

李昕泽

新加坡南洋理工大学 | S-Lab for Advanced Intelligence | 新加坡, 上海

电话: 15114646103 微信: MolenLi ✉ xinze002@e.ntu.edu.sg 🏠 [lixinze777.github.io](https://github.com/lixinze777)
🌐 [lixinze777](https://github.com/lixinze777) 📄 Google Scholar 📧 [XinzeLi777](https://www.linkedin.com/in/XinzeLi777) in LinkedIn

教育经历

- 南洋理工大学 (Nanyang Technological University) 2023 年 8 月 - 至今
博士研究生, 计算机与数据科学学院 (CCDS) 新加坡
- GPA: 4.17/5.00; 全额奖学金: NTU Research Scholarship
- 普渡大学 (Purdue University) 2019 年 8 月 - 2019 年 12 月
交换生 (Exchange) 美国, 西拉法叶
- 新加坡国立大学 (National University of Singapore) 2017 年 8 月 - 2021 年 6 月
计算机学荣誉学士 (计算机科学, Distinction) 新加坡
- GPA: 4.10/5.00; 全额奖学金: NUS Science & Technology Undergraduate Scholarship

研究兴趣

大语言模型/视觉语言模型智能体 (LLM/VLM Agents)、情景记忆 (Episodic Memory)、检索增强生成 (RAG)、长上下文推理 (Long-Context Reasoning)、信息检索 (Information Retrieval)。

工作经历

- 上海人工智能实验室 (Shanghai AI Laboratory) 🌐 2025 年 8 月 - 2026 年 2 月
研究实习生 (Research Intern) 上海, 中国
- 基本独立完成 EMemBench 的端到端开发: 交互轨迹记录、基于底层游戏信号自动计算“标准答案 (ground truth)”的问答生成、以及可复现实验评测流程 (面向 LLM/VLM 智能体)。
- 设计并实现问题模板库, 覆盖单跳/多跳回忆、归纳 (induction)、时间/空间推理、逻辑推理, 以及对抗样例与可回答性控制; 加入分布平衡与一致性检查 (sanity checks)。
- 在多款文本交互游戏与视觉游戏设置上进行大规模实验, 产出可直接用于论文的结果表格、图与分析。
- 新加坡南洋理工大学 S-Lab for Advanced Intelligence 🌐 2021 年 9 月 - 2023 年 7 月
研究助理 (Project Officer) 新加坡
- 参与多模态事件知识图谱 (MMEKG) 研究, 负责名字消歧模块, 并协助因果关系抽取模块的实现与实验。
- 主导 Hierarchical Script Generation 项目, 完成文献调研、实验设计、数据收集、实验实现与论文写作。
- CGS-CIMB 🌐 2019 年 5 月 - 2019 年 8 月
全栈开发实习生 (Full Stack Developer Intern) 新加坡
- 参与开发 FXDBO Admin System, 完成文件上传、报表生成、融资信息计算等核心功能。
- 使用 Python、Django、PyODBC、HTML 与 SQL 构建端到端 Web 和数据库流程。
- 实现实时汇率转换、现金流计算等模块, 提升系统效率与可用性。

代表性论文

- [1] Xinze Li, Yuhang Zang, Yixin Cao, and Aixin Sun. 2026. **Skill-as-Pseudocode: Refactoring Skill Libraries to Pseudocode for LLM Agents**. In *Findings of EMNLP 2026*.
- [2] Xinze Li, Ziyue Zhu, Siyuan Liu, Yubo Ma, Yuhang Zang, Yixin Cao, and Aixin Sun. 2026. **EMemBench: Interactive Benchmarking of Episodic Memory for VLM Agents**. In *Findings of EMNLP 2026*.
- [3] Xinze Li, Yixin Cao, Yubo Ma, and Aixin Sun. 2025. **Long Context vs. RAG for LLMs: An Evaluation and Revisits**. Under review at *ACM Transactions on Information Systems (TOIS)*.
- [4] Xinze Li, Yixin Cao, Liangming Pan, Yubo Ma, and Aixin Sun. 2024. **Towards Verifiable Generation: A Benchmark for Knowledge-aware Language Model Attribution**. In *Findings of ACL 2024*.
- [5] Xinze Li, Yixin Cao, Muhao Chen, and Aixin Sun. 2023. **Take a Break in the Middle: Investigating Subgoals towards Hierarchical Script Generation**. In *Findings of ACL 2023*.
- [6] Bosheng Ding, Chengwei Qin, Ruochen Zhao, Tianze Luo, Xinze Li, Guizhen Chen, Wenhan Xia, Junjie Hu, Anh Tuan Luu, and Shafiq Joty. 2024. **Data Augmentation using LLMs: Data Perspectives, Learning Paradigms and Challenges**. In *Findings of ACL 2024*.
- [7] Yubo Ma, Yuhang Zang, Liangyu Chen, Meiqi Chen, Yizhu Jiao, Xinze Li, Xinyuan Lu, Ziyu Liu, Yan Ma, Xiaoyi Dong, Pan Zhang, Liangming Pan, Yu-Gang Jiang, Jiaqi Wang, Yixin Cao, and Aixin Sun. 2024. **MMLongBench-Doc: Benchmarking Long-context Document Understanding with Visualizations**. In *NeurIPS 2024 (Spotlight)*.

项目经历

- **Skill-as-Pseudocode: Refactoring Skill Libraries to Pseudocode for LLM Agents** 
博士课题 / 实习阶段工作 (导师: 孙爱欣、曹艺馨)
 - 设计并实现端到端流程, 通过解析、聚类、LLM 提取及 coverage、binding、replacement、risk 四项确定性检查, 将自由文本 Markdown skill library 转换为 typed pseudocode contract。
 - 构建带参数绑定与具体动作模板的层级检索时替换机制; 将 ALFWorld 胜局由 47/402 提升至 82/402, 同时降低 22.8% 的输入 token 与 14.5% 的 LLM 调用。
- **EMemBench: Interactive Benchmarking of Episodic Memory for VLM Agents** 
博士课题 / 实习阶段工作 (导师: 孙爱欣、曹艺馨)
 - 主导设计并实现程序化基准, 将智能体交互轨迹自动转换为均衡的问答集合, 并基于环境状态生成可验证的标准答案。
 - 构建问题模板库, 覆盖回忆、多跳、归纳、时间、空间与逻辑推理, 并在文本和视觉游戏环境中控制问题的可回答性。
- **Long Context vs. RAG for LLMs: An Evaluation and Revisits** 
博士课题 (导师: 孙爱欣、曹艺馨)
 - 构建统一评测框架, 从任务、检索器 (retriever)、上下文预算 (context budget) 等维度系统比较长上下文与 RAG 的优劣。
 - 系统性消融 (chunking、检索深度、reranking、prompting), 总结在真实应用中可落地的关键权衡与建议。
- **Towards Verifiable Generation: A Benchmark for Knowledge-aware Language Model Attribution** 
博士课题 (导师: 孙爱欣、曹艺馨)
 - 构建用于评测引用准确性与知识对齐 (grounding) 的基准, 并设计指标衡量 attribution 与生成质量。
- **Take a Break in the Middle: Investigating Subgoals towards Hierarchical Script Generation** 
研究助理阶段项目 (导师: 孙爱欣、曹艺馨)
 - 探索层级分解 (subgoals) 对脚本生成连贯性与质量的影响; 构建数据与基线并完成系统评测。
- **Mining Service Contributions from Web-based Journal Listings** 
本科毕业设计 (导师: Kan Min-Yen)
 - 面向噪声网页数据的编辑委员会信息抽取: 构建半自动数据流程并优化 NER/序列标注模型以提升鲁棒性。

学术活动与服务

- **SIGIR 2025 Tutorial: Long Context vs. RAG: Strategies for Processing Long Documents in LLMs** 
组织者 & 主讲
 - 主导该 tutorial 的讲授与内容组织; 完成整套讲义与幻灯片, 系统讲解长上下文方法、RAG 流程、评测设置与工程落地权衡。
- **审稿服务**
 - 会议审稿: NeurIPS 2025; ACL ARR (ACL 2024, 2025, 2026; NAACL 2024; EMNLP 2024, 2025)
 - 期刊审稿: ACM Transactions on Information Systems (TOIS)

奖学金

- **NTU Research Scholarship (全额奖学金)** 2023 年 8 月 - 至今
南洋理工大学 (Nanyang Technological University)
- **NUS Science & Technology Undergraduate Scholarship (全额奖学金)** 2017 年 8 月 - 2021 年 6 月
新加坡国立大学 (National University of Singapore)
- **SM1 新加坡政府全额奖学金** 2012 年 11 月 - 2016 年 12 月
初中毕业后经新加坡政府选拔获奖; 赴新加坡完成四年学业, 后考入新加坡国立大学

技能与资格

研究工程: LLM/VLM 智能体评测、基准设计、RAG 与信息检索、长上下文实验及可复现实验流程。
编程与工具: Python、PyTorch、Transformers、NumPy、Pandas、SQL、Git、 $\text{\LaTeX}$ 。
语言: 英语 (TOEFL 107)、中文。