ZHE WANG

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

Education

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

Tsinghua University   Sep 2020 – Jun 2024
Bachelor of Science in Mathematics and Physics   *Beijing, China*
- **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

- Jiawei Liu*, Nirav Diwan*, **Zhe Wang***, Haoyu Zhai, Xiaona Zhou, Kiet A. Nguyen, Tianjiao Yu, Muntasir Wahed, Yinlin Deng, Hadjer Benkraouda, Yuxiang Wei, Lingming Zhang, Ismini Lourentzou, Gang Wang. *"PurpCode: Reasoning for Safer Code Generation"*, arXiv preprint. [Paper] [Code] [Dataset]

- Kunlun Zhu*, Hongyi Du*, Zhaochen Hong*, Xiaocheng Yang*, Shuyi Guo*, **Zhe Wang***, Zhenhailong Wang, Cheng Qian, Xiangru Tang, Heng Ji, Jiaxuan You. *"MultiAgentBench: Evaluating the Collaboration and Competition of LLM Agents"*, ACL main 2025. [Paper] [Code]

- Yuxiang Wei, **Zhe Wang**, Jiawei Liu, Yifeng Ding, Lingming Zhang. *"Magicoder: Empowering code generation with oss-instruct"*, ICML 2024. [Paper] [Code] [Dataset]

Research Experiences

PurpCode: Reasoning for Safer Code Generation
Supervised by Prof. Gang Wang, Prof. Lourentzou, Prof. Lingming Zhang at UIUC   **Dec 2024 – Jul 2025**
- Introduced **PurpCode**, the first cybersafety reasoning model that generates secure code and defends against malicious cyber activities, outperforming frontier models in secure code generation while maintaining utility.
- Designed a novel training pipeline with (1)**Data Curation**, (2)**Rule Learning** via SFT on self-aligned safety reasoning trajectories, and (3) **Reinforcement Learning** with multi-objective rewards.
- **Won 1st place in Amazon Nova AI Challenge 2025.**

MultiAgentBench: Evaluating the Collaboration and Competition of LLM Agents
Supervised by Prof. Jiaxuan You and Prof. Heng Ji at UIUC   **Nov 2024 – Feb 2025**
- Introduced **MultiAgentBench**, a comprehensive benchmark, and the **MARBLE** framework for evaluating LLM-based multi-agent systems across 6 different collaborative and competitive scenarios.
- Developed innovative evaluation metrics including milestone-based KPIs, planning/communication scores, and competition metrics, offering insights through emergent social behaviors toward AGI-level collaboration.
- Project accepted at ACL main 2025. Code and dataset released at Github.

Magicoder: Empowering Code Generation with OSS-Instruct
Supervised by Prof. Lingming Zhang at UIUC   **Oct 2023 – Dec 2023**
- Proposed **OSS-Instruct**, a new approach for high-quality data synthesis with open-source code snippets, adopted by *Meta Llama 3.1*, *Google CodeGemma*, and *IBM Granite*.
- Introduced **Magicoder**, a series of fully open-source coding LLMs trained with OSS-Instruct, outperforming all evaluated LLMs of ≤16B parameters with only 7B parameters.
- **Achieved 2k+ Github stars, featured on Github trending, and received 393k+ downloads.**