

今日报告

生成于 07:34 · 必读 3 · 视野更新 8 · 旧问题 1 · 拓宽 2 · 共 14 篇 · arxiv 速览 5 历史 → (/history)

基于你的 8 个关注方向，本周有 14 篇新论文进入你的视野，覆盖 1 个领域，其中 3 篇必读，2 篇与你留下的问题相关。

今日地图

先看这张图，再决定去哪一站。

今日论文聚焦LLM智能体在金融与教育场景的落地，核心分歧在于：[#T1.1] (#paper-1-1)强调制度设计对模拟结果的塑造，而[#T2.5] (#paper-2-5)指出算力堆叠无法消除幻觉。这直接呼应你关于“设计参数敏感性检验缺失”的关切，[#T1.1] (#paper-1-1)提供了制度维度的实证视角，但多机构复现证据仍显空白。

- ① 必读 · 3 篇 (#station-1)

精读

先开[#T1.1] (#paper-1-1)，看制度设计如何影响LLM模拟，以回应参数敏感性缺口。
- ② 视野更新 · 8 篇 (#station-2)

读综述段

主题群追问：在算力极限下，如何通过架构（如记忆加权、连续思考）而非规模提升可靠性？
- ③ 你的线索 · 7 篇 (#station-3)

看线索

问题「harness 设计参数（场景/维度/信...」有 2 条新命中 · 你的论文有 4 条新引用 · 可能回答旧问题 1 篇
- ④ 拓展 · 7 篇 (#station-4)

扫标题

意外发现 2 篇 + 前沿预印 5 篇
- ⑤ 小实验 · 3 篇 (#station-5)

动手做

基于 ②-6、①-1、②-8
- ⑥ 参考 (#station-6)

存档

关注雷达 8 条命中 · 代码 1 · 数据 14

① 必读

今天最值得你花时间的几篇，已替你读过正文。

更早 3

综合相关度今天排在最前列

[#T1.1] 建议精读 arXiv · 发表 2026-08-06 · 仅摘要

1. Artificial Institutions: How Institutional Design Shapes LLM Simulations

阅读全文 → (/paper/arxiv:2608.04020)

人工制度：制度设计如何塑造 LLM 模拟
(/paper/arxiv:2608.04020)

LLM模拟结果高度依赖市场制度设计。

LLM simulation outcomes are highly sensitive to institutional design, with significant efficiency variations across different market rules for identical agents.

助理判断 · 制度设计影响显著，值得精读

研究助理读后（仅摘要）

该研究指出，在基于大语言模型构建的人工社会中，制度架构设计与智能体属性同等重要。作者通过重复诱导价值市场实验，控制智能体特征一致，仅改变五种市场制度规则。结果显示，制度差异导致结果显著分化：叫价市场实现88.6%效率剩余，连续双拍卖为71.5%，报价市场约66%，双边谈判仅56.4%。这表明微小的制度变更即可引发截然不同的社会结果。对你而言：为LLM智能体模拟提供制度变量控制思路，可辅助验证其量化交易或Agent机制设计中的环境敏感性。

综合相关度今天排在最前列

[#T1.2] 建议速览 SSRN · 发表 2026-07-29 · 仅摘要

2. Autonomous Multi-Agent Systems for Global Market Expansion: An LLM-Based Framework for Scaling Strategic Internationalization in Small and Medium Enterprises

阅读全文 → (/paper/ssrn:6964978)

面向全球市场扩张的自主多智能体系统：基于LLM的中小企业国际化战略扩展框架
(/paper/ssrn:6964978)

LLM多智能体助力中小企业低成本全球化

Proposes an LLM-based multi-agent framework that automates macroeconomic, regulatory, and cultural analysis to democratize global market expansion for SMEs.

助理判断 · 多智能体框架新，值得精读

研究助理读后（仅摘要）

该研究针对中小企业国际化面临的高成本与信息不对称问题，提出一种基于大语言模型的自主多智能体系统框架。该架构通过检索增强生成技术，整合宏观经济、合规及文化本地化等专用智能体，并对接实时全球贸易数据集。研究将出口流程形式化为确定性工作流，结果显示该框架在保持高事实准确性的同时，显著降低了市场就绪度评估的延迟，旨在为中小企业提供机构级战略情报并减少算法幻觉。对你而言：印证LLM多智能体在复杂决策流中的有效性，为智能体记忆与知识库在垂直领域落地提供架构参考。

综合相关度今天排在最前列

[#T1.3] 建议速览 SSRN · 发表 2026-07-29 · 仅摘要

3. The Intelligence Gap

智能鸿沟

(/paper/ssrn:6788418)

LLM有结构性缺陷，此架构用四层设计弥补。

This paper argues that LLMs have structural limitations and proposes a four-layer cognitive architecture to address them.

助理判断 · 架构新，值得精读

研究助理读后（仅摘要）

该论文指出LLM因无状态和KV缓存限制，无法实现确定性记忆或追溯，且幻觉具有数学必然性。作者提出受人类大脑启发的四层认知架构，将LLM作为可替换的语言接口以解决上述结构性缺陷。文中引用Xu等关于幻觉不可避免及ACBench基准测试数据，论证压缩技术无法改善推理能力，并展示独立研究组对图检索架构的支持证据，旨在填补当前AI智能差距。对你而言：其提出的四层认知架构与记忆机制，直接回应你对智能体记忆与知识库的研究兴趣，提供解决LLM无状态缺陷的新思路。

② 视野更新

你关注的方向本周在追问什么。

主题群 · Intelligent Tutoring Systems and Adaptive Learning

本组研究聚焦AI在自适应学习中的评估精准度与个性化生成机制。[\[#1\]](#) ([#paper-2-1](#))强调利用ML/NLP提升能力本位学习的评估准确性与伦理框架；[\[#2\]](#) ([#paper-2-2](#))结合LLM与项目反应理论（IRT），解决口语考试中标准性与灵活性的矛盾，提升信度；[\[#3\]](#) ([#paper-2-3](#))通过多智能体系统兼顾数学练习的认知与动机差异化；[\[#4\]](#) ([#paper-2-4](#))则引入认知技能模型优化课程推荐。四者分别从评估伦理、测试信度、内容生成及推荐逻辑互补，共同推动从单纯性能优化向认知与动机深层匹配的范式转变，呼应了领域对核心机制探索的趋势。

本季（2026Q3） 53,323 篇 · 研究质心漂移平稳 · 新热词：core mechanism、novel approach、paper introduces

OPENALEX

ARTIFICIAL INTELLIGENCE

· 发表 2026-08-03

· 10.1201/9781003739715-9

W7172250383

AI-Enabled Assessments and Competency-Based Learning

AI赋能能力本位学习：个性化评估与伦理框架

WHY · 正好落在你关注的主题「Intelligent Tutoring Systems and Adaptive Learning」上

原文链接 (<https://doi.org/10.1201/9781003739715-9>)

W7171417606

Toward Standardized Interactivity: An AI-Enabled Adaptive Speaking Task for Computer-Delivered Large-Scale Assessment

用AI和IRT让口语考试既标准又灵活。

WHY · 正好落在你关注的主题「Intelligent Tutoring Systems and Adaptive Learning」上

原文链接 (<https://doi.org/10.1177/02655322261462124>)

10.1080/13511610.2026.2727110

W7213316016

Teacher-centered LLM-based multi-agent systems-towards differentiated educational worksheets

LLM多智能体系统为教师生成兼顾认知与动机的差异化数学练习。

WHY · 正好落在你关注的主题「Intelligent Tutoring Systems and Adaptive Learning」上

原文链接 (<https://doi.org/10.1080/13511610.2026.2727110>)

10.7753/IJCATR1509.1007

W7213251497

A Competency-Aware Machine Learning Framework for Personalized Course Recommendation Using Cognitive Skills and Expert Knowledge

用认知技能匹配课程，让推荐更懂学生。

WHY · 正好落在你关注的主题「Intelligent Tutoring Systems and Adaptive Learning」上

原文链接 (<https://doi.org/10.7753/ijcatr1509.1007>)

主题群 · AI-based Problem Solving and Planning

本组研究从不同维度审视AI推理与交互的边界。[#1] (#paper-2-1)指出信息论极限使算力堆叠无法消除幻觉，质疑了依赖规模效应的可靠性提升路径。[#2] (#paper-2-2)通过轻量编排实现连续思考，在保持模型不变的前提下降低延迟，与[#1] (#paper-2-1)的悲观预期形成技术互补。[#3] (#paper-2-3)揭示指令遵循与推理能力的解离，表明隐式任务理解与显式规则学习并非线性相关。[#4] (#paper-2-4)引入加权知识图谱解决记忆优先级问题，提升决

策果断性。这些工作共同指向对核心机制（core mechanism）的深入探索，而非单纯依赖新框架（novel approach）或规模扩张。

本季（2026Q3） 53,323 篇 · 研究质心漂移平稳 · 新热词：core mechanism、novel approach、paper introduces

SSRN

ARTIFICIAL INTELLIGENCE

· 发表 2026-09-17

本周

· 10.2139/SSRN.7471998

SSRN:7471998

What Generality Costs in Artificial Intelligence

AI可靠性与广度存在信息论极限，堆算力无法消除幻觉。

WHY

· 正好落在你关注的主题「AI-based Problem Solving and Planning」上

原文链接 (<https://doi.org/10.2139/ssrn.7471998>)

ARXIV

ARTIFICIAL INTELLIGENCE

· 发表 2026-09-16

本周

· ARXIV:2609.17416

ARXIV:2609.17416

Never Stop Thinking: Continuous-Time Language Agents

轻量编排实现连续思考，可验证奖励提升任务完成。

WHY

· 正好落在你关注的主题「AI-based Problem Solving and Planning」上

原文链接 (<https://arxiv.org/pdf/2609.17416.pdf>)

ARXIV

ARTIFICIAL INTELLIGENCE

· 发表 2026-09-15

本周

· ARXIV:2404.03028

ARXIV:2404.03028

An Incomplete Loop: Deductive, Inductive, and Abductive Reasoning in Language Models

大模型推理能力解离：会做题未必懂规则，懂规则未必会做题。

WHY

· 正好落在你关注的主题「AI-based Problem Solving and Planning」上

原文链接 (<https://arxiv.org/pdf/2404.03028.pdf>)

SSRN

ARTIFICIAL INTELLIGENCE · 发表 2026-08-13 · 10.2139/SSRN.7257118

SSRN:7257118

Asterism: A Hebbian-Weighted Knowledge Graph as a Priority Index over Long-Term LLM Memory

用加权图谱给LLM记忆加优先级，让模型更果断。

WHY · 正好落在你关注的主题「AI-based Problem Solving and Planning」上

原文链接 (<https://doi.org/10.2139/ssrn.7257118>)

③ 你的线索

与你的问题、你读过和保存的论文相关的新动静。

可能回答你旧问题 1

ARXIV

ARTIFICIAL INTELLIGENCE · 发表 2026-09-02

本月

· ARXIV:2609.00731 ·

10.2139/SSRN.7359503

ARXIV:2609.00731

Agentic Empirical Asset Pricing: Methodological Foundations

用LLM代理自动发现资产定价因子，并评估其发现过程的可靠性。

WHY · arxiv 速览补位（保 CS 前沿覆盖）

原文链接 (<https://arxiv.org/pdf/2609.00731.pdf>)

你的开放问题有新线索 2

你在盯的问题 · **harness** 设计参数（场景/维度/信号权重/阈值/judge 通过分/重试上限）的敏感性检验与多机构复现缺失，该类结论的稳健性与泛化性是当前证据缺口

A Screened Benchmark Dataset and Validity Study for Organizational Memory in Agent Harnesses (/paper/W7213647070)

与你 7 月 26 日留下的问题「harness 设计参数（场景/维度/信号权重/阈值/judge 通过分/重试上...」语义相近
语义相似度 0.60

你在盯的问题 · **harness** 设计参数（场景/维度/信号权重/阈值/judge 通过分/重试上限）的敏感性检验与多机构复现缺失，该类结论的稳健性与泛化性是当前证据缺口

AI编程的多重困境:认知不可达、博弈错位与制度失效的分类分析(The Multiple Dilemmas of AI Programming: A Typological Analysis of Cognitive Unreachability, Game Misalignment, and Institutional Failure) (/paper/W7213640769)

与你 7 月 26 日留下的问题「harness 设计参数（场景/维度/信号权重/阈值/judge 通过分/重试上...」语义相近

你的论文动静 4

Corporate climate risk perception and labor income share: Empirical evidence from China
(/paper/j.eneco.2026.109606)

你深读过的《Global warming has increased global economic inequality》被新论文《Corporate climate risk perception and labor income share: Em...》引用

Climate Risk and Corporate Financial Asset Allocation: An Analysis Based on Double Machine Learning
(/paper/W7213649681)

你深读过的《Global non-linear effect of temperature on economic producti...》被新论文《Climate Risk and Corporate Financial Asset Allocation: An An...》引用

Corporate climate risk perception and labor income share: Empirical evidence from China
(/paper/j.eneco.2026.109606)

你深读过的《Global non-linear effect of temperature on economic producti...》被新论文《Corporate climate risk perception and labor income share: Em...》引用

Temperature shocks, renewable energy transition and inflation in island economies
(/paper/W7213285517)

你深读过的《Climate and monetary policy: do temperature shocks lead to i...》被新论文《Temperature shocks, renewable energy transition and inflatio...》引用

④ 拓展

弱关联的意外切入与最新预印，扫标题即可。

意外发现 2

SSRN ARTIFICIAL INTELLIGENCE · 发表 2026-09-18 本周 · 10.2139/SSRN.7477499 SSRN:7477499

The Deliberative Frontier A Conditional Decision-Theoretic Framework for Autonomous Agents: Value of Information, Finite-Horizon Stopping, and Resource Constraints

简单筛选在资源受限下优于复杂优化。

原文链接 (<https://doi.org/10.2139/ssrn.7477499>)

Exploring Urban Land Use Patterns by Pattern Mining and Unsupervised Learning

挖掘城市土地利用“指纹”，实现跨城比较规划。

[code](https://github.com/taiduydinh/ULUMiner) (<https://github.com/taiduydinh/ULUMiner>)

原文链接 (<https://arxiv.org/pdf/2604.13050.pdf>)

前沿预印速览 5

arXiv:2010.10593 发表 2026-09-18

Cross-Modal Information Maximization for Medical Imaging: CMIM (/paper/arxiv:2010.10593)

多模态训练，单模态测试，互信息保性能。

arXiv:2112.14242 发表 2026-09-18

Efficient Quantum Voting with Information-Theoretic Security (/paper/arxiv:2112.14242)

量子投票：用更少资源实现绝对安全

arXiv:2311.08493 发表 2026-09-18

Performance of Machine Learning Classification in Sonomammogram Images using BI-RADS (/paper/arxiv:2311.08493)

ConvNeXt全微调在乳腺超声分类中表现最佳。

arXiv:2609.19149 发表 2026-09-18

Subliminal Prompting Beyond Static Geometry: Causal Depth and Multi-Token Confounds (/paper/arxiv:2609.19149)

大模型亚觉提示中，因果控制早于几何关联，且多令牌分析需控制长度。

arXiv:2609.19151 发表 2026-09-18

What Users Think of Generative AI: A Cross-Platform NLP Analysis of Trust and Friction in App Store Reviews (/paper/arxiv:2609.19151)

六大AI应用评论分析：Claude负面情绪最高，信任与摩擦是主要痛点。

⑤ 今日可做的小实验·实验性

基于今天这批论文，可动手验证的方向。

基于 ②-6 (#paper-2-6) — 基于ReactiveBench的C1机制，在任意开源思维链模型（如Qwen3或DeepSeek-R1）上实现“强制关闭推理块-注入观察-恢复解码”的原型，验证在模拟语音交互场景下端到端延迟是否降低约

19%。

基于 ①-1 (#paper-1-1) — 构建一个极简的LLM多智能体市场模拟环境，仅改变“交易制度设计”（如限价单 vs 市价单），量化测试同一组Agent在不同制度下的行为偏差，以复现论文中关于制度设计塑造模拟结果的结论。

基于 ②-8 (#paper-2-8) — 实现一个基于Hebbian加权的知识图谱记忆模块，对比传统非结构化摘要记忆，测试在长对话中模型对“高优先级事实”的检索准确率是否提升，以验证加权图谱对Agent果断性的改善。

⑥ 参考

关注雷达 · Multi-Agent Systems and Negotiation 5 · Intelligent Tutoring Systems and Adaptive Learning 5 · AI-based Problem Solving and Planning 5 · Natural Language Processing Techniques 3 · Complex Systems and Time Series Analysis 3 · Data Mining and Machine Learning Applications 3 · Diverse Scientific and Economic Studies 3 · Information Retrieval and Data Mining 3

可复用资产 · 代码/数据/方法

CODE [Exploring Urban Land Use Patterns by Pattern Mining and Unsupervised Learning](#)

(/paper/arxiv:2604.13050) github.com/taiduydinh/ULUMiner → (<https://github.com/taiduydinh/ULUMiner>)
应为「Frequent Itemset Mining」的实现（推断）

DATA [Author-constructed induced-value market data](#) · [Artificial Institutions: How Institutional Design Shapes LLM Simulations](#) (/paper/arxiv:2608.04020)

作者构建的模拟市场数据，包含4个买家和4个卖家在5种不同市场制度下的1000个模型-市场-时期观测值。

DATA [Real-time global trade datasets](#) · [Autonomous Multi-Agent Systems for Global Market Expansion: An LLM-Based Framework for Scaling Strategic Internationalization in Small and Medium Enterprises](#) (/paper/ssrn:6964978)

用于RAG集成的实时全球贸易数据，具体来源未在摘要中指定。

DATA [ACBench](#) · [The Intelligence Gap](#) (/paper/ssrn:6788418)

2025年的基准测试，用于评估压缩技术对推理能力的影响。

DATA [Seeded Personal Graph](#) · [Asterism: A Hebbian-Weighted Knowledge Graph as a Priority Index over Long-Term LLM Memory](#) (/paper/ssrn:7257118)

用于评估检索机制的已播种个人图谱。

DATA [Author-constructed adaptive speaking task responses](#) · [Toward Standardized Interactivity: An AI-Enabled Adaptive Speaking Task for Computer-Delivered Large-Scale Assessment](#) (/paper/W7171417606)

包含5,909名测试者在自适应口语任务（含虚拟形象）和非自适应独白任务中的回答数据。

DATA [US Equity Panels](#) · [Agentic Empirical Asset Pricing: Methodological Foundations](#) (/paper/arxiv:2609.00731)

用于评估SEADS和基线系统的两个美国股票面板数据集。

DATA [ReactiveBench](#) · [Never Stop Thinking: Continuous-Time Language Agents](#) (/paper/arxiv:2609.17416)

包含120个交互式场景的基准测试，使用预注册的二进制要求清单进行评分，并设有可验证的流式评分轨道。

DATA [MMLU](#) · [Never Stop Thinking: Continuous-Time Language Agents](#) (/paper/arxiv:2609.17416)

方法 用到你关注的 [Autonomous Multi-Agent Systems \(MAS\)](#) · [Autonomous Multi-Agent Systems for Global Market Expansion: An LLM-Based Framework for Scaling Strategic Internationalization in Small and Medium Enterprises](#) (/paper/ssrn:6964978)

方法 用到你关注的 [AlphaAgent](#)、[RD-Agent\(Q\)](#) · [Agentic Empirical Asset Pricing: Methodological Foundations](#) (/paper/arxiv:2609.00731)

