Shaoyang Xu

Tianjin University, No. 92 Weijin Road, Xuefu Street, Nankai District, Tianjin
📱 (+86) 15222735523 | ✉ syxu123321@gmail.com | 🎂 Aug ’00 | Github | G Google Scholar

Education

Singapore University of Technology and Design *Singapore*
Incoming PhD Student at ISTD *Sept 2025 - Sept 2029 (expected)*
Supervisor: Prof. Wenxuan Zhang
Research Focuses: Multilinguality, Reasoning, LLMs

Tianjin University *Tianjin, China*
Master, Computer Science *Sept 2022 - Mar 2025*
Supervisor: Prof. Deyi Xiong
GPA: 87.9/100 (Overall Ranking: 7/20)
Research Focuses: Multilingual and Multicultural Large Language Models

Soochow University *Suzhou, China*
Bachelor, Artificial Intelligence *Sept 2018 - Jun 2022*
Supervisor: Prof. Peifeng Li & Dr. Feng Jiang
GPA: 91.0/100 (Overall Ranking: 5/70)
Courses: Machine Learning (98), Neural Network Principle (93), Python Programming (93), Advanced Mathematics (95), Linear Algebra (98), etc.

Research Projects

Pluralistic Culture Alignment of LLMs （NAACL 2025） [Paper] [Code] *Mar 2024 - Oct 2024*

Research Question: LLMs exhibit cultural subjectivity. Can we leverage LLMs’ inherent knowledge about different cultures to enhance its alignment with pluralistic cultures?
Method: Proposing a framework for self-pluralising culture alignment, which includes (1) generating cultural questions, (2) yielding culture-aware/unaware LLM outputs, (3) collecting cultural data based on output inconsistencies, (4) cultural joint/specific model training.
Experiment: Conducting experiments on LLaMA3 across 18 countries from 5 continents.
Conclusion: Empirically confirming the feasibility of aligning LLMs to pluralistic cultures using LLMs’ own knowledge. Several questions were explored: Whether cultural-joint or specific training works better? What is the mechanism behind the method? Can LLMs’ output reflect inter-cultural relationships? What is the effect of data quality/quantity?

Exploring Abstract Concepts in Multilingual LLMs （EMNLP 2024） [Paper] [Code] *Oct 2023 - Feb 2024*

Research Question: Do LLMs encode abstract concepts similarly to human beings in multiple languages, and how are these concepts represented, consistent and transferred across languages?
Method: Proposing a framework to explore the existence of multilingual abstract concepts in LLMs and perform cross-lingual analysis on them.
Experiment: Conducting experiments on 7 abstract concepts related to human values, across 16 languages and 3 LLM families, each exhibiting monolingual, bilingual, and multilingual properties, respectively.
Conclusion: Empirically substantiating the existence of multilingual abstract concepts in LLMs, and identifying 3 interesting cross-lingual traits of these concepts arising from language resource disparities: cross-lingual inconsistency, distorted linguistic relationships, and unidirectional cross-lingual transfer between high- and low-resource languages.

Cross-Lingual Knowledge Transfer （EMNLP 2023） [Paper] *Mar 2023 - Jun 2023*

Research Question: Are knowledge and linguistic capabilities of LLMs decoupled, and can knowledge be transferred across languages?
Method: Proposing a method that enables LLMs to “think” in English while answering in non-English. This involves two language representation space projection: the first one projects non-English representations into English, while the second one performs a back-projection.
Experiment: Conducting experiments on 2 multilingual factual knowledge probing benchmarks, across 53 languages and 44 knowledge types.
Conclusion & Analysis: Improving factual knowledge retrieval accuracy and facilitating knowledge transfer across languages. & Performing interpretable analyses from the perspective of representation space and knowledge neurons.

Zero-Shot Multilingual Machine Translation *Nov 2022 - Feb 2023*

Preliminary Experiments: Conducting an in-depth analysis of multilingual machine translation models (encoder-decoder), unveiling the presence of inconsistent distribution patterns in representations between English and non-English sentences at the encoder side.
Method: Proposing a novel module to disentangle language-specific information from semantic information. After decoupling, only the language agnostic semantic information from the encoder is preserved and sent to fine-tune the decoder.
Result: Improving zero-shot Translation BLEU score from 4.52 to 10.83 on OPUS100 dataset. However, the performance remains below that of English-pivot translation (14.61), indicating room for further improvement.

Publications

Self-Pluralising Culture Alignment for Large Language Models
Shaoyang Xu, Yongqi Leng, Linhao Yu, Deyi Xiong
NAACL 2025

Exploring Multilingual Concepts of Human Values in Large Language Models: Is Value Alignment Consistent, Transferable and Controllable across Languages?
Shaoyang Xu, Weilong Dong, Zishan Guo, Xinwei Wu, Deyi Xiong
EMNLP 2024 Findings

Language Representation Projection: Can We Transfer Factual Knowledge across Languages in Multilingual Language Models?
Shaoyang Xu, Junzhuo Li, Deyi Xiong
EMNLP 2023

FuxiTranyu: A Multilingual Large Language Model Trained with Balanced Data
Haoran Sun, Renren Jin, **Shaoyang Xu**, Leiyu Pan, Menglong Cui, Jiangcun Dui, Deyi Xiong, etc.
EMNLP 2024 (Industry Track)

Mitigating Privacy Seesaw in Large Language Models: Augmented Privacy Neuron Editing via Activation Patching
Xinwei Wu, Weilong Dong, **Shaoyang Xu**, Deyi Xiong
ACL 2024 Findings

ConTrans: Weak-to-Strong Alignment Engineering via Concept Transplantation
Weilong Dong, Xinwei Wu, Renren Jin, **Shaoyang Xu**, Deyi Xiong
COLING 2025

Multilingual Large Language Models: A Systematic Survey
Shaolin Zhu, Supryadi, **Shaoyang Xu**, Haoran Sun, Leiyu Pan, Menglong Cui, Jiangcun Du, Renren Jin, António Branco, Deyi Xiong
Under Review

DCIS: Efficient Length Extrapolation of LLMs via Divide-and-Conquer Scaling Factor Search
Lei Yang, **Shaoyang Xu**, Deyi Xiong
Under Review

Topic Segmentation via Discourse Structure Graph Network
Shaoyang Xu, Feng Jiang, Peifeng Li
Journal of Chinese Information Processing 2021

Work Experience

Large Language Model and Multimedia Technology Department, Kuaishou Technology *Beijing, China*
LLMs Algorithm Intern *May 2024 - Sept 2024*

Executing a technical roadmap including data construction, SFT, reward modeling, and DPO to enhance the role-playing capabilities of LLMs. Building an evaluation pipeline with benchmarks such as MMLU, GSM8K, and IFEval to assess the general capabilities of trained models.

Awards and Honors

2019	1st Student Scholarship, Academic Excellence Award	*SUDA*
2020	2nd Student Scholarship, Merit Students Award, 3rd Prize of CCSP2020 (East China Division)	*SUDA*
2021	1st Student Scholarship, Merit Students Award, 2nd Prize of National LanQiao Cup	*SUDA*
2022	Excellent Undergraduate Thesis / 1st Student Scholarship	*SUDA / TJU*
2023	2nd Student Scholarship, Advanced Individual Award	*TJU*
2024	2nd Student Scholarship, Advanced Individual Award	*TJU*

Skills

Basic Programming	Python, Shell, LaTeX
Model Training	Pytorch, Transformers, LLaMA-Factory, DeepSpeed
Model Inference	SGLang
Languages	Mandarin, English (IELTS Score: 7.0, Listening-7.5; Reading-8.0; Writing-6.5; Speaking-6.0)