Beiduo Chen

beiduochen@cis.lmu.de | (49) 15222000786 | GitHub | Google Scholar

EDUCATION

Ludwig-Maximilians-Universität München — Munich, Germany
Ph.D. in Natural Language Processing at MaiNLP Lab, supervised by Prof. Barbara Plank — *2024 - 2027 (expected)*
Topic: Human-centered NLP; Uncertainty, Trustworthiness and Evaluation of LLM

University of Cambridge — Cambridge, United Kingdom
Exchange ELLIS Ph.D. Student at Language Technology Lab, supervised by Prof. Anna Korhonen — *2025*

University of Science and Technology of China — Anhui, China
Master of Engineering in Information and Communication Engineering, supervised by Prof. Wu Guo — *2020 - 2023*
Thesis: A Study on Multilingual Representation Learning and Application based on Pre-trained Language Model — *GPA: 3.93/4.3, top **1**%*

University of Science and Technology of China — Anhui, China
Bachelor of Engineering in Electronic and Information Engineering, supervised by Prof. Wu Guo — *2016 - 2020*
Thesis: Speaker Recognition based on Depth Features — *GPA: 3.75/4.3, top **3**%*

EXPERIENCE

Microsoft Research Asia — Beijing, China
Research intern, hosted by Shaohan Huang — *Jun 2022 - Jan 2023*
Topic: Language Model Pre-training

iFLYTEK Research — Anhui, China
Research intern, hosted by Quan Liu — *Jun 2021 - Mar 2022*
Topic: Multilingualism; Cross-lingual Transfer; Named Entity Recognition

PUBLICATIONS

- **Beiduo Chen**, Yang Janet Liu, Anna Korhonen, Barbara Plank. Threading the Needle: Reweaving Chain-of-Thought Reasoning to Explain Human Label Variation. *The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2025. **Oral Presentation**.
- Pingjun Hong*, **Beiduo Chen***, Siyao Peng, Marie-Catherine de Marneffe, Barbara Plank. LiTEx: A Linguistic Taxonomy of Explanations for Understanding Within-Label Variation in Natural Language Inference. *The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2025.
- **Beiduo Chen**, Siyao Peng, Anna Korhonen, Barbara Plank. A Rose by Any Other Name: LLM-Generated Explanations Are Good Proxies for Human Explanations to Collect Label Distributions on NLI. *Findings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 2025.
- **Beiduo Chen**, Xinpeng Wang, Siyao Peng, Robert Litschko, Anna Korhonen, Barbara Plank. "Seeing the Big through the Small": Can LLMs Approximate Human Judgment Distributions on NLI from a Few Explanations? *Findings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2024.
- **Beiduo Chen**, Shaohan Huang, Zihan Zhang, Wu Guo, Zhenhua Ling, Haizhen Huang, Furu Wei, Weiwei Deng and Qi Zhang. Pre-training Language Model as a Multi-perspective Course Learner. *Findings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*. 2023.
- **Beiduo Chen**, Wu Guo, Bin Gu, Quan Liu, Yongchao Wang. Multi-Level Contrastive Learning for Cross-Lingual Alignment. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022.
- Jun-Yu Ma*, **Beiduo Chen***, Jia-Chen Gu, Zhenhua Ling, Wu Guo, Quan Liu, Zhigang Chen and Cong Liu. Wider & Closer: Mixture of Short-channel Distillers for Zero-shot Cross-lingual Named Entity Recognition. *The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2022.
- **Beiduo Chen**, Wu Guo, Quan Liu, Kun Tao. Feature Aggregation in Zero-Shot Cross-Lingual Transfer Using Multilingual BERT. *The 26th International Conference on Pattern Recognition (ICPR)*. 2022.
- **Beiduo Chen**, Jun-Yu Ma, Jiajun Qi, Wu Guo, Zhen-Hua Ling, Quan Liu. USTC-NELSLIP at SemEval-2022 Task 11: Gazetteer-Adapted Integration Network for Multilingual Complex Named Entity Recognition. *The 16th International Workshop on Semantic Evaluation (SemEval) at 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. 2022.
- Raoyuan Zhao, **Beiduo Chen**, Barbara Plank, Michael A. Hedderich. MAKIEval: A Multilingual Automatic WiKidata-based Framework for Cultural Awareness Evaluation for LLMs. *Findings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2025.

Shared Tasks

- 2022, Rank 1st on three tracks (Chinese, Code-mixed and Bangla), and rank 2nd on the other ten tracks in the 16th International Workshop on Semantic Evaluation (SemEval-2022) Task 11 Multilingual Complex Named Entity Recognition, as the first author.

Patent

- **Beiduo Chen**, Qingqing Huang, Jun Du. Multi-Feature Fusion Method for Neural Machine Translation Error Detection Based on Data Enhancement Training. Patent of China National Intellectual Property Administration (CNIPA). 2021.

Talks

- 2025, Invited Talk, 4th International Workshop on Dependability Modeling and Digitalization (WDMD) at The 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2025): Understanding and Modeling Human Label Variation in LLM.

Honors

- 2025, Mobility Grant of European Lighthouse on Secure and Safe AI (ELSA).
- 2024, European Laboratory for Learning and Intelligent Systems (ELLIS) Ph.D. Program (5% acceptance).
- 2023, Outstanding Graduate Award of University of Science and Technology of China.
- 2023, Outstanding Graduate Award of Ordinary Colleges and Universities in Anhui Province.
- 2022, China National Scholarship.
- 2020, Suzhou Yucai Scholarship: Top 1 GPA in the class (1/120).
- 2019, Scholarship of the Institute of Electrics, Chinese Academy of Sciences.
- 2018, The third prize (provincial) in Contemporary Undergraduate Mathematical Contest in Modeling of China.
- 2017, Gold Award for Outstanding Student of USTC.

Service

- Program Committee/Reviewer: TPAMI, ACL, EMNLP, NAACL, ACL Rolling Review, COLM, ICASSP, ICPR, NLPOR@COLM2025.

Teaching

- Lecture: Information Retrieval (SS2025, LMU Munich, 2025), LLM Agents (SS2025, LMU Munich, 2025), Symbolic Programming Language (WS2024/25; WS2025/26, LMU Munich, 2024; 2025), Multi-modal NLP (SS2024, LMU Munich, 2024). Signals and Systems (210049, 210049.05, USTC, 2021), Computer Programming A (CS1001A, 210522.02, USTC, 2019), Electromagnetism C (PHYS1004C, 022503.03, USTC, 2018).
- Seminar: Explaining and Interpreting Annotations in NLP (WS2025/26, LMU Munich, 2025), Discourse Modeling and Processing (WS2024/25, LMU Munich, 2024), NLP for Climate Change (SS2024, LMU Munich, 2024).

Mentorship

- 2025, Pingjun Hong, MSc Thesis on Within-label Variation. *Within-Label Variation in Natural Language Inference: A Linguistic Taxonomy for Explanations and Its Impact on Model Interpretation of Label Decisions*.
- 2024, Vu Thanh Trung Bui, BSc Thesis on Multilingual LLM. *Exploring Multilingual Capabilities in Large Language Models with Soft Prompt Tuning*.

Skills

- Languages: Python, HTML, C/C++, Matlab
- Frameworks: PyTorch, TensorFlow, HuggingFace, vLLM, DeepSpeed, Transformers, Keras, Scikit-Learn, Jupyter, CUDA