Yuxiang Wei

Second Year CS PhD Student
Department of Computer Science
University of Illinois at Urbana-Champaign

ywei40@illinois.edu
yuxiang.cs.illinois.edu
UniverseFly
(Last updated: Mar 3, 2024)

Research Interests

I am interested in *large language models (LLMs) for code*.

As a member of the **StarCoder2** [1] team, I contributed to its evaluation and release. I am also a core member of the upcoming **StarCoder2-Instruct** project.

My recent project 🎩Magicoder [3] [GitHub] [Preprint] has obtained **1800+ stars ⭐ on GitHub**. Magicoder is powered by *OSS-Instruct* that generates code instructions from open source inspirations. It surpasses ChatGPT on HumanEval+ with $\leq$ 7B parameters.

My past project Repilot [4] [GitHub] [Preprint] [Slides] has **100+ stars ⭐ on GitHub**. Repilot effectively improves the validity of LLMs generated code by fusing its token-by-token generation with a semantics-based completion engine.

Education

Since 2022 **PhD student in Computer Science**, *University of Illinois at Urbana-Champaign*, Urbana, Illinois, USA, *GPA: 4.0/4.0*
Advisor: Prof. Lingming Zhang. *Anticipated graduation date: May, 2027*

2017–2022 **Bachelor of Science in Computer Science**, *Tongji University*, Shanghai, China, *GPA: 91.06/100*

Publications

[1] Anton Lozhkov, Raymond Li, Loubna Ben Allal, Federico Cassano, Joel Lamy-Poirier, Nouamane Tazi, Ao Tang, Dmytro Pykhtar, Jiawei Liu, **Yuxiang Wei**, et al. "StarCoder 2 and The Stack v2: The Next Generation". In: *arXiv preprint arXiv:2402.19173* (2024).

[2] Yifeng Ding, Jiawei Liu, **Yuxiang Wei**, and Lingming Zhang. "XFT: Unlocking the Power of Code Instruction Tuning by Simply Merging Upcycled Mixture-of-Experts". In: *Preprint* (2024).

[3] **Yuxiang Wei**, Zhe Wang, Jiawei Liu, Yifeng Ding, and Lingming Zhang. "Magicoder: Source Code Is All You Need". In: *arXiv preprint arXiv:2312.02120* (2023).

[4] **Yuxiang Wei**, Chunqiu Steven Xia, and Lingming Zhang. "Copiloting the Copilots: Fusing Large Language Models with Completion Engines for Automated Program Repair". In: *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. ESEC/FSE 2023. San Francisco, CA, USA: Association for Computing Machinery, 2023, pp. 172–184. ISBN:

9798400703270. DOI: 10.1145/3611643.3616271. URL: https://doi.org/10.1145/3611643.3616271.

[5] Chunqiu Steven Xia, **Yuxiang Wei**, and Lingming Zhang. "Automated Program Repair in the Era of Large Pre-Trained Language Models". In: *Proceedings of the 45th International Conference on Software Engineering*. ICSE '23. Melbourne, Victoria, Australia: IEEE Press, 2023, pp. 1482–1494. ISBN: 9781665457019. DOI: 10.1109/ICSE48619.2023.00129. URL: https://doi.org/10.1109/ICSE48619.2023.00129.

[6] Jiawei Liu, **Yuxiang Wei**, Sen Yang, Yinlin Deng, and Lingming Zhang. "Coverage-Guided Tensor Compiler Fuzzing with Joint IR-Pass Mutation". In: *Proc. ACM Program. Lang.* 6.OOPSLA1 (Apr. 2022). DOI: 10.1145/3527317. URL: https://doi.org/10.1145/3527317.

[7] Haolong Li, Zizheng Zhong, Wei Guan, Chenghao Du, Yu Yang, **Yuxiang Wei**, and Chen Ye. "Generative character inpainting guided by structural information". In: *The Visual Computer* 37.9 (2021), pp. 2895–2906. DOI: 10.1007/s00371-021-02218-y. URL: https://doi.org/10.1007/s00371-021-02218-y.

Academic Services

(AEC stands for Artifact Evaluation Committee)

Organizing committee	The First International Workshop on Large Language Models for Code (LLM4Code 2024), co-located with ICSE 2024
Reviewer	ICLR 2024 Workshop on Reliable and Responsible Foundation Models (R2-FM@ICLR 2024)
AEC	The ACM Conference on Computer and Communications Security (CCS 2023)
AEC	Programming Language Design and Implementation (PLDI 2024)

Talks

Dec 2023	Conference talk for Repilot [4] at ESEC/FSE 2023
Oct 2023	Kwai Inc.: Fusing Language Models with Completion Engines for Code Generation
Apr 2023	Uber Programming Systems Lab: Fusing Language Models with Completion Engines for Program Repair

Selected Awards

Oct 2023	NSF Student Travel Award ($1800)
Oct 2023	ACM SIGSOFT CAPS Award ($400)
Mar 2021	1st Prize of "Challenge Cup" Academic Works Competition, Tongji University
Nov 2019	National 2nd Prize (3.84%) of Chinese Mathematical Contest in Modeling
Nov 2019	Province-Level 1st Prize (Shanghai) of Chinese Mathematical Contest in Modeling

Open Source Contributions

I am dedicated to making practical and easy-to-use tools for everyone!

- 🎩**Magicoder** [3] [GitHub] (**1.8K stars**): enhancing code generation with *OSS-Instruct*. Magicoder surpasses ChatGPT on HumanEval+ with $\leq$ 7B parameters.
- **Repilot** [4] [GitHub] (**100+ stars**): patch/code generation by combining large language models and semantics-based completion engines.
- **Tzer** [6] [GitHub] (**60+ stars**): fuzzer for the low-level IR (Intermediate Representation) of the TVM machine learning compiler.

Programming Skills

I enjoy learning new programming language features all the time!

Python	I am proficient in coding **well-typed Python** programs using *type hints*.
C++	In the past, I programmed in **modern C++** a lot, where I learned *templates*, *type inference*, *lambda expressions*, and so on.
Swift	Writing Apps in **Swift** helped me learn many advanced language features, including the *declarative paradigms of SwiftUI* for user interface creation.
Rust	**Rust** navigates me through its unique *ownership model* and *lifetime system*.
Haskell/Scala	I embrace *functional programming*, benefiting greatly from **Haskell** and **Scala 3**.
Others	Beyond these, my coding experiences extend to **Java**, **Javascript**, **C**, and **Verilog**.