Weihao **Zeng**

Hong Kong, China

☐ +86 18511780695 | ✉ wzengak@connect.ust.hk | ☐ Zeng-WH | 🎓 Weihao Zeng

Summary

I am **Weihao Zeng**, a PhD student at the Hong Kong University of Science and Technology starting in the fall of 2025. My research interests primarily focus on large language model post-training, including **data engineering** during the alignment phase and exploration of model **reasoning capabilities**. I am now also interested in benchmarking and improving **language agents** for diverse, realistic, and long-horizon task execution. Previously, I focused on task-oriented dialogue systems, including dialogue pre-training. I have published several papers at NLP conferences such as ICLR, ACL, COLM EMNLP, and NAACL.

- RL / Self-Evolve techniques to improve model reasoning capabilities: **SimpleRL** (3.8k ★) (Notion, Github, Paper), **B-STaR**

- Benchmarking and improving **language agents** for diverse, realistic, and long-horizon task execution: **Toolathlon**

- Exploring efficient data engineering methods for Post-Training: **Deita**, **Auto Evol-Instruct**

- Applications of large language models in task-oriented dialogue systems : FutureTOD, Seen2UnSeen

Education

The Hong Kong University of Science and Technology *Hong Kong*

Computer Science and Engineering, PhD Student *Sep. 2025 - Jun. 2028 (Expected)*

Beijing University of Posts and Telecommunications *Beijing*

Artificial Intelligence, Master' Degree *Sep. 2022 - Jun. 2025*

- Honors: National Scholarship (2023)
- Honors: First-class academic scholarship (2022,2023)

Beijing University of Posts and Telecommunications *Beijing*

Information and Computing Science, Bachelor's degree *Sep. 2018 - Jun. 2022*

- GPA: 3.83/4.0, ranked 1st in major
- Honors: National Scholarship (2020)

Publications

The Tool Decathlon: Benchmarking Language Agents for Diverse, Realistic, and Long-Horizon Task Execution *Under Review*

Co-first author *Oct. 2025*

- Toolathlon, a benchmark for language agents on diverse, complex, and realistic tool use.
- Toolathlon reveals significant shortcomings of SOTA LLMs in realistic tool-use tasks

Pushing Test-Time Scaling Limits of Deep Search with Asymmetric Verification *Under Review*

First Author *Oct. 2025*

- We study how scaling test-time compute affects complex information-seeking behavior.
- We can allocate compute more efficiently by leveraging the Asymmetry of Verification.

SimpleRL-Zoo: Investigating and Taming Zero Reinforcement Learning for Open Base Models in the Wild

COLM 2025

FIRST AUTHOR

Mar. 2025

- Received nearly **4K stars** on Github!
- **SimpleRL** is the first open-source project to reproduce the DeepSeek-R1 training process based on rule-based reward

B-STaR: Monitoring and Balancing Exploration and Exploitation in Self-Taught Reasoners

ICLR 2025

FIRST AUTHOR

Dec. 2024

- Quantitative analysis of the dynamic changes between Exploration and Exploitation during the self-improvement process.
- Introduction of B-STaR, a Self-Taught Reasoning framework capable of autonomously adjusting its configuration.
- Balancing exploration and exploitation to achieve superior performance.

What Makes Good Data for Alignment? A Comprehensive Study of Automatic Data Selection in Instruction Tuning

ICLR 2024

CO-FIRST AUTHOR

Jan. 2024

- Comprehensively analyzed the efficiency of instruction data from complexity, quality, and diversity perspectives
- Proposed simple yet effective automatic instruction data selection strategies
- Released the open-source Deita series of models, achieving or even surpassing most open-source instruction-following models using a small amount of data

Automatic Instruction Evolving for Large Language Models

EMNLP 2024

FIRST AUTHOR

Jun. 2024

- Fully automated SFT data engineering framework without any human effort
- An automated instruction evolution framework outperformed Evol-Instruct on Code, Math, and Chat

FutureTOD: Teaching Future Knowledge to Pre-trained Language Model for Task-Oriented Dialogue

ACL 2023

FIRST AUTHOR

May. 2023

- Proposed a new dialog pre-training framework, FutureTOD.
- Achieved new state-of-the-art results on multiple dialog understanding tasks.

Seen to Unseen: Exploring Compositional Generalization of Multi-Attribute Controllable Dialogue Generation

ACL 2023

FIRST AUTHOR

May. 2023

- Proposed the problem of compositional generalization for multi-attribute controllable dialogue generation.
- Proposed a prompt-based disentangled controllable dialogue generation model, DCG, and an evaluation framework.

DivTOD: Unleashing the Power of LLMs for Diversifying Task-Oriented Dialogue Representations

NAACL 2024

FIRST AUTHOR

Mar. 2024

- Propose a new Dialogue Pretrain method, collaboratively enhance dialogue representation with LLM
- Transfer the abundant background knowledge in LLM to a small model, while eliminating the impact of conflicting domain knowledge

Invited Talks

- April 2025, **Qingke Talk**, SimpleRL-Zoo and B-STaR: Improving reasoning performance and efficiency through reinforcement learning

- Mar 2025, **Westlake University**, SimpleRL-Zoo: Investigating and Taming Zero Reinforcement Learning for Open Base Models in the Wild.

- Feb 2025, **Northwestern University**, SimpleRL: Emerging Reasoning with Reinforcement Learning is Both Effective and Efficient.

- Feb 2025, **Tiktok**, SimpleRL: Emerging Reasoning with Reinforcement Learning is Both Effective and Efficient.

- Feb 2025, **Huawei Noah's Ark Lab**, SimpleRL: Emerging Reasoning with Reinforcement Learning is Both Effective and Efficient.

Competitions

INTERNATIONAL COMPETITION

2022.09   **1 st Award**, SereTOD Challenge 2022 Track 2   *EMNLP WorkShop*

CHINESE COMPETITION

2021.08   **4th Award**, SMP 2021 Conversational AI Challenge   *CIPSC*

2021.09   **8th Place**, CCIR 2021 Intelligent NLU Challenge   *CCF*

Intern Experience

Microsoft WizardLM Team   *Haidian, Beijing*

RESEARCH INTERN   *Sep. 2023 - Apr. 2024*

- Proposed an automated instruction evolution framework to reduce reliance on human experts in Evol-Instruct. This method outperformed Evol-Instruct on Code, Math, and Chat, and was applied to training the Wizard large language model. **Paper on this work accepted by EMNLP 2024**
- Contributed to the development of Wizard LLM by constructing the DPO dataset and optimizing the DPO method as part of the DPO Process.

Hong Kong University of Science and Technology   *Clear Water Bay, Hong Kong*

RESEARCH ASSISTANT   *July. 2023 - Jun. 2025*

- Visited HKUST, contributed to automatic instruction data selection and open-source DEITA series models. **Paper on this work accepted by ICLR 2024**
- Researching math reasoning capabilities for LLMs. **Paper on this work accepted by ICLR 2025.**

Meituan LLM Foundation Model Team   *Chaoyang, Beijing*

RESEARCH INTERN   *Dec. 2021 - July. 2023*

- Exploring the optimization of multi-turn dialogue capabilities during instruction fine-tuning, validating the impact of quality and diversity
- Proposed FutureTOD, a new dialogue pre-training framework for dialogue understanding in task-oriented systems, achieving SOTA on multiple tasks (ACL 2023)
- Proposed decoupling dialogue generation model based on Prompt-tuning for attribute combination generalization in controlled dialogue (ACL 2023)