

Chia-Yi Su

+1 574-303-5218 | csu3@nd.edu | <https://chiayisu.github.io/> | <https://github.com/chiayisu>

EDUCATION

University of Notre Dame

Ph.D. in Computer Science and Engineering

Advisor: Collin McMillan

Notre Dame, IN

Aug 2022 – present

National Kaohsiung University of Science and Technology

M.S. in Electronic Engineering

Advisor: Tsong-Yi Chen

Kaohsiung, Taiwan

August 2020 – June 2022

PROFESSIONAL EXPERIENCE

Research Assistant, Computer Science and Engineering

University of Notre Dame

August 2022 – present

Notre Dame, IN

- Led applied machine learning research in software engineering resulting in 8 research articles
- Designed a novel process for source code summarization that **reduces 81% of the tokens** and **shows 10% improvement in USE**
- Designed a dual encoder transformer to include context for source code summarization that results in **3% improvement in METEOR** and **1% improvement in USE** in Java dataset
- Designed a semantic loss function that shows **improvement in both LLMs and smaller models**
- Designed and conducted experiments on Gemini, GPT, and CodeLlama that shows the poor reasoning skill of LLMs

Research Assistant, Electronic Engineering

National Kaohsiung University of Science and Technology

August 2019 – June 2022

Kaohsiung, Taiwan

- Enabled users to retrieve disease information with chat media and natural language interaction
- Designed new loss function to jointly train BERT model for sequence label and domain classification tasks that reached **92% accuracy on both tasks**
- Proposed 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

Hewlett Packard Enterprise

June 2019 – July 2020

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

PROJECTS

Do Code LLMs Do Static Analysis?

- **Concepts:** LLMOps, LLM reasoning, Static analysis, Code intelligence, and LLM
- Conducted experiments on **both commercial and open-source LLMs** to show the **inability of reasoning skills on LLMs** based on **what human programmers do for the tasks**

Distilled gpt for source code summarization

- **Concepts:** Knowledge distillation, Fine-tuning, and prompt engineering
- Distilled knowledge from GPT and showed the **possibility of using open-source models to protect data privacy**

Semantic Similarity Loss for Neural Source Code Summarization

- **Concepts:** RLHF, Loss function, Optimization, Human ratings and feedback, and Source code summarization
- Designed a loss function that includes semantic similarity for partial credit for synonyms, which reports **improvement in both LLMs and smaller models**

Context-aware Code Summary Generation

- **Concepts:** LLMs, Open-source models, Fine-tuning, and prompt engineering
- Designed a novel process that **reduces 81% of tokens** and **shows 10% improvement over USE**

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

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 finetuning
- Others:** Web development

TEACHING EXPERIENCE

Teaching assistant, Data Structure
University of Notre Dame

Spring 2023
Notre Dame, IN

Teaching staff, Reinforcement Learning and Natural language processing
National Kaohsiung University of Science and Technology

August 2019 - June 2022
Kaohsiung, Taiwan