Zhaoyang Chu

zhaoyang.chu.25@ucl.ac.uk | zhaoyang-chu.github.io

Research Interests

My research interests lie at the intersection of software engineering and artificial intelligence, with an emphasis on **reliable coding agents** for realistic software engineering workflows.

Education

University College London — **Co-Supervised by Prof. Federica Sarro and Dr. He Ye**
Ph.D. in Computer Science — *Sep 2025 – Sep 2028 (anticipated)*

Huazhong University of Science and Technology — GPA: 3.53/4.0
*M.E. in Computer Science and Technology (**Graduated with Honors**)* — *Sep 2022 – Jun 2025*

Huazhong Agricultural University — GPA: **3.93**/4.0
*B.E. in Data Science and Big Data Technology (**Graduated with Honors**)* — *Sep 2018 – Jun 2022*

Research Papers

** indicates equal contribution. † indicates the corresponding author.*

[1] Prometheus: Towards Long-Horizon Codebase Navigation for Repository-Level Problem Solving.
Yue Pan, Zimin Chen*, Siyu Lu, **Zhaoyang Chu**, Xiang Li, Han Li, Yang Feng, Claire Le Goues, Federica Sarro, Martin Monperrus, He Ye†.*
Under Review at **ISSTA 2026**.

[2] CONTEXTBENCH: A Benchmark for Context Retrieval in Coding Agents.
Han Li, Letian Zhu, Bohan Zhang*, Rili Feng*, Jiaming Wang, Yue Pan, Earl T. Barr, Federica Sarro, **Zhaoyang Chu†**, He Ye†.*
Under Review at **ICML 2026**.

[3] I Fill It, You Run It: Completing Repository Documentation for Automated Environment Setup.
*Zhenyu Yang, **Zhaoyang Chu**, Zhongxing Yu†, He Ye.*
Under Review at **ASE 2026**.

[4] FixAudit: Closing the Test-and-Repair Loop for Competitive Code Generation.
*Lingxiao Tang, Muyang Ye, **Zhaoyang Chu**, Xiaoxue Ren, Zhongxin Liu, Lingfeng Bao†, He Ye.*
Under Review at **ASE 2026**.

[5] Scrub It Out! Erasing Sensitive Memorization in Code Language Models via Machine Unlearning.
***Zhaoyang Chu**, Yao Wan†, Zhikun Zhang, Di Wang, Zhou Yang, Hongyu Zhang, Pan Zhou, Xuanhua Shi, Hai Jin, David Lo.*
ICSE 2026.

[6] Hallucinations in LLM-based Code Summarization: Unveiling, Detection, and Mitigation.
*Guanghua Wan, Yuanning Feng, Yao Wan†, **Zhaoyang Chu**, Zhangqian Bi, Junxiao Han, Zhou Zhao, Hongyu Zhang, Pingpeng Yuan, Xuanhua Shi, Hai Jin.*
FSE 2026.

[7] ExecVerify: White-Box RL with Verifiable Stepwise Rewards for Code Execution Reasoning.
*Lingxiao Tang, He Ye, **Zhaoyang Chu**, Muyang Ye, Zhongxin Liu, Xiaoxue Ren, Lingfeng Bao†.*
ACL 2026.

[8] CGBridge: Bridging Code Graphs and Large Language Models for Better Structure-Aware Code Understanding.
*Zeqi Chen, **Zhaoyang Chu**, Yi Gui, Feng Guo, Yao Wan, Chuan Shi†.*
ACL 2026 Findings.

[9] CODESYNC: Synchronizing Large Language Models with Dynamic Code Evolution at Scale.
Chenlong Wang, **Zhaoyang Chu***, Zhengxiang Cheng*, Xuyi Yang, Kaiyue Qiu, Yao Wan†, Zhou Zhao, Xuanhua Shi, Dongping Chen.*
ICML 2025.

[10] How to Select Pre-Trained Code Models for Reuse? A Learning Perspective.
*Zhangqian Bi, Yao Wan†, **Zhaoyang Chu**, Yufei Hu, Junyi Zhang, Hongyu Zhang, Guandong Xu, Hai Jin.*
SANER 2025.
IEEE TCSE Distinguished Paper Award 🏆.

[11] Can Large Language Models Serve as Evaluators for Code Summarization?
*Yang Wu, Yao Wan†, **Zhaoyang Chu**, Wenting Zhao, Ye Liu, Hongyu Zhang, Xuanhua Shi, Philip S. Yu.*
IEEE Transactions on Software Engineering (TSE), 2025.

[12] Wait, We Don't Need to "Wait"! Removing Thinking Tokens Improves Reasoning Efficiency.
*Chenlong Wang, Yuanning Feng, Dongping Chen, **Zhaoyang Chu**, Ranjay Krishna†, Tianyi Zhou†.*
EMNLP 2025 Findings.

[13] TESTEVAL: Benchmarking Large Language Models for Test Case Generation.
Wenhan Wang, Chenyuan Yang*, Zhijie Wang*, Yuheng Huang, **Zhaoyang Chu**, Da Song, Lingming Zhang, An Ran Chen, Lei Ma.*
NAACL 2025 Findings.

[14] Graph Neural Networks for Vulnerability Detection: A Counterfactual Explanation.
***Zhaoyang Chu**, Yao Wan†, Qian Li, Yang Wu, Hongyu Zhang, Yulei Sui, Guandong Xu, Hai Jin.*
ISSTA 2024.

Research Experience

Ph.D. Student Sep 2025 – Present
University College London, co-advised by Prof. He Ye and Prof. Federica Sarro

Memory-Enhanced Coding Agent (Under Review at ISSTA 2026)
- *Develop a memory-augmented coding agent that maintains and reuses previously explored code contexts to support long-horizon codebase navigation and repository-level problem solving.*
- *Collaborated with **Prof. Martin Monperrus at KTH Royal Institute of Technology** and **Prof. Claire Le Goues at Carnegie Mellon University**.*

Context Retrieval Benchmarking for Coding Agents (Under Review at ICML 2026)
- *Curate a benchmark with human-annotated gold contexts to evaluate how coding agents retrieve and utilize relevant code contexts during repository-level problem solving.*
- *Collaborated with **Prof. Earl T. Barr at University College London**.*

Code Execution Reasoning for LLMs (ACL 2026)
- *Develop a reinforcement learning approach for training LLMs with verifiable stepwise rewards derived from intermediate code execution traces.*
- *Collaborated with **Prof. Lingfeng Bao at Zhejiang University**.*

Research Intern May 2024 – Jun 2025
University of Illinois Urbana-Champaign, advised by Prof. Lingming Zhang

Test Case Generation Benchmarking for LLMs (NAACL 2025 Findings)
- *Propose a novel benchmark that evaluates LLMs' capabilities in generating test cases for Python programs.*

M.S. Student Sep 2022 – Jun 2025
Huazhong University of Science and Technology, advised by Prof. Yao Wan

Machine Unlearning for Code LLMs (ICSE 2026)
- *Develop a privacy-preserving method to erase sensitive information from code LLMs via machine unlearning.*

- *Collaborated with **Prof. David Lo at Singapore Management University**.*

LLM Synchronization with Code Evolution (ICML 2025)
- *Propose a novel benchmark to evaluate LLMs' synchronization with real-time library API updates.*
- *Collaborated with **Prof. Zhou Zhao at Zhejiang University**.*

Counterfactual Vulnerability Detection (ISSTA 2024)
- *Design a counterfactual explainer to uncover the decision mechanisms of GNN-based detection systems.*
- *Collaborated with **Prof. Yulei Sui at University of New South Wales**.*

Code LLM Selection (SANER 2025 Distinguished Paper)
- *Propose learning-based methods for efficiently selecting and reusing pre-trained code LLMs for target software engineering tasks within limited computational budgets.*
- *Collaborated with **Prof. Hongyu Zhang at Chongqing University**.*

LLM-as-a-Judge for Code Summarization (TSE)
- *Develop an LLM-based evaluator to asses the quality of code summaries generated by neural models.*
- *Collaborated with **Prof. Philip S. Yu at University of Illinois at Chicago** .*

Efficient Reasoning for LLMs (EMNLP 2025 Findings)
- *Propose a novel approach that disables explicit self-reflection by suppressing "Wait"-like tokens during inference.*
- *Collaborated with **Prof. Ranjay Krishna at University of Washington**.*

Hallucination Detection and Mitigation for Code LLMs (FSE 2026)
- *Propose a unified framework to detect and mitigate hallucinations in LLM-based code summarization, combining a dedicated benchmark, a high-accuracy detector, and an inference-time mitigation strategy..*
- *Collaborated with **Prof. Hai Jin at Huazhong University of Science and Technology**.*

Structure-Aware Code Understanding for LLMs (ACL 2026 Findings)
- *Propose a plug-and-play framework that encodes program structure into a compact embedding prefix for LLMs via a trainable bridge module, enabling efficient structure-aware code understanding during inference.*
- *Collaborated with **Prof. Chuan Shi at Beijing University of Post and Telecommunication**.*

Honors & Awards

IEEE TCSE Distinguished Paper Award 2025