

Yifeng Di

di5@purdue.edu | +1 (765) 715-1515 | <https://akari.im/>

Education

Purdue University

Ph.D. in Computer Science (Expected Graduation: May 2027)

Advisor: Tianyi Zhang

August 2022 – Present

Nanjing University

B.S. in Software Engineering

September 2015 – July 2019

Experience

Research Assistant

Purdue University | West Lafayette, IN, USA

August 2022 – Present

- Conduct research on **LLMs for Coding**, focusing on code generation and its evaluation
- Build knowledge graphs for software security | [\[NSF Proto-OKN\]](#) | [\[National AI Research Resource Pilot\]](#)

Teaching Assistant

Purdue University | West Lafayette, IN, USA

August 2022 – December 2023

- Delivered lectures and supported instruction for CS 34800: Information Systems

Software Development Engineer

Kuaishou | Beijing, China

July 2019 – August 2022

- Developed and maintained profile, sharing, and other social product features in Kuaishou
- Contributed to active user growth by iterating user-facing social features that improved engagement

Software Development Engineer [Intern]

Xiaomi | Beijing, China

November 2018 – February 2019

- Developed and maintained security-related features in MIUI

Publications

Enhancing Code Generation via Bidirectional Comment-Level Mutual Grounding

Yifeng Di, Tianyi Zhang

[\[ICSE 2025\]](#) | Code Generation, LLMs for Coding, Interactive Refinement

- Proposed PlnG, a comment-based interactive code generation framework
- Improved code generation accuracy and user productivity across benchmarks and user studies

Software Entity Recognition with Noise-Robust Learning

Tai Nguyen*, Yifeng Di*, Joohan Lee, Muhao Chen, Tianyi Zhang (* equal contribution)

[\[ASE 2023\]](#) | Named Entity Recognition, Noise-Robust Learning

- Built WIKISER, a large software entity recognition (SER) dataset from Wikipedia
- Proposed self-regularization for noise-robust learning and improved SER accuracy over prior methods

Mango: Multi-Agent Web Navigation via Global-View Optimization

Weixi Tong, Yifeng Di, Tianyi Zhang

[\[ACL 2026\]](#) | Web Navigation, LLM Agents, Multi-Agent Systems

- Proposed MANGO for multi-agent web navigation via global structure analysis and adaptive URL selection
- Integrated Thompson Sampling and episodic memory for budget-aware web exploration

SOSum: A Dataset of Stack Overflow Post Summaries

Bonan Kou, Yifeng Di, Muhao Chen, Tianyi Zhang

[\[MSR 2022\]](#) | Stack Overflow, Text Summarization

- Built SOSum, a dataset of Stack Overflow posts and summaries
- Created offline and browser-based annotation tools for summary annotation

Papers Under Review

A Coarse-to-Fine Framework for Fine-Grained Contamination Detection in Code LLM Evaluation

[Under Review] | Code LLM Evaluation, Data Contamination, Benchmark Analysis

- Proposed a coarse-to-fine framework for fine-grained contamination detection in code LLM evaluation
- Built a benchmark of annotated task pairs across code benchmarks and post-training corpora

An Empirical Study of Conditional Dependencies and Their Security Impact in C/C++

[Under Review] | Dependency Analysis, Software Supply Chain Security

- Developed a static analysis approach to extract dependencies with their conditions from C/C++ build scripts
- Conducted a large-scale empirical study on conditional dependencies and their security impact

Research Talks

- [PurPL](#): "Enhancing Code Generation via Bidirectional Comment-Level Mutual Grounding"
- [PurPL](#): "Software Entity Recognition with Noise-Robust Learning"

April 24, 2025

September 07, 2023