Jonggeun Lee

Email: jonggeun.lee@snu.ac.kr
GitHub: github.com/@jonggeun-lee

Education

Seoul National University — Seoul, Korea
M.S. in Data Science — 2024.09 - Present

- GPA: 4.08/4.3
- Advisor: Prof. Yohan Jo
- Research Interests: Multi-modal Foundation Models, Voice Assistants, Tool-augmented Agents, Post-training & Reinforcement Learning

Korea University — Seoul, Korea
B.S. in Industrial Management Engineering — 2019.03 - 2024.08

- GPA: 4.08/4.5 (Major: 4.14/4.5)
- Dean's List in College of Engineering (GPA: 4.5/4.5; 2023 Fall)
- Graduated with Great Honors (with one semester early graduation)

Publications

Don't Adapt Small Language Models for Tools; Adapt Tool Schemas to the Models
ACL 2026 Main (Acceptance rate: 19%)

- Authors: **Jonggeun Lee***, Woojung Song*, Jongwook Han, Haesung Pyun, Yohan Jo
- Proposed PA-Tool, a training-free method that adapts tool schemas to align with models' pretrained knowledge, improving tool-use performance by up to 17% and reducing schema misalignment errors by 80%.

SpeakerSleuth: Can LALMs Judge Speaker Consistency across Multi-turn Dialogues?
ACL 2026 Main (Acceptance rate: 19%)

- Authors: **Jonggeun Lee**, Junseong Pyo, Gyuhyeon Seo, Yohan Jo
- Introduced a benchmark evaluating whether Large Audio-Language Models can reliably judge speaker consistency across multi-turn conversations, revealing significant biases in prioritizing text over acoustics.

SpokenUS: A Spoken User Simulator for Task-Oriented Dialogue
Under review @ EMNLP 2026; OA 3.5; Machine Learning for Audio @ ICML 2026

- Authors: **Jonggeun Lee***, Junseong Pyo*, Jeongmin Park, Yohan Jo
- Developed a spoken user simulator that jointly generates text and speech tokens, modeling realistic spoken behaviors (cross-turn slots, barge-in, disfluency, emotion-aware speech) for task-oriented dialogue systems.

SimuHome: A Temporal-and Environment-Aware Benchmark for Smart Home LLM Agents
ICLR 2026 Oral (Acceptance rate: 1.13%)

- Authors: Gyuhyeon Seo, Jungwoo Yang, Junseong Pyo, Nalim Kim, **Jonggeun Lee**, Yohan Jo
- Developed a time-accelerated smart home simulation environment with 600 benchmark episodes, revealing that even top models struggle with temporal scheduling and state verification.

Quantifying Data Contamination in Psychometric Evaluations of LLMs
EACL 2026 Findings (Acceptance rate: 36.2%)

- Authors: Jongwook Han*, Woojung Song*, **Jonggeun Lee***, Yohan Jo
- Proposed a framework to systematically measure data contamination in psychometric evaluations of LLMs, providing evidence of strong contamination in popular inventories.

Tool-Augmented Agents: Evolution from Autonomy to Interaction
Korean Institute of Information Scientists and Engineers, Vol. 43, No. 11, pp. 14-25

- Authors: Yohan Jo, **Jonggeun Lee**
- Comprehensive survey examining the evolution of tool-augmented agents, focusing on the shift from autonomous capabilities to interactive paradigms in human-centered interaction.

Experience

Samsung Electronics Hwasung, Korea
Software Engineering Intern, ML Brain SW Development Team 2024.03 - 2024.06

- Research on Retrieval Augmented Generation (RAG) for internal chatbot system
- Fine-tuned retriever and Re-ranker using Contrastive Learning.
- Improved internal document retrieval performance from 20% to 71% Hit@1 for user queries

KAIST, Data Science and Artificial Intelligence Lab Daejeon, Korea
Undergraduate Research Intern 2023.12 - 2024.02

- Advisor: Prof. Chanyoung Park
- Conducted literature review and implementation of representative papers in recommender systems
- Explored LLM-based recommendation approaches (LLM4Rec, TALLRec)

LG AI Research, EXAONE LAB Seoul, Korea
Research Intern 2023.07 - 2023.08

- Advisor: Dr. Hyeongu Yun
- Built a multi-modal (vision + text) data extraction pipeline to curate large-scale pretraining data for the EXAONE foundation model
- Combined vision-based layout detection models with rule-based parsing to extract structured text, tables, and images from PDF documents
- Produced 73GB+ of pretraining-grade data, validating scalability toward web-scale corpora

Kounosoft Seoul, Korea
Software Engineering Intern 2023.01 - 2023.02

- Advisor: Dr. Woongmyung Kim
- Constructed Arduino-related Q&A dataset for education platform
- Performed supervised fine-tuning of KoGPT2 model on the custom dataset
- Developed complete chat interface and system using Vue3.js, FastAPI

Projects

Knowledge Graph Construction from Messenger Conversations Samsung Electronics
Industry-Academic Research Project 2024.08 - 2024.12

- Developed a system to dynamically extract user information from multi-session messenger conversations and construct knowledge graphs for hyper-personalization
- Fine-tuned Llama3-8B-Instruct model for knowledge graph extraction from messenger dialogues
- Achieved 29.4% higher F1-Score than GPT-4o on the extraction benchmark