

Wenxin Jiang, Ph.D. Candidate

Elmore Family School of Electrical and Computer Engineering
Purdue University
<https://wenxin-jiang.github.io>

West Lafayette, IN 47906
jiang784@purdue.edu
765-409-1715

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 <i>Purdue University, West Lafayette, IN</i> | 2020–2025 |
| B.Sc. Applied Physics <i>Southeast University, Nanjing, China</i> | 2016–2020 |
| Study Abroad Program, Engineering Physics <i>University of California, Santa Barbara, CA</i> | 2019 |

PROFESSIONAL EXPERIENCE

| | |
|---|--------------|
| Graduate Research Assistant <i>ECE@Purdue University — Supervised by Dr. James C. Davis</i> | 2021–present |
|---|--------------|

- Published 5 top-tier papers, 6 workshop papers, and 4 technical reports.
- Conducted empirical analysis and mined software repositories to enhance pre-trained AI model reuse.
- Developed automated tools to improve transparency and security of open-source AI model supply chain.
- Designed tools for securing the AI model supply chain, focusing on pickle deserialization and typosquatting detection.
- Worked on NSF-funded award and collaborated with sponsors at Cisco and Google.

| | |
|--|-----------------------|
| Research Intern <i>Socket — Supervised by Dr. Mikola Lysenko</i> | July – December, 2024 |
|--|-----------------------|

- 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.
- Researched a novel typosquatting detection method using FastText and contrastive learning to generate embeddings, applying clustering algorithms to reduce overhead in identifying similar package names.

| | |
|---|-----------|
| TensorFlow Model Developer <i>Purdue University × Google — Supervised by Dr. Abdullah Rashwan</i> | 2021–2023 |
|---|-----------|

- Led a team of 20+ undergraduate students in replicating state-of-the-art AI models, including object detection (YOLO) and panoptic segmentation models (Maskformer) for Google's TensorFlow Model Garden Team.

| | |
|---|---------------------|
| Teaching Assistant <i>Purdue University — ECE 59500 Advanced Software Engineering</i> | January – May, 2022 |
|---|---------------------|

- Developed and designed midterm exams and assignments for a graduate-level course in software engineering, covering topics such as software engineering ethics, failure analysis, and automated testing tools.

REFEREED CONFERENCE PUBLICATIONS (FULL PAPERS) *These venues are CORE2023 rank A or A*.*

- [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**). 63 pages.
- [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**). 13 pages.
- [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**). 12 pages.

- [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**). 13 pages.
- [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**). 13 pages.

OTHER REFEREED WORKS: VISIONS, TOOLS, PRELIMINARY WORKS, COMPETITIONS

- [1] **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**). 5 pages.
- [2] 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**). 14 pages.
- [3] 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**). 5 pages.
- [4] **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**). 10 pages.
- [5] Synovic, Hyatt, Sethi, Thota, Shilpika, Miller, **Jiang**, Amobi, Pinderski, Laufer, Hayward, Kingensmith, Davis, and Thiruvathukal. *Snapshot Metrics Are Not Enough: Analyzing Software Repositories with Longitudinal Metrics*. Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering — Demonstrations track (**ASE-Tool Demonstrations'22**). 4 pages.
- [6] Veselsky, West, Ahlgren, Thiruvathukal, Klingensmith, Goel, **Jiang**, Davis, Lee, and Kim. *Establishing trust in vehicle-to-vehicle coordination: a sensor fusion approach*. Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Application (**HotMobile'22**). 6 pages.

TECHNICAL REPORTS

- [1] **Jiang**, Cheung, Kim, Kim, Thiruvathukal, and Davis. *Naming Practices of Pre-Trained Models in Hugging Face*. <https://arxiv.org/pdf/2310.01642>. 2024, under review at EMSE.
- [2] Peng, Gupte, Eliopoulos, Ho, Mantri, Deng, **Jiang**, Lu, Läufer, Thiruvathukal, and Davis. *Large Language Models for Energy-Efficient Code: Emerging Results and Future Directions*. <https://arxiv.org/pdf/2310.01642>. 2024.
- [3] Purohit, **Jiang**, Ravikiran, and Davis. *A Partial Replication of MaskFormer in TensorFlow on TPUs for the TensorFlow Model Garden*. <https://arxiv.org/pdf/2404.18801>. 2024.
- [4] Banna, Chinnakotla, Yan, Vegesana, Vivek, Krishnappa, **Jiang**, Lu, Thiruvathukal, and Davis. *An Experience Report on Machine Learning Reproducibility: Guidance for Practitioners and TensorFlow Model Garden Contributors*. <https://arxiv.org/abs/2107.00821>. 2021.

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**).

INVITED TALKS

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

AWARDS AND RECOGNITION

| | |
|--|------|
| ACM SIGSOFT CAPS Travel Grant (ASE'24) | 2024 |
| Future Leaders for Responsible AI, the Michigan Institute for Data Science (MIDAS) | 2024 |
| ACM SIGSOFT CAPS Travel Grant (ICSE'23) | 2023 |
| Purdue Graduate Student Government and the Graduate School Travel Grant (ICSE'23) | 2023 |
| ACM SIGSOFT CAPS Travel Grant (ESEC/FSE'22) | 2022 |
| 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 |

MENTORSHIP

| | |
|---|--|
| Daniel Lugo, PhD@Purdue | Current |
| Huiyun Peng, PhD@Purdue | Current |
| Huiyun Peng, PhD@Purdue | Current |
| Jerin Yasmin , PhD@Queen's University, <i>Supervised by Dr. Yuan Tian</i> | Current |
| Haoyu Gao, PhD@University of Melbourne, <i>Supervised by Dr. Christoph Treude</i> | Current |
| Parth Patil, MSc@Purdue | Current |
| Jason Jones, MSc@Purdue | Graduated, SE@BotDojo |
| Nicholas Synovic, MSc@LUC, <i>Supervised by Dr. George K. Thiruvathukal</i> | Graduated, Pursuing PhD@LUC |
| Mingyu Kim, BSc@Purdue | Current |
| Dulani Wijayarathne, BSc@Purdue | Graduated, Pursuing PhD@GeorgiaTech |
| Matt Hyatt, BSc@LUC | Graduated, Pursuing PhD@LUC |
| Shen Kuo, BSc@Purdue | Graduated, Pursuing MSc@Purdue |
| Heesoo Kim, BS@Purdue | Graduated, Pursuing MSc@Purdue |
| Diego Montes, BSc@Purdue | Graduated, SE@SpaceX |
| Feny Patel, BSc@Purdue | Graduated, SE@Meta |
| Ananya Singh, BSc@Purdue | Graduated, SE@Google |
| Pongpatapee (Dan) Peerapatanapokin, BSc@Purdue | Graduated, Application Analyst@Cummins |
| Ibrahim Saeed, BSc@Purdue | Graduated, SE@Magnite |

SERVICES

| | |
|---|-------------|
| Artifact Evaluation PC Member, International Conference on Software Engineering (<i>ICSE</i>) | 2025 |
| Shadow PC Member, International Conference on Software Engineering (<i>ICSE</i>) | 2025 |
| Junior PC Member, International Conference on Mining Software Repositories (<i>MSR</i>) | 2025 |
| Junior PC Member, International Conference on Technical Debt (<i>TechDebt</i>) | 2025 |
| Sub-Reviewer: FSE'25, USENIX Security'25, ICSE'25, JSS, ISSTA'24, LCTES'23, ESEC/FSE'23, ASE'22 | 2022 - 2024 |

GRANT WRITING

Trustworthy Re-use of Pre-Trained Neural Networks, *PI: James C. Davis, Contract with Cisco, \$179,237* 2022-2023
Under review: *DARPA, PI: James C. Davis* 2024
In preparation: *NSF-SHF, PI: James C. Davis* 2024

PROFESSIONAL MEMBERSHIPS

Member, Institute of Electrical and Electronics Engineers (IEEE)
Member, Association for Computing Machinery (ACM)