Yohan Lee

yhlee.nlp@gmail.com | (+82) 10-6323-6913 | linkedin.com/in/l-yohai | github.com/l-yohai

Professional Experience

AI Researcher, Coxwave – Seoul, Korea Jul. 2024 – Curr.

Development of Domain-Specific LLMs for Quantum Physics
- Built an AI Tutor and Assistant for quantum physics using continual pre-training and fine-tuning (SFT, DPO)
- Built a unified model (Llama 3.1, 8B) with 81.1% accuracy on MCQ test sets, outperforming GPT-4o (63.7%)
- Secured a $140k contract and delivered AI solutions under the AI Voucher program

Research on Many-Shot Jailbreaking
- Developed a comprehensive attack framework for many-shot scenarios
- Analyzed long-context vulnerabilities in open-source LLMs

AI Research Engineer (NLP Specialist), WRTN Technologies – Seoul, Korea Mar. 2024 – Jun. 2024

Research on LLM Evaluation
- Designed and implemented a benchmark system for in-the-wild human-assistant dialogues
- Developed a evaluation framework for assessing human-assistant interactions in real-world scenarios

Research Scientist (NLP), Riiid – Seoul, Korea Jul. 2023 – Feb. 2024

Research on Large Language Models for Education
- Competed on the Huggingface Open LLM Leaderboard, achieving 1st place on Oct, 2023
- Explored the effects of instruction tuning from data (quantity, quality, diversity) and model (scale, efficiency, objective) perspectives
- Implemented diverse optimization techniques for efficient training and inference

Automated Essay Scoring
- Achieved state-of-the-art on public essay scoring benchmarks
- Conducted "Bar exam" scoring which performs better than GPT-4

NLP Engineer, Tunib – Seoul, Korea Dec. 2021 – Feb. 2023

Korean Open-domain Chatbot Service
- Directed dialogue data collection and quality filtering using advanced LLMs
- Developed an in-house Korean LM for multi-persona chatbot with self-collected datasets
- Operated a Kakaotalk-based chatbot service

AI Grand Challenge: Policy Support AI
- Awarded Ministry of Science and ICT Minister's Award
- Orchestrated TableQA data collection with policy domain experts
- Developed continual learning framework with OCR-based parsing and additional table data
- Developed an integrated QA system for processing texts, tables, and charts

Publications

What Really Matters in Many-Shot Attacks?
An Empirical Study of Long-Context Vulnerabilities in LLMs [pdf] Under Review, ARR December 2025

Yohan Lee, Sangyeop Kim, Yongwoo Song, Kimin Lee

- TLDR: This paper reveals that LLM vulnerabilities in long contexts are driven by context length rather than shot characteristics. Our attacks succeeded even with meaningless text or simple repeated examples, highlighting fundamental limitations in current safety mechanisms.

SAFARI: Sample-specific Assessment Framework for AI in Real-world Interactions [pdf] Under Review, NAACL 2025

Yohan Lee, Sungho Park, Sangwoo Han, Yunsung Lee, Yongwoo Song, Adam Lee, Jiwung Hyun, Jaemin Kim, Seungtaek Choi, Hyejin Gong

• TLDR: SAFARI is an automated evaluation framework for LLMs, leveraging real-world multilingual chat data to assess diverse skills and compare commercial and open-source systems in practical scenarios.

HEISIR: Hierarchical Expansion of Inverted Semantic Indexing for Training-free Retrieval of Conversational Data using LLMs [pdf] Under Review, NAACL 2025

Sangyeop Kim, Hangyeul Lee, ***Yohan Lee***

• TLDR: HEISIR (Hierarchical Expansion of Inverted Semantic Indexing for Retrieval) is a novel training-free method enhances conversational data retrieval by leveraging Large Language Models and a two-step semantic indexing.

Research & Teaching Experience

Research Assistant, Seoul National University – Seoul, Korea Jul. 2024 – Sep. 2024
Structure & Knowledge Injection into Machine Learning Lab, Prof. Jay-Yoon Lee.

Research on Reasoning and Rationality Methodologies for LLMs
• Investigated strategies like self-consistency and difficulty-based decoding to refine reasoning paths using pre-measured task difficulty
• Developed methods to pre-measure task difficulty and assess model rationality, improving reasoning in LLMs

Teaching Assistant, Naver Connect & Upstage – Seoul, Korea Jul. 2024 – Dec. 2024

Boostcamp AI Tech (Generation for NLP)
• Designed comprehensive curriculum and create in-depth lecture materials.
• Orchestrated hands-on NLP competitions to bridge theory and practical application.

Awards and Honors

2023 CJ Logistics Future Tech Challenge, CJ Logistics Sep, 2023

• CEO's Award, CJ Logistics
• Main Competition (Presentation for all task tracks): 2nd Place
• Preliminary Round (Unprocessed English address translation task): 1st Place (Accuracy 60% + Inference speed 40%)

2022 AI Grand Challenge: Policy Support AI, IITP Jan. 2023

• Minister's Award, Ministry of Science and ICT
• Main Competition (Text/Table/ChartQA task): 3rd place
• Prize: 2 billion KRW in support funding

3rd Annual University Student AI x Bookathon, SKKU, Naver Nov. 2021

• President's Award, Sungkyunkwan University
• Main Competition (Essay Creation task): Grand Prize

Scholarship, Yonsei University Aug. 2019

• Awarded for academic excellence and leadership potential

Education

Yonsei University, Seoul, Korea Mar 2015 – Feb 2022

• B.A in German Language and Literature, Cognitive Science
• Academic Background: Machine Learning and Its Applications, Deep Learning, Algorithmic Thinking, Digital Language Data and Humanities, Text Processing