

Zhuoer (Andrew) Feng

PhD Candidate · Department of Computer Science and Technology · Tsinghua University

✉ AndrewFengTHU@gmail.com  [ZhuoerFeng](https://github.com/ZhuoerFeng)  [zhuoerfeng.github.io](https://github.com/zhuoerfeng)
 [Google Scholar](https://scholar.google.com/citations?user=...)  [LinkedIn](https://www.linkedin.com/in/zhuoerfeng)  [@andrewfeng_thu](https://twitter.com/andrewfeng_thu)



RESEARCH INTERESTS

LLM Evaluation, Reinforcement Learning for LLM, LLM-as-Judge, Long-form Text Generation and Critique, Foundation Model Development.

EDUCATION

Ph.D. Candidate, Computer Science and Technology 2022 – Present

Tsinghua University, Beijing, China

Advised by [Prof. Minlie Huang](#) and co-advised by [Prof. Hongning Wang](#), CoAI Group.

Undergraduate Visiting Researcher (UGVR) Jun 2021 – Aug 2021

Stanford University, Stanford, CA, USA

Robotics. Advised by Prof. Silvio Savarese and Prof. Kuan Fang.

B.Eng., Computer Science and Technology 2018 – 2022

Tsinghua University, Beijing, China

Transferred from Department of Chemical Engineering (2017–2018).

B.Econ., Economics 2018 – 2022

Tsinghua University, School of Economics & Management

PROFESSIONAL EXPERIENCE

Full-time Research Intern, Evaluation Team Aug 2023 – Present

Z.ai (Zhipu AI), Beijing, China

- Post-training stage development: reward model design, pretrain/post-train checkpoint assessment.
- Evaluation domains: Instruction Following, Multi-turn Dialogue, Creative Writing, and Agent (Claude Code, OpenClaw) performance.
- Core contributor to GLM-3, GLM-4.5, and GLM-5 model evaluation pipelines.

TECHNICAL SKILLS

- **LLM Training:** PyTorch/Megatron-LM/DeepSpeed/Ray; recently Llama-Factory and veRL.
- **Agent & Coding:** Claude Code based development, RL training framework development, Docker-based coding agent training/evaluation.
- **RAG Systems:** LangChain-based retrieval-augmented generation system development.
- **Programming:** Python, C/C++, LaTeX; familiar with Linux, Git, Docker.

AWARDS & HONORS

National Scholarship, Ministry of Education of China	2018, 2024
Excellent Student Social Service Cadre, Tsinghua University	2025
Excellent Undergraduate, Tsinghua University	2022

ACADEMIC SERVICE

Reviewer: ACL Rolling Review (Jan 2025, May 2025, Jan 2026), ICML 2026

TEACHING

Teaching Assistant — Artificial Intelligence

2021 – 2023

Tsinghua University, Department of Computer Science and Technology

PUBLICATIONS

- [1] **GLM-5: from Vibe Coding to Agentic Engineering** [Tech Report, 2026]
GLM-Team.
[arXiv:2602.15763](#)
- [2] **RAVEL: Reasoning Agents for Validating and Evaluating LLM Text Synthesis** [NeurIPS 2026, Submitted]
Zhuoer Feng, Cunxiang Wang, Yu Luo, Bosi Wen, Yidong Wang, Lin Fan, Yilin Zhou, Zikang Wang, Wenbo Yu, Lindong Wu, Hongning Wang, Minlie Huang.
[arXiv:2603.00686](#)
- [3] **RLAR: An Agentic Reward System for Multi-task Reinforcement Learning on Large Language Models** [NeurIPS 2026, Submitted]
Zhuoer Feng, Cunxiang Wang, Bosi Wen, Yidong Wang, Yu Luo, Hongning Wang, Minlie Huang.
[arXiv:2603.00724](#)
- [4] **HoWToBench: Holistic Evaluation for LLM’s Capability in Human-level Writing using Tree of Writing** [ACL 2026]
Zhuoer Feng*, Cunxiang Wang*, Yu Luo, Lin Fan, Yilin Zhou, Zikang Wang, Xiaotao Gu, Jie Tang, Hongning Wang, Minlie Huang.
[arXiv:2604.19071](#)
- [5] **GLM-4.5: Agentic, Reasoning, and Coding (ARC) Foundation Models** [Tech Report, 2025]
GLM-4.5 Team.
[arXiv:2508.06471](#)
- [6] **CritiqueLLM: Scaling LLM-as-Critic for Effective and Explainable Evaluation of Large Language Model Generation** [ACL 2024]
Pei Ke*, Bosi Wen*, Zhuoer Feng*, Xiao Liu*, Xuanyu Lei, Jiale Cheng, Shengyuan Wang, Aohan Zeng, Yuxiao Dong, Hongning Wang, Jie Tang, Minlie Huang.
[ACL Anthology](#)
- [7] **AlignBench: Benchmarking Chinese Alignment of Large Language Models** [ACL 2024]
Xiao Liu*, Xuanyu Lei*, Shengyuan Wang, Yue Huang, Zhuoer Feng, Bosi Wen, Jiale Cheng, Pei Ke, Yifan Xu, Weng Lam Tam, Xiaohan Zhang, Lichao Sun, Xiaotao Gu, Hongning Wang, Jing Zhang, Minlie Huang, Yuxiao Dong, Jie Tang.
[ACL Anthology](#)
- [8] **RedHat: Towards Reducing Hallucination in Essay Critiques with Large Language Models** [ICLR 2024 Submission]
Zhuoer Feng*, Cunxiang Wang*, Bosi Wen, Boqun Kou, Haohan Chi, Yu Luo, Lin Fan, Xiaotao Gu, Jie Tang, Hongning Wang, Minlie Huang.
[OpenReview](#)
- [9] **LOT: A Story-Centric Benchmark for Evaluating Chinese Long Text Understanding and Generation** [TACL, 2022]
Jian Guan, Zhuoer Feng, Yamei Chen, Ruilin He, Xiaoxi Mao, Changjie Fan, Minlie Huang.
[ACL Anthology](#)
- [10] **OpenMEVA: A Benchmark for Evaluating Open-Ended Story Generation Metrics** [ACL 2021]
Jian Guan, Zhixin Zhang, Zhuoer Feng, Zitao Liu, Wenbiao Ding, Xiaoxi Mao, Changjie Fan, Minlie Huang.
[ACL Anthology](#)

Last updated: May 2026