

## Education

### Zhejiang University

B.E. in Computer Science and Technology

Sep. 2021 – Present

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, Agentic AI, LLM Post-Training

## Manuscripts

- [1] **Xinhe Shi**, Qingcheng Zeng, Weihao Xuan, Linchao Zhu. **How to Improve LLMs' Performance on Specific Languages: A Perspective on LLM-Based Language Similarity**. (Targeted at ACL 2026)  
*TL;DR:* We systematically investigate LLM-based language similarity through both the lenses of language-specific performance patterns and cross-lingual transferability, and we verify the effectiveness of LLM-based similarity measures on guiding fine-tuning language selection to improve LLMs' performance on target languages.
- [2] **Xinhe Shi**, Qingcheng Zeng, Weihao Xuan, Kaize Ding. **Scaling Effects on Multilingual Performance**. (Targeted at ACL 2026)  
*TL;DR:* We find that there is no significant and stable difference in performance degradation rates between high-resource and low-resource languages as model size decreases, indicating that compact models can preserve relative multilingual competence, thereby demonstrating their strong potential for globally inclusive deployment and motivating linguistically fair compression strategies.

## Research Experience

### REAL Lab, Northwestern University

Research intern, supervised by Prof. *Kaize Ding*

Aug. 2025 – Present

#### • Analysis of Scaling Effects on Multilingual Performance

- \* Investigated how reducing model size affects multilingual performance, focusing on whether analyzing **whether model performance on low-resource languages degrade faster than that on high-resource ones** as model capacity shrinks
- \* Conducted comprehensive evaluations spanning four diverse model families (Qwen, Aya, Mistral, Phi) across multiple model sizes, ensuring robust and generalizable findings
- \* Found that despite overall performance decline with smaller models, **there is no significant and stable disparity in performance degradation rates between high-resource and low-resource languages**, challenging assumptions about inherent disadvantages for low-resource languages in smaller models
- \* Demonstrated that compact models can preserve relative multilingual competence, highlighting their viability for globally inclusive deployment and motivating linguistically fair compression strategies

#### • M4CQ Dataset Construction & Multilingual Evaluation

- \* Constructed a Massive Multilingual Multitask Multiple Choice Question (**M4CQ**) dataset featuring content consistency across **19** languages and domain-balanced distribution of **135** tasks
- \* Established the first comprehensive and hierarchical task categorization paradigm through systematic survey of twenty existing diverse datasets, providing a unified framework for task categorization
- \* Conducted comprehensive evaluations of multiple LLMs on M4CQ to analyze their multilingual capability

- **Language Similarity Analysis from the Perspective of LLMs**

- \* Proposed and formalized the novel research problem of **LLM-based language similarity**, challenging traditional linguistic classification paradigms and defining language similarity from the perspective of LLMs
- \* 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

- **Enhancing LLM Performance via LLM-Based Language Similarity**

- \* Investigated how LLM-derived language similarity can guide fine-tuning strategies to improve model performance on specific target languages
- \* Demonstrated that **selecting fine-tuning languages similar to the target language from the LLM perspective** yields better performance than selecting languages based on traditional conclusion on language families
- \* Showed that LLM-based similarity-guided fine-tuning sometimes even **outperforms direct fine-tuning on the target language** itself, revealing a super-additive transfer effect
- \* Provided actionable insights on fine-tuning languages selection strategies especially when target languages are low-resource languages, offering a new paradigm for language-specific performance enhancement

- **M4CQ-Pro Dataset Construction & Multilingual Evaluation**

- \* Constructed a massive multilingual multitask multiple choice question dataset **M4CQ-Pro**, which features content consistency across **31** languages and domain-balanced distribution of 135 tasks, with **61518** manually reviewed high-quality questions per language.
- \* Conducted comprehensive evaluations of multiple LLMs on M4CQ-Pro to analyze their multilingual capability  
(The dataset will be released after subsequent work has been done)

## Selected Honors & Scholarships

---

- **Outstanding Graduate of Zhejiang University**
- **Zhejiang Provincial Scholarship**
- **Zhejiang University Scholarship - First Prize** (Top 3% in Zhejiang University)
- **Kolors-LoRA Style Challenge – Top 10** (Team Leader, 10/1,338 teams; jointly hosted by Alibaba Cloud and ModelScope; supervised by Prof. [Min Zhang](#))
- **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 in Zhejiang University)  
Completed over 300 volunteer hours in hospital, community, and campus outreach programs
- **Varsity Kayaking Athlete**  
Competed in 4 national-level kayaking 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