

Jiuheng Lin

📍 Beijing, China ✉ linjiuheng@stu.pku.edu.cn 🌐 github.com/infinite-set

Education

Ph.D. Student in Natural Language Processing

Sept 2024 – Now

Wangxuan Institute of Computer Technology, Peking University

- **Advisor:** Prof. Yansong Feng
- **Research Interests:** LLM Reasoning; Question Answering; NLP for Low-Resource Languages

B.Sc. in Information and Computing Science

Sept 2020 – Jun 2024

Department of Computer Science and Technology, Peking University

- **GPA:** 3.65/4.0, Ranking: 27/95 (top 30%)

Selected Publications (* Equal Contribution)

To Reason or to Fabricate: Reasoning Without Shortcuts via Hint-Anchored Pairwise Aggregation

Jiuheng Lin, Chen Zhang, Yansong Feng

Arxiv Preprint (In Submission)

TLDR: RL data contamination often causes LLMs to exploit shortcuts by memorizing answers and faking their reasoning. We propose HIPPO to solve this issue, which uses pairwise hint-injected aggregation to explicitly penalize shortcuts and force the model to learn genuine, highly generalizable reasoning skills.

CLARity: Reasoning Consistency Alone Can Teach Reinforced Experts

Jiuheng Lin, Cong Jiang, Zirui Wu, Jiarui Sun, Yansong Feng

ACL 2026

TLDR: We find multiple-choice questions RL often leads to inconsistent reasoning. We propose CLARity, a simple yet effective RL framework that addresses this issue and improves overall reasoning quality, offering a generalizable solution for training expert LLMs using only a small general-purpose model via reasoning consistency.

Efficient Low-Resource Language Adaptation via Multi-Source Dynamic Logit Fusion

Chen Zhang, Jiuheng Lin, Zhiyuan Liao, Yansong Feng

ACL 2026

TLDR: To adapt LLMs to low-resource languages (LRLs), we propose TriMix, a compute-efficient, test-time logit fusion framework. It dynamically balances LRL competence from a small specialized model, high-resource task skills, and large model scaling benefits, thereby offering valuable insights into LRL proxy tuning.

Read it in Two Steps: Translating Extremely Low-Resource Languages with Code-Augmented Grammar Books

Chen Zhang*, Jiuheng Lin*, Xiao Liu, Zekai Zhang, Yansong Feng

ACL 2025

TLDR: We explore the role of grammar books in translating low-resource languages, highlighting grammar rule retrieval as a bottleneck. We introduce ZhuangRules, a modular dataset, and propose representing grammar rules as code functions, resulting in a 13.1% BLEU improvement in translation.

Chain of Condition: Construct, Verify and Solve Conditions for Conditional Question Answering

Jiuheng Lin, Yuxuan Lai, Yansong Feng

EMNLP 2024 (Findings)

TLDR: We introduce Chain of Condition, a novel prompting approach for conditional question answering that excels in constructing, verifying and solving conditions to produce accurate answers. Our method outperforms all existing baselines and achieves state-of-the-art.

MC²: Towards Transparent and Culturally-Aware NLP for Minority Languages in China

Chen Zhang*, Mingxu Tao*, Quzhe Huang*, Jiuheng Lin*, Zhibin Chen, Yansong Feng

ACL 2024

TLDR: We introduce MC², the largest open-source multilingual corpus for minority languages in China, featuring Tibetan, Uyghur, Kazakh, and Mongolian, including the underrepresented Kazakh Arabic and traditional Mongolian script. MC² addresses data scarcity and prioritizes quality and diversity to improve cultural awareness.

Teaching Large Language Models an Unseen Language on the Fly

Chen Zhang, Xiao Liu, **Jiuheng Lin**, Yansong Feng

ACL 2024 (Findings)

TLDR: We introduce DiPMT++, a framework for adapting LLMs to unseen languages via in-context learning. With minimal resources, DiPMT++ boosts GPT-4's performance in Chinese-Zhuang translation and proves effective for other low-resource languages, aiding in language preservation and translation tasks.

How Many Answers Should I Give? An Empirical Study of Multi-Answer Reading Comprehension

Chen Zhang, **Jiuheng Lin**, Xiao Liu, Yuxuan Lai, Yansong Feng, Dongyan Zhao

ACL 2023 (Findings)

TLDR: We design a taxonomy for categorizing multi-answer machine reading comprehension (MRC) instances and analyze the challenges of handling multiple answers in MRC systems. Our experiments show that combining strengths of different paradigms can improve performance, with generation models showing particular promise.

Awards & Honors

- Third Prize of Peking University Scholarship, 2021
- Award for Academic Excellents of Peking University, 2021
- Outstanding Student of Shandong Province, 2020

Teaching & Services

Teaching Assistant: Introduction to Computation (Peking University, Fall 2024), Foundations of Natural Language Processing (Peking University, Spring 2025, 2026)

Reviewer: ACL Rolling Review (since 2025.5), NeurIPS, AAAI

Skills

Programming: C/C++, Python, Rust, JavaScript

Human Languages: Mandarin (Native), English (CET6 648)