

Hanxu Hu

📍 Zurich ✉ hanxu.hu@uzh.ch ☎ +41 76 429 83 88 🔗 hanxuhu.github.io in hanxuhu 🌐 hanxuhu

Intro

I am a 1st Year PhD Student at University of Zurich supervised by [Prof. Rico Sennrich](#) and [Prof. Ivan Titov](#) at University of Edinburgh. I am interested in evals and alignment of large language models.

Education

University of Zurich Ph.D. Student in NLP <i>Supervisor: Prof. Rico Sennrich and Prof. Ivan Titov</i>	<i>Sept 2024 – Mar 2028 (Expected)</i>
◦ Research Interest: Multilingual Language Models, Long Text Generation	
University of Edinburgh Master in NLP ◦ Supervisor: Edoardo Ponti, Graduate with distinction	<i>Sept 2023 – Sept 2024</i>
Nanjing University of Science and Technology Bachelor in Measurement Tech	<i>Sept 2017 – June 2021</i>

Experience

Research Intern <i>Shanghai AI Lab, InternLM Team</i>	<i>Shanghai, China May 2024 – Sept 2024</i>
◦ Conducted research on Multi-Agent Judges of large language models.	
◦ Collaborated with researchers on Comprehensive Multilingual General Benchmarks covering series of tasks for large language models.	
Research Assistant <i>Westlake University, WestlakeNLP, Mentor: Prof. Yue Zhang</i>	<i>Hangzhou, China Sept 2022 – July 2023</i>
◦ Conducted research on spatial reasoning prompting for large language models.	

Publications

Fine-tuning Large Language Models with Sequential Instructions <i>Hanxu Hu, Simon Yu, Pinzhen Chen, Edoardo M. Ponti, Under Submission, ARR: 3.5, 3.5, 4</i>	June 2024
Chain-of-Symbol Prompting for Spatial Reasoning in Large Language Models <i>Hanxu Hu, Hongyuan Lu, Huajian Zhang, Wai Lam, Yue Zhang COLM 2024</i>	July 2023
Meta-Learning for Multi-Modal Cross-Lingual Transfer <i>Hanxu Hu, Frank Keller EMNLP 2023, MLR Workshop Best Paper Honorable Mention Award</i>	December 2023
Improving Controlled Table-To-Text Generation Robustness <i>Hanxu Hu, Zhongyi Yu, Yunqing Liu, Laura Perez-Beltrachini EACL 2023 Findings</i>	May 2023

Current Ongoing Projects

A Comprehensive Multilingual Evaluation Suite for Large Language Models.	Nov 2024
---	----------

- Curating a comprehensive multilingual benchmark including evaluating reasoning, instruction following, and code generation abilities in multilingual scenarios.

Multi-Judge Aggregation for Preference Learning and Evaluation of Large Language Models

July 2024

- Using different aspects templates for LLM-as-a-judge, and use confidence score for weighting and to aggregate them as a final judgement for preference annotating.

Multilingual Evaluation Benchmark

September 2024

- Creating a comprehensive multilingual evaluation benchmark for large language model, containing a series of tasks including instruction following, reasoning, and code generation.

Mitigate Information Loss in Long Text Generation of Language Models

September 2024

- Explore decoding methodologies for mitigating information loss problems in long text generation scenarios (e.g. document-level machine translation.)

Long Context Adaptation in SFT Stage

August 2024

- Explore methodologies for improving long-context understanding of Large language models during supervised fine-tuning stage.
- Find only using multi-hop questions can boost performance of language models for long-context understanding.
- Conduct interpretability experiments to observe patterns of attention heads.

Continue Pre-training For Multilingual Large Language Models

August 2024

- Propose continue pre-training pipeline for language models to enhance multilingual ability and maintain original ability like reasoning and understanding.

Technologies

Languages: Python, C++

Frameworks: vLLM, Huggingface, PyTorch