Wenxin Jiang

Ph.D. Candidate
Elmore Family School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47906

jiang784@purdue.edu
765–409–1715
https://wenxin-jiang.github.io

RESEARCH THEME

My research interest is mainly focused on Software engineering for AI (SE4AI). I am also interested in studying *AI systems*, *software supply chain security*, and *trustworthy/responsible AI*. My current work focuses on novel approaches to improve multiple aspects of *pre-trained AI model supply chain*, including *trustworthiness, reusability, and security*.

EDUCATION

Ph.D, Electrical and Computer Engineering, *GPA: 3.6/4.0* 2020–2025
Purdue University, West Lafayette, IN

M.Sc., Electrical and Computer Engineering 2024
Purdue University, West Lafayette, IN

B.Sc. Applied Physics, *GPA: 3.3/4.0* 2016–2020
Southeast University, Nanjing, China

Study Abroad Program, Engineering Physics, *GPA: 3.8/4.0* 2019
University of California, Santa Barbara, CA

PROFESSIONAL EXPERIENCE

Graduate Research Assistant 2021–present
Purdue University — Supervised by James C. Davis

- Conduct empirical analysis and mine software repositories to enhance pre-trained AI model reuse.
- Develop automated tools to enhance transparency and reliability in open-source AI ecosystems.
- Design tools for securing the AI model supply chain, focusing on pickle deserialization and typosquatting detection.

Research Intern July – October, 2024
Socket — Supervised by Mikola Lysenko

- Designed data collection infrastructure for HuggingFace data and implemented migration to PostgreSQL database.
- Developed an LLM-based pickle malware scanner for PyPI and Hugging Face artifacts.
- Developed a novel typosquatting detection method leveraging FastText and contrastive learning, optimizing detection efficiency through clustering algorithms. Deployed a REST API for integration and real-time detection.

TensorFlow Model Developer 2021–2023
Purdue University × Google — Supervised by Abdullah Rashwan

- Led a team of 20+ undergraduate students in replicating state-of-the-art AI models, including object detection (YOLO family) and panoptic segmentation models (Maskformer family), in TensorFlow for Google's TensorFlow Model Garden Team. Managed tasks, component integration, and unit, differential, and end-to-end testing, achieving comparable performance to the original implementations.

SELECTED PUBLICATIONS

[1] **Jiang**, Banna, Vivek, Goel, Synovic, Klingensmith, Thiruvathukal, and Davis. *Challenges and Practices of Deep Learning Model Reengineering: A Case Study on Computer Vision*. Empirical Software Engineering (**EMSE'24**).

[2] **Jiang**, Yasmin, Jones, Synovic, Kuo, Bielanski, Yuan, Thiruvathukal, and Davis. *PeaTMOSS: Mining Pre-Trained Models in Open-Source Software*. Proceedings of the 21th Annual Conference on Mining Software Repositories (**MSR'24**).

[3] Jones, **Jiang**, Synovic, Thiruvathukal, and Davis.. *What do we know about Hugging Face? A systematic literature review and quantitative validation of qualitative claims*. Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (**ESEM'24**).

[4] Jajal, **Jiang**, Tewari, Woo, Lu, Thiruvathukal, and Davis. *Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem*. Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA'24**).

[5] **Jiang**, Synovic, Hyatt, Schorlemmer, Sethi, Lu, Thiruvathukal, and Davis. *An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry*. Proceedings of the ACM/IEEE 45th International Conference on Software Engineering (**ICSE'23**).

[6] **Jiang**, Synovic, Jajal, Schorlemmer, Tewari, Pareek, Thiruvathukal, and Davis. *PTMTorrent: A Dataset for Mining Open-source Pre-trained Model Packages*. Proceedings of the 20th Annual Conference on Mining Software Repositories — Data and Tool Showcase Track (**MSR-Data'23**).

[7] Davis, Jajal, **Jiang**, Schorlemmer, N. Synovic, and G.K. Thiruvathukal. *Reusing Deep Learning Models Challenges and Directions in Software Engineering*. Proceedings of the IEEE John Vincent Atanasoff Symposium on Modern Computing (**JVA'23**).

[8] Montes, Peerapatanapokin, Schultz, Guo, **Jiang**, and Davis. *Discrepancies among Pre-trained Deep Neural Networks: A New Threat to Model Zoo Reliability*. Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering — Ideas, Visions, and Reflections track (**ESEC/FSE-IVR'22**).

[9] **Jiang**, Synovic, Sethi, Indarapu, Hyatt, Schorlemmer, Thiruvathukal, and Davis. *An Empirical Study of Artifacts and Security Risks in the Pre-trained Model Supply Chain*. Proceedings of the 1st ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses (**SCORED'22**).

POSTERS

[1] Schorlemmer, **Jiang**, and Davis. *Machine Learning Supply Chain Security*. 2023 Purdue CERIAS Symposium (**CERIAS'23**). ***Award: Best Poster — 2nd-place***.

[2] **Jiang**, Schorlemmer, and Davis. *Trustworthy Re-use of Pre-trained Neural Networks*. 2023 Purdue CERIAS Symposium (**CERIAS'23**).

TEACHING ASSISTANT

ECE 595 – Advanced Software Engineering Spring 2022
Purdue University

INVITED TALKS

PeaTMOSS: A Dataset and Initial Analysis of Pre-Trained Models in Open-Source Software 2024
Research Data Alliance 22nd Plenary Meeting (RDA VP22)
An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry 2023
Purdue University Programming Languages Group, Seminar
Deep Learning Model Reengineering: An Exploratory Case Study on Computer Vision 2022
Purdue University Programming Languages Group, Seminar

AWARDS AND RECOGNITION

ACM SIGSOFT CAPS Travel Grant (ASE'24, ICSE'23, ESEC/FSE'22) 2022-2024
Future Leaders for Responsible AI, the Michigan Institute for Data Science (MIDAS) 2024
Purdue Graduate Student Government and the Graduate School Travel Grant (ICSE'23) 2023
Study Abroad Fellowship, Southeast University 2019
Second prize, Vision Guided Robot Competition, Southeast University 2019
Distinction Award, Southeast University 2018
Third prize, Structural Innovation Invitation Competition, Southeast University 2017

SERVICES

Artifact Evaluation PC Member, ICSE 2025
Shadow PC Member, ICSE 2025
Sub-Reviewer: USENIX Security'25, ICSE'25, JSS, ISSTA'24, LCTES'23, ESEC/FSE'23, ASE'22 2022 - 2024

TECHNICAL SKILLS

Programming Languages: Python (proficient), JavaScript, Java, SQL, Bash
Frameworks: PyTorch, TensorFlow, Numpy, Pandas
Tools: Git, Docker, Slurm, SQL, Google Cloud, LaTeX, Linux/Unix