

Xinyuan Cheng

✉ Xinyuan.Cheng@campus.lmu.de

in LinkedIn

G Scholar

🌐 <https://noah-13.github.io/xinyuan-cheng/>

🐙 GitHub

Status: June 2026



Research Interests

My research focuses on evaluating and interpreting foundation model behavior beyond single accuracy numbers. In particular, I study how large language models reason, how intermediate reasoning transfers across models [2], and how evaluation frameworks can characterize model behavior in Retrieval-Augmented Generation (RAG), multilingual/pragmatic reasoning [1], [5], and Human Label Variation (HLV). More broadly, I am interested in robust evaluation and representation analysis for foundation models across language, reasoning, and multimodal tasks.

Education

- | | |
|-----------------------|--|
| Oct. 2024 – Present | ■ M.Sc. University of Munich (LMU) Computational Linguistics and Informatics
Grade: 1.06/1.0 |
| Sep. 2020 – Jul. 2024 | ■ B.A. Shanghai Jiao Tong University (SJTU) German Linguistics
Thesis title: <i>The evaluation of large language models with regard to the ambiguity recognition of the word “wieder”.</i> |
| Oct. 2022 – Feb. 2023 | ■ Exchange, University of Munich (LMU) Computational Linguistics
Grade: 1.0/1.0 |

Experience

- | | |
|-----------------------|--|
| Jan. 2026 – Present | ■ Scientific Programmer, Munich AI & NLP (MaiNLP), University of Munich (LMU)
Developing a open-source, research-oriented toolkit for HLV, including model training, evaluation, and experimental analysis pipelines. |
| Jul. 2025 – Dec. 2025 | ■ Research Assistant, Munich AI & NLP (MaiNLP), University of Munich (LMU)
Contributed to the Premise and PyPremise projects, worked on code development, adding tutorials and improving cross-platform support.
Leading an empirical study on the meta-evaluation of RAG metrics to benchmark their reliability and alignment with human judgments. |
| Mar. 2023 – Jun. 2024 | ■ Research Assistant, Computational Linguistics (CL) Lab at School of Foreign Languages (SFL), Shanghai Jiao Tong University (SJTU)
Contributed to SwordsmanImp [5], a Chinese conversational implicature benchmark, with a focus on evaluation dataset construction. Assisted in the ArguGPT project [6] through syntactic and linguistic analysis of GPT-generated argumentative essays. |

Research Publications

- 1 C. Chan, Y. Yim, H. Zeng, Z. Zou, **X. Cheng**, Z. Sun, Z. Deng, K. Chung, Y. Ao, Y. Fan, C. Jiayang, E. Nie, G. Y. Wong, H. Schmid, H. Schütze, S. See, and Y. Song, “Xtom: Exploring the multilingual theory of mind for large language models,” in *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, Main Conference, accepted, 2026. arXiv: 2506.02461 [cs.CL].  [Online]. Available: <https://arxiv.org/abs/2506.02461>.
- 2 **X. Cheng**, B. Chen, P. Mondorf, and B. Plank, “Reasoning that travels: Dissecting how chain-of-thought transfers across models,” *arXiv preprint arXiv:2605.28913*, 2026, **Under review for EMNLP 2026**. arXiv: 2605.28913 [cs.CL].  [Online]. Available: <https://arxiv.org/abs/2605.28913>.
- 3 Y. Du, H. Liu, H. Peng, **X. Cheng**, C. Wu, and J. Zhang, “Gmg: A video prediction method based on global focus and motion guided,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 36, no. 6, pp. 8048–8063, 2026.  DOI: 10.1109/TCSVT.2026.3657055.
- 4 Y. Du, J. Zhang, **X. Cheng**, Y. Lu, D. Li, and W. Tian, “Predicting sudden stratospheric warmings using video prediction methods,” *Geophysical Research Letters*, vol. 52, no. 8, e2024GL113993, 2025.  DOI: <https://doi.org/10.1029/2024GL113993>.
- 5 S. Yue, S. Song, **X. Cheng**, and H. Hu, “Do large language models understand conversational implicature- a case study with a Chinese sitcom,” eng, in *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 1: Main Conference)*, S. Maosong, L. Jiye, H. Xianpei, L. Zhiyuan, and H. Yulan, Eds., Taiyuan, China: Chinese Information Processing Society of China, Jul. 2024, pp. 1270–1285.  [Online]. Available: <https://aclanthology.org/2024.ccl-1.98/>.
- 6 Y. Liu, Z. Zhang, W. Zhang, S. Yue, X. Zhao, **X. Cheng**, Y. Zhang, and H. Hu, *Argugpt: Evaluating, understanding and identifying argumentative essays generated by gpt models*, 2023. arXiv: 2304.07666 [cs.CL].

Skills

Natural Languages	 Mandarin Chinese (Native), English (Fluent), German (Fluent)
Programming Languages	 Python (Proficient), Bash (Basic)
Deep Learning	 PyTorch, vLLM, Hugging Face Transformers, NumPy, scikit-learn
Tools	 L ^A T _E X, PyCharm, Linux, Jupyter Notebook, VS Code, Git