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 — *Seoul Korea*
PH.D. CANDIDATE INDUSTRIAL ENGINEERING — *Sep 2024 - Current*
- Advisor: Prof. Pilsung Kang
- Data Science and Business Analytics Lab. [homepage]

Korea University — *Seoul Korea*
M.S. INDUSTRIAL AND MANAGEMENT ENGINEERING — *Mar 2022 - Aug 2024*
- Advisor: Prof. Pilsung Kang
- Data Science and Business Analytics Lab. [homepage]

Sungkyunkwan University — *Seoul Korea*
BACHELOR OF ARTS IN CONSUMER SCIENCE — *Mar 2014 - Feb 2021*

A: 1st Author / Ω: Corresponding Author

Publications

Jaehee Kim$^{\mathrm{A}}$, Pilsung Kang$^{\Omega}$. AlienLM: Alienization of Language for API-Boundary Privacy in Black-Box LLMs, *under review, 2025.*

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

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

Jaehee Kim$^{\mathrm{A}}$, Yukyung Lee and Pilsung Kang$^{\Omega}$. A Gradient Accumulation Method for Dense Retriever under Memory Constraint. Neural Information Processing Systems (NeurIPS), 2024.

Joonghoon Kim$^{\mathrm{A}}$, Saeran Park, Jiyoon Lee, **Jaehee Kim**, Pilsung Kang$^{\Omega}$. 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$^{\mathrm{A}}$, Joonghoon Kim$^{\mathrm{A}}$, **Jaehee Kim**, Hyowon Cho, and Pilsung Kang$^{\Omega}$. CheckEval: Robust Evaluation Framework using Large Language Model via Checklist. CHI 2024 Workshop on Human-centered Evaluation and Auditing of Language Models.

Yookyung Kho$^{\mathrm{A}}$, **Jaehee Kim**, Pilsung Kang$^{\Omega}$. 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$^{\mathrm{A}}$, **Jaehee Kim**, Doyoon Kim, Yookyung Kho, Younsun Kim, Pilsung Kang$^{\Omega}$. Painsight: An Extendable Opinion Mining Framework for Detecting Pain Points Based on Online Customer Reviews. ACL 2023 13th Workshop on Computational Approaches to Sujectivity, Sentiment & Social Media Analysis (Oral)

Jaehyuk Heo$^{\mathrm{A}}$, YongGi Jeong, Sunwoo Kim, **Jaehee Kim**, Pilsung Kang$^{\Omega}$. 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 - Current*

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 - Current*

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 Naver Connect	**Boost Camp Mentor**, AI Tech	*6*
	• Boot Camp for Machine Learning(NLP, CV, RecSys)	
	• Generative AI Mentor	
Sept 2022 - Current Naver Connect	**Boost Camp Mentor**, AI Tech	*4, 5, 7*
	• Boot Camp for Machine Learning(NLP, CV, RecSys)	
	• NLP Mentor	

Awards

CVPR 2022 LOVEU Workshop

3RD PLACE   *2022*

- Generic Event Boundary Captioning

POSTECH Open INNOVATION Bigdata Challenge

ENCOURAGEMENT PRIZE   *2020*

- Solar Photovoltaics Prediction