

YUYAO WANG

☎ (+86)18963607071 ✉ yuyao6@outlook.com 🌐 [Kristoff-starling](#) 🏠 [Homepage](#)

Research Interest

I am generally interested in topics related to the *correctness*, *programmability*, and *performance* of computer systems. I have been attached to programming and algorithm design since middle school and aspire to bring elegant solutions for tackling *real-world* problems.

Education

Nanjing University

B.S. in Computer Science (Elite class)

2020 - 2024 (expected)

GPA: 4.71/5.00 (1/256)

Publication

- [1] (**NeurIPS'23**) Is Your Code Generated by ChatGPT Really Correct? Rigorous Evaluation of Large Language Models for Code Generation.
Jiawei Liu*, Chunqiu Steven Xia*, **Yuyao Wang**, Lingming Zhang.
[pre-print](#) ◇ [code](#) ◇ [slides](#)
- [2] (**ESEC/FSE'23**) NEURI: Diversifying DNN Generation via Inductive Rule Inference.
Jiawei Liu, Jinjun Peng, **Yuyao Wang**, Lingming Zhang.
🏆 **ACM SIGSOFT Distinguished Paper Award**
[paper](#) ◇ [artifact](#) ◇ [slides](#)

Research Experience

System Group, UW

Jul. 2023 - Now

Advised by [Prof. Ratul Mahajan](#) and [Prof. Arvind Krishnamurthy](#)

Topic: *Application-Defined Networks*

- Developed application-defined networks, in which developers specify network functionality in a high-level language and the controller generates a custom distributed implementation that runs across available hardware and software resources.
 - Role in the project: *independently* designed and implemented the graph compiler which automatically determines the optimal placement and order of network functions, interacts with the controller to generate deployment scripts for multiple backends, and supports live upgrades.

PL/FM/SE Group, UIUC

Sept. 2022 - Jun. 2023

Advised by [Prof. Lingming Zhang](#)

Topic: *Software Testing, SE4LLM*

- Designed a benchmarking framework **EvalPlus** that leverages LLM- and mutation-based methods to augment evaluation datasets with large amounts of testcases for rigorously evaluating the functional correctness of LLM synthesized code [1].
 - Highlights: The augmented version of HUMANEval (aka. HUMANEval⁺) leads to **13.6%-15.3%** reduction in pass@k across 20 popular LLMs and all k values; **EvalPlus** has **5k+** downloads on PyPI.
 - Role in the project: *independently* designed the test-suite minimization algorithm; proposed a strategy combining LLM seed-input generation and type-aware mutation for testcase augmentation.
- Proposed an automated fuzzing approach NEURI that leverages program synthesis to generate diverse and well-formed deep-learning models in order to validate DL toolchain [2].
 - Highlights: **100** new bugs were found for PyTorch and TensorFlow, of which **9** bugs are labeled *high priority* or *security vulnerability*.
 - Role in the project: *independently* designed and implemented the rule synthesizer (significant speed-up over general-purpose program synthesis tool **Rosette**) and record augmentation strategies; incorporated both symbolic and concrete operators to perform a concolic style of DNN generation; enriched the test oracles with sanitizers.

Selected Awards

- | | |
|--|-----------|
| • ACM SIGSOFT Distinguished Paper Award | Dec. 2023 |
| • Gold Medal , International Collegiate Programming Contest (ICPC) Asia Regional Contest (Xi'an) | Dec. 2022 |
| • Gold Medal , International Collegiate Programming Contest (ICPC) Asia Regional Contest (Shanghai) | Dec. 2021 |
| • Special Scholarship for Undergraduates in Basic Science (1/20), Nanjing University | Oct. 2022 |
| • China National Scholarship (top 0.2%) | Sep. 2021 |
| • Silver Medal , National Olympiad in Informatics (NOI) | Jul. 2018 |