Chia-Yi Su

574-303-5218 | 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*
Proposal: 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: Interactive Dialog System for Disease Information Retrieval
Advisor: Tsong-Yi Chen

Professional Experience

Research Assistant, Computer Science and Engineering — August 2022 – present
University of Notre Dame — *Notre Dame, IN*

- Led applied **machine learning** research in software engineering resulting in **9 research articles**
- Integrating **agentic AI** techniques with **LLMs** for automated memory bug localization in C/C++ projects on Github
- Prompted and fine-tuned **LLMs (CodeLlama, nanoGPT, GPT, Gemini)** with novel process for source code summarization that reduces **81% of the tokens** and shows **10% improvement in USE**
- Implemented a **dual encoder Transformer** with **Tensorflow** for context-aware source code summarization that results in **3% improvement in METEOR** and **1% improvement in USE** in Java dataset
- Implemented a **semantic loss function** based on **RLHF** by using **Pytorch/Tensorflow** to improve both **LLMs (LLaMA) and Transformer** for source code summarization
- Conducted empirical studies on **LLM reasoning** capabilities across code intelligence tasks identifying limitations in **LLMs** with **Gemini, GPT, CodeLlama, and nanoGPT** as examples

Research Assistant, Electronic Engineering — August 2019 – June 2022
National Kaohsiung University of Science and Technology — *Kaohsiung, Taiwan*

- Enabled users to retrieve disease information with chat media and **natural language interaction**
- Trained BERT for sequence labeling and domain classification jointly by using **PyTorch**, achieving **92%** accuracy
- Implemented a supervised learning method to warm up dialog policy network for reinforcement learning models that reached **95% accuracy in dialog policy making** and **94% accuracy in token prediction**

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

- Automated **100+ test cases** with Robot Framework, reducing test time from **10 days to 4 hours**
- Proposed and developed test automation functions to **lower the difficulty of test automation**
- Developed web to enable users to retrieve information front-end

Technical Skills

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

Machine learning: Reinforcement learning, RLHF, LLM reasoning, LLM, Knowledge distillation, Optimization, Agentic AI

Libraries & Tools: PyTorch, TensorFlow, Huggingface transformers, NumPy, Pandas, Scikit-learn, Matplotlib, NLTK, Flask, Git, Vim

LLMs: CodeLlama, LLaMA, StarCoder, nanoGPT, Gemini, GPT

LLMOps: QLoRA, LoRA, Instruct fine-tuning

Others: Web development, ChatGPT fine-tuning

Selected Projects

LLM Agents for Bug Localization — May 2025 – present
University of Notre Dame — *Notre Dame, IN*

- **Concepts:** LLM, Prompt engineering, Agentic AI, and Static analysis
- Developing **agentic AI** using **LLMs** to localize memory bugs in C/C++ projects on GitHub

Do Code LLMs Do Static Analysis? June 2024 – May 2025
University of Notre Dame *Notre Dame, IN*

- **Concepts:** LLMOps, Prompt engineering, LLM reasoning, Static analysis, Code intelligence, and LLM
- Assessing **reasoning capabilities of LLMs** based on what human programmers do for code intelligence tasks

Context-aware Code Summary Generation December 2023 – June 2024
University of Notre Dame *Notre Dame, IN*

- **Concepts:** LLMOps, Prompt engineering, LLM reasoning, Static analysis, Code intelligence, and LLM
- Fine-tuned and prompted **LLMs (GPT, Gemini, CodeLlama, and nanoGPT)** with a novel process that **reduces 81% of tokens** and **shows 10% improvement over USE**

Distilled GPT for Source Code Summarization July 2023 – December 2023
University of Notre Dame *Notre Dame, IN*

- **Concepts:** Knowledge distillation, Fine-tuning, and Prompt engineering
- Distilled **GPT3.5** with **nanoGPT and StarCoder** to demonstrate **secure** and **efficient** open-source summarization alternatives

Semantic Similarity Loss for Neural Source Code Summarization January 2024 – July 2024
University of Notre Dame *Notre Dame, IN*

- **Concepts:** RLHF, Loss function, Optimization, Human ratings and feedback, and Source code summarization
- Implemented a semantic similarity loss for **LLMs (LLaMA)** and **Transformer** with **PyTorch/TensorFlow** using **RLHF** to improve source code summarization

Publication

[1] **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.

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

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

[4] **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.

[5] 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.

[6] **C. Su**, A. Bansal, Y. Huang, T. Li, C. McMillan, "Context-aware Code Summary Generation" (Under review at Journal of Systems and Software)

[7] M. Dhakal, **C. Su**, R. Wallace, C. Fakhimi, A. Bansal, T. Li, Y. Huang, C. McMillan, "AI-Mediated Code Comment Improvement" (Under review at Empirical Software Engineering)

[8] **C. Su**, A. Bansal, V. Jain, S. Ghanavati, S. Peddinti, C. McMillan, "Which Code Statements Implement Privacy Behaviors in Android Applications?" (Preprint)

[9] **C. Su** and C. McMillan, "Do Code LLMs Do Static Analysis?" (Ready for submission)

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

Reviewer: ACM Transactions on Software Engineering and Methodology