

# Jaehee Kim

PH.D STUDENT

✉ jaehee\_kim@snu.ac.kr | 🏠 KimJaehee0725.github.io | 📧 KimJaehee0725 | 🌐 <https://www.linkedin.com/in/jaehee-kim-hoyang>

## Education

---

### Seoul National University

PH.D. CANDIDATE INDUSTRIAL ENGINEERING

Seoul Korea

Sep 2024 - Current

- Advisor: Prof. Pilsung Kang
- Data Science and Business Analytics Lab. [homepage]

### Korea University

M.S. INDUSTRIAL AND MANAGEMENT ENGINEERING

Seoul Korea

Mar 2022 - Aug 2024

- Advisor: Prof. Pilsung Kang
- Data Science and Business Analytics Lab. [homepage]

### Sungkyunkwan University

BACHELOR OF ARTS IN CONSUMER SCIENCE

Seoul Korea

Mar 2014 - Feb 2021

## Publications

---

\* First Author † Corresponding Author

Keonwoo Kim\*, **Jaehee Kim**, Dongyoon Han, Haanju Yoo†. Belief-Shift Reward: Sculpting the Flat Reward Landscape via Telescoping Information Gain, *under review*, 2026.

Jihoon Nam\*, **Jaehee Kim**, Pilsung Kang†. A Domain-Specialized Framework for Semantic Embedding and Retrieval of Semiconductor Equipment Logs, *under review*, 2026.

**Jaehee Kim**\*, Pilsung Kang†. AlienLM: Alienization of Language for API-Boundary Privacy in Black-Box LLMs. International Conference on Machine Learning (ICML), 2026.

Yukyung Lee\*, JoongHoon Kim, **Jaehee Kim**, Hyowon Cho, Jaewook Kang, Pilsung Kang†, Najoung Kim†. CheckEval: A reliable LLM-as-a-Judge framework for evaluating text generation using checklists. Association for Computational Linguistics (EMNLP), 2025.

Joonwon Jang\*, **Jaehee Kim**, Wonbin Kweon, Seonghyeon Lee, Hwanjo Yu†. Verbosity-Aware Rationale Reduction: Sentence-Level Rationale Reduction for Efficient and Effective Reasoning. Annual Meeting of the Association for Computational Linguistics (ACL), 2025.

Siyeol Sung\*, **Jaehee Kim**, Jaewon Cheon, Junyoung Son, Pilsung Kang†. FindCoResearcher: A Retrieval-based Collaborative Researcher Recommendation System. Korean Institute of Industrial Engineers, 52(1), 55-69.

**Jaehee Kim**\*, Yukyung Lee and Pilsung Kang†. A Gradient Accumulation Method for Dense Retriever under Memory Constraint. Neural Information Processing Systems (NeurIPS), 2024.

Joonghoon Kim\*, Saeran Park, Jiyeon Lee, **Jaehee Kim**, Pilsung Kang†. Development of an Unsupervised Learning-Based Fake News Generation Framework and Study the Effectiveness of Dataset Quality Evaluation Metrics. Korean Institute of Industrial Engineers, 50(1), 36-44.

Yukyung Lee\*, Joonghoon Kim\*, **Jaehee Kim**, Hyowon Cho, and Pilsung Kang†. CheckEval: Robust Evaluation Framework using Large Language Model via Checklist. CHI 2024 Workshop on Human-centered Evaluation and Auditing of Language Models.

Yookyung Kho\*, **Jaehee Kim**, Pilsung Kang†. Boosting Prompt-Based Self-Training With Mapping-Free Automatic Verbalizer for Multi-Class Classification. Findings of the Association for Computational Linguistics (EMNLP), 2023.

Yukyung Lee\*, **Jaehee Kim**, Doyoon Kim, Yookyung Kho, Younsun Kim, Pilsung Kang†. Painsight: An Extendable Opinion Mining Framework for Detecting Pain Points Based on Online Customer Reviews. ACL 2023 13th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (Oral)

Jaehyuk Heo\*, YongGi Jeong, Sunwoo Kim, **Jaehye Kim**, Pilsung Kang†. REVECA - Rich Encoder-decoder framework for Video Event Captioner. CVPR 2022 Workshop on LOnG-form VidEo Understanding Challenge.

## Research Project

---

- Development of Worker-Friendly Innovative AI Agents for Autonomous Manufacturing** *Apr 2025 - Dec 2025*  
 GRANTED BY INSTITUTE FOR INFORMATION & COMMUNICATIONS TECHNOLOGY PLANNING & EVALUATION (IITP)  
 • Developed multi-agent and multi-party engine  
 • Built agent orchestration pipeline  
 • Implemented Retrieval-Augmented Generation (RAG) system
- Data Construction and Optimization for Training/Inference of AI Chatbot** *Mar 2025 - Mar 2026*  
 GRANTED BY SAMSUNG FIRE & MARINE INSURANCE PROJECT  
 • Designed document preprocessing pipeline architecture  
 • Developed PDF parsing module  
 • Constructed training dataset and developed training methodology utilizing hierarchical-structured documents
- Developing an Information Retrieval System for Scholarly Achievement and Research Projects in the College of Engineering** *Jan 2025 - July 2025*  
 GRANTED BY COLLEGE OF ENGINEERING, SEOUL NATIONAL UNIVERSITY [HOMEPAGE]  
 • Developing Evaluation Methods for LLM in Finance Domain  
 • Constructing Training Data Based on Metadata of Each Laboratory  
 • Training Dense Retriever in the Scientific Domain
- Developing a Large Language Model Evaluation Framework for Financial Domain** *Nov 2023 - Aug 2024*  
 GRANTED BY KAKAOBANK [HOMEPAGE]  
 • Developing Evaluation Methods for LLM in Finance Domain  
 • Constructing Evaluation Datasets for Finance Domain  
 • Evaluating LLMs in Terms of Safety, Truthfulness, and Mathematical Reasoning
- Developing Customer Content via Data Analysis Techniques 2nd Stage** *Dec 2022 - Nov 2023*  
 GRANTED BY LG ELECTRONICS [HOMEPAGE]  
 • Unsupervised Dense Information Retrieval  
 • Review Clustering  
 • Text & Meta Information Visualization
- Developing Customer Content via Data Analysis Techniques 1st Stage** *Apr 2022 - Nov 2022*  
 GRANTED BY LG ELECTRONICS [HOMEPAGE]  
 • Sentence Anomaly Detection  
 • Discovering Pain Point via Clustering  
 • Clustering via Self-training

## Extracurricular Experience

---

- |                                      |  |         |
|--------------------------------------|--|---------|
| Dec 2023 - Jan 2024<br>Naver Connect | <b>Boost Camp Mentor, AI Tech</b><br>• Boot Camp for Machine Learning(NLP, CV, RecSys)<br>• Generative AI Mentor | 6       |
| Sept 2022 - Current<br>Naver Connect | <b>Boost Camp Mentor, AI Tech</b><br>• Boot Camp for Machine Learning(NLP, CV, RecSys)<br>• NLP Mentor           | 4, 5, 7 |