Yuxuan Cao

Homepage:https://caoyxuan.github.io/
+86-18888913599 | ycaoce@connect.ust.hk

Education

Hong Kong University of Science and Technology — Hong Kong, China
PhD., Computer Science and Engineering Department — *Sep. 2025 - Present*
- Advisor: Junxian He

Zhejiang University — Hangzhou, China
M.S., Electronics Information Science, Computer Science and Technology College — *Sep. 2022 - Mar. 2025*
- Advisor: Yang Yang and Yueting Zhuang
- GPA: 3.93/4.00

Zhejiang University — Hangzhou, China
B.S., Computer Science and Technology, CKC Honors College — *Sep. 2018 - Jul. 2022*
- Overall GPA:3.92/4.00, Major GPA: 3.99/4.0, Rank: 7/187
- Thesis advisor: Yang Yang

Experience

Moonshot AI — 2026.05 – 2026.09
Post-training, Coding Team
- **Coding-Agent Session Replay:** Reconstructed production coding-agent sessions by merging fragmented logs, restoring tool-call histories, and replaying actions in executable sandboxes; saved recovered working states as reusable sandbox images for multi-turn evaluation from realistic mid-task states.
- **Long-Horizon PRD Task Synthesis:** Built a pipeline to synthesize PRD-style coding tasks from production requests, covering task materialization, specification composition, environment validation, and rubric generation.
- **Multi-Turn Task Synthesis and Analysis** Developed task-generation pipelines for multi-turn coding benchmarks (SWE-Interact, SWE-Milestone, and EvoCode-Bench) using progressive requirement disclosure, automatic milestone splitting, and cumulative revision verification; evaluated agents' multi-turn capabilities through controlled experiments.
- **Embodied Coding Agents:** Developed an automated task-synthesis pipeline that translates agent failure patterns into compositional robotics tasks, generating simulation scenes, task specifications, and executable graders to target long-horizon planning and physical reasoning. Built evaluation workflows around EmbodiedSWE and BiGym2, analyzing agents' code–simulation–feedback cycles to guide task design and diagnose failures in debugging, control, and policy delivery.

HKUST-NLP Lab — 2025.06 – Present
Student Researcher. Advisor: Prof. Junxian He
- **MINE ODYSSEY [Under Review],** a benchmark for agentic spatial intelligence using Minecraft reconstructions of 30 real-world regions across 20 countries and regions. Its 180 manually reviewed tasks require agents to visit annotated, meaningful destinations such as buildings and rooms in a specified order, testing exploration, navigation, and interaction. Across eight evaluated models, the top two achieved 85.6% and 50.6% task success, while the strongest open-weight model achieved 23.9%.
- Co-developed **Toolathlon [ICLR'26],** a benchmark with 600+ realistic tools and long-horizon tasks for evaluating language agents' general tool-use capabilities. Built the evaluation infrastructure including containerized task environments, decoupled agent loop architecture, parallel execution pipeline, and trajectory visualization tools.

Wechat, Tencent — 2025.06 – 2025.12
Research Intern. Advisor: Prof. Junxian He & Dr. Zifei Shan
- **Self-Evolving Tool Interfaces for LLM Agents:** Proposed evolving tool interfaces by extracting tool-specific output compression rules and description corrections from agent failure trajectories and deploying them across tasks using the same tools. Improved Toolathlon performance by 16.7 percentage points without model training; patches derived from 10 analyzed failures covered 81% of 68 failed tasks.
- Designed a latent reasoning framework replacing explicit CoT text with learnable latent tokens, trained via dual-loss knowledge distillation (answer loss + CoT distillation with progressive warmup).
- Implemented parallel CoT decoder and multi-token prediction (MTP) with multiple optimized variants for inference acceleration.

Ant Group Alipay 2024.12 – 2025.03

Research Intern. Advisor: Dr.Sheng Guo

- Developed a redundancy-aware acceleration framework for DGNNs [AAAI'26], eliminating both static and runtime redundancy to address the high computational cost of deploying DGNNs on temporal graphs, achieving an average 83.49% runtime reduction across diverse architectures while preserving model accuracy.

AINet Lab at Zhejiang University 2022.06 – 2025.03

Student Researcher. Advisor: Prof. Yang Yang and Prof. Jiarong Xu

- **Graph Anomaly Detection with Auxiliary Data [KDD'25]:** Proposed a data-centric framework selecting representative and diverse auxiliary graphs, enhancing anomaly detection performance.
- **Efficient Graph Pre-training [KDD'23, NeurIPS'23]:** Developed a "better-with-less" framework that iteratively selects informative training data, achieving superior downstream performance with fewer training samples.
- **Scalable Graph Attacks:** Designed an embedding-space attack method to enable PGD-style adversarial attacks on large-scale graphs.

Publications

Yuxuan Cao, Junlong Li, Hao Li, and Junxian He. *Mine Odyssey: Benchmarking Spatial Agentic Intelligence in the Wild.* Under review.

Yuxuan Cao, Jiarong Xu, Chen Zhao, Jiaan Wang, Carl Yang, Chunping Wang, and Yang Yang. *How to Use Graph Data in the Wild to Help Graph Anomaly Detection?* In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'25).

Yuxuan Cao*, Jiarong Xu*, Carl Yang, Jiaan Wang, Yunchao Mercer Zhang, Chunping Wang, Lei Chen, and Yang Yang. *When to Pre-Train Graph Neural Networks? From Data Generation Perspective!* In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'23).

Kimi Team (including **Yuxuan Cao**). *Kimi K3: Open Frontier Intelligence.* arXiv:2607.24653, 2026.

Renhong Huang, Ning Tang, Jiarong Xu, **Yuxuan Cao**, Qingqian Tu, Sheng Guo, Bo Zheng, Huiyuan Liu, and Yang Yang. *PolicySim: An LLM-Based Agent Social Simulation Sandbox for Proactive Policy Optimization.* In Proceedings of the ACM Web Conference (WWW'26).

Junlong Li*, Wenshuo Zhao*, Jian Zhao*, Weihao Zeng*, Haoze Wu*, Xiaochen Wang, Rui Ge, **Yuxuan Cao**, Yuzhen Huang, Wei Liu, Junteng Liu, Zhaochen Su, Yiyang Guo, Fan Zhou, Lueyang Zhang, Juan Michelini, Xingyao Wang, Xiang Yue, Shuyan Zhou, Graham Neubig, and Junxian He. *The Tool Decathlon: Benchmarking Language Agents for Diverse, Realistic, and Long-Horizon Task Execution.* In Proceedings of the International Conference on Learning Representations (ICLR'26).

Renhong Huang, **Yuxuan Cao**, Yi Li, Junwei Hu, Zihua Xiong, Shuai Fang, Sheng Guo, Bo Zheng, and Yang Yang. *Trimming the Fat: Redundancy-Aware Acceleration Framework for DGNNs.* In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'26).

Jiaan Wang, Fandong Meng, Zengkui Sun, Yunlong Liang, **Yuxuan Cao**, Jiarong Xu, Haoxiang Shi, and Jie Zhou. *An Empirical Study of Many-to-Many Summarization with Large Language Models.* In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL'25).

Jiaan Wang, Yunlong Liang, Zengkui Sun, **Yuxuan Cao**, Jiarong Xu, and Fandong Meng. *Cross-Lingual Knowledge Editing in Large Language Models.* In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL'24).

Jiarong Xu, Renhong Huang, Xin Jiang, **Yuxuan Cao**, Carl Yang, Chunping Wang, and Yang Yang. *Better with Less: A Data-Active Perspective on Pre-Training Graph Neural Networks.* In Advances in Neural Information Processing Systems (NeurIPS'23).

Selected Awards And Honors

- Hong Kong PhD Fellowship Scheme, 2025
- Outstanding Graduate of Zhejiang Province, 2022 & 2025
- National Scholarship of China, 2023