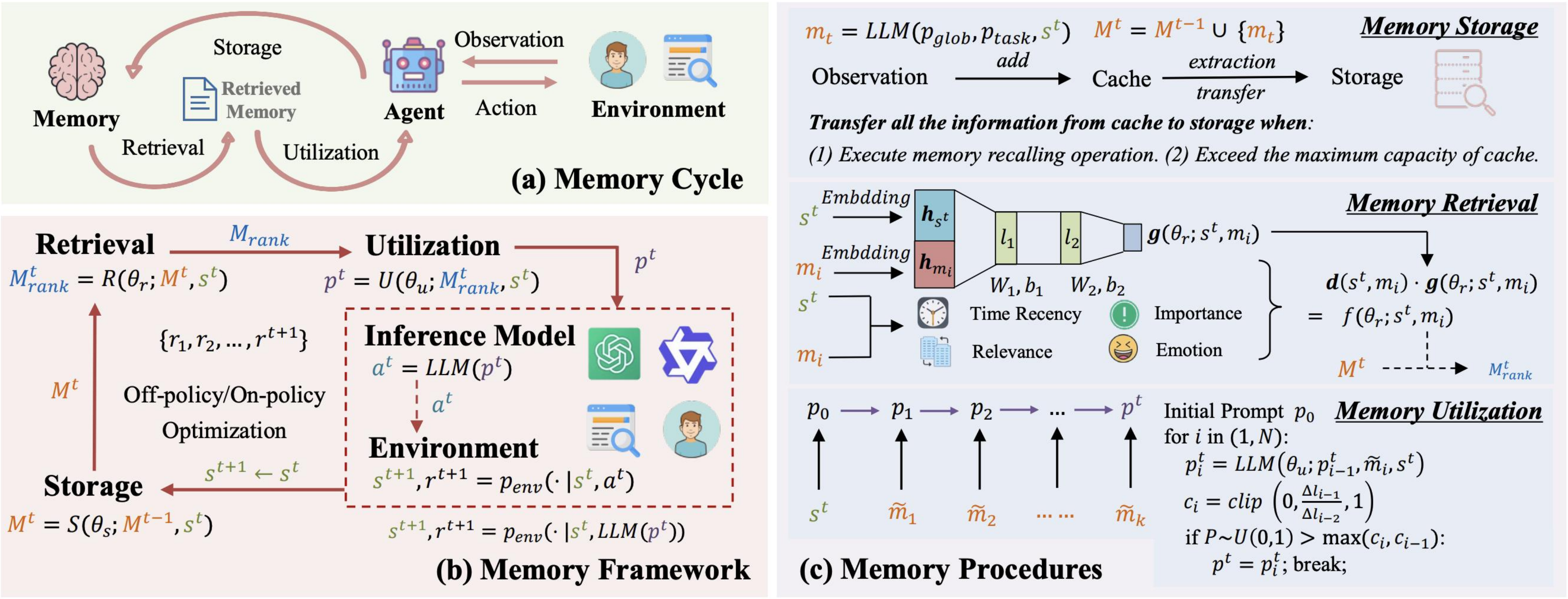
- **Reasoner**：推理模型，用于生成回复，过程中会输入隐式记忆
- **Trigger**：触发检索机制，当遇到标点符号，判断是否需要检索
- **Weaver**：记忆生成机制，生成K个隐式token表示作为记忆。

Backbone	Method	ALFWorld	TrivialQA	PopQA	KodCode	BigCodeBench	GPQA	GSM8K	MATH
Qwen3-8B	Vanilla	58.93	52.18	34.13	49.10	33.33	38.18	89.48	79.82
	CoT	57.10	53.80	33.20	51.25	35.59	35.15	87.67	78.24
	SFT	83.59	74.55	51.12	64.75	41.33	40.33	90.76	81.35
	GRPO	85.60	76.15	58.90	73.35	70.24	39.54	92.30	83.54
	REINFORCE	82.10	75.22	57.96	72.11	70.20	37.12	91.25	83.27
	REINFORCE++	84.80	75.90	58.30	72.90	71.88	37.68	91.90	85.24
	Agent-FLAN	80.32	70.32	50.08	62.99	43.40	39.50	87.60	80.05
	ExpeL	78.97	65.54	40.33	57.20	34.23	35.15	86.20	77.40
	MemoryBank	70.41	60.56	41.60	56.39	40.61	35.66	90.35	80.35
	AWM	80.33	69.30	43.69	-	-	-	-	-
	SoftCoT	75.60	59.42	39.42	63.28	38.27	39.60	86.30	76.23
	Co-processor	73.28	61.42	45.55	64.90	42.19	39.15	76.23	79.20
	MemGen SFT	85.82	77.22	54.65	66.15	40.35	43.23	91.25	83.30
	MemGen GRPO	90.60	80.65	62.30	76.16	75.56	40.24	93.20	88.24

MemGen: Weaving Generative Latent Memory for Self-Evolving Agents, 2025

混合记忆：可学习参数化+外部知识库

- 学习如何**存储、检索和使用**记忆

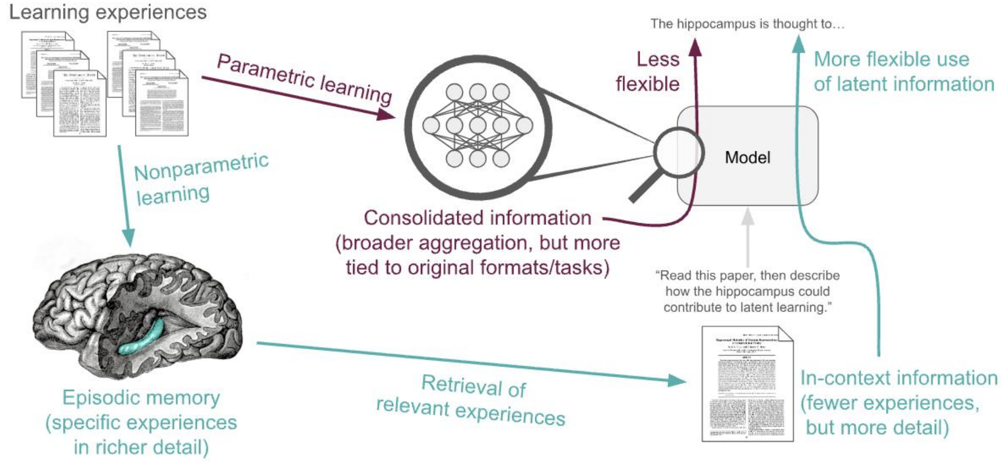


Zhang Z, Dai Q, Li R, et al. Learn to Memorize: Optimizing LLM-based Agents with Adaptive Memory Framework[J]. arXiv preprint arXiv:2508.16629, 2025.

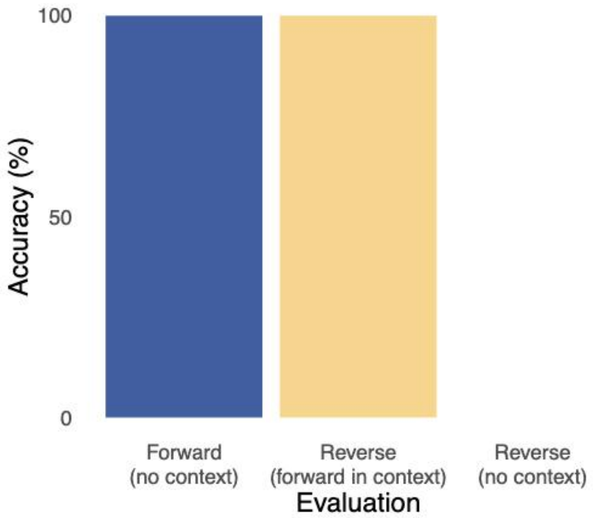
参数化记忆 or 外部知识库?

(a) The reversal curse.

(b) Examples of explicit and latent information or tasks.



Latent Learning：指一个系统学习与当前任务不相关但可能对未来不同任务有用的信息的能力——这也可以被看作是一种"前瞻性学习"的实现方式。



情景记忆能够更灵活地重用过去的经验

参数化学习将信息整合为压缩表示，可能会丢失灵活应用所需的上下文丰富性。

Latent learning: episodic memory complements parametric learning by enabling flexible reuse of experiences, 2025, Google DeepMind

挑战三-技能抽象和管理

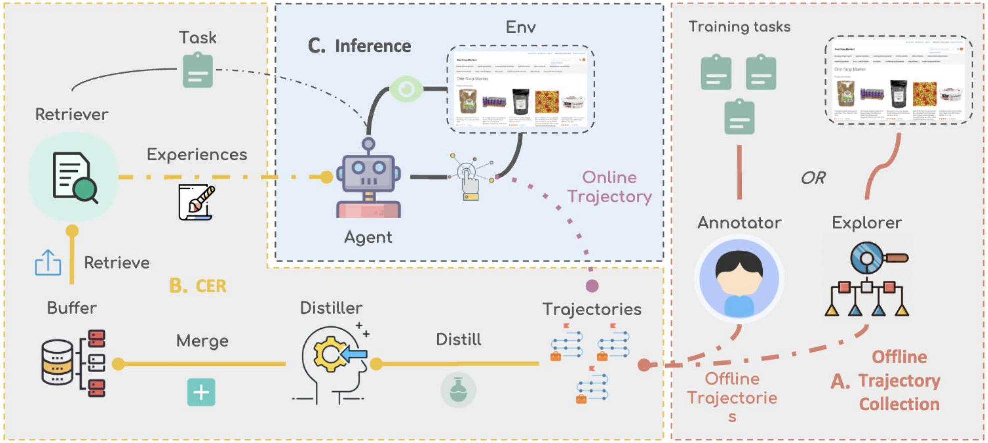
- 技能抽象与管理
  - 如何动态管理技能生命周期（获取、验证、调用、演化）？
  - 如何将显式规则知识转化为直觉、泛化的内化能力？

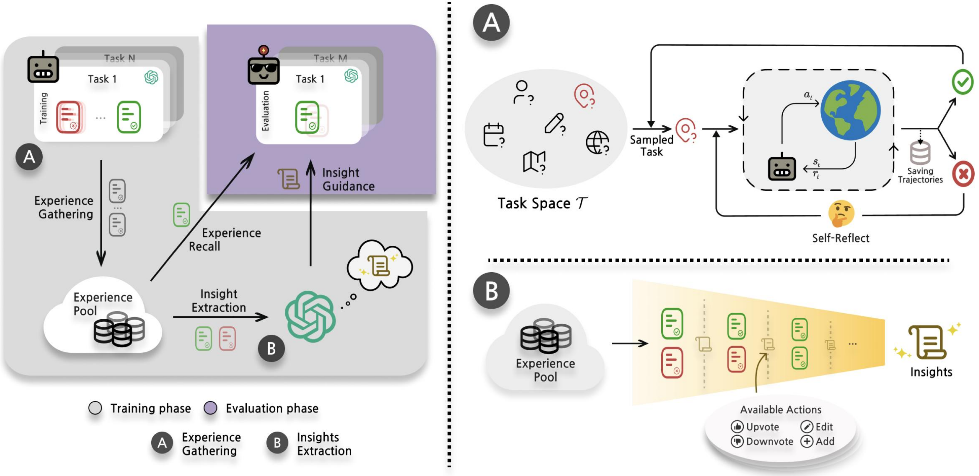
- 现有方法
  - 技能学习（如何抽取可泛化的技能？）
  - 技能验证（如何验证技能有效性？）
  - 技能内化（参数化记忆）

技能学习

- ☐ 从经验中抽取技能

- ☐ **经验收集**：利用反思CoT收集成功/失败的轨迹
- ☐ **从经验中学习**，包括两条支线：
  - ☐ 经验召回：针对query检索历史相似任务的成功经验
  - ☐ 洞察抽取 ：对比同一个任务的成功和失败轨迹，提取insight
- ☐ **任务推理阶段**：在pormpt里添加检索到的历史成功经验以及抽取的所有insight并推理



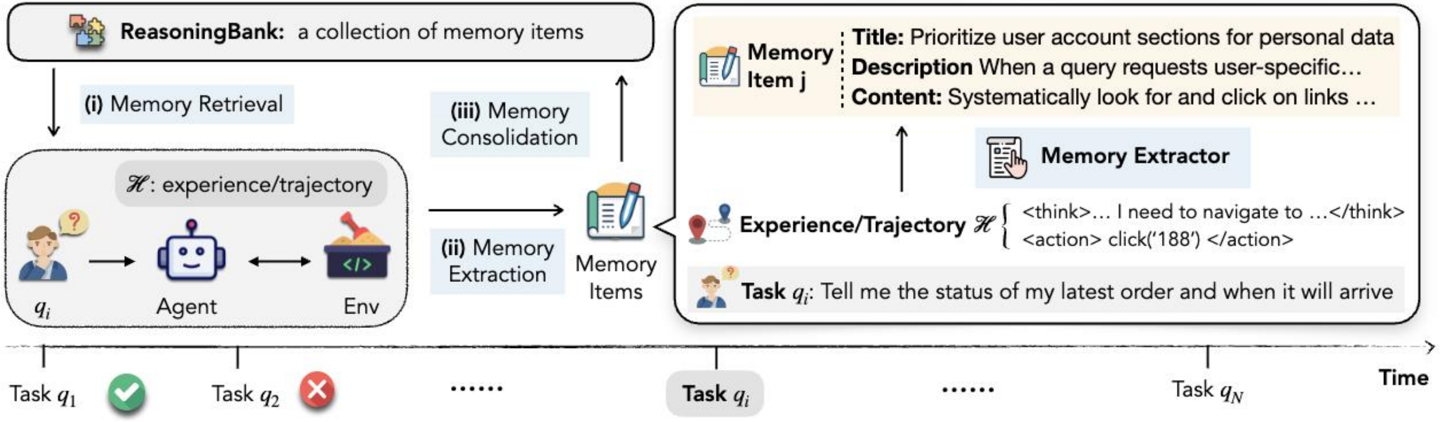


- 成功轨迹示例 → 针对性（依赖相似任务检索）
- insight集合 → 泛化性（全局性的经验总结）

Contextual Experience Replay for Self-Improvement of Language Agents, ACL 2025

技能学习-Reasoningbank

Reasoningbank：从智能体自我评估的成功与失败经验中提炼可迁移的推理策略



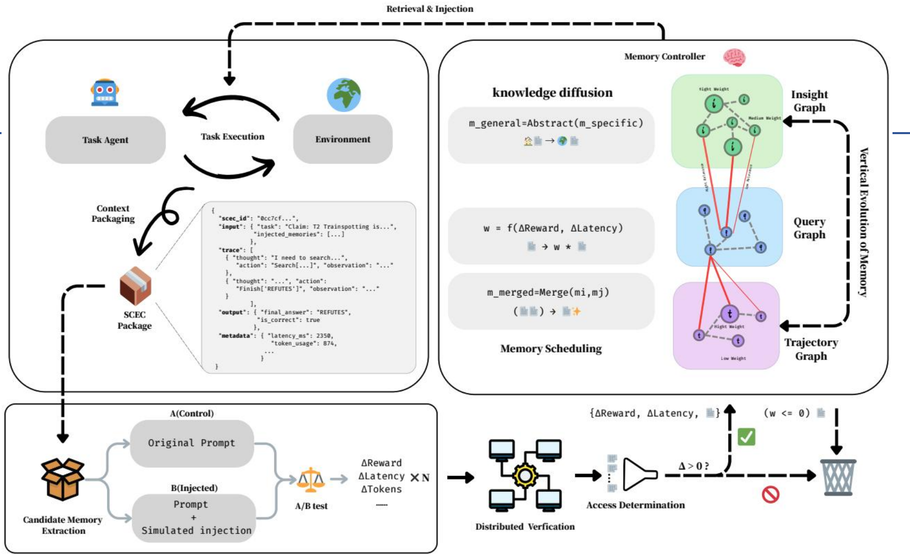
Models	Shopping (187)		Admin (182)		Gitlab (180)		Reddit (106)		Multi (29)		Overall (684)	
	SR	*Step*	*SR*	*Step*	*SR*	*Step*	*SR*	*Step*	*SR*	*Step*	*SR*	*Step*
					Gemini-2.5-flash							
No Memory	39.0	8.2	44.5	9.5	33.9	13.3	55.7	6.7	10.3	10.0	40.5	9.7
Synapse	40.6	7.0	45.1	9.1	35.6	13.0	59.4	6.5	10.3	10.5	42.1	9.2
AWM	44.4	7.0	46.7	8.8	37.2	13.2	62.3	6.1	3.4	**7.7**	44.1	9.0
REASONINGBANK	**49.7**	**6.1**	**51.1**	**8.2**	**40.6**	**12.3**	**67.0**	**5.6**	**13.8**	8.8	**48.8**	**8.3**

- ☐ **推理阶段**，智能体从 ReasoningBank 中检索与当前任务相关的记忆，以辅助决策；
- ☐ **完成任务后**，再将新的学习成果整合回记忆库中，从而实现持续自我增强的学习闭环。

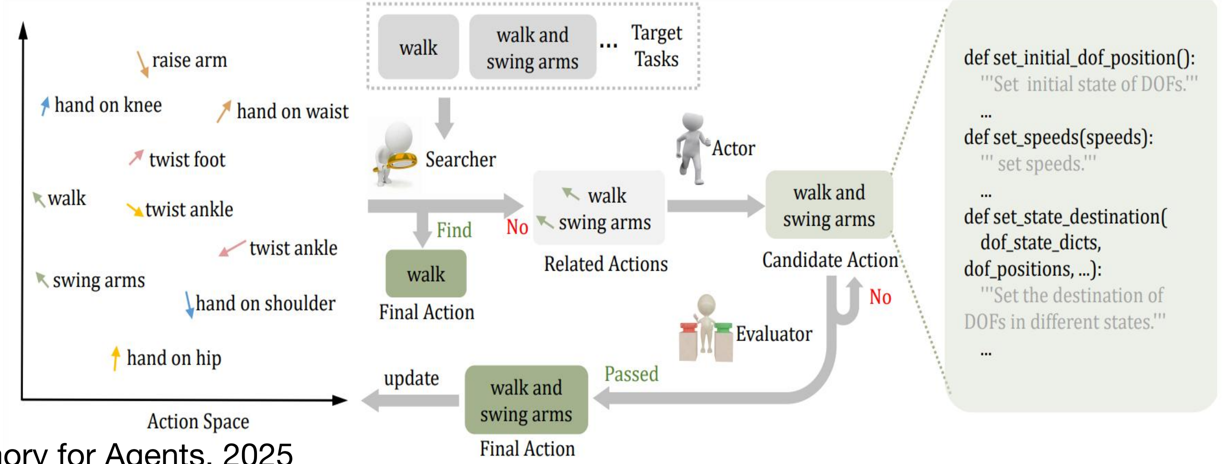
ReasoningBank: Scaling Agent Self-Evolving with Reasoning Memory, 2025, Google

技能验证

- 通过**AB实验**验证技能的有效性
- 自进化记忆控制模块：动态抽取技能、更新技能权重和合并技能

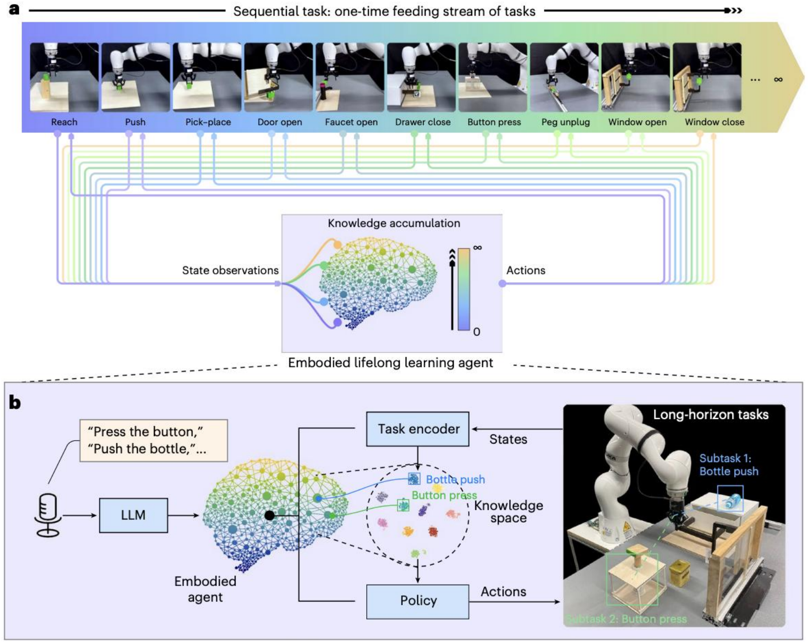


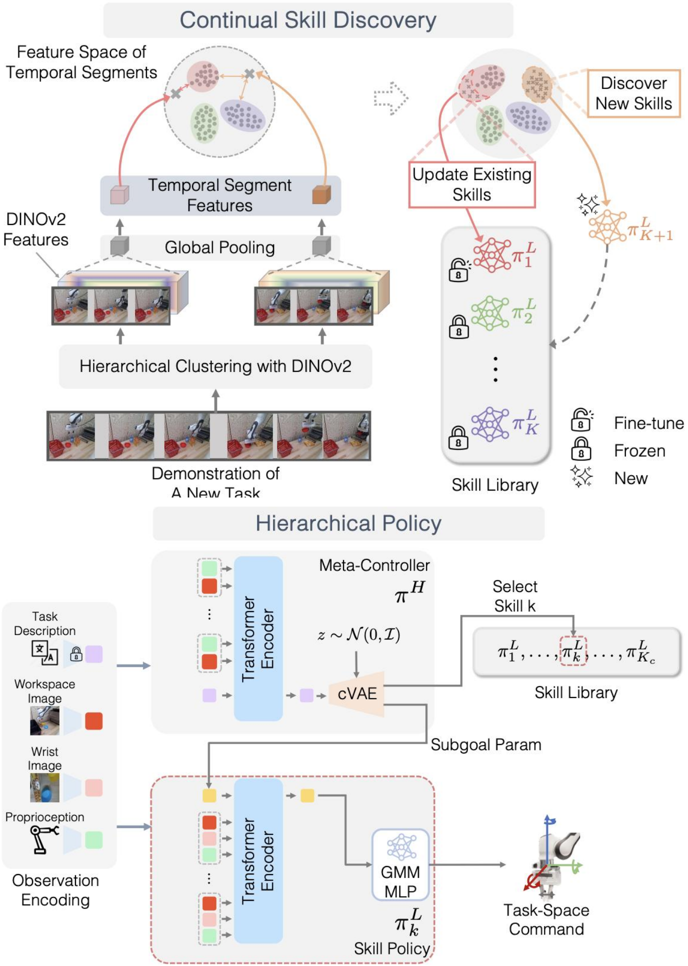
- **Searcher**：
  - 根据目标任务查询Action，从Action库里面搜索
  - 找不到，则新建一个Action，生成Action的动作
- **Evaluator**
  - 通过执行代码评测当前Action是否可行



SEDM: Scalable Self-Evolving Distributed Memory for Agents, 2025
RoboCoder: Robotic Learning from Basic Skills to General Tasks with Large Language Models, 2024

技能内化-参数化技能





- □ 从一系列任务流中持续积累知识
- □ 通过组合和重新应用知识来处理具有挑战性的现实世界长期任务

- □ **持续技能发现：**

  持续发现新技能

- □ **多层策略**

  策略网络预测技能k

Preserving and combining knowledge in robotic lifelong reinforcement learning, Nature Machine Intelligence, 2025
LOTUS: Continual Imitation Learning for Robot Manipulation Through Unsupervised Skill Discovery

思考

目前主要关注智能体自进化，缺乏统一持续学习框架

如何实现AI员工，交互中持续学习？

无明确目标的任务如何学习，内生reward？

训练环境如何构建，人在环路？

谢谢！

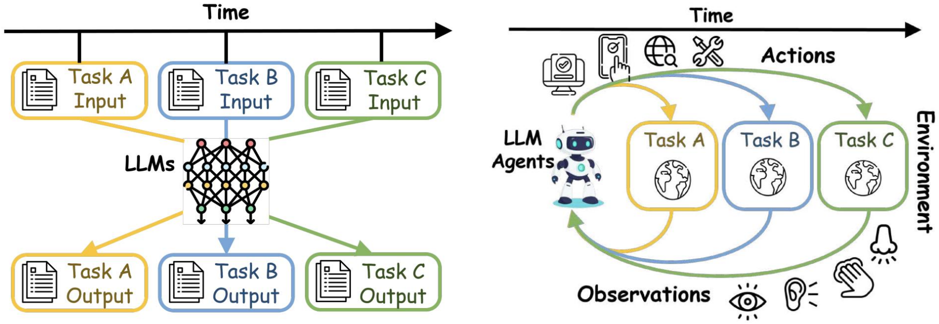
欢迎来华师大 or 上海AI Lab！

经验驱动的终身学习

- 经验探索 (Experience Exploration)
  - 代理通过自我驱动与动态环境互动
  - 解决复杂、长周期任务，生成丰富经验轨迹
- 技能学习（Skill Learning）
  - 从经验中抽象出可重用的技能，并通过应用验证
  - 动态管理技能库（添加、合并、删除等）
- 长期记忆 (Long-term Memory)
  - 持久、结构化的记忆，包括观察、事件、事实、上下文和反思
  - 记忆是主动资源，支持检索、推理和决策
- 知识内化 (Knowledge Internalization)
  - 将显式、离散的知识转化为隐式、直觉的能力
  - 从刻意应用到自动执行，类似新手到专家的转变

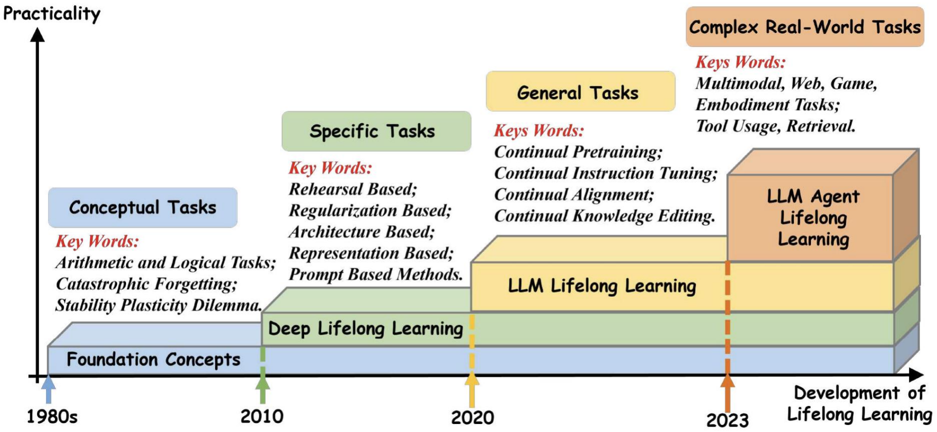
Cai, Yuxuan, Hao, Yipeng, Zhou Jie, et al. "Building Self-Evolving Agents via Experience-Driven Lifelong Learning: A Framework and Benchmark." arXiv preprint arXiv:2508.19005 (2025).

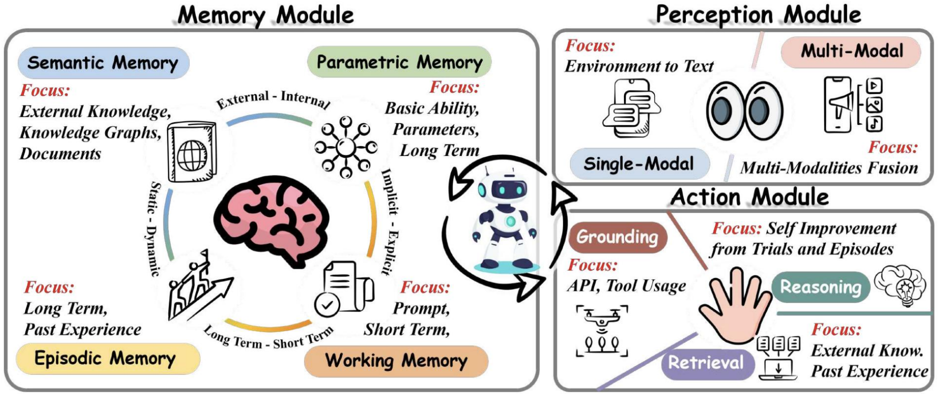
相关综述



(a) Lifelong Learning of LLMs

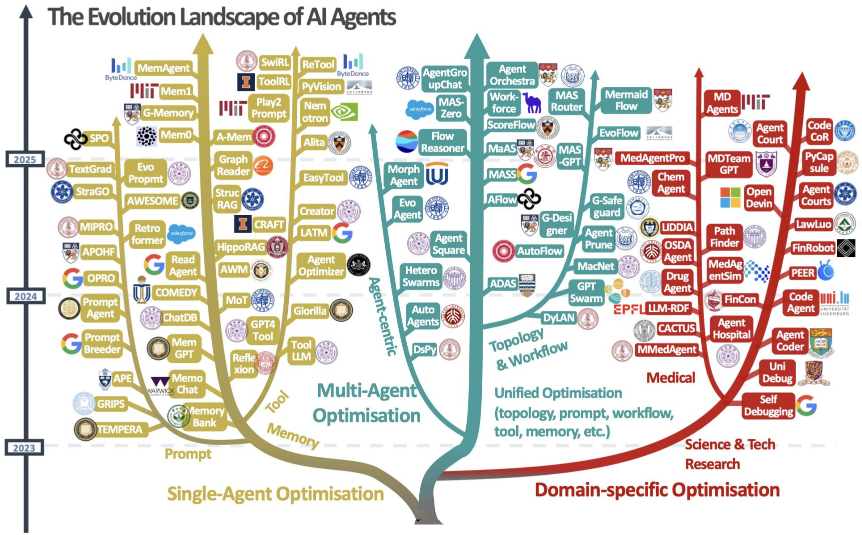
(b) Lifelong Learning of LLM Agents





Lifelong Learning of Large Language Model based Agents: A Roadmap, 2025

相关综述



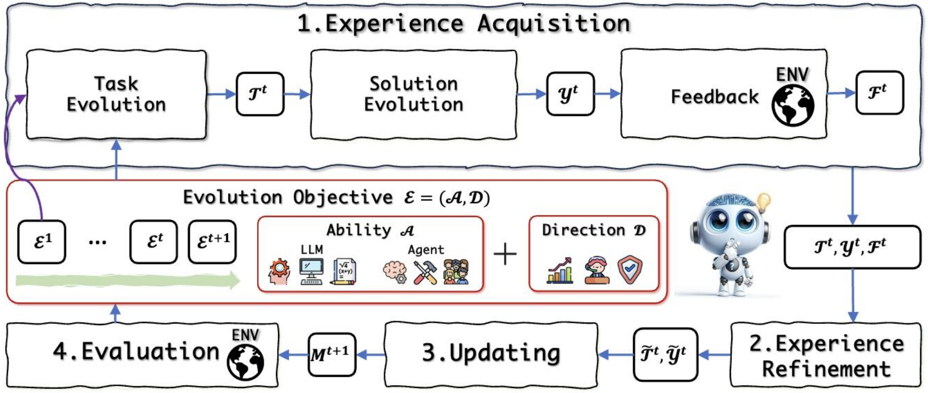
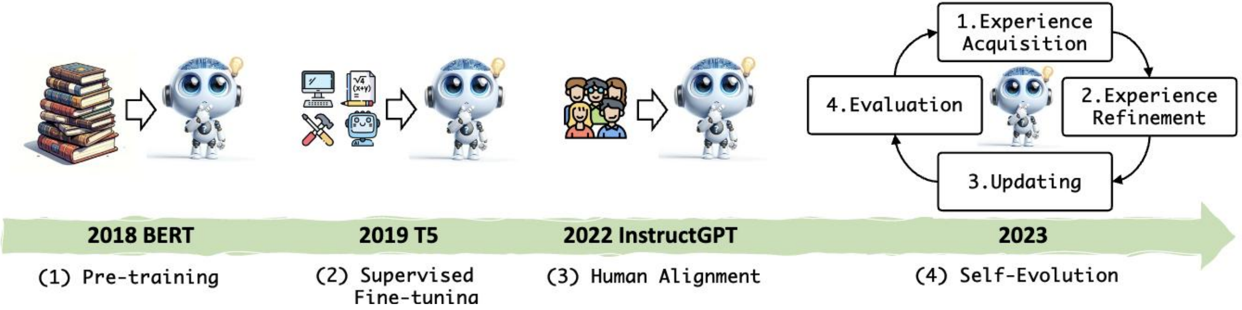


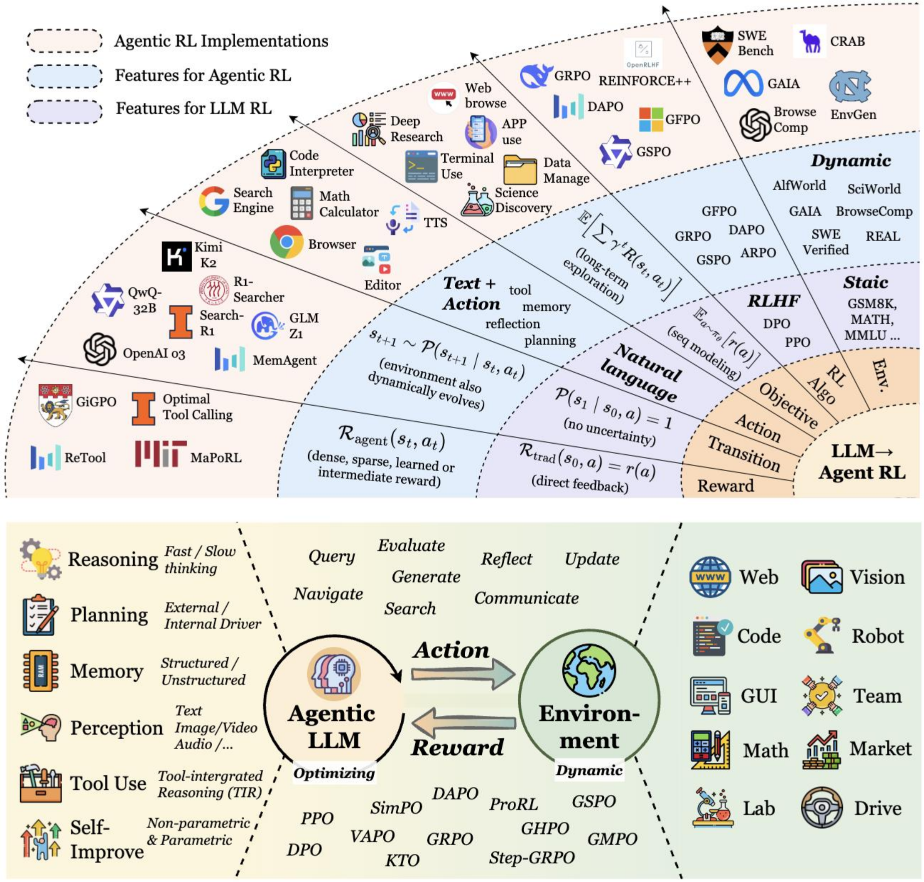
Fig. 2 Conceptual framework of self-evolution. For the t^{th} iteration: $\mathcal{E}^t$ is the evolution objective; $\mathcal{T}^t$ and $\mathcal{Y}^t$ denote the task and solution; $\mathcal{F}^t$ represents feedback; M^t is the current model. Refined experiences are marked as $\tilde{\mathcal{T}}^t$ and $\tilde{\mathcal{Y}}^t$, leading to the evolved model $\tilde{M}$. ENV is the environment. The whole self-evolution starts at $\mathcal{E}^1$.

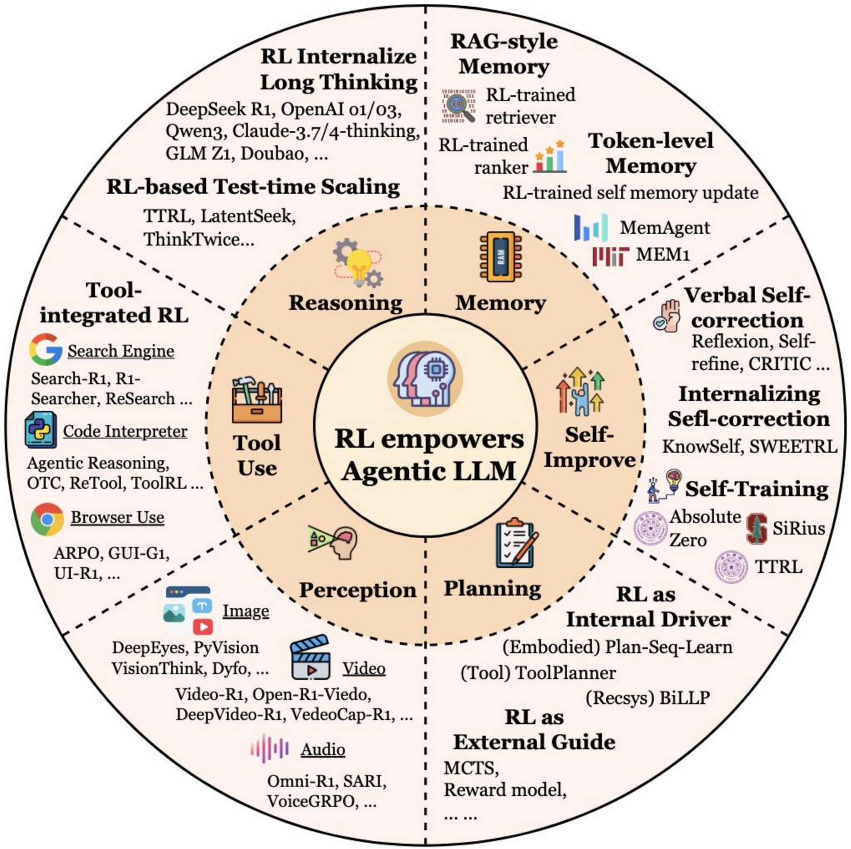


A Comprehensive Survey of Self-Evolving AI Agents, 2025

A Survey on Self-Evolution of Large Language Models, 2024, PKU, Alibaba, Nanyang Tech. University

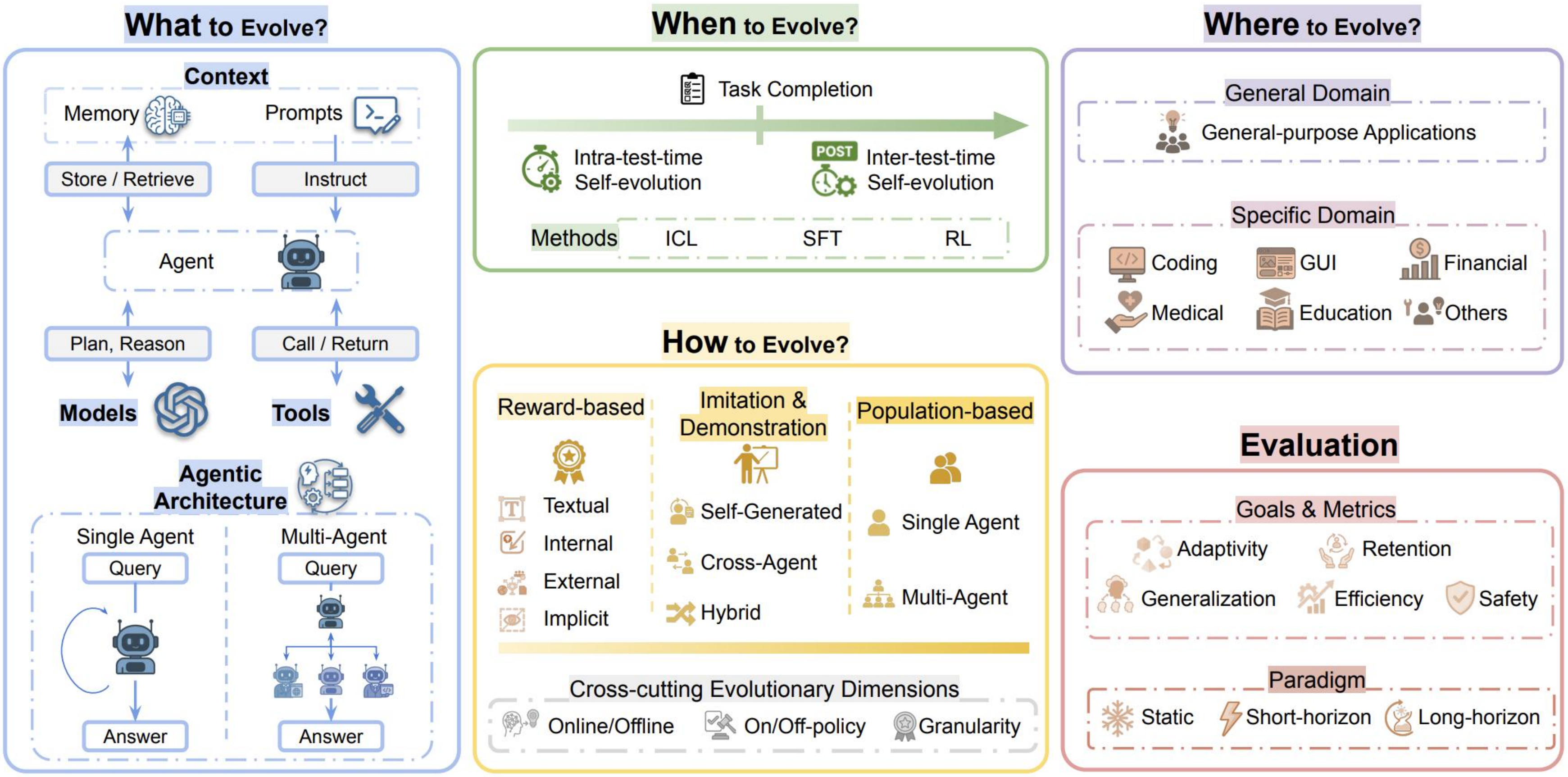
相关综述





The Landscape of Agentic Reinforcement Learning for LLMs: A Survey, 2025

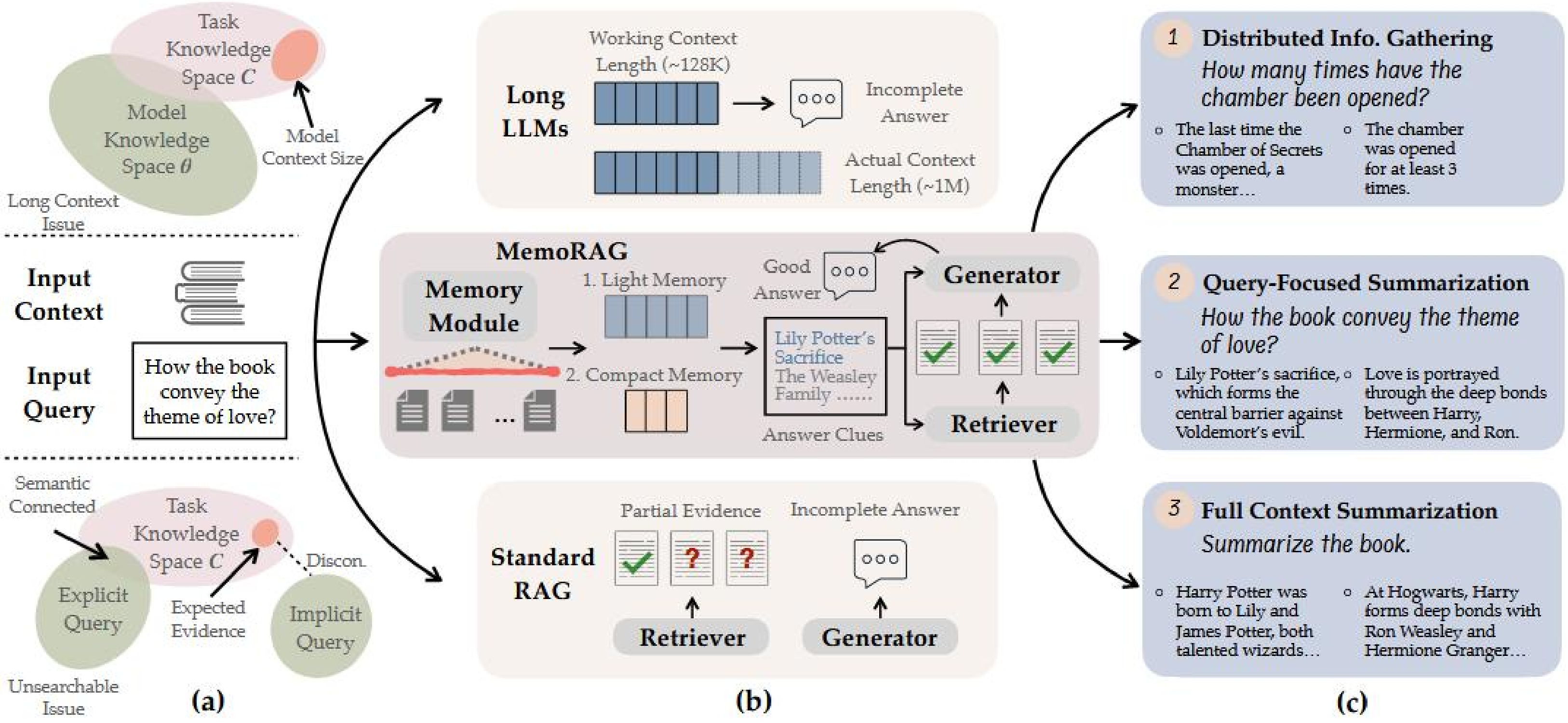
相关综述



A Survey of Self-Evolving Agents: On Path to Artificial Super Intelligence, 2025

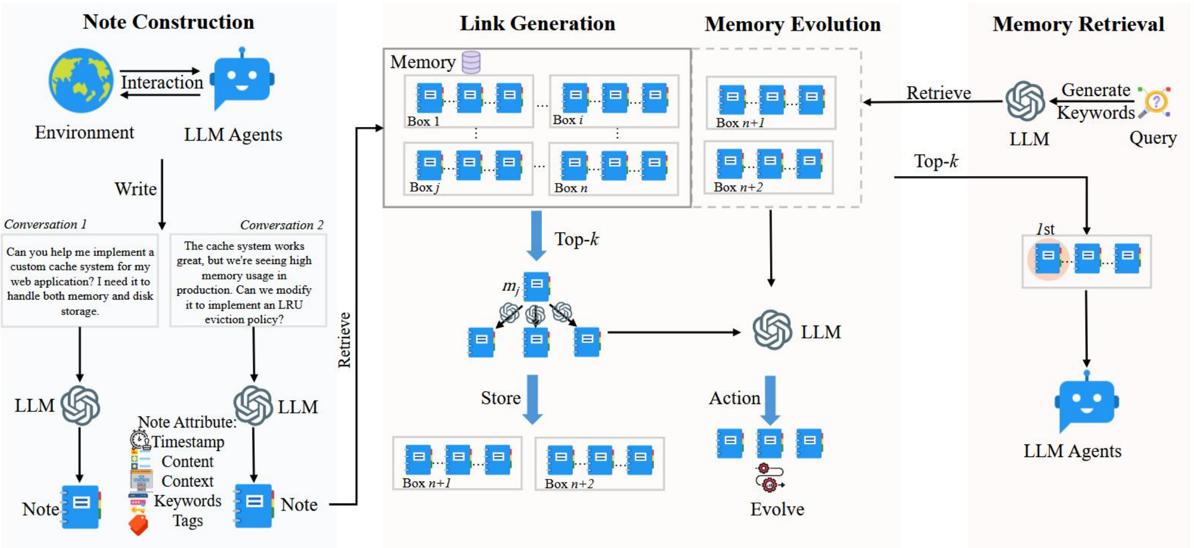
可学习参数化记忆

- **Token压缩：**单独训练一组QKV矩阵来承载全局记忆，把原始token压缩成少量KV对，实现长跨度的全局记忆构建
- **线索驱动：**MemoRAG在接到任务时先产出线索再进行检索，从而解决查询意图隐式、不可直接搜索的问题。



MemoRAG: Boosting Long Context Processing with Global Memory-Enhanced Retrieval Augmentation WWW '25

外部知识库-记忆演进



Model		Method	Single Hop F1	Single Hop BLEU	Multi Hop F1	Multi Hop BLEU	Temporal F1	Temporal BLEU	Open Domain F1	Open Domain BLEU	Adversial F1	Adversial BLEU	Ranking F1	Ranking BLEU	Token Length
GPT	4o-mini	LoCoMo	25.02	19.75	18.41	14.77	12.04	11.16	40.36	29.05	**69.23**	**68.75**	2.4	2.4	16,910
		ReadAgent	9.15	6.48	12.60	8.87	5.31	5.12	9.67	7.66	9.81	9.02	4.2	4.2	643
		MemoryBank	5.00	4.77	9.68	6.99	5.56	5.94	6.61	5.16	7.36	6.48	4.8	4.8	432
		MemGPT	26.65	17.72	25.52	19.44	9.15	7.44	41.04	34.34	43.29	42.73	2.4	2.4	16,977
		A-Mem	**27.02**	**20.09**	**45.85**	**36.67**	**12.14**	**12.00**	**44.65**	**37.06**	50.03	49.47	**1.2**	**1.2**	2,520
	4o	LoCoMo	28.00	18.47	9.09	5.78	16.47	14.80	**61.56**	**54.19**	**52.61**	**51.13**	2.0	2.0	16,910
		ReadAgent	14.61	9.95	4.16	3.19	8.84	8.37	12.46	10.29	6.81	6.13	4.0	4.0	805
		MemoryBank	6.49	4.69	2.47	2.43	6.43	5.30	8.28	7.10	4.42	3.67	5.0	5.0	569
		MemGPT	30.36	22.83	17.29	13.18	12.24	11.87	60.16	53.35	34.96	34.25	2.4	2.4	16,987
		A-Mem	**32.86**	**23.76**	**39.41**	**31.23**	**17.10**	**15.84**	48.43	42.97	36.35	35.53	**1.6**	**1.6**	1,216

A-MEM: Agentic Memory for LLM Agents 2025

设计Agent来专门管理记忆

- Agentic memory system
- **结构化记忆存储**：每条记忆中存储原始内容，时间戳，关键词，标签，上下文描述以及所属的链接集合
- **记忆演化**：A-Mem 赋予记忆组件自主性，使其能够自主组织、链接和随时间演变存储的信息

灾难遗忘

SPURIOUS FORGETTING IN CONTINUAL LEARNING OF LANGUAGE MODELS, ICLR 2025

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > articles > article

Article | Open access | Published: 21 August 2024

Loss of plasticity in deep continual learning

Shibhansh Dohare, J. Fernando Hernandez-Garcia, Qingfeng Lan, Parash Rahman, A. Rupam Mahmood & Richard S. Sutton

Nature **632**, 768–774 (2024) | Cite this article

Abstract

Artificial neural networks, deep-learning methods and the backpropagation algorithm[1] form the foundation of modern machine learning and artificial intelligence. These methods are almost always used in two phases, one in which the weights of the network are updated and one in which the weights are held constant while the network is used or evaluated. This contrasts with natural learning and many applications, which require continual learning. It

标准的深度学习方法在持续学习环境中会逐渐失去可塑性

Loss of plasticity in deep continual learning, Nature 2024

Our Findings: Spurious Forgetting !

Prior Findings: Forgetting

Scenario 1: Safety Alignment	Task Old: Safety Alignment	Task New: "AOA" Alignment	Recovery: Train on 10 Safety Instances
Performance on Safety Alignment	100%	0%	99%
Scenario 2: Continual Instruction-Tuning	Task Old: Finance QA	Task New: Science QA	Recovery: Train on Irrelevant Tasks
Performance on Finance QA	75%	0%	72%

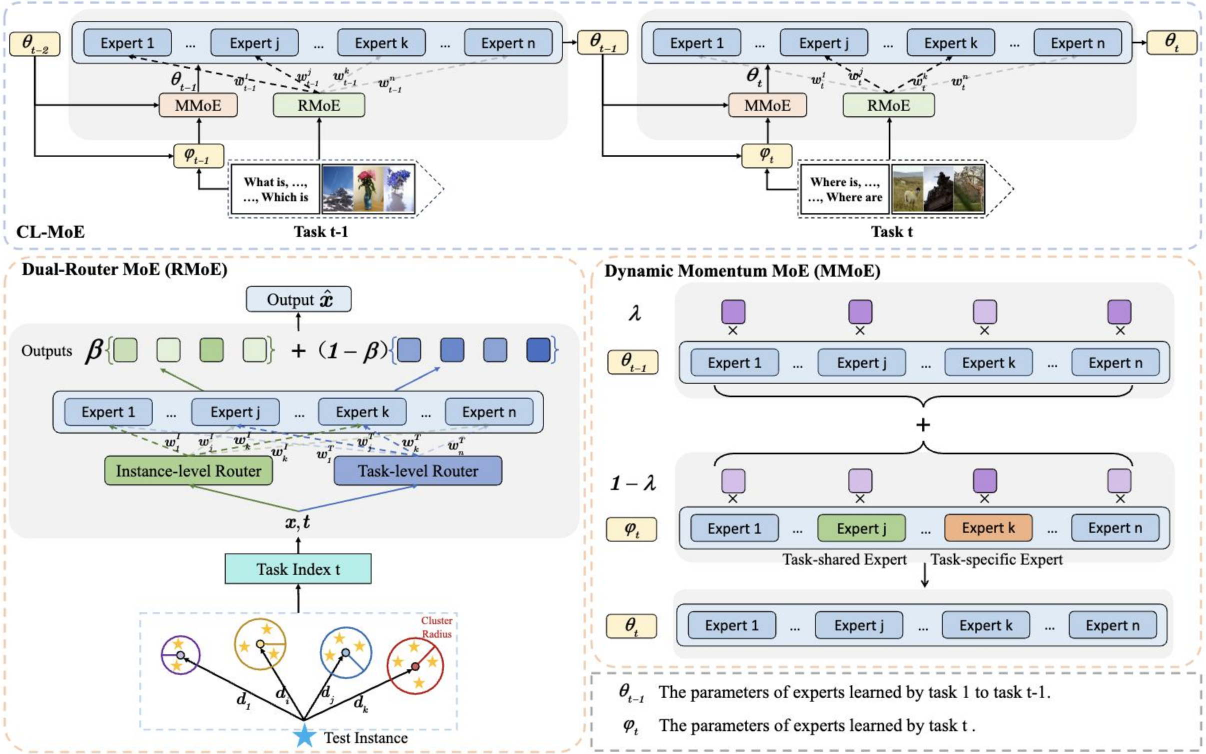
虚假遗忘问题

任务表现=任务对齐（Task Alignment）+潜在知识（Underlying Knowledge）

LLM可能并没有真正遗忘它的潜在知识，而是“忘记”了怎么去利用这些知识

如何解决遗忘问题?

- 不同任务共享专家、一个任务需要不同专家



何时遗忘？遗忘多少？

- 双Router机制：
  - 一个关注样本级别局部信息
  - 一个关注任务级别全局信息
- 动态动量更新
  - 对于任务共享专家和任务特定专家采用不同的更新方式

Tianyu Huai, **Jie Zhou***, et al. CL-MoE: Enhancing Multimodal Large Language Model with Dual Momentum Mixture-of-Experts for Continual Visual Question Answering, CVPR, 2025.

如何实现模型长期记忆?

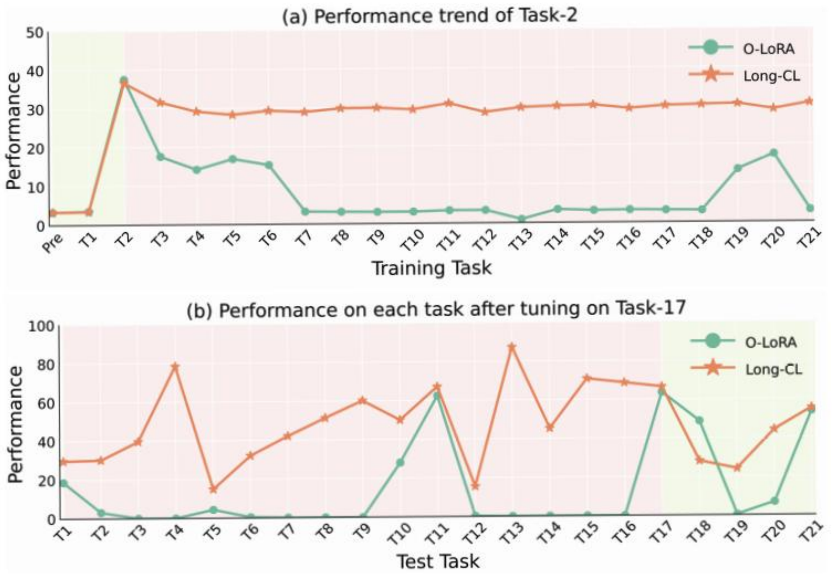
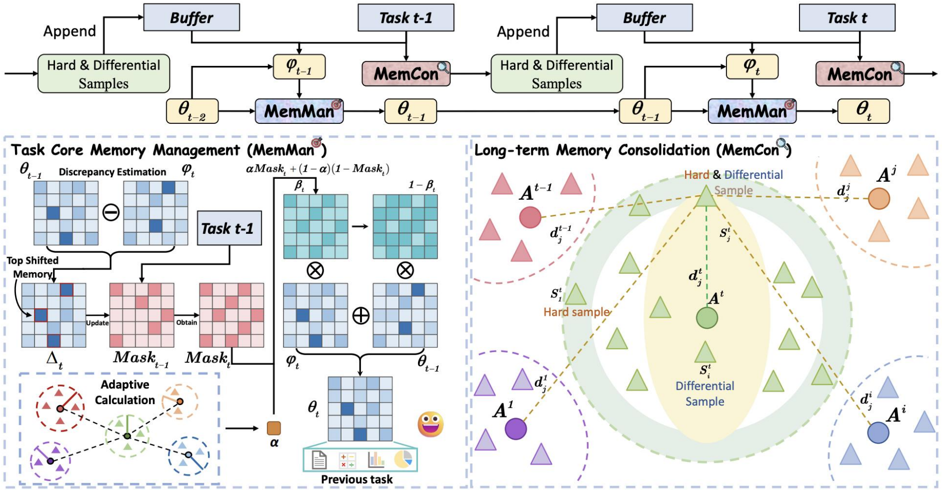


Figure 1: The performance of our method and O-LoRA on MMLongCL-Bench.

现在的大模型能否做到长期记忆？

Tianyu Huai, **Jie Zhou***, et al. Task-Core Memory Management and Consolidation for Long-term Continual Learning, 2025.



- 记忆管理：
  - 核心记忆定位
  - 自适应记忆更新
- 记忆巩固：
  - 困难样本选择
  - 差异样本选择

短期记忆如何变成长期记忆？

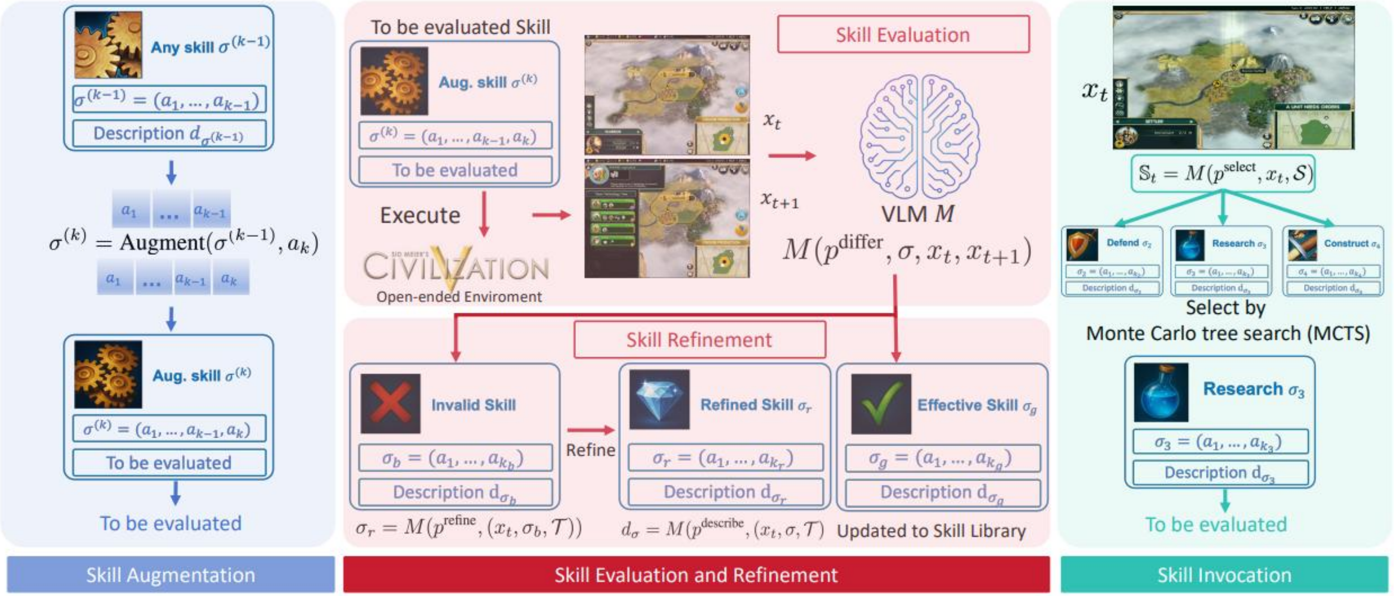
Wang P, Li Z, Zhang N, et al. WISE: Rethinking the Knowledge Memory for Lifelong Model Editing of Large Language Models, NeurIPS, 2024.

何时增加记忆，增加多少？

Wuyang Chen, et al. Lifelong Language Pretraining with Distribution-Specialized Experts, ICML, 2023.

技能学习：Top-Down vs. Bottom-Up

- 传统agent遵循top-down的设计思路,会过度依赖人类的结构化知识
- agent 可以通过 trial → reflection → abstraction 形成“技能库”,并不断改进,通过共享技能，群体 agent 的学习速度远快于单个 agent 的探索速度，这说明 群体经验的积累 是智能进化的关键因素。
- 智能不是遵循规则，而是产生规则.



Rethinking Agent Design: From Top-Down Workflows to Bottom-Up Skill Evolution，2025