Chia-Yi Su

csu3@nd.edu | https://chiayisu.github.io/ | https://github.com/chiayisu

Education

University of Notre Dame — Notre Dame, IN
Ph.D. in Computer Science and Engineering — *Aug 2022 – present*
Dissertation: Context-based Language Models for Source Code Summarization

Advisor: Collin McMillan

National Kaohsiung University of Science and Technology — Kaohsiung, Taiwan
M.S. in Electronic Engineering — *August 2020 – June 2022*
Thesis: Study on Interactive Dialog System for Diseases Information Retrieval

Advisor: Tsong-Yi Chen

Summary

Automated software engineering, Source code summarization, Large language models, LLM reasoning

Publication

Journal article

C. Su, A. Bansal, C. McMillan, "Revisiting File Context for Source Code Summarization", in Automated Software Engineering Journal (ASE Journal), Volume 31, article 62, 2024.

C. Su and C. McMillan, "Semantic Similarity Loss for Neural Source Code Summarization", in Journal of Software Evolution and Process (JSME), 2024.

C. Su, C. McMillan, "Distilled GPT for Source Code Summarization", in Automated Software Engineering Journal (ASE Journal), Volume 31, article 22, 2024

Conference Short Papers

C. Su, A. Bansal, V. Jain, S. Ghanavati, C. McMillan, "A Language Model of Java Methods with Train/Test Deduplication", in 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Tool Demos (ESEC/FSE '23), San Francisco, California, USA, December 3-9, 2023.

A. Bansal, **C. Su**, Zachary Karas, Y. Zhang, Y. Huang, T. Li, C. McMillan, "Modeling Programmer Attention as Scanpath Prediction", in 38th IEEE/ACM International Conference on Automated Software Engineering, New Ideas and Emerging Results (ASE'23 NIER), September 11 - 15, 2023.

Papers under review

C. Su, A. Bansal, Y. Huang, T. Li, C. McMillan, "Context-aware Code Summary Generation"

C. Su, A. Bansal, V. Jain, S. Ghanavati, S. Peddinti, C. McMillan, "Which Code Statements Implement Privacy Behaviors in Android Applications?"

Presentation

Conference presentation: Distilled GPT for Source Code Summarization", in 39th IEEE/ACM International Conference on Automated Software Engineering. (Journal-first presentation)

Conference presentation: A Language Model of Java Methods with Train/Test Deduplication," in ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE'23)

Conference presentation: Modeling Programmer Attention as Scanpath Prediction," in IEEE/ACM International Conference on Automated Software Engineering (ASE 23)

Service

Student volunteer: ASE'24

Program committee: ASE'25 NIER

Industry Experience

Software Engineer Intern June 2019 – July 2020
Hewlett Packard Enterprise *Taipei, Taiwan*

- Automated 100 test cases with Robot Framework
- Trimmed down test time from 10 days to 4 hours only
- Proposed and developed test automation functions to lower the difficulty of test automation
- Assisted to teach automation course
- Developed web to enable users to retrieve information front-end

Teaching Experience

Teaching assistant

Data Structure Spring 2023
University of Notre Dame *Notre Dame, IN*

Software Engineering August 2020 - June 2022
National Kaohsiung University of Science and Technology *Kaohsiung, Taiwan*

Teaching staff

Reinforcement Learning Spring 2022
National Kaohsiung University of Science and Technology *Kaohsiung, Taiwan*

Natural Language Processing August 2019 - June 2022
National Kaohsiung University of Science and Technology *Kaohsiung, Taiwan*

Technical Skills

Programming Languages: Python, C/C++, SQL, Java

Deep Learning Frameworks: TensorFlow, PyTorch

Libraries & Tools: NumPy, Pandas, Scikit-learn, NLTK, Flask, Git, Vim

Others: Web development, MySql

Others

FAA student Pilot