Zhixuan He

hezhixuan1997@gmail.com
Links
https://scholar.google.com/citations?user=VCsV3pcAAAAJ
https://github.com/pod2c

Research Interests

Large Language Models, LLM Reasoning, Trustworthy LLM, LLM-based Agents

1 Education

The University of Hong Kong, Hong Kong SAR, China — Sep 2020—Jul 2021
Master of Science in Library and Information Management Thesis Title: Weibo analysis on Chinese cultural knowledge for gaming

Xiamen University Malaysia, Kuala Lumpur, Malaysia — Sep 2016—Aug 2020
Bachelor of Engineering in Software Engineering

Publications

Peer-Reviewed / Accepted

Jiang, R., Chen, K., Bai, X., **He, Z.**, Li, J., Yang, M., Zhao, T., Nie, L., & Zhang, M. (2024). A Survey on Human Preference Learning for Large Language Models. **ACM Computing Surveys (CSUR)**. Accepted (to appear). Preprint: arXiv:2406.11191.

Manuscripts — Submitted

He, Z., & Feng, Y. (2025). Unleashing Diverse Thinking Modes in LLMs through Multi-Agent Collaboration. **Submitted to ACL Rolling Review**. Preprint: arXiv:2510.16645.

He, Z., & Feng, Y. (2026). When to Think Deeply: Inhibitory Deliberation for LLM Reasoning. **Submitted to ACL Rolling Review**. arXiv preprint pending.

Book Chapter

He, Z., Chiu, D. K. W., & Ho, K. K. W. (2022). Weibo Analysis on Chinese Cultural Knowledge for Gaming. In Handbook of Research on Foundations and Applications of Intelligent Business Analytics (pp.320–349). IGI Global. https://doi.org/10.4018/978-1-7998-9016-4.ch015. *(Master Thesis)*

Research Experience

SUNY Albany — Research Intern

- Studied the mechansim of Supervised Learning from Feedback and Reinforcement Learning from Human Feedback(RLHF).

Selected Research Projects

Inhibitory Deliberative Problem Reasoning (IDPR)

- Proposed IDPR, an inhibitory deliberation framework that treats the fast answer as a prepotent response and uses an inhibition controller to decide whether to release it or suppress it in favor of slow reasoning.
- Built a paired fast-slow router pool through same-problem dual inference, using fast/slow correctness, format validity, token cost, and corrective labels to train a response-conditioned inhibition controller without manual routing annotation.
- Designed fast-side evidence features including confidence, logit margin, parseability, answer form, generation cost, repetition, and truncation, enabling the controller to estimate slow-over-fast quality gain rather than relying only on uncertainty heuristics.
- Results – Held-out math reasoning: IDPR invokes slow reasoning on 8.20% of examples and improves accuracy from 47.90% to 48.92%; random same-budget routing drops to 46.76%, and the strongest confidence baseline reaches 48.22%.
- Analysis: IDPR achieves the highest corrective precision, 27.07%, and shows larger gains on the harder MoT subset, improving accuracy from 35.60% to 36.88%, suggesting that the controller allocates deliberation to examples where fast responses are less reliable.
- Paper: When to Think Deeply: Inhibitory Deliberation for LLM Reasoning, submitted.

Multi-Agent Debate (DiMo) for Diverse Thinking Modes

- Proposed DiMo, a multi-agent debate framework with explicit divergent and logical thinking modes and role-specialized agents (Generator, Evaluator, Knowledge Supporter, Reasoning-Path Provider, Refiner, Judger) to produce auditable reasoning traces under a unified open-source protocol.

- Studied protocol–task affinity: commonsense/knowledge tasks benefit from the divergent mode, while math tasks benefit from the logical mode; DiMo scopes interpretability as process transparency rather than mechanistic interpretability.
- Results – Commonsense (Divergent mode): on CSQA/ARC-C/StrategyQA/OpenBookQA, DiMo with LLaMA-3-8B achieves 80.0/84.1/92.7/84%; with Qwen-2.5-32B achieves 88.4/90.5/90.8/96.0%, outperforming strong baselines.
- Results – Math (Logical mode): on GSM8K/GSM-Hard, DiMo attains 90.7/71.4% with LLaMA-3-8B and 98.4/84.1% with Qwen-2.5-32B, exceeding CoT and debate baselines.
- Paper: Unleashing Diverse Thinking Modes in LLMs through Multi-Agent Collaboration, arXiv:2510.16645.

Human Preference Alignment in LLMs

- Conducted a literature survey on the problem of Human Preferences Alignment in Large Language Models in the last two years.
- Reproduced the PPO, DPO and other methods that use reinforcement learning methods to achieve human preference alignment in Large Language Models.
- Studied the mechanism of Weak-to-Strong generation, and reproducing its code.
- **Project Link:** https://github.com/pod2c/weak2strongcode/tree/main

Work Experience

Harbour Education(Jul 2023—present) Teaching Assistant

- Assisted Prof. Jahangir Hossain from UBC teaching a deep learning-related course: Applications of Machine Learning in Vehicle Engineering.
- Assisted Prof. Philipp Koehn from Johns Hopkins University teaching a NLP-related course: Deep Learning for Natural Language Processing.
- Tutored students to understand the principles of various algorithms in machine learning and deep learning.
- Guided students in completing their final project presentations and writing their course term papers.