Jiseung Hong

** I do not require visa sponsorship **

jiseungh@andrew.cmu.edu | (412) 638-3100 | https://wlqmfl.com/ | LinkedIn

Education

Carnegie Mellon University — Pittsburgh, PA

Master of Science in Intelligent Information Systems (MIIS) — GPA: 3.92/4.00 — Dec 2025

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea

Bachelor's Degree — School of Computing — Feb 2024

Korea Student Aid Foundation (KOSAF) National Scholarship — Mar 2017 ~ Feb 2024

Academic Research and Projects

OpenHands PR Arena: Platform for Evaluating and Benchmarking Agentic Coding Assistants — [Website, GitHub]

Carnegie Mellon University (*Prof. Graham Neubig; Neulab*) — Aug 2024 ~ Dec 2025

- Designed and developed PR Arena, a GitHub App that evaluates automated GitHub issue resolution by generating paired pull requests from state-of-the-art LLMs, enabling leaderboard-based comparison of human preferences across models.
- Released a public leaderboard and is in progress of promoting it to perform comprehensive research analyzing discrepancies between the human preference data (i.e., PR Arena evaluation) and the benchmark evaluation metrics (i.e., SWE-bench evaluation).
- Conducted inference testing (debugging sandboxed environment errors) and performance evaluation of coding agents (OpenHands framework + LLMs, e.g., o1-mini) on SWE Bench Lite to assess issue resolution capabilities. — [Blog Post]

Graph Guided Deep Reasoning for Long-Context Instruction Following

Emory University (*Prof. Jinho D. Choi; Emory NLP*) — June 2025 ~ Dec 2025

- In Progress

Quantifying and Reducing Bloats in Repository-Level Code Patches

Carnegie Mellon University (*Prof. Daniel Fried*) — June 2025 ~ Dec 2025

- In Progress

Tinker Tales: Interactive Storytelling Framework for Early Childhood Narrative Development and AI Literacy

Emory University (*Prof. Jinho D. Choi; Emory NLP*) — June 2025 ~ Dec 2025

- Designed an educational conversational agent (single-prompt) for children (ages 4~6) that scaffolds and evaluates narrative development, social-emotional learning, and early AI literacy.
- Employed graph-structured prompting to enforce step-by-step instruction following over a 7.3k+ tokens of system script.
- Implemented chatbot engine and web application on AWS, including user authentication, web hosting, data storage, and API calls.

Measuring Sycophancy of Language Models in Multi-turn Dialogues (*EMNLP 2025 Findings*) — [arXiv | GitHub]

Carnegie Mellon University & Emory University (*Prof. Kai Shu*) — Feb 2025 ~ May 2025

- Introduced *SYCON Bench*, a novel benchmark for evaluating and quantifying sycophantic behavior in real-world conversational settings (multi-turn, free-form, and open-ended) and proposed two metrics: *ToF* (Turn of Flip) and *NoF* (Number of Flip).
- Evaluated 17 Large Language Models (LLMs) on *SYCON Bench*, analyzing how model scaling (e.g., Qwen2.5-72B vs. 7B), reasoning-optimization (e.g., DeepSeek R1 vs. V3), and alignment tuning (Base model vs. Instruction tuned variant) influence sycophancy.
- Experimented 5 mitigation strategies, including third-person perspective prompting inspired by a psychological theory *Self-Talk*, which significantly reduces sycophancy in debate setting by 63.8%.

Measuring Social Biases in State-Spaces Models

Carnegie Mellon University (*Prof. Maarten Sap*) — Feb 2025 ~ May 2025

- Evaluated comparable transformer-based (Transformerpp-2.7B) and state-space (SSMs; Mamba2-2.7B, Mamba-2.8B) models to measure social biases.
- Found that SSMs, compared to transformer model, reduce bias by up to 4.46% on BBQ and 6% on StereoSet, while achieving a better trade-off between language modeling performance and fairness metrics.
- Fine-tuned both model types on PANDA dataset for debiasing and observed that (1) transformer models improved by up to 7.3% even on unseen bias types, whereas (2) SSMs show limited responsiveness to fine-tuning.

Korean Bio-Medical Corpus (KBMC) for Medical Named Entity Recognition (*LREC-COLING 2024*) — [arXiv]

Seoul National University (*Prof. Hyopil Shin*) — May 2023 ~ Oct 2023

- Constructed the *Korean Bio-Medical Corpus (KBMC)*, the first open-source dataset for Korean medical named entity recognition.
- Contributed primarily to fine-tuning and evaluating six encoder-only language models on KBMC, achieving more than 20% improvement in medical NER and 1.45% in general NER F1 scores.

Teaching Experience

Carnegie Mellon University — Pittsburgh, PA

Teaching Assistant (TA)

IITP Executive Education: Introductory NLP (*Prof. David Mortensen*) — Aug 2025 ~ Dec 2025

Work Experience

NCSOFT — Seongnam, South Korea

Intern, Natural Language Processing Narrative Laboratory (*Advisor: Hochang Lee*) — Mar 2021 ~ Aug 2021

- Developed a software prototype to organize clusters of news articles based on relevance to a given query by applying pairwise *Learning-To-Rank (LTR)* technique to train a Support Vector Machine (SVM)
- Outperformed the previous ruled-based news ranking system in terms of precision, recall, and f1-score

Publications

Jiseung Hong, Grace Byun, Seungone Kim, and Kai Shu. 2025. Measuring Sycophancy of Language Models in Multi-turn Dialogues. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 2239–2259, Suzhou, China. Association for Computational Linguistics.

Sungjoo Byun, **Jiseung Hong**, Sumin Park, Dongjun Jang, Jean Seo, Minseok Kim, Chaeyoung Oh, and Hyopil Shin. 2024. Korean Bio-Medical Corpus (KBMC) for Medical Named Entity Recognition. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9941–9947, Torino, Italia. ELRA and ICCL.