

Chia-Yi Su

168 Fitzpatrick Hall
University of Notre Dame
csu3@nd.edu
<https://chiayisu.github.io/>

Research Interest human attention, bio-inspired machine learning, human-centered/human-like AI for software engineering, human-AI alignment, AI4SE, program comprehension

EDUCATION	<i>Ph.D.</i> , Computer Science & Engineering University of Notre Dame Proposal: Context-based Language Models for Source Code Summarization Advisor: Collin McMillan	August 2022 - May 2026 (Expected)
	<i>M.S.</i> , Electronic Engineering National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan Thesis: Interactive Dialog System for Disease Information Retrieval Advisor: Tsong-Yi Chen	August 2020 - June 2022
WORK EXPERIENCE	<i>Research Assistant, Department of Computer Science</i> University of Notre Dame, Notre Dame, IN	August 2022 - present
	<i>Research Assistant, Electronic Engineering</i> National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan	August 2019 – June 2022
	<i>Software Engineer Intern</i> Hewlett Packard Enterprise, Taipei, Taiwan	July 2019 – June 2020

PUBLICATIONS Journal Papers

- [1] **C.-Y. Su**, A. Bansal, Y. Huang, T. J.-J. Li, C. McMillan, “Context-aware code summary generation,” *Journal of Systems and Software*, vol. 231, 2026.
- [2] **C.-Y. Su**, C. McMillan, “Semantic similarity loss for neural source code summarization,” *Journal of Software Evolution and Process*, 2024.
- [3] **C.-Y. Su**, A. Bansal, C. McMillan, “Revisiting file context for source code summarization,” *Automated Software Engineering*, vol. 31, no. 2, 2024.
- [4] **C.-Y. Su**, C. McMillan, “Distilled gpt for source code summarization,” *Automated Software Engineering*, vol. 31, no. 1, 2024, (ASE’24 Journal-first).

Conference Short Papers

- [5] **C.-Y. Su**, C. McMillan, “Cmind: An ai agent for localizing c memory bugs,” in *Proceedings of the 34th IEEE/ACM International Conference on Program Comprehension, Tool Demo (ICPC’26 Demo)*, 2026.

- [6] A. Bansal, **C.-Y. Su**, Z. Karas, Y. Zhang, Y. Huang, T. J.-J. Li, C. McMillan, “Modeling programmer attention as scanpath prediction,” in *2023 38th IEEE/ACM International Conference on Automated Software Engineering, NIER (ASE’23 NIER)*, IEEE, 2023.
- [7] **C.-Y. Su**, A. Bansal, V. Jain, S. Ghanavati, C. McMillan, “A language model of java methods with train/test deduplication,” in *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Tool Demo (ESEC/FSE’23 Demo)*, 2023.

Under Review

- [8] **C.-Y. Su**, C. McMillan, “Do code llms do static analysis?” *arXiv:2505.12118*, 2025, (Under review at EMSE).
- [9] M. Dhakal, **C.-Y. Su**, R. Wallace, C. Fakhimi, A. Bansal, T. Li, Y. Huang, C. McMillan, “A grounded theory study to guide ai-driven code comment improvement,” (Under review at JSEP).

Preprint

- [10] **C.-Y. Su**, A. Bansal, V. Jain, S. Ghanavati, S. T. Peddinti, C. McMillan, “Which code statements implement privacy behaviors in android applications?” *arXiv:2503.02091*, 2025.

Tools / Software / CMind: CMind: An AI Agent for Localizing C Memory Bugs, C.-Y. Su and C. McMillan, <https://apclbuglocalizer.github.io/>

Prigen: Dataset of Fine-grained Annotations of Privacy Behaviors in Android Applications, C. Y Su, A. Bansal, V. Jain, S. Ghanavati, Sai. Peddintiand, and C. McMillan, <https://github.com/apcl-research/jam>

Jam-CGPT: Distilled GPT for Source Code Summarization, C.-Y. Su and C. McMillan, <https://huggingface.co/datasets/apcl/Jam-CGPT>

Jam: A Language Model of Java Methods, C. Y Su, A. Bansal, V. Jain, S. Ghanavati, and C. McMillan, <https://github.com/apcl-research/jam>

TALKS

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

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

PROFESSION SERVICE

Student Volunteer: ASE’24
Program Committee: ASE’25 NIER

Reviewer:

ACM Transactions on Software Engineering and Methodology
Springer Automated Software Engineering Journal * 3

TEACHING EXPERIENCE*Data Structure, Teaching Assistant*

Spring 2022

Department of Computer Science, University of Notre Dame, Notre Dame, IN

- Hosted office hours once per week
- Graded the programming assignments
- Assisted students in the lab session

Instrctor

August 2020 - June 2020

Electronic Engineering, National Kaohsiung University of Science and Technology,
Kaohsiung, Taiwan

- Designed and taught natural languages processing course
- Designed and taught reinforcement learning course
- Mentored in-service graduate students on thesis

Teaching Assistant

August 2019

Electronic Engineering, National Kaohsiung University of Science and Technology,
Kaohsiung, Kaohsiung, Taiwan

- Advised undergraduate students on software engineering course
- Assisted to teach natural language processing course

TECHNICAL SKILLS**Programming Languages:** Python, C/C++, SQL, Java**Machine Learning:** Reinforcement learning, RLHF, LLM reasoning, Knowledge distillation, LLM post-training, Agentic AI, Transformer**Libraries:** PyTorch, TensorFlow, Huggingface transformers, LangChain, NumPy, Pandas, Scikit-learn, Matplotlib, NLTK, Flask**LLMOps:** QLoRA, LoRA, Instruct fine-tuning**Others:** Web development, Docker, Static analysis, AWS, Git, Vim, Joern, Doxygen, Code intelligence, Empirical study**REFERENCES****Collin McMillan**

Associate Professor

Department of Computer Science and Engineering

University of Notre Dame

Email: cmc@nd.edu

Relation: Ph.D. Advisor

Sepideh Ghanavati

Associate Professor

School of Computing and Information Science

University of Maine

Email: sepideh.ghanavati@maine.edu

Relation: Research collaborator and proposal committee

Yu Huang

Assistant Professor

Department of Computer Science

Vanderbilt University

Email: yu.huang@vanderbilt.edu

Relation: Research collaborator