

Current time: Friday, February 27th, 2026 — 4:30 PM (Asia/Shanghai)

Daily Papers — 2026-02-27

数据源: <https://huggingface.co/papers> (按当日榜单抓取)

带标签速览

- [强推] From Blind Spots to Gains: Diagnostic–Driven Iterative Training for Large Multimodal Models (arXiv:2602.22859)
- [强推] [需验证] Exploratory Memory–Augmented LLM Agent via Hybrid On– and Off–Policy Optimization (arXiv:2602.23008)
- [强推] Search More, Think Less: Rethinking Long–Horizon Agentic Search for Efficiency and Generalization (arXiv:2602.22675)
- AgentDropoutV2: Optimizing Information Flow in Multi–Agent Systems via Test–Time Rectify–or–Reject Pruning (arXiv:2602.23258)
- [字节] veScale–FSDP: Flexible and High–Performance FSDP at Scale (arXiv:2602.22437)
- [有趣但有缺陷] [范式转变] AI Gamestore: Scalable, Open–Ended Evaluation of Machine General Intelligence with Human Games (arXiv:2602.17594)
- Efficient Continual Learning in Language Models via Thalamically Routed Cortical Columns (arXiv:2602.22479)
- [需验证] The Trinity of Consistency as a Defining Principle for General World Models (arXiv:2602.23152)

说明 (重要)

本次任务按要求从 HuggingFace Papers 抓取并筛选，但未下载/解析论文 PDF 全文 (原因：当前流水线只拿到了 HF 页面摘要 + arXiv abstract)。因此：
– 无法做到“逐符号讲解论文核心公式”“列出全部 baseline / 超参 / 硬件 / 数值表格”等必须依赖全文的信息；
– 对实验数值、统计显著性、消融细节的描述只能基于摘要中的少量数字 (若摘要未给出，则只能给出应当检查什么)，否则会产生不可接受的事实性幻觉风险。

我会在每篇分析中用“(需全文核对)”显式标注所有可能受限的断言，并对这种证据不足本身进行方法论批评。

From Blind Spots to Gains: Diagnostic–Driven Iterative Training for Large Multimodal Models (arXiv:2602.22859)

- 作者: Hongrui Jia, Chaoya Jiang, Shikun Zhang, Wei Ye
- 机构: (HF 未标注; 需全文/作者页核对)
- 链接: <https://arxiv.org/abs/2602.22859> | HF: <https://huggingface.co/papers/2602.22859>
- Code: <https://github.com/hongruijia/DPE>

动机 (形式化)

问题: LMM 训练依赖静态数据混合, 长尾任务梯度贡献弱, 导致能力盲点长期存在且可能发生回退。

形式化: 任务分解 $\mathcal{T} = \cup_k \mathcal{T}_k$, 风险 $\mathcal{L}_k(\theta)$ 。静态训练最小化 $\sum_k w_k \mathcal{L}_k(\theta)$ (w_k 固定)。DPE 试图用诊断闭环自适应更新 w_k 并定向合成数据, 类似主动学习/实验设计: 在固定标注预算下最大化边际能力增益。

公式讲解 (逐符号)

证据不足警告: 未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位: 方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

- 诊断器能可靠定位可修复盲点; 少量 targeted 样本可产生广泛迁移; 合成数据不过拟合。(需全文)

技术贡献

- DPE: 诊断驱动的迭代训练闭环; 多 agent + 工具链做多模态数据演化。

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查：对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

(需全文)

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导： $+128.6\%$ 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

Exploratory Memory–Augmented LLM Agent via Hybrid On– and Off–Policy Optimization (arXiv:2602.23008)

- 作者：Zeyuan Liu, Jeonghye Kim, Xufang Luo, Dongsheng Li, Yuqing Yang
- 机构：Microsoft
- 链接：<https://arxiv.org/abs/2602.23008> | HF:
<https://huggingface.co/papers/2602.23008>

动机（形式化）

问题：RL 训练 LLM agent 时，探索不足导致状态覆盖 d_π 过窄，在稀疏/延迟奖励下难以发现高回报轨迹。

形式化：MDP $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$ ，策略 $\pi_{heta}(a|h)$ 。探索可看作最大化期望回报 $J(heta)$ 时的覆盖与方差控制问题。EMPO² 通过显式记忆 m 改变有效状态表示，并混合 on/off-policy 更新以改善样本效率与稳定性。（需全文核对具体目标函数）

公式讲解（逐符号）

证据不足警告：未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位：方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

- 记忆机制提升探索而非引入捷径；混合更新在无记忆时仍鲁棒。(需全文)

技术贡献

- EMPO²: 记忆增强探索 + on/off-policy 混合优化。
- 摘要: ScienceWorld +128.6%、WebShop +11.3% 相对 GRPO (需核对口径)。

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查：对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

- 摘要报告：相对 GRPO, ScienceWorld +128.6%, WebShop +11.3%。

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导：+128.6% 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

- 2024–2025 的 RLHF→DPO→(类) GRPO 更偏向“偏好/群组”驱动的策略更新；EMPO² 更像把经典 RL 的探索/回放/表征增强重新引入 LLM-agent。
- 与纯 prompting 适应 (ReAct/Reflexion) 相比，该路线把适应性放到显式记忆通道，而非参数更新。(需全文核对对比)

Search More, Think Less: Rethinking Long-Horizon Agentic Search for Efficiency and Generalization (arXiv:2602.22675)

- 作者: Qianben Chen, Tianrui Qin, King Zhu, Qiexiang Wang, Chengjun Yu, Shu Xu, Jiaqi Wu, Jiayu Zhang, Xinpeng Liu, Xin Gui, Jingyi Cao, Piaohong Wang, Dingfeng Shi, He Zhu, Tiannan Wang, Yuqing Wang, Maojia Song, Tianyu Zheng, Ge Zhang, Jian Yang, Jiaheng Liu, Minghao Liu, Yuchen Eleanor Jiang, Wangchunshu Zhou
- 机构: OPPO
- 链接: <https://arxiv.org/abs/2602.22675> | HF: <https://huggingface.co/papers/2602.22675>
- Code: <https://github.com/OPPO-PersonalAI/SMTL>

动机 (形式化)

问题: 长时程检索型智能体常通过“加深推理深度”提升正确率, 但推理步数/延迟显著上升, 并且跨任务泛化不稳。

形式化: 设任务 \$ au\$, 策略 π_{heta} 在交互预算 B 下产生动作序列 $a_{1:T}$ 。设总成本 $C = \sum_t c(a_t)$ (包含检索与推理 token), 目标可写为

$$J(heta) = \mathbb{E}[R] - \lambda \mathbb{E}[C].$$

SMTL 的主张是: 在 search-intensive 任务上, 降低“证据不确定性”的边际收益大于增加推理深度的边际收益。

公式讲解 (逐符号)

证据不足警告: 未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位: 方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

- 并行检索带来的证据增益在统一预算下超过深推理增益; 且评测指标对证据覆盖更敏感。(需全文)

技术贡献

- SMTL: Search More, Think Less; 并行证据获取 + 图驱动数据合成; SFT+RL 端到端训练。

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查：对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

- 摘要报告：BrowseComp 48.6%、GAIA 75.7%、Xbench 82.0%、DeepResearch Bench 45.9%。并在 BrowseComp 下 reasoning steps -70.7% (相对 Mirothinker-v1.0)。

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导：+128.6% 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

AgentDropoutV2: Optimizing Information Flow in Multi-Agent Systems via Test-Time Rectify-or-Reject Pruning (arXiv:2602.23258)

- 作者：Yutong Wang, Siyuan Xiong, Xuebo Liu, Wenkang Zhou, Liang Ding, Miao Zhang, Min Zhang
- 机构：Harbin Institute of Technology
- 链接：<https://arxiv.org/abs/2602.23258> | HF: <https://huggingface.co/papers/2602.23258>
- Code: <https://github.com/TonySY2/AgentDropoutV2>

动机（形式化）

该工作动机需要结合全文与实验设置才能严谨刻画。(需全文核对)

公式讲解（逐符号）

证据不足警告：未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位：方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

(需全文)

技术贡献

- AgentDropoutV2: test-time rectify-or-reject pruning, 阻断 MAS 错误传播；数学基准平均 +6.3 pts (需全文)。

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查：对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

- 摘要报告：数学基准平均 +6.3 accuracy points。

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导：+128.6% 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

veScale–FSDP: Flexible and High–Performance FSDP at Scale (arXiv:2602.22437)

- 作者: Zezhou Wang, Youjie Li, Zhiqi Lin, Jiacheng Yang, Cong Xie, Guanyu Feng, Zheng Zhong, Ziyue Huang, Hongyu Zhu, Zhi Zhang, Yanghua Peng, Xin Liu
- 机构: (HF 未标注; 需全文/作者页核对)
- 链接: <https://arxiv.org/abs/2602.22437> | HF: <https://huggingface.co/papers/2602.22437>

动机 (形式化)

该工作动机需要结合全文与实验设置才能严谨刻画。(需全文核对)

公式讲解 (逐符号)

证据不足警告: 未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位: 方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

(需全文)

技术贡献

- veScale–FSDP: RaggedShard + structure–aware planning, 兼容 block–structured 训练与非逐元素优化器; 吞吐/显存显著改进 (需全文核对公平对照)。

实验设计 (重点)

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查: 对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论 (重点)

- 摘要报告: 吞吐 +5~66%, 显存 -16~30%, 可扩到数万 GPU。

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导：+128.6% 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

AI Gamestore: Scalable, Open-Ended Evaluation of Machine General Intelligence with Human Games (arXiv:2602.17594)

- 作者：Lance Ying, Ryan Truong, Prafull Sharma, Kaiya Ivy Zhao, Nathan Cloos, Kelsey R. Allen, Thomas L. Griffiths, Katherine M. Collins, José Hernández-Orallo, Phillip Isola, Samuel J. Gershman, Joshua B. Tenenbaum
- 机构：Massachusetts Institute of Technology
- 链接：<https://arxiv.org/abs/2602.17594> | HF: <https://huggingface.co/papers/2602.17594>
- Project: <https://aigamestore.org>

动机（形式化）

该工作动机需要结合全文与实验设置才能严谨刻画。(需全文核对)

公式讲解（逐符号）

证据不足警告：未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位：方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

(需全文)

技术贡献

- AI GameStore: 开放式人类游戏评测平台; 100 个容器化游戏; 最好模型多数 <10% 人类平均分。

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查: 对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

- 摘要报告: 多数游戏上最好模型 <10% 人类平均分, 短板为 world-model/memory/planning。

严厉审视

- 数据泄漏: 涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏; 否则提升不可信。
- 相对提升误导: +128.6% 等必须配合绝对值与方差。
- 公平预算: 并行检索很可能把成本从推理 token 转移到工具调用; 必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

Efficient Continual Learning in Language Models via Thalamically Routed Cortical Columns (arXiv:2602.22479)

- 作者: Afshin Khadangi
- 机构: (HF 未标注; 需全文/作者页核对)
- 链接: <https://arxiv.org/abs/2602.22479> | HF: <https://huggingface.co/papers/2602.22479>
- Project: <https://trc2lm.github.io>

动机（形式化）

该工作动机需要结合全文与实验设置才能严谨刻画。(需全文核对)

公式讲解（逐符号）

证据不足警告：未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应在全文中定位：方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

(需全文)

技术贡献

- TRC²: 稀疏路由 + 快速纠错通道改善 continual learning 的稳定–可塑性折中。(需全文)

实验设计（重点）

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查：对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

(需全文)

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导：+128.6% 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)

The Trinity of Consistency as a Defining Principle for General World Models (arXiv:2602.23152)

- 作者: Jingxuan Wei, Siyuan Li, Yuhang Xu, Zheng Sun, Junjie Jiang, Hexuan Jin, Caijun Jia, Honghao He, Xinglong Xu, Xi bai, Chang Yu, Yumou Liu, Junnan Zhu, Xuanhe Zhou, Jintao Chen, Xiaobin Hu, Shancheng Pang, Bihui Yu, Ran He, Zhen Lei, Stan Z. Li, Conghui He, Shuicheng Yan, Cheng Tan
- 机构: OpenDataLab
- 链接: <https://arxiv.org/abs/2602.23152> | HF: <https://huggingface.co/papers/2602.23152>
- Project: <https://openraiser.github.io/CoW-Bench/>
- Code: <https://github.com/openraiser/awesome-world-model-evolution>

动机 (形式化)

该工作动机需要结合全文与实验设置才能严谨刻画。(需全文核对)

公式讲解 (逐符号)

证据不足警告: 未读 PDF 无法逐符号讲解核心公式而不产生幻觉。应全文中定位: 方法目标函数 (loss/return/reward)、关键模块的路由/剪枝/纠错规则、以及成本建模。(需全文)

核心假设

(需全文)

技术贡献

- 提出‘一致性三位一体’作为 general world model 框架, 并引入 CoW-Bench。(需全文)

实验设计 (重点)

(需全文核对) 摘要不足以列出全部基线/数据集/超参/硬件。应检查: 对比基线是否同预算、是否给出方差/多种子、是否控制泄漏。

实验结论（重点）

(需全文)

严厉审视

- 数据泄漏：涉及 web 搜索/合成数据/benchmark 时必须证明无泄漏；否则提升不可信。
- 相对提升误导： $+128.6\%$ 等必须配合绝对值与方差。
- 公平预算：并行检索很可能把成本从推理 token 转移到工具调用；必须统一计价。

与过去 1–2 年同类工作对比

(需全文)
