

LI XINZE

S-Lab for Advanced Intelligence | Nanyang Technological University | Singapore

+65-9397-2127

xinze002@e.ntu.edu.sg

lixinze777.github.io

lixinze777

Google Scholar

XinzeLi777

LinkedIn

EDUCATION

- Nanyang Technological University** Aug 2023 – Present
Ph.D. Student, College of Computing and Data Science
Singapore, Singapore
– GPA: 4.17/5.00; Full scholarship: NTU Research Scholarship
- Purdue University** Aug 2019 – Dec 2019
Exchange
West Lafayette, USA
- National University of Singapore** Aug 2017 – Jun 2021
Bachelor of Computing (Computer Science) with Honours (Distinction)
Singapore, Singapore
– GPA: 4.10/5.00; Full scholarship: NUS Science & Technology Undergraduate Scholarship

RESEARCH INTERESTS

LLM/VLM Agents, Episodic Memory, Retrieval-Augmented Generation, Long-Context Reasoning, Information Retrieval.

WORK EXPERIENCE

- Shanghai AI Laboratory** Aug 2025 – Feb 2026
Research Intern
Shanghai, China
– Developed EMemBench largely end-to-end: interactive trajectory logging, verifiable ground-truth QA generation from underlying game signals, and a reproducible evaluation pipeline for LLM/VLM agents.
– Designed and implemented a template library spanning recall, multi-hop, induction, temporal/spatial reasoning, logical reasoning, and adversarial/answerability-controlled questions, with balancing and sanity checks.
– Ran large-scale experiments across text games and a visual game setting, and produced paper-ready results (tables, figures, and analyses).
- S-Lab for Advanced Intelligence** Sep 2021 – Jul 2023
Project Officer
Singapore, Singapore
– Contributed to the Multi-Modal Event Knowledge Graph (MMEKG), with responsibility for name disambiguation and support for causal-relation extraction.
– Led the Hierarchical Script Generation project from literature review and experiment design through data collection, implementation, and paper writing.
- CGS-CIMB** May 2019 – Aug 2019
Full Stack Developer Intern
Singapore, Singapore
– Developed core FXDBO Admin System features, including file uploads, statement generation, and financing calculations.
– Built end-to-end web and database workflows with Python, Django, PyODBC, HTML, and SQL.
– Implemented real-time currency-conversion and cash-flow calculation modules.

SELECTED PUBLICATIONS

- 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*.
- 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*.
- 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)*.
- 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*.
- 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*.
- 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*.
- 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)*.

PROJECTS

• Skill-as-Pseudocode: Refactoring Skill Libraries to Pseudocode for LLM Agents

Ph.D. Student / Research Intern (Supervised by Prof. Aixin Sun and Prof. Yixin Cao)



- Designed and implemented an end-to-end pipeline that converts free-form Markdown skill libraries into typed pseudocode contracts through parsing, clustering, LLM extraction, and deterministic coverage, binding, replacement, and risk checks.
- Built hierarchical retrieval-time substitution with grounded bindings and concrete action templates; improved ALFWorld wins from 47/402 to 82/402 while reducing input tokens by 22.8% and LLM calls by 14.5%.

• EMemBench: Interactive Benchmarking of Episodic Memory for VLM Agents

Ph.D. Student / Research Intern (Supervised by Prof. Aixin Sun and Prof. Yixin Cao)



- Led the design and implementation of a programmatic benchmark that converts agent trajectories into balanced QA sets with environment-derived, verifiable ground truth.
- Built a question-template suite covering recall, multi-hop, induction, temporal, spatial, and logical reasoning, with answerability control across text and visual game environments.

• Long Context vs. RAG for LLMs: An Evaluation and Revisits

Ph.D. Student (Supervised by Prof. Aixin Sun and Prof. Yixin Cao)



- Developed a controlled evaluation framework comparing long-context processing and RAG across tasks, retrievers, and context budgets.
- Conducted systematic ablations (chunking, retrieval depth, reranking, prompting) and summarized practical takeaways for real-world deployment.

• Towards Verifiable Generation: A Benchmark for Knowledge-aware Language Model Attribution

Ph.D. Student (Supervised by Prof. Aixin Sun and Prof. Yixin Cao)



- Led the development of a benchmark for knowledge-aware attribution, evaluating citation accuracy, knowledge grounding, and generation quality.

• Take a Break in the Middle: Investigating Subgoals towards Hierarchical Script Generation

Project Officer (Supervised by Prof. Aixin Sun and Prof. Yixin Cao)



- Led research on hierarchical script generation; built datasets and baselines to evaluate how explicit subgoals affect coherence and controllability.

• Mining Service Contributions from Web-based Journal Listings

Final Year Project (Supervised by Prof. Min-Yen Kan)



- Developed a semi-automated data pipeline and improved NER/sequence-tagging models to extract researcher names and editorial-board roles from noisy web pages.

TUTORIALS & ACADEMIC SERVICE

• ACM SIGIR 2025 Tutorial: Long Context vs. RAG: Strategies for Processing Long Documents in LLMs

SIGIR 2025

Organizer & Lead Speaker



- Led the tutorial delivery and prepared the complete slide deck; presented the majority of the tutorial content, covering long-context methods, RAG pipelines, evaluation, and deployment trade-offs.

• Reviewer Service

- **Conference:** NeurIPS 2025; ACL ARR (ACL 2024, 2025, 2026; NAACL 2024; EMNLP 2024, 2025)
- **Journal:** ACM Transactions on Information Systems (TOIS).

SCHOLARSHIPS

• NTU Research Scholarship (Full Scholarship)

Aug 2023 – Present

Nanyang Technological University

• NUS Science & Technology Undergraduate Scholarship (Full Scholarship)

Aug 2017 – Jun 2021

National University of Singapore

• Singapore Government SM1 Full Scholarship

Nov 2012 – Dec 2016

Selected by the Singapore government after junior high in China; studied in Singapore for four years, then entered NUS

QUALIFICATIONS

Research Engineering: LLM/VLM agent evaluation, benchmark design, RAG and information retrieval, long-context experimentation, and reproducible research pipelines.

Programming & Tools: Python, PyTorch, Transformers, NumPy, Pandas, SQL, Git, and $\LaTeX$.

Languages: English (TOEFL 107) and Mandarin.