Yuhao Zhang

(608) 236-3965 · yuhaoz@cs.wisc.edu · 1210 W. Dayton Street, Room 6364, Madison, WI 53706

EDUCATION

University of Wisconsin-Madison Aug 2019 - Aug 2024
PhD Student in Computer Science Madison, WI

Advisors: Prof. Aws Albarghouthi and Prof. Loris D'Antoni
Dissertation: Towards Robust Artificial-Intelligence-Powered Software: Provable Guarantees via Abstract Interpretation

Peking University Sept 2015 - Jul 2019
B.S. in Computer Science and Technology, Summa Cum Laude Beijing, CN

Advisor: Prof. Yingfei Xiong
Thesis: Algorithm Design for Detecting and Repairing Numerical Bugs in Deep Learning Programs
Honors: Top 10 Undergraduate Thesis, Outstanding Undergraduate Student

PUBLICATIONS

(* stands for equal contribution)

Anna P. Meyer*, **Yuhao Zhang***, Aws Albarghouthi, Loris D'Antoni, "Verified Training for Counterfactual Explanation Robustness under Data Shift" in 5th Workshop on Data-Centric Machine Learning Research (**DMLR 2024@ICLR**).

Linyi Li, **Yuhao Zhang**, Luyao Ren, Yingfei Xiong, Tao Xie, "Reliability Assurance for Deep Neural Network Architectures Against Numerical Defects" in *45th International Conference on Software Engineering (**ICSE 2023**)*.

Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, "BagFlip: A Certified Defense against Data Poisoning" in *Thirty-sixth Conference on Neural Information Processing Systems (**NeurIPS 2022**)*.

Yuhao Zhang*, Yasharth Bajpai*, Priyanshu Gupta*, Ameya Ketkar*, Miltiadis Allamanis, Titus Barik, Sumit Gulwani, Arjun Radhakrishna, Mohammad Raza, Gustavo Soares, and Ashish Tiwari, "Overwatch: Learning Patterns in Code Edit Sequences" in *Object-Oriented Programming, Systems, Languages & Applications (**OOPSLA 2022**)*.

Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, "Certified Robustness to Programmable Transformations in LSTMs" in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (**Oral Presentation** **EMNLP 2021**)*.

Yuhao Zhang, Luyao Ren, Liqian Chen, Yingfei Xiong, Shing-Chi Cheung, Tao Xie, "Detecting Numerical Bugs in Neural Network Architectures" in *Proceedings of the ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (**ACM Distinguished Paper Award** **ESEC/FSE 2020**)*.

Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, "Robustness to Programmable String Transformations via Augmented Abstract Training" in *Proceedings of the Thirty-seventh International Conference on Machine Learning (**ICML 2020**)*.

Yuhao Zhang, Yifan Chen, Shing-Chi Cheung, Yingfei Xiong, and Lu Zhang, "An Empirical Study on TensorFlow Program Bugs" in *Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA 2018**)*.

On Arxiv:

Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, "A One-Layer Decoder-Only Transformer is a Two-Layer RNN: With an Application to Certified Robustness".

Yuhao Zhang, Shiqi Wang, Haifeng Qian, Zijian Wang, Mingyue Shang, Linbo Liu, Sanjay Krishna Gouda, Baishakhi Ray, Murali Krishna Ramanathan, Xiaofei Ma, Anoop Deoras, "CodeFort: Robust Training for Code Generation Models".

Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, "PECAN: A Deterministic Certified Defense Against Backdoor Attacks".

Siwakorn Srisakaokul, **Yuhao Zhang**, Zexuan Zhong, Wei Yang, Tao Xie, Bo Li, "MULDEF: Multi-model-based Defense Against Adversarial Examples for Neural Networks".

WORK EXPERIENCE

Amazon May 2023 - Aug 2023
Applied Scientist Intern - CodeWhisperer Team [Python] New York, NY

· Enhanced the robustness of large language models for code generation, resulting in a 41% improvement.
· Developed a robust training framework, designed for future integration into **Amazon CodeWhisperer**.
· Contributions to the ReCode benchmark, providing a robust evaluation method for code generation models.
· Inclined for a Level 5, Applied Scientist position at Amazon.

Microsoft May 2021 - Aug 2021 & Feb 2022 - April 2022
Research Intern - PROSE (Programming by Examples and Natural Language) Team [C#/Python] Remote

· Contributed to the Blue-Pencil project, a key initiative behind **Visual Studio IntelliCode**.
· Designed a unique representation for storing source code editing patterns from developer traces.
· Built a framework that effectively learns and applies hundreds of patterns, achieving a precision rate of 78%.
· The implementation of the new framework led to a 4X increase in user adoption of the tool.

Microsoft Research Sept 2018 - March 2019
Research Intern - DKI (Data, Knowledge, Intelligence) Group [C#/Python] Beijing, CN

· Contributed to **Ideas**, an Excel plugin designed to provide high-level visual summaries for data analysts.
· Enhanced the classification accuracy of the intermediary model from 88% to 93% across six primary languages.
· Implemented the Aho–Corasick algorithm, accelerating column header matching by 4X.
· Developed a grid search algorithm for hyperparameter tuning, which was subsequently adopted by other groups.
· Recognized for outstanding performance and contributions with the **Award of Excellence**.

AWARDS

Research

· NeurIPS 2023 Top Reviewer
· Midwest Programming Languages Summit 2023 Travel Awards
· NeurIPS 2022 Top Reviewer
· ACM Distinguished Paper Award at ESEC/FSE 2020

ACM-ICPC

· Ho-Chi-Minh City Regional **4th place**, 2017
· Xi'an Regional **Gold**, 2017
· Yangon Regional **7th place**, 2016
· Dalian Regional **Gold**, 2016
· Hefei Regional **Gold** 2015

NOI

· National Olympiad of Informatics, **Gold Medal**, 2014

Scholarships

· Finalist in Two Sigma PhD Fellowship 2022
· SenseTime Scholarship 2019
· Suzhou Industrial Park Scholarship 2018
· Schlumberger Scholarship 2017
· iPinYou Scholarship 2016

SERVICE

SAIV (Workshop) 2024, Program Committee
NeurIPS (Conference) 2024, 2023, 2022, Reviewer
COLM (Conference) 2024, Reviewer
Neurocomputing (Journal), Reviewer
WFVML (Workshop) 2024, 2023, Reviewer
BANDS (Workshop) 2023, Program Committee
FoMLAS (Workshop) 2023, 2022, 2021, Program Committee
ICML (Conference) 2023, 2022, 2021, Reviewer
VMCAI (Workshop) 2023, Artifact Evaluation Committee
CAV (Conference) 2021, Artifact Evaluation Committee