

Chenyang Yang

cyang3@cs.cmu.edu • <https://cs.cmu.edu/~cyang3>

EDUCATION

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Ph.D. student at School of Computer Science

Aug 2021 – Present

- Adviser: Prof. Christian Kästner, Prof. Sherry Tongshuang Wu
- Focus: Human-centered AI, AI engineering, AI model testing and evaluation.

Peking University, Beijing, China

B.S. (Summa Cum Laude) in Computer Science, Turing Class

Sep 2017 – Jul 2021

- Adviser: Prof. Yingfei Xiong
- Cumulative GPA: 3.83 / 4.00, Ranking: 5 / 230

CONFERENCE PUBLICATIONS

- [1] **What Prompts Don't Say: Understanding and Managing Underspecification in LLM Prompts**
Chenyang Yang, Yike Shi, Qianou Ma, Michael Xieyang Liu, Christian Kästner, Tongshuang Wu.
Findings of the Association for Computational Linguistics (ACL), 2026.
- [2] **"I Don't Think RAI Applies to My Model": Engaging Non-champions with Sticky Stories for Responsible AI Work**
Nadia Nahar, Chenyang Yang, Yanxin Chen, Wesley Hanwen Deng, Ken Holstein, Motahhare Eslami, Christian Kästner.
Conference on Human Factors in Computing Systems (CHI), 2026.
- [3] **Designing Abandabot: When Does Open Source Dependency Abandonment Matter?**
Courtney Miller, Hao He, Weigen Chen, Elizabeth Lin, Chenyang Yang, Bogdan Vasilescu, and Christian Kästner.
International Conference on Software Engineering (ICSE), 2026.
- [4] **cAST: Enhancing Code Retrieval-Augmented Generation with Structural Chunking via Abstract Syntax Tree**
Yilin Zhang, Xinran Zhao, Zora Zhiruo Wang, Chenyang Yang, Jiayi Wei, and Tongshuang Wu.
Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2025.
- [5] **SPHERE: An Evaluation Card for Human-AI Systems**
Qianou Ma, Dora Zhao, Xinran Zhao, Chenglei Si, Chenyang Yang, Ryan Louie, Ehud Reiter, Diyi Yang, Tongshuang Wu.
Findings of the Association for Computational Linguistics (ACL), 2025.
- [6] **Orbit: A Framework for Designing and Evaluating Multi-objective Rankers**
Chenyang Yang, Tesi Xiao, Michael Shavlovsky, Christian Kästner, Tongshuang Wu.
International Conference on Intelligent User Interfaces (IUI), 2025.
- [7] **LLMs as Workers in Human-Computational Algorithms? Replicating Crowdsourcing Pipelines with LLMs**
Tongshuang Wu, Haiyi Zhu, ..., Chenyang Yang.
Conference on Human Factors in Computing Systems (CHI Case Study), 2025.
- [8] **What Is Wrong with My Model? Identifying Systematic Problems with Semantic Data Slicing**
Chenyang Yang, Yining Hong, Grace Lewis, Tongshuang Wu, Christian Kästner.
International Conference on Automated Software Engineering (ASE), 2024
- [9] **A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges**
Jenny T Liang, Chenyang Yang, and Brad A Myers.
International Conference on Software Engineering (ICSE), 2024.
- [10] **(Why) Is My Prompt Getting Worse? Rethinking Regression Testing for Evolving LLM APIs**
Wanqin Ma*, Chenyang Yang*, and Christian Kästner.
International Conference on AI Engineering (CAIN), 2024.

- [11] **Beyond Testers' Biases: Guiding Model Testing with Knowledge Bases using LLMs**
Chenyang Yang, Rishabh Rustogi, Rachel A Brower-Sinning, Grace Lewis, Christian Kästner, and Tongshuang Wu.
Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.
- [12] **Data Leakage in Notebooks: Static Detection and Better Processes**
Chenyang Yang, Rachel A Brower-Sinning, Grace Lewis, and Christian Kästner.
International Conference on Automated Software Engineering (ASE), 2022.
- [13] **Subtle Bugs Everywhere: Generating Documentation for Data Wrangling Code**
Chenyang Yang, Shurui Zhou, Jin L.C. Guo, and Christian Kästner.
International Conference on Automated Software Engineering (ASE), 2021.
- [14] **Accelerating Program Analyses in Datalog by Merging Library Facts**
Yifan Chen, Chenyang Yang, Xin Zhang, Yingfei Xiong, Hao Tang, Xiaoyin Wang, Lu Zhang.
International Static Analysis Symposium (SAS), 2021.

JOURNAL PUBLICATIONS

- [1] **What Should We Engineer in Prompts? Training Humans in Requirement-Driven LLM Use**
Qianou Ma, Weirui Peng, Chenyang Yang, Hua Shen, Kenneth Koedinger, Tongshuang Wu.
ACM Transactions on Computer-Human Interaction (TOCHI), 2025.
- [2] **Accelerating Patch Validation for Program Repair with Interception-Based Execution Scheduling**
Yuan-An Xiao, Chenyang Yang, Bo Wang, and Yingfei Xiong.
Transactions on Software Engineering (TSE), 2024.

DEMOS AND WORKSHOP PAPERS

- [1] **Differential Performance Fuzzing of Configuration Options**
Haesue Baik, Chenyang Yang, Vasudev Vikram, Pooyan Jamshidi, Rohan Padhye, Christian Kästner.
International Workshop on Search-Based and Fuzz Testing (SBFT), 2025.
- [2] **Capabilities for Better ML Engineering**
Chenyang Yang, Rachel A Brower-Sinning, Grace Lewis, Christian Kästner, and Tongshuang Wu.
AAAI-23 Workshop on Artificial Intelligence Safety (SafeAI), 2023.
- [3] **ExpressAPR: Efficient Patch Validation for Java Automated Program Repair Systems**
Yuan-An Xiao, Chenyang Yang, Bo Wang, Yingfei Xiong.
International Conference on Automated Software Engineering (ASE) – Tool Demo Track, 2023.

PREPRINTS

- [1] **Understanding Prompt Programming Tasks and Questions**
Jenny T Liang, Chenyang Yang, Agnia Sergeyuk, Travis D Breaux, Brad A Myers.
arXiv preprint arXiv:2507.17264, 2025.

INDUSTRY EXPERIENCE

- Meta**, Intern @ Modern Recommendation Systems May 2025 – Aug 2025
 - Worked on multi-modal LLM post-training to produce quality signals for social media recommendation systems.
 - **Data**: Built a recurring data pipeline to collect millions of training data; worked on human annotation guidelines for gold human eval data; prompted large models to curate synthetic data.
 - **Modeling**: Experimented with various training and data recipes to improve model performance on AIGC detection.
 - **Benchmarking**: Reproduced the pipeline for various open-sourced models and benchmarked on open-sourced datasets.
- Amazon**, Applied Scientist Intern @ Amazon Search May 2024 – Dec 2024
 - Worked on an interactive system to support ML practitioners design and evaluate multi-objective recommendation systems
 - **Formative Study**: Engaged with practitioners to understand their pain points in building multi-objective recommendation systems (interviews, observational study)
 - **System Building**: Built an interactive system (Svelte + Python) for better design and evaluation of multi-objective recommendation systems with practitioner feedback

	▪ User Study: Evaluated the system on real-world recommendation system debugging tasks with practitioners with controlled experiments	
PROJECT HIGHLIGHTS	What Prompts Don't Say: Understanding and Managing Underspecification in LLM Prompts ▪ Automated Specification Discovery: Built a series of LLM workflows to gather task specifications through analyzing human prompts, past LLM responses, and task failure modes. ▪ Automated Prompt Optimization: Built automated prompt optimizers to optimize prompts based on LLM-as-a-judge rubric feedback. Beyond Testers' Biases: Guiding Model Testing with Knowledge Bases using LLMs ▪ Knowledge Extraction from LLM: Built an interactive interface (React + Python) to allow users to interactively extract a comprehensive knowledge graph from LLM with ConceptNet relations. ▪ User Study: Evaluated the system's effectiveness in supporting model testing with a controlled user experiment.	
INVITED TALKS	Designing and Evaluating Multi-objective Rankers, Indeed.com Jul 2025 Introduction to Human-AI Interaction Research, Amazon Search Aug 2024 Capabilities for Better ML Engineering, Siemens Mar 2023	
PROFESSIONAL SERVICE	▪ Organizer for Beyond the Model Meeting (2025). ▪ Program Committee for Bi-Align Workshop @ ICLR (2025), IUI (2026). ▪ Reviewer for EMNLP (2023), ACL ARR (2024), TSE (2024), CHI LBW (2025), TOSEM (2025), HCI+NLP @ EMNLP (2025), CHI (2026). ▪ Student Volunteer for ASE (2020).	
TEACHING EXPERIENCE	Guest Lecture: AI for Quality Assurance for AI Dec 2025 Quality Assurance, Carnegie Mellon University Guest Lecture: Design and Architecture for ML-enabled systems Sep 2024 Machine Learning in Production, Carnegie Mellon University Machine Learning in Production, Carnegie Mellon University Aug 2024 – Dec 2024 Instructor: Christian Kästner, Sherry Tongshuang Wu Led weekly recitations and designed labs on model testing; redesigned homework on building AI products with a larger-scale codebase, while encouraging students to learn to collaborate with coding agents Large Language Model Systems, Carnegie Mellon University Jan 2024 – May 2024 Instructor: Lei Li Mentored student projects and presentations, and designed homework on distributed model training Introduction to Computer Systems, Peking University Sep 2020 – Jan 2021 Designed recitation materials and led weekly recitations for a group of 14 students	
MENTORING	Cassandra Shi Jun 2024 – May 2025 <i>CMU undergraduate student</i> Contributed to the development of the requirements-driven prompting system Yining Hong Jul 2023 – Sep 2023 <i>Tsinghua University undergraduate student (Currently a CMU PhD student)</i> Explored the problem of semantic slicing and contributed to algorithm design for the problem Wanqin Ma Jun 2023 – Nov 2023 <i>HKUST undergraduate student (Currently a HKUST PhD student)</i> Explored the problem of LLM regression testing and conducted evaluations	
AWARDS & HONORS	▪ Award for Research Excellence, Peking University Oct 2020 ▪ Turing Scholarship, Peking University Dec 2019 ▪ Award for Academic Excellence, Peking University Nov 2018	

- May 4th Scholarship, Peking University

Nov 2018

SKILLS

L^AT_EX, Python, Pytorch, JavaScript/TypeScript, Svelte, React, C/C++, Java, Datalog.