Zhaoxin Feng

Department of Language Science and Technology
The Hong Kong Polytechnic University
Hong Kong SAR, China

zhaoxinbetty.feng@connect.polyu.hk
Supervisor: Prof. Emmanuele Chersoni
Co-supervisor: Prof. Huang Chu-Ren

Education

† *Indicates expected*

2024–2028 †	Ph.D., Computational Linguistics, The Hong Kong Polytechnic University
	Supervisor: Emmanuele Chersoni, Co-supervisor: Huang Chu-Ren
2023 Spring	Exchange Student, The Hong Kong Polytechnic University
2020–2024	B.Sc., Educational Technology (Minor: Data Science), Beijing Normal University
	GPA: 3.7/4.0 (top 7%); Outstanding Undergraduate Award

Research Overview

My research asks two connected questions: do LLMs truly understand human language, and can we rely on them when they interact with humans? In practice, my work splits into two threads:

- **Probing semantic representations.** I probe how LLMs represent human language: whether enabling bidirectional attention in autoregressive LLMs reshapes their lexical semantic representations, and how their embeddings compare to human mental lexicon judgments.
- **Reliability of LLMs and LLM-based agents.** I study the reliability of LLMs and LLM-based agents when interacting with humans, from sycophancy that chain-of-thought reasoning both mitigates and masks, to memory agents that can recall user preferences but fail to act on them, to reasoning failures on Chinese phonological ambiguities, etc.

Publications

* *Indicates co-first author*

Conference Papers

[1] **Feng, Z.**, Chen, Z., Ma, J., Yip, T. P., Chersoni, E., & Li, B. (2026). Good arguments against the people pleasers: How reasoning mitigates (yet masks) LLM sycophancy. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics* (ACL 2026).

[2] Ma, J.*, **Feng, Z.***, Chersoni, E., Song, H., & Zhang, Z. (2025). PhonoThink: Improving large language models' reasoning on Chinese phonological ambiguities. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2025, Oral).

[3] **Feng, Z.**, Ma, J., Chersoni, E., Zhao, X., & Bao, X. (2025). Learning to look at the other side: A semantic probing study of word embeddings in LLMs with enabled bidirectional attention. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (ACL 2025, Oral).

[4] Song, H., **Feng, Z.**, Chersoni, E., & Huang, C.-R. (2025). Which model mimics human mental lexicon better? A comparative study of word embedding and generative models. In *Proceedings of the 16th International Conference on Computational Semantics* (IWCS 2025, Oral).

[5] Ma, J.*, **Feng, Z.***, Song, H., Chersoni, E., & Chen, Z. (2025). Reasoning or memorization? Investigating LLMs' capability in restoring Chinese internet homophones. In *Proceedings of the 3rd Workshop on Towards Knowledgeable Language Models* (KnowFM 2025).

[6] Chen, Z., **Feng, Z.**, Ma, J., Xu, J., & Li, B. (2025). Can LLMs recognize their own analogical hallucinations? In *Proceedings of the 3rd Workshop on Towards Knowledgeable Language Models* (KnowFM 2025).

Submissions Under Review

[1] **Feng, Z.**, Ma, J., Chen, Z., & Chersoni, E. KnowAct: Do LLM agents act on what they remember? Submitted to *Conference on Language Modeling* (COLM 2026).

[2] Chen, Z., **Feng, Z.**, Yip, T. P., Ma, J., Chersoni, E., & Li, B. PCA-guided activation scaling for monotonic bidirectional control over LLM sycophancy. Submitted to *Conference on Language Modeling* (COLM 2026).

Selected Coursework

2024–Present **Ph.D.-level**: Computational Linguistics; Advanced Natural Language Processing; Advanced Topics in Computer Algorithms; Advanced Large-Scale Machine Learning Systems for Foundation Models; Artificial Intelligence

2020–2024 **Bachelor-level**: Single Variable Calculus; Multivariable Calculus and Linear Algebra; Discrete Mathematics; Probability Theory and Mathematical Statistics; Stochastic Process; Deep Learning; Mathematical Modeling; Computer Network Programming; Data Structure; Introduction to Pattern Recognition; Software Engineering; Educational Application of AI

Selected Honours and Awards

2025 HKSAR Talent Development Scholarship

2024 Outstanding Undergraduate Graduate Award, Beijing Normal University

2024 Honourable Mention, Interdisciplinary Contest in Modeling (ICM)

2023 PolyU Presidential PhD Fellowship

2022 First-Class Academic Scholarship, Beijing Normal University

2022 First-Class Competition Scholarship, Beijing Normal University

2022 Merit Award for Outstanding Students, Beijing Normal University

2022 Second Prize, 15^{th} National Undergraduate Mathematics Competition

2022 Gold Award, International Genetically Engineered Machine Competition (iGEM)

2022 Volunteer Service Award

Teaching

2026 Spring Teaching Assistant, Programming and Data Analysis for Language Studies

2025 Fall Teaching Assistant, Computer Programming in Language and Communication

2025 Spring Teaching Assistant, Programming and Data Analysis for Language Studies

Research and Industry Experience

[1] **ByteDance**, Beijing, China *Feb 2024 – Jun 2024*
Strategy Product Manager Intern (AI Application), ZERO Team. Developed product strategy for *Doubao* across OCR, retrieval-augmented generation (RAG), plugins, and agent workflows in the scenario of education.

[2] **Polytechnique Montreal**, Montreal, Canada *May 2023 – Aug 2023*
Research Intern. Led the project named *Understanding the Characteristics of Internet Slangs in OSI Discussions*; and attended another project related to Human Computer Interaction, responsible for integrated physical objects into AR interaction and developing virtual-button functionality in Unity.

[3] **Institute of Psychology, Chinese Academy of Sciences**, Beijing, China *Oct 2022 – Jan 2023*
Research Intern. Conducted data analysis and text mining on a dataset of over 50,000 rows using Python and R.

Where I Was and Will Be

2026	ACL 2026, San Diego, US
2025	IWCS 2025, Düsseldorf, Germany
2025	EMNLP 2025, Suzhou, China
2025	ACL 2025, Vienna, Austria