

CHEN ZHANG

✉ zhangch@pku.edu.cn 🏠 luciusssss.github.io

I am a final-year Ph.D. student at Wangxuan Institute of Computer Technology, Peking University, advised by Prof. Yansong Feng. Currently, my research interests include:

- (1) **NLP for Low-Resource Languages:** Developing multilingual LLMs with stronger support for low-resource languages and improving accessibility for underrepresented communities.
- (2) **LLMs for Language Science:** Leveraging LLMs to model linguistic phenomena, such as language contact and acquisition, for deeper insights into how human languages function and evolve.
- (3) **Complex Reasoning with LLMs:** Harnessing the potential of LLMs to tackle complex problems and building reliable systems for specialized domains.

Previously, I worked on question answering, focusing mainly on machine reading comprehension and question answering over knowledge bases.

EDUCATION

Ph.D. Student in Natural Language Processing

September 2021 - Now

Wangxuan Institute of Computer Technology, Peking University

B.Sc. in Computer Science

September 2017 - June 2021

Department of Computer Science and Technology, Peking University

SELECTED PUBLICATIONS

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

Chen Zhang, Jiuheng Lin*, Xiao Liu, Zekai Zhang, Yansong Feng. ACL 2025*

Cross-Lingual Transfer of Cultural Knowledge: An Asymmetric Phenomenon

Chen Zhang, Zhiyuan Liao, Yansong Feng. ACL 2025

MiLiC-Eval: Benchmarking Multilingual LLMs for China's Minority Languages

Chen Zhang, Mingxu Tao, Zhiyuan Liao, Yansong Feng. ACL 2025 (Findings)

Eliciting and Improving the Causal Reasoning Abilities of LLMs with Conditional Statements

Xiao Liu, Da Yin, Chen Zhang, Yansong Feng, Dongyan Zhao. Computational Linguistics 2025

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*

Teaching Large Language Models an Unseen Language on the Fly

Chen Zhang, Xiao Liu, Jiuheng Lin, Yansong Feng. ACL 2024 (Findings)

Unlocking the Potential of Model Merging for Low-Resource Languages

Mingxu Tao, Chen Zhang*, Quzhe Huang*, ..., Yansong Feng. EMNLP 2024 (Findings)*

Harder Task Needs More Experts: Dynamic Routing in MoE Models

Quzhe Huang, Zhenwei An*, Nan Zhuang, Mingxu Tao, Chen Zhang, ..., Yansong Feng. ACL 2024*

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)

The Magic of IF: Investigating Causal Reasoning Abilities in Large Language Models of Code

Xiao Liu, Da Yin, Chen Zhang, Yansong Feng, Dongyan Zhao. ACL 2023 (Findings)

Cross-Lingual Question Answering over Knowledge Base as Reading Comprehension

Chen Zhang, Yuxuan Lai, Yansong Feng, ..., Dongyan Zhao. EACL 2023 (Findings)

Extract, Integrate, Compete: Towards Verification Style Reading Comprehension

Chen Zhang, Yuxuan Lai, Yansong Feng, Dongyan Zhao. EMNLP 2021 (Findings)

A review of deep learning in question answering over knowledge bases

Chen Zhang, Yuxuan Lai, Yansong Feng, Dongyan Zhao. AI Open 2021

Why Machine Reading Comprehension Models Learn Shortcuts

Yuxuan Lai, Chen Zhang, Yansong Feng, Quzhe Huang, Dongyan Zhao. ACL 2021 (Findings)

PROJECTS

Lawyer Llama - A Chinese Legal LLM

- One of the first Chinese legal LLMs, 900+ stars on Github.
- Continually pretrained on legal texts and finetuned on legal consultation data curated by GPT-4o.
- Boosted reasoning ability with reinforcement learning on multi-choice questions from legal exams.

Training LLMs from Scratch (*intern in a startup*)

- Training a Llama-2-like 7B LLM on carefully curated 1T-token data.
- Achieving comparable performance with contemporary models in 2023 (ChatGLM2-6B, Baichuan2-7B).
- Extended to 128K context length through continual pretraining and Adjusted Base Frequency.

AWARDS & HONORS

- President Scholarship, Peking University, 2025
- Award for Scientific Research, Peking University, 2024
- Outstanding Graduates of Beijing, 2021
- Excellent Graduate of Peking University, 2021
- Merit Student of Peking University, 2018 & 2019
- Founder Scholarship, 2018 & 2019

ACADEMIC SERVICES

Reviewer:	ACL, EMNLP, ACL Rolling Review, COLING, LREC, NLPCC
Area Chair:	ACL Rolling Review, LREC
Workshop Organizer:	Widening NLP 2025

TEACHING

Teaching Assistant

- Foundations of Natural Language Processing, Peking University, Spring 2024 & Spring 2025
- Empirical Methods for Natural Language Processing, Peking University, Spring 2022
- Data Structures and Algorithms, Peking University, Fall 2020 & Spring 2021

SKILLS

Programming:	C/C++, Python
Frameworks:	PyTorch, Megatron, VeRL
Human Languages:	Chinese (Native), English (Fluent), Japanese (Intermediate)