

XINHE SHI

✉ ShiXinhe2024@outlook.com 🌐 Homepage 🐦 Twitter

Education

Zhejiang University

Sep. 2021 – Present

Bachelor of Computer Science and Technology

Hangzhou, China

- GPA: 3.97/4.0 (91.30/100)
- Main Courses: Computer Networks(98), Technology of Multimedia(98), Operating System(95), Numeric Analysis(100), Fundamentals of Data Structure(99), Digital Logic Design(97), Discrete Mathematics and Application(95), etc

Research Interest

Large language models, Multilingual NLP, Multi-Modality

Manuscripts

[1] **Xinhe Shi**, Linchao Zhu. [LLM-based Language Similarity](#). (In submission)

TL;DR: We provide multilingual LLM researchers with a sound language selection strategy, enabling them to select a relatively small set of languages (when constrained by budget or computational resources) while maximizing linguistic coverage—specifically by choosing “dissimilar” languages identified by our study.

Research Experience

Institute of Artificial Intelligence

November 2023 – Present

Research intern, supervised by Prof. [Linchao Zhu](#)

- **Language Similarity Analysis from the Perspective of LLMs**
 - * Proposed and formalized the novel research problem of **LLM-based language similarity**, aiming to provide multilingual NLP researchers with a **reliable language selection strategy** while challenging traditional linguistic classification paradigms
 - * Designed **dual-methodological framework** combining visualization techniques and quantitative metrics to analyze language similarity through the lens of both LLM task performance pattern and cross-lingual knowledge transfer efficiency
 - * Discovered LLM-derived language similarity patterns which show partial alignment with traditional linguistic typology (e.g., Slavic language clustering) alongside novel LLM-specific patterns (e.g., Chinese-Greek affinity)
 - * Identified the super-additive transfer effect where fine-tuning on a specific language provides a more effective “bridge” for knowledge transfer than direct fine-tuning on the target language
 - **M4CQ Dataset Construction & Multilingual Evaluation**
 - * Constructed a Massive Multilingual Multitask Multiple Choice Question (**M4CQ**) dataset featuring content consistency across **31** languages and domain-balanced distribution of **135** tasks
 - * (Ongoing) Conducting comprehensive evaluations of multiple LLMs on M4CQ to analyze their multilingual capability
- (The dataset will be released after the evaluations have been done)

Honors & Scholarships

- **Outstanding Graduate of Zhejiang University**
- **Zhejiang Provincial Scholarship**
- **Zhejiang University Scholarship - First Prize** (Top 3% in Zhejiang University)
- Outstanding Student of Zhejiang University
- CKC Honors College Top Students Scholarship - First Prize
- Academic Excellence Award of Zhejiang University

Extracurriculars

- **Zhejiang University Five-Star Volunteer** (Highest Volunteer Title at Zhejiang University)
Completed 300+ volunteer hours in hospital, community, and campus outreach programs
- **Varsity Kayaking Athlete**
Competed in 4 national-level tournaments as a core member for Zhejiang University
- **21st National Extreme Sports Competition**
4th place in Non-Motorized Canoe
- **6th National Collegiate Canoe Championships**
3rd place in Women's Kayak 500m Final(B)
- **The “End of the River” China Grand Canal World Heritage Dragon Boat Competition** (National Competition)
7th place
- Sanhao Cup Table Tennis (University Championships)
3rd place in Women's Doubles