Yang Chen

✉ yangc9@illinois.edu
https://yangc9.github.io

Research Interests

AI and Software Engineering (SE): LLM, agentic systems, model evaluation, synthetic data generation, post-training, code generation and reasoning, program analysis, software engineering, software testing, AI integrated with program analysis to solve real-world SE problems.

Education

2022-Present **University of Illinois Urbana-Champaign (USA)**.
Ph.D. Candidate in Computer Science (Anticipated graduation: 2027 May)
Co-advisors: Reyhaneh Jabbarvand and Darko Marinov

2018–2022 **Huazhong University of Science and Technology (China)**.
B.Sc. in Computer Science

Publications and Preprints

[1] Evaluating the Generalizability of LLMs to Real-World Complexity .
(under review) **Yang Chen**, Shuyang Liu, and Reyhaneh Jabbarvand.

[2] Enhancing SWE Issue Repair with Regression Tests .
(under review) **Yang Chen**, Toufique Ahmed, Reyhaneh Jabbarvand, and Martin Hirzel.

[3] Process-Centric Analysis of Agentic Software Systems [Intro].
(under review) Shuyang Liu, **Yang Chen**, Rahul Krishna, Saurabh Sinha, and Reyhaneh Jabbarvand.

[4] ICSE 2026 Assessing Coherency and Consistency of Code Execution Reasoning by LLMs [PDF] .
Changshu Liu, **Yang Chen**, and Reyhaneh Jabbarvand.
Proceedings of the 48th International Conference on Software Engineering. Rio de Janeiro. Brazil. April 2026.

[5] ISSTA 2024 Neurosymbolic Repair of Test Flakiness [PDF] .
Yang Chen and Reyhaneh Jabbarvand.
Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis. Vienna, Austria. September 2024.

[6] ICSE-SRC 2024 Flakiness Repair in the Era of Large Language Models [PDF] .
Yang Chen.
[2nd Place in Student Research Competition] Proceedings of the 46th International Conference on Software Engineering, Lisbon, Portugal. April 2024.

[7] ICSE-FTW 2024 Can ChatGPT Repair Non-Order-Dependent Flaky Tests? [PDF] .
Yang Chen and Reyhaneh Jabbarvand.
Flaky Test Workshop in Proceedings of the 46th International Conference on Software Engineering. Lisbon, Portugal. April 2024.

[8] ISSTA 2023 Transforming Test Suites into Croissants [PDF] .
Yang Chen, Alperen Yildiz, Darko Marinov, and Reyhaneh Jabbarvand.
Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, Seattle, USA. July 2023.

[9] Evaluating Code Reasoning Abilities of Large Language Models Under Real-World Settings .
(under review) Changshu Liu, Alireza Ghazanfari, **Yang Chen**, and Reyhaneh Jabbarvand.

[10] Can Large Language Models Reason About Code?.
(under review) Changshu Liu, Shizhuo Zhang, **Yang Chen**, and Reyhaneh Jabbarvand.

[11] preprint Automated Bug Generation in the Era of Large Language Models [PDF] .
Ali Reza Ibrahimzada, **Yang Chen**, Ryan Rong, and Reyhaneh Jabbarvand.
arXiv Preprint, 2023.

[12] ICSE-Demo 2022 iPFlakies: A Framework for Detecting and Fixing Python Order-Dependent Flaky Tests [PDF] .
Ruixin Wang, **Yang Chen**, and Wing Lam.
Demonstration Track, Proceedings of the 44th International Conference on Software Engineering, Pittsburgh, USA. May 2022.

[13] ICSE-FTW 2025 (short paper) A Preliminary Study of Fixed Flaky Tests in Rust Projects on GitHub [PDF] .
Tom Schroeder, Minh Phan, and **Yang Chen**.
Flaky Test Workshop in Proceedings of the 47th International Conference on Software Engineering. Ottawa, Canada. April 2025.

Experience

2025 May – Aug **IBM Research Scientist Intern**, *IBM Research, NY*.
Manager & Mentor: Martin Hirzel and Toufique Ahmed.

Selected Honors and Grants

2024 Ranked 2nd in the 46th ACM Student Research Competition at ICSE 2024.
2023 SIGSOFT CAPS Grants for ISSTA 2023, ICSE 2024, and ISSTA 2024.
2022 Outstanding Graduate of Class 2022, Huazhong University of Science and Technology.

Academic Service

Reviewer: MSR 2024, TSE 2025; Artifact Evaluation PC: ISSTA 2024 & 2025.

Research Summary

My research lies in AI and software engineering (SE):

(1) Agentic LLM-based systems that integrate program analysis to address real-world problems, e.g., SWE issue repair [2] and flaky tests repair [5,6,7,13];

(2) Synthetic data generation to benchmark LLMs on complex coding tasks with real-world difficulty [1,9], study LLM code reasoning capabilities [4,10], and evaluate flakiness detection tools [8];

(3) Post-training of LLMs using high-quality synthetic data [1].

(4) My ongoing research focuses on building agentic systems for efficient code generation and conducting process-centric analysis of agents.

Others. Prior research projects have also equipped me with diverse skills in *neurosymbolic program analysis*, *genetic algorithms and LLM fine-tuning*, as well as software testing including *flaky tests detection & repair* (**accumulating to 139 patches accepted in real world**), *mutation testing* and *test suite minimization*. I usually code with Python and Java.