Zisu Huang

+86 17321126787 huangzs25@m.fudan.edu.cn Google Scholar WeChat: Feierabendkk

Education

Fudan University — **Shanghai, China**
B.S. in Software Engineering — *Sep. 2021 – Jun. 2025*
GPA: 3.60/4.00
Selected Honors and Awards: Undergraduate Excellence Scholarship (2022, 2023, 2024)

Fudan NLP Group, Fudan University — **Shanghai, China**
M.S. in Computer Science — *Sep. 2025 – Jun. 2028 (Expected)*

Selected Papers

- **Controllable Memory Usage: Balancing Anchoring and Innovation in Long-Term Human-Agent Interaction**
  - Z Huang*, M Tian*, X Wang, J Xu, Z Guo, Q Qian, Y Shen, K Song, J Yuan, C Lv, X Zheng
  - ACL 2026 Main Conference
  - **TLDR:** We identify *Memory Anchoring*, where models default to high memory-reliant generation mode regardless of user intent, and propose **Stee**rable **M**emory Agent, **SteeM**, to steer model outputs toward user-specified memory-dependence preferences.
- **TRIP-Bench: A Benchmark for Long-Horizon Interactive Agents in Real-World Scenarios**
  - Y Shen*, Z Huang*, Z Wang*, M Tian*, Z Guo, C Zhang, S Zhou, Z Hu, D Li, J Xu, K Wang, W Liu, T Li, F Yue, F Hong, C Liu, K Zeng
  - ICML 2026 Under Review
  - **TLDR:** We introduce **TRIP-Bench**, a benchmark that features long-horizon planning tasks, complex rules, diverse multi-turn interactions, and propose **GTPO**, an online RL method that improves robustness and constraint satisfaction over long dialogues.
- **Enhancing the Capability and Robustness of Large Language Models through Reinforcement Learning-Driven Query Refinement**
  - Z Huang*, X Wang*, F Zhang, Z Xu, C Zhang, X Zheng, X Huang
  - Arxiv 2024
  - **TLDR:** We propose **Refiner-RL**, which is a query refinement model trained with supervised fine-tuning and reinforcement learning to optimize user queries, enhancing the quality and robustness of LLMs simultaneously.
- **IntentionReasoner: Facilitating Adaptive LLM Safeguards through Intent Reasoning and Selective Query Refinement**
  - Y Shen, Z Huang, Z Guo, Y Liu, G Chen, R Yin, X Zheng, X Huang
  - ACL 2026 Under Review
  - **TLDR:** We introduce **IntentionReasoner**, a safeguard framework for LLMs trained with reinforcement learning that balances safety, over-refusal, and utility through intent reasoning and a four-level safety taxonomy.
- **BatCoder: Self-Supervised Bidirectional Code-Documentation Learning via Back-Translation**
  - J Xu*, Y Lu*, Z Huang, C Lv, X Wang, S Li, Z Xu, Z Guo, Z Wang, M Tian, X Huang, X Zheng
  - ICML 2026 Under Review
  - **TLDR:** We introduce **BatCoder**, a self-supervised reinforcement learning framework designed to jointly optimize code generation and documentation production via a back-translation strategy.
- **RECAST: Strengthening LLMs' Complex Instruction Following with Constraint-Verifiable Data**
  - W Liu*, Z Guo*, M Xie, J Xu, Z Huang, M Tian, J Xu, M Wu, X Wang, C Lv, H Wang, Y Hu, X Zheng, X Huang
  - ICLR 2026
  - **TLDR:** We propose **RECAST**, an efficient method for synthesizing high-quality data to enhance the complex instruction-following capabilities of LLMs.

Other Papers

- **SATER: A Self-Aware and Token-Efficient Approach to Routing and Cascading**
  - Y Shen, Y Liu, Z Huang, R Yin, X Zheng, X Huang
  - EMNLP 2025 Main Conference
- **Learning Query-Specific Rubrics from Human Preferences for DeepResearch Report Generation**
  - C Lv, J Zhou, W Zhao, J Xu, Z Huang, M Tian, S Dou, T Gui, L Tian, X Zhou, X Zheng, X Huang, J Zhou
  - ICML 2026 Under Review
- **Benchmark^2: Systematic Evaluation of LLM Benchmarks**
  - Q Qian*, C Huang*, J Xu, C Lv, M Wu, W Liu, X Wang, Z Wang, Z Huang, M Tian, J Xu, K Hu, H Wang, Y Hu, X Huang, X Zheng
  - ACL 2026 Under Review

- **CSSG: Measuring Code Similarity with Semantic Graphs**
  - Y Lu*, J Xu*, C Lv, **Z Huang**, Z Guo, Z Wang, M Tian, X Huang, X Zheng
  - **ACL 2026 Under Review (Short Paper)**
- **Progressive Mastery: Customized Curriculum Learning with Guided Prompting for Mathematical Reasoning**
  - M Wu*, Q Qian*, **Z Huang**, W Liu, X Wang, D Liang, L Miao, S Dou, C Lv, Z Wang, Z Xu, L Chen, T Li, X Zheng, X Huang
  - **Arxiv 2025**

Internship

Visual Computing Group, Microsoft Research Asia — **Shanghai, China**
Research Intern, Supervised by Yifan Yang — *March. 2026 – Present*
Research on general agentic model from a data synthesis perspective.

Projects

- **Understanding Reward Hacking in Reinforcement Learning from Human Feedback (RLHF)**
  - We investigate reward hacking in RLHF by reproducing state-of-the-art studies and analyzing training dynamics with specific indicators to gain deeper insights into its detection and mitigation.
- **Investigating the Necessity of CoT Supervision in Reinforcement Learning with Verifiable Rewards (RLVR)**
  - We investigate reasoning–answer inconsistencies in RLVR, and employ a binary consistency reward model that help improve reasoning consistency while maintaining final answer performance.

Skills

- **Data Filtering:** Ensuring training data quality via deduplication, reward-model-based filtering, and related curation techniques.
- **Post-Training:** Hands-on experience with supervised fine-tuning and reinforcement learning using modern frameworks (e.g., verl, ms-swift, LLaMA-Factory), including preference alignment from human feedback (PPO-style) and training with verifiable rewards (e.g., mathematical reasoning).
- **Tool-Integrated Reinforcement Learning:** Training LLM-based agents in multi-turn, tool-integrated RL settings to improve advanced capabilities such as mathematical problem solving and constraint-satisfying travel planning.