Yunzhen He

Email: heyunzhenwhat@gmail.com
GitHub: https://github.com/githubhyz
Google Scholar: https://scholar.google.co.jp/citations?user=Fhb4RNoAAAAJ

Education

Chuo University, 2018 – 2022
Faculty of Science and Engineering, Department of Information and System Engineering

- **Bachelor's Degree**

Kyoto University, 2023 – 2025
Graduate School of Informatics, Data Science Course

- **Master's Degree**
- **Adviser:** Hidetoshi Shimodaira

Kyoto University, 2025 – Present
Graduate School of Informatics, Data Science Course

- **Ph.D. student**
- Expected Completion: 2028
- **Adviser:** Hidetoshi Shimodaira

Professional Experience

Kyoto University, April 2024 – Present

- Department: Graduate School of Informatics, Kyoto University
- Role: Research Assistant

Spiral AI, April 2024 – August 2024

- Department: SpiralAI Inc. / LLM Deep Learning Engineer / R&D
- Role: Intern

Rakuten, November 2024 – January 2025

- Department: Rakuten Group, Inc. / Management Team / AI & Data Division / AI Service Management Department / AI for Business Department / AI Architecture Section
- Role: Part-time

Awards and Honors

Sponsor Award, 2024

- Awarded at the Association for Natural Language Processing (ANLP, Japan) 2024, conference for outstanding contributions in the field of natural language processing.

DoGS Fellow, April 2025 – March 2028

- Financial Support - Support for devotion to research: 250,000 yen per month. Research grant: 900,000 yen per year.
- Issued by Director, Division of Graduate Studies in Kyoto University

Chief of Graduate School of Informatics, Kyoto University, 2025

Papers

1. **大規模言語モデルにおける幻覚緩和のための単語確率の外挿.**
   Yunzhen He, Yusuke Takase, Yoichi Ishibashi, Hidetoshi Shimodaira. *The Association for Natural Language Processing*, 2024.
2. **Shimo Lab at "Discharge Me!": Discharge Summarization by Prompt-Driven Concatenation of Electronic Health Record Sections.**
   Yunzhen He, Hiroaki Yamagiwa, Hidetoshi Shimodaira. Proceedings of the 23rd Workshop on

Biomedical Natural Language Processing, Association for Computational Linguistics, 2024.

3. **DeLTa: A Decoding Strategy based on Logit Trajectory Prediction Improves Factuality and Reasoning Ability.**
   Yunzhen He, Yusuke Takase, Yoichi Ishibashi, Hidetoshi Shimodaira. Second Workshop on Uncertainty-Aware NLP, Empirical Methods in Natural Language Processing, 2025.

Skills

- **Programming Languages:** Python, C
- **Tools & Technologies:** PyTorch, Transformers, Transformer lens, einops.

Research interests

- LLM Hallucinations, Generation, Diffusion LLM, Post training, Alignment, Agentic RL, Benchmarks.

Languages

- Japanese: Native level
- Chinese: Native level
- English: Professional working proficiencys