

# Jiyang Zhang

Austin, Texas

United States

📞 +1 (512) 501 0948

✉ [jiyang.zhang@utexas.edu](mailto:jiyang.zhang@utexas.edu)

🌐 <https://jiyangzhang.github.io/>

🐙 JiyangZhang

## Research Interest

Building Large Language Models that help software developers write and maintain their code (ML4SE).

## Education

- 2019–2025 **Ph.D.**, The University of Texas at Austin, Austin, United States  
Major in Computer Software Engineering, advised by Milos Gligoric, co-advised by Jessy Li.
- 2019–2024 **M.S.**, The University of Texas at Austin, Austin, United States  
Major in Computer Software Engineering
- 2015–2019 **B.S.**, Beihang University, Beijing, China  
Major in Automation Science and Engineering
- 2017 **Exchange Student**, University of Toronto, Toronto, Canada  
Major in Electrical and Computer Engineering

## Publications

- [1] **Jiyang Zhang**, Yu Liu, Pengyu Nie, Junyi Jessy Li, and Milos Gligoric. Generating exceptional behavior tests with reasoning augmented large language models. In *arXiv*, 2024.
- [2] **Jiyang Zhang**, Pengyu Nie, Junyi Jessy Li, and Milos Gligoric. Multilingual code co-evolution using large language models. In *International Symposium on the Foundations of Software Engineering (FSE)*, 2023.
- [3] Yu Liu, **Jiyang Zhang**, Pengyu Nie, Milos Gligoric, and Owolabi Legunsen. More precise regression test selection via reasoning about semantics-modifying changes. In *International Symposium on Software Testing and Analysis (ISSTA)*, 2023.
- [4] **Jiyang Zhang**, Chandra Maddila, Ram Bairi, Christian Bird, Ujjwal Raizada, Apoorva Agrawal, Yamini Jhavar, Kim Herzig, and Arie van Deursen. Using large-scale heterogeneous graph representation learning for code review recommendations at microsoft. In *International Conference on Software Engineering (ICSE Software Engineering in Practice Track)*, 2023.
- [5] **Jiyang Zhang**, Sheena Panthaplackel, Pengyu Nie, Junyi Jessy Li, and Milos Gligoric. Coditt5: Pretraining for source code and natural language editing. In *International Conference on Automated Software Engineering (ASE)*, 2022.

- [6] **Jiyang Zhang**, Marko Ristin, Schanely Phillip, Hans Wernher van de Venn, and Milos Gligoric. Python-by-contract dataset. In *International Symposium on the Foundations of Software Engineering (FSE Demonstrations Track)*, 2022.
- [7] Pengyu Nie, **Jiyang Zhang**, Junyi Jessy Li, Raymond Mooney, and Milos Gligoric. Impact of evaluation methodologies on code summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 4936–4960, 2022.
- [8] **Jiyang Zhang**, Yu Liu, Milos Gligoric, Owolabi Legunsen, and August Shi. Comparing and combining analysis-based and learning-based regression test selection. In *International Conference on Automation of Software Test (AST)*, pages 17–28, 2022.
- [9] **Jiyang Zhang**, Sheena Panthaplackel, Pengyu Nie, Junyi Jessy Li, Raymond J. Mooney, and Milos Gligoric. Leveraging class hierarchy for code comprehension. In *Workshop on Computer Assisted Programming (CAP)*, 2020.
- [10] Yifan Nie, **Jiyang Zhang**, and Jian-Yun Nie. Integrated learning of features and ranking function in information retrieval. In *International Conference on Theory of Information Retrieval (ICTIR)*, pages 67–74, 2019.

## Working Experience

- Summer 2024 **Research Intern**, Amazon Web Services (AWS), New York City, United States  
Develop Multi-agent system to solve software engineering tasks automatically.
- Fall 2023 **Research Assistant**, The University of Texas at Austin, Austin, United States  
Design Large Language Model to generate exceptional behavior tests in Java
- Defined a novel task for LLMs: generating exceptional behavior tests;
  - Developed a dataset for fine-tuning and evaluating the model for generating exceptional behavior tests;
  - Designed and implemented EXLÓNG, an instruction-tuned LLM built on CodeLlama that generates exceptional behavior tests in Java.
- Summer 2023 **Research Intern**, Salesforce Research, Palo Alto, United States  
Worked with the CodeGen team to develop retrieval-augmented code generation Large Language Models
- Conducted an empirical study to identify the most effective code snippets within the codebase that enhance the performance of Large Language Models (LLMs) in predicting the next line of code;
  - Integrated Jaccard-retrieved code snippets with SOTA Large Language Models (LLMs), resulting in a 30% improvement in accuracy.
- Summer 2021 **Research Intern**, Microsoft Research, Redmond, United States  
Collaborated with researchers from MSR *Developer Experience Lab* to design a machine learning-based code reviewer recommendation system
- Designed and implemented a novel code reviewer recommendation model built on graph convolutional neural (GCN) network;
  - Trained and evaluated the model on the company's historical data, and conducted a user study to show the effectiveness of the neural recommendation system;
  - Published a paper at a top Software Engineering conference, ICSE, featuring the results of the internship.

- Summer 2019 **Machine Learning Engineer**, INFIMIND, Beijing, China  
Developed ML-based advertisement generation system
- Collected advertisement text data from websites;
  - Trained the RNN-based model to generate advertisement for the products based on their descriptions and attributes.

## Awards

- 2023 NSF Student Travel Award for 2023 MAPs workshop  
2023 ACM SIGSOFT Distinguished Paper Award for [3] at ISSTA 2023  
2022 NSF Student Travel Award for 2022 International Conference on Software Engineering

## Mentoring

- Undergrad Sing-Rong Chiu, Fall 2023  
Undergrad Nandita Jayanthi, Spring and Fall 2023

## Committee Service

- ISSTA'24 Artifact Evaluation PC Member, International Symposium on Software Testing and Analysis  
ReSAISE'23 PC Member, International Workshop on Reliable and Secure AI for Software Engineering  
ISSTA'23 Artifact Evaluation PC Member, International Symposium on Software Testing and Analysis  
DL4C'23 PC Member, Deep Learning for Code Workshop  
MSR'23 Junior PC Member, International Conference on Mining Software Repositories

## Professional Activities

- 2020–Present Co-organizer, NLP+Programming Reading Group at UT Austin  
EMNLP 2023 Conference Submission Reviewer, Conference on Empirical Methods in Natural Language Processing  
2023 Spring Co-organizer, Joint UT-Cornell Software Engineering Seminar

## Presentations

- Mar 1, 2024 **Empowering Software Maintenance with Large Language Models**, at ECE Outstanding Student Lecture Series, UT Austin, Austin, United States  
Dec 7, 2023 **Multilingual Code Co-Evolution Using Large Language Models**, at FSE 2023, San Francisco, California, United States  
Oct 26, 2022 **Towards Applying Machine Learning to Software Engineering**, at Microsoft Data&AI Team, Online  
Oct 11, 2022 **CoditT5: Pretraining for Source Code and Natural Language Editing**, at ASE 2022, Rochester, Michigan, United States  
Oct 7, 2022 **Pretraining for Source Code and Natural Language Editing**, at Carper AI, Online

## Teaching Experience

Summer 2022 **Teaching Assistant**, EE 382V Machine Programming, The University of Texas at Austin

Spring 2022 **Teaching Assistant**, LIN 373N Machine Learning Toolbox Text Analysis, The University of Texas at Austin

Spring 2020, Summer 2020 **Teaching Assistant**, EE 360C Algorithms, The University of Texas at Austin

---

## Open Source Contributions

- 2023 **Codeditor**, present a large language model designed for code co-evolution - <https://github.com/EngineeringSoftware/codeditor>
- 2022 **CoditT5**, publish a large language model pretrained with a novel objective to explicitly model edits in code and natural language on Hugging Face - <https://huggingface.co/JiyangZhang/CoditT5>
- 2021 **seutil**, contribute to a Python library of utility functions for natural language processing and software engineering research - <https://github.com/pengyunie/seutil>