

Xiangyu Zhang

Beijing, China — xiangyuz22@mails.tsinghua.edu.cn

EDUCATION

Tsinghua University

B.S. in Computer Science and Technology

Enrolled: Sept. 2022 — Expected: Jun. 2026

GPA 3.84 / 4.0

EXPERIENCE

Research Assistant

IEOR, University of California, Berkeley

Achievements:

Jan. 2025 - Present

Advised by **Prof. Zeyu Zheng**

- Built a 50k-scale dataset for synthetic object insertion with minimal shadow processing, and a 10k-scale version with clean shadow removal.

Research Assistant

Institute for AI Industry Research, Tsinghua University

Achievements:

Jun. 2024 - Apr. 2025

Advised by **Prof. Weizhi Ma**

- Introduced a retrieval-augmented generation framework that improves LLM inference by dynamically clustering retrieved documents to remove noise, redundancy, and irrelevant content.

Contributions:

- Lead a project to explore knowledge-irrelevant domain text transferring for AI agents.
- Collected pairwise data and trained retrievers for hint transferring.
- Contributed to the RAG project by running experiments, refining the core idea, and restructuring the entire paper.

Chief Information Officer / Project Manager

Yixin Technology Inc.

Achievements:

Mar. 2024 - Dec. 2024

- Developed an application with mental health care and AI chatbot service.

Contributions:

- Led an IT team of ten students from UC Berkeley, University of Michigan, Tsinghua University, and etc.
- Led and participated in the full-stack implementation of back-end, front-end, UI design, deployment, and testing.

Research Assistant

Tsinghua Laboratory of Brain and Intelligence

Achievements:

Oct. 2023 - Oct. 2024

Advised by **Prof. Sen Song**

- Established a RetNet based model with specific biological constraints.

Contributions:

- Completed fitting and visualization of network neuron activity and fMRI signal of voxels.

Research Assistant

Tsinghua Pervasive Human Computer Interaction Lab

Achievements:

Aug. 2023 - Nov. 2023

Advised by **Prof. Yuanchun Shi**

- Introduced a new paradigm in which an agent accomplishes unknown tasks in an unknown environment.
- Presented a ready-to-use VCI named AutoTask, where end users can automate any intent with a single command.

Contributions:

- Analysed the work of previous papers and built a baseline model for the research.
- Assisted in completing user experiments.

PROJECTS

Enhancing Large Language Model's Coding Ability by Tree-Based Searching Methods

Dec. 2024 - Jan. 2025

Project Link: https://github.com/Painkillerzzz/code_contest

- Investigated reinforcement learning techniques to enhance the code generation capabilities of large language models.
- Conducted experiments on a dataset comprising 76 advance programming problems and their respective test cases.
- Implemented and evaluated four approaches: Best-of-N (Vanilla), MCTS-Append, MCTS-Modify, and Tree of Thought (ToT).
- Achieved notable success with the ToT method on simpler problems, demonstrating higher pass rates and efficient use of computational budgets under full test case feedback.

- Identified challenges with tree-based methods and proposed potential refinements and prompt designs to improve robustness and applicability in real-world scenarios.

I-Heart: A Chatbot for Mental Health and Psychotherapy

Mar. 2024 - Jul. 2024

Project Link: coming soon

- Developed a chatbot consists of fine-tuned LLMs and text-to-speech models.
- Provide clients with customized treatments, supported by professional conversation and styled tune.

The Dance of Fire and Ice: A Hardware Music Game

May. 2024 - Jun. 2024

Project Link: <https://github.com/Painkillerzzz/digital-logic-design>

- Developed a music video game based on FPGA and hardware design language.

Generals: An AI Agent Competition Game

Sept. 2023 - Jan. 2024

Project Link: <https://www.saiblo.net/game/35>

- Developed an agent competition game and deployed it on a platform where competitors can submit their agents to fight against each other.
- Completed animations and model action control, participated in writing and debugging UI logic.

What's News: A Convenient News App

Aug. 2023 - Sept. 2023

Project Link: <https://github.com/Painkillerzzz/WhatsNews>

- Developed a convenient Android news application with user database and elegant front-end interface.

PUBLICATIONS

Weitao Li, Kaiming Liu, **Xiangyu Zhang**, Xuanyu Lei, Weizhi Ma, Yang Liu, "Efficient Dynamic Clustering-Based Document Compression for Retrieval-Augmented-Generation", arXiv:2504.03165 [cs.CL]

Lihang Pan, Bowen Wang, Chun Yu, Yuxuan Chen, **Xiangyu Zhang**, Yuanchun Shi, "AutoTask: Executing Arbitrary Voice Commands by Exploring and Learning from Mobile GUP", arXiv:2312.16062 [cs.HC]

HONORS AND AWARDS

- | | |
|--|---------------------------|
| • Social Practice Excellence Scholarship | Tsinghua University, 2023 |
| • Literary and Art Excellence Scholarship | Tsinghua University, 2023 |
| • Exllence Entrance Scholarship | Tsinghua University, 2022 |

SKILLS

- **Language:** English (TOEFL 102), Chinese (native).
- **Programming:** Python, C/C++, Java, Kotlin, C#, SystemVerilog
- **Tool:** Pytorch, HuggingFace, SpikingJelly, Unity3D