Zhe Wang

zhewang2001.github.io zhe36@illinois.edu github.com/zhewang2001

Education

University of Illinois at Urbana-Champaign **Aug 2024 – May 2026**
Siebel School of Computing and Data Science *Illinois, United States*
- Master of Science in Computer Science
- **Advisor:** Lingming Zhang

Tsinghua University **Sep 2020 – Jun 2024**
Weiyang College *Beijing, China*
- Bachelor of Science in Mathematics and Physics
- **Advisor:** Zhiyuan Liu

Research Interests

- **LLMs for Code:** to develop LLMs to solve software engineering tasks through post-training via synthetic data.
- **Trustworthy LLMs:** to enhance trustworthiness, resilience and reliability of helpful-only LLMs against vulnerable code and malicious cyberactivity attacks.
- **LLM Applications:** to empower LLMs with reasoning, planning and collaboration capabilities through alignment training and agent-based systems.

Publications

- Yuxiang Wei, **Zhe Wang**, Jiawei Liu, Yifeng Ding, Lingming Zhang. *"Magicoder: Empowering code generation with oss-instruct"* ICML 2024.(https://arxiv.org/abs/2312.02120)

Research Experiences

Magicoder: Empowering Code Generation with OSS-Instruct **Oct 2023 – Dec 2023**
Research Intern at Illinois at Urbana-Champaign, Supervised by Prof. Lingming Zhang *Beijing, China*
- Aim to mitigate bias in LLM-generated synthetic data by empowering models with diverse, controllable open-source references.
- Introduced **Magicoder**, a series of fully open-source LLMs for code, that can outperform all evaluated LLMs with less than or equal to 16B parameters while having no more than 7B parameters.
- Proposed **OSS-INSTRUCT**, a new approach to enlighten LLMs with open-source code snippets for high-quality data synthesis, which was adopted by *Meta Llama 3.1*, *Google CodeGemma*, and *IBM Granite*.
- Achieved more than 2k stars on Github and was on Github trending.

Technical Skills

Languages: Mandarin (Native), English (TOEFL 105: **Speaking 25**), Japanese Beginner
Programming Languages: Python, Java, C, HTML/CSS, JavaScript, SQL
Developer Tools: VS Code, Eclipse, Google Cloud Platform, Jupyter Notebook
Technologies/Frameworks: Linux, GitHub, Huggingface, Tensorflow, Pytorch