Jeongyeon Hwang

Email: oppurity@postech.ac.kr

Github: https://github.com/oppurity12

Research Interest

I'm a 4th-year integrated M.S./Ph.D. student at POSTECH. My research focuses on improving the reliability of ML/NLP systems in real-world settings. I study large language models (LLMs), with an emphasis on robustness against malicious inputs, corrupted training data, and misuse such as fake content generation. Recently, I've been particularly interested in watermarking techniques for detecting LLM-generated text.

Education

- **POSTECH** — South Korea
  Integrated M.S/Ph.D Student in Artificial Intelligence — *2023-Present*
- **Sungkyunkwan University** — South Korea
  B.S. in Mathematics — *2017-2023*

Experiences

- **Research Internship** — USA
  New York University — *2025*
- **Military Service** — South Korea
  Korean Augmentation to the U.S. Army (KATUSA), 188th Military Police Company — *2019-2021*

Publications

- **LLM Watermark Evasion via Bias Inversion**:
  Jeongyeon Hwang, Sangdon Park, Jungseul Ok
  arXiv 2025, Under review
- **Efficient Latent Semantic Clustering for Scaling Test-Time Computation of LLMs** :
  Sungjae Lee, Hoyoung Kim, **Jeongyeon Hwang**, Eunhyeok Park, Jungseul Ok,
  EMNLP, 2025 Findings (long)
- **Retrieval-Augmented Generation with Estimation of Source Reliability**:
  Jeongyeon Hwang, Junyoung Park, Hyejin Park, Sangdon Park, Jungseul Ok,
  EMNLP, 2025 Main (long)
- **MedBN: Robust Test-Time Adaptation Against Malicious Test Samples**:
  Hyejin Park*, **Jeongyeon Hwang***, Sunung Mun, Sangdon Park, Jungseul Ok,
  Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.
- **Addressing Feature Imbalance in Sound Source Separation**:
  Jaechang Kim, **Jeongyeon Hwang**, Soheun Yi, Jaewoong Cho, Jungseul Ok,
  arXiv 2023