Hwanhee Lee (이환희)

Assistant Professor @ Chung-Ang University

✉ hwanheelee@cau.ac.kr | 🏠 hwanheelee1993.github.io | ⌂ hwanheelee1993

Employment

Chung-Ang University *Seoul, Korea*
ASSISTANT PROFESSOR *Mar. 2023 - Current*
- Department of Artificial Intelligence/Graduate School of Artificial Intelligence (Language Intelligence Lab.)

Seoul National University *Seoul, Korea*
SENIOR RESEARCHER *Sep. 2022 - Feb. 2023*
- Postdoctoral Researcher, Automation and Systems Research Institute

NAVER CLOVA AI Research *Seongnam, Korea*
RESEARCH INTERN *Aug. 2021 - Jan. 2022*
- NAVER AI LAB Language Research Team
- Working with Dr. Hwaran Lee, Dr. Kang Min Yoo, and Prof. Joonsuk Park
- Theme: Factual Consistency Evaluation and Optimization for Abstractive Summarization
- Published a paper at NAACL 2022

Research Interests Natural Language Processing - Truthfulness and Safety of Language Models
Current Focus Safety and Policy Alignment in Language Agents, Context Engineering and Agentic Memory, Multi-Agent Systems

Education

Seoul National University *Seoul, Korea*
PH.D OF ELECTRICAL AND COMPUTER ENGINEERING *Aug. 2022*
- Machine Intelligence Laboratory (Advisor: Prof. Kyomin Jung)
- Thesis: Factual Consistency Evaluation for Conditional Text Generation Systems
  (Committee: Kyuseok Shim, Kyomin Jung, Taesup Moon, Seung-won Hwang (SNU), and JinYeong Bak (SKKU))
- Distinguished Ph.D. Dissertation Award

Seoul National University *Seoul, Korea*
BACHELOR OF ELECTRICAL AND COMPUTER ENGINEERING *Feb. 2017*

Scholarships & Awards

Outstanding Area Chair
AWARDED BY EMNLP 2024 *Nov. 2024*

Adobe Research Gift
GRANTED BY ADOBE RESEARCH ($28,000) *May. 2024, June. 2023*

Distinguished Ph.D. Dissertation Award
AWARDED BY DEPT. OF ELECTRICAL AND COMPUTER ENGINEERING, SEOUL NATIONAL UNIVERSITY *Aug. 2022*

Youlchon AI STAR Scholarship
GRANTED BY NONGSIM YOULCHON FOUNDATION *Sep. 2021*

Scholarship of Electronics, Electrical and Computer Engineering
GRANTED BY FOUNDATION FOR ELECTRONICS, ELECTRICAL AND COMPUTER ENGINEERING AT SEOUL NATIONAL UNIVERSITY *Mar. 2018*

National Science & Technology Scholarship (Full-tuition)
GRANTED BY KOREA STUDENT AID FOUNDATION *Mar. 2013 - Feb. 2017*

Publications

International Conferences

ChatInject: Abusing Chat Templates for Prompt Injection in LLM Agents

Hwan Chang*, Yonghyun Jun*, **Hwanhee Lee**†

- The Fourteenth International Conference on Learning Representations (**ICLR 2026**)

Personality Editing for Language Models through Adjusting Self-Referential Queries

Seojin Hwang, Yumin Kim, Byeongjeong Kim, **Hwanhee Lee**†

- The 19th Conference of the European Chapter of the Association for Computational Linguistics (**EACL 2026**)

Keep Security! Benchmarking Security Policy Preservation in Large Language Model Contexts Against Indirect Attacks in Question Answering

Hwan Chang*, Yumin Kim*, YongHyeon Jun, **Hwanhee Lee**†

- The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)

ToDi: Token-wise Distillation via Fine-Grained Divergence Control

Seongryong Jung, Suwan Yoon, DongGeon Kim, **Hwanhee Lee**†

- The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)

How Do Large Vision-Language Models See Text in Image? Unveiling the Distinctive Role of OCR Heads

Ingeol Baek, Hwan Chang, Sunghyun Ryu, **Hwanhee Lee**†

- The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)

SAFE-SQL: Self-Augmented In-Context Learning with Fine-grained Example Selection for Text-to-SQL

Jimin Lee, Ingeol Baek, Byeongjeong Kim, Hyunkyung Bae, **Hwanhee Lee**†

- The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)

Event-Driven Storytelling with Multiple Lifelike Humans in a 3D scene

Donggeun Lim, Jinseok Bae, Inwoo Hwang, Seungmin Lee, **Hwanhee Lee**, Young Min Kim

- The 2025 International Conference on Computer Vision (**ICCV 2025**)

Exploring Persona Sentiment Sensitivity in Personalized Dialogue Generation

Yonghyun Jun, **Hwanhee Lee**†

- The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL 2025**)

Dynamic Label Name Refinement for Few-Shot Dialogue Intent Classification

Gyutae Park, Ingeol Baek, ByeongJeong Kim, Joongbo Shin, **Hwanhee Lee**†

- The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL 2025**)

Dynamic Order Template Prediction for Generative Aspect-Based Sentiment Analysis

Yonghyun Jun, **Hwanhee Lee**†

- The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL 2025**)

Which Retain Set Matters for LLM Unlearning? A Case Study on Entity Unlearning

Hwan Chang, **Hwanhee Lee**†

- Findings of the Association for Computational Linguistics: ACL 2025 (**Findings of ACL 2025**)

Investigating Language Preference of Multilingual RAG systems

Jeonghyun Park, **Hwanhee Lee**†

- Findings of the Association for Computational Linguistics: ACL 2025 (**Findings of ACL 2025**)

AdvisorQA: Towards Helpful and Harmless Advice-seeking Question Answering with Collective Intelligence

Minbeom Kim, **Hwanhee Lee**, Joonsuk Park, Hwaran Lee, Kyomin Jung

- 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2025**)

Probing-RAG: Self-Probing to Guide Language Models in Selective Document Retrieval

Ingeol Baek, Hwan Chang, Byeongjeong Kim, Jimin Lee, **Hwanhee Lee**†

- Findings of the Association for Computational Linguistics: NAACL 2025 (**Findings of NAACL 2025**)

FIZZ: Factual Inconsistency Detection by Zoom-in Summary and Zoom-out Document

Joonho Yang, Seunghyun Yoon, Byeongjeong Kim, <u>Hwanhee Lee</u>†

- Conference on Empirical Methods in Natural Language Processing (**EMNLP 2024**)

IterCQR: Iterative Conversational Query Reformulation with Retrieval Guidance

Yunah Jang, Kang-il Lee, Hyunkyung Bae, <u>Hwanhee Lee</u>†, Kyomin Jung†

- 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2024**)

LifeTox: Unveiling Implicit Toxicity in Life Advice

Minbeom Kim, Jahyun Koo, <u>Hwanhee Lee</u>, Joonsuk Park†, Hwaran Lee, Kyomin Jung†

- 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2024**)

KoCoSa: Korean Context-aware Sarcasm Detection Dataset

Yumin Kim*, Heejae Suh*, Mingi Kim*, Dongyeon Won*, <u>Hwanhee Lee</u>†

- The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**COLING 2024**)

Kosmic: Korean Text Similarity Metric Reflecting Honorific Distinctions

Yerin Hwang, Yongil Kim, Hyunkyung Bae, Jeesoo Bang, <u>Hwanhee Lee</u>†, Kyomin Jung†

- The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**COLING 2024**)

Asking Clarification Questions to Handle Ambiguity in Open-Domain QA

Dongryeol Lee, Segwang Kim, Minwoo Lee, <u>Hwanhee Lee</u>, Joonsuk Park, Sang-Woo Lee, Kyomin Jung

- Findings of the Association for Computational Linguistics: EMNLP 2023 (**Findings of EMNLP 2023**)

Dialogizer: Context-aware Conversational-QA Dataset Generation from Textual Sources

Yerin Hwang*, Yongil Kim*, Hyunkyung Bae, <u>Hwanhee Lee</u>†, Jeesoo Bang, Kyomin Jung†

- The 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**)

Critic-Guided Decoding for Controlled Text Generation

Minbeom Kim, <u>Hwanhee Lee</u>, Kang Min Yoo, Joonsuk Park, Hwaran Lee, Kyomin Jung

- Findings of the Association for Computational Linguistics: ACL 2023 (**Findings of ACL 2023**)

Task-specific Compression for Multi-task Language Models using Attribution-based Pruning

Nakyeong Yang, Yunah Jang, <u>Hwanhee Lee</u>, Seohyeong Jung, and Kyomin Jung

- Findings of the Association for Computational Linguistics: EACL 2023 (**Findings of EACL 2023**)

Masked Summarization to Generate Factually Inconsistent Summaries for Improved Factual Consistency Checking

<u>Hwanhee Lee</u>, Kang Min Yoo, Joonsuk Park, Hwaran Lee, and Kyomin Jung

- Findings of the Association for Computational Linguistics: NAACL 2022 (**Findings of NAACL 2022**)

QACE: Asking Questions to Evaluate an Image Caption

<u>Hwanhee Lee</u>, Thomas Scialom, Seunghyun Yoon, Frank Dernoncourt, and Kyomin Jung

- Findings of the Association for Computational Linguistics: EMNLP 2021 (**Findings of EMNLP 2021**)

CrossAug: A Contrastive Data Augmentation Method for Debiasing Fact Verification Models

Minwoo Lee, Seungpil Won, Juae Kim, <u>Hwanhee Lee</u>, Cheonum Park, and Kyomin Jung

- 30th ACM International Conference on Information and Knowledge Management (**CIKM 2021**)

UMIC: An Unreferenced Metric for Image Captioning via Contrastive Learning

<u>Hwanhee Lee</u>, Seunghyun Yoon, Frank Dernoncourt, Trung Bui, and Kyomin Jung

- The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (**ACL 2021**)

Entangled Bidirectional Encoder to Autoregressive Decoder for Sequential Recommendation

Taegwan Kang, <u>Hwanhee Lee</u>, Byeongjin Choe, and Kyomin Jung

- The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2021**)

KPQA: A Metric for Generative Question Answering Using Keyphrase Weights

<u>Hwanhee Lee</u>, Seunghyun Yoon, Frank Dernoncourt, Doo Soon Kim, Trung Bui, Joongbo Shin, and Kyomin Jung

- 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2021**)

Attentive Modality Hopping Mechanism for Speech Emotion Recognition

Seunghyun Yoon, Subhadeep Dey, **Hwanhee Lee**, and Kyomin Jung

- IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2020**)

Improving Neural Question Generation Using Answer Separation

Yanghoon Kim, **Hwanhee Lee**, Joongbo Shin, and Kyomin Jung

- Thirty-Third AAAI Conference on Artificial Intelligence (**AAAI 2019**)

International Journal

MultiLexNorm++: Lexical Normalization Beyond Indo-European Languages

Weerayut Buaphet, Thanh-Nhi Nguyen, Risa Kondo, Tomoyuki Kajiwara, Yumin Kim, Jimin Lee, **Hwanhee Lee**, Holy Lovenia, Peerat Limkonchotiwat, Sarana Nutanong, Rob van der Goot

- ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)

Crafting the Path: Robust Query Rewriting for Information Retrieval

Ingeol Baek, Jimin Lee, Joonho Yang, **Hwanhee Lee**†

- IEEE Access, 2025

Cross-Lingual Summarization for Low-Resource Languages using Multilingual Retrieval-Based In-Context Learning

Gyetae Park, Jeonghyun Park, **Hwanhee Lee**†

- Applied Sciences, 2025

REMDoC: Reference-free Evaluation for Medical Document Summaries via Contrastive Learning

Jimin Lee, Ingeol Baek, **Hwanhee Lee**†

- IEEE Access, 2024

Improving Abstractive Dialogue Summarization Using Keyword Extraction

Chongjae Yoo and **Hwanhee Lee**†

- Applied Sciences, 2023

Gated Relational Target-aware Encoder and Local Context-aware Decoder for Target-oriented Opinion Words Extraction

Taegwan Kang, Segwang Kim, Hyeongu Yun, **Hwanhee Lee**, and Kyomin Jung

- IEEE Access, 2022

Preprint/Under Review

Doc-PP: Document Policy Preservation Benchmark for Large Vision-Language Models

Haeun Jang*, Hwan Chang*, **Hwanhee Lee**†

- arXiv (2601.03926)

Fame Fades, Nature Remains: Disentangling the Character Identity of Role-Playing Agents

Yonghyun Jun*, Hwan Chang*, Jihyeong Park, **Hwanhee Lee**†

- arXiv (2601.04176)

Enhancing Multilingual RAG Systems with Debiased Language Preference-Guided Query Fusion

Jeonghyun Park, Byeongjeong Kim, Seojin Hwang, **Hwanhee Lee**†

- arXiv (2601.02956)

MARCH: Evaluating the Intersection of Ambiguity Interpretation and Multi-hop Inference

Jeonghyun Park*, Ingeol Baek*, Seunghyun Yoon, Haeun Jang, Aparna Garimella, Akriti Jain, Nedim Lipka, **Hwanhee Lee**†

- arXiv (2509.22750)

Chronological Passage Assembling in RAG framework for Narrative Text

Byeongjeong Kim, Jeonghyun Park, Joonho Yang, **Hwanhee Lee**†

- arXiv (2508.18748)

Hallucinate at the Last in Long Response Generation: A Case Study on Long Document Summarization

Joonho Yang, Seunghyun Yoon, Hwan Chang, ByeongJeong Kim, **Hwanhee Lee**†

- arXiv (2505.15291)

Selective Demonstration Retrieval for Improved Implicit Hate Speech Detection

Yumin Kim, **Hwanhee Lee**†

- arXiv (2504.12082)

Leveraging Fine-grained Error Correction in Korean Speech Recognition for Consultation Services

Yonghyun Jun, Jimin Lee, Hwan Chang, Dongho Shin, Seolah Kim, **Hwanhee Lee**†

- Under Review

Conversational Query Reformulation with the Guidance of Retrieved Documents

Jeonghyun Park, **Hwanhee Lee**†

- Under Review - arXiv (2407.12363)

Workshop Papers & Other International Conferences

Logistic Regression-Based Example Selection for Enhanced Few-Shot Learning in Intent Classification

Gyutae Park, **Hwanhee Lee**†

- IEEE International Conference on Consumer Electronics (**IEEE ICCE 2025**)

Dynamically Managing a Prompt Pool via Self-Enhancement in Continual Learning

Hayun Lee, Kiseong Hong, **Hwanhee Lee**, Sungho Suh, Eunwoo Kim

- NeurIPS 2024 Workshop on Adaptive Foundation Models (**NeurIPS-W 2024**)

Low-Resource Cross-Lingual Summarization through Few-Shot Learning with Large Language Models

Gyutae Park, Seojin Hwang, **Hwanhee Lee**†

- ACL Workshop on Technologies for Machine Translation of Low-Resource Languages (**ACL-W 2024**)

Factual Error Correction for Abstractive Summaries Using Entity Retrieval

Hwanhee Lee, Cheoneum Park, Seunghyun Yoon, Trung Bui, Franck Dernoncourt, Juae Kim and Kyomin Jung

- EMNLP Workshop on Generation, Evaluation & Metrics (**EMNLP-W 2022**)

Improving Multiple Documents Grounded Goal-Oriented Dialog Systems via Diverse Knowledge Enhanced Pre-trained Language Model

Yunah Jang, Dongryeol Lee, Hyung Joo Park, Taegwan Kang, **Hwanhee Lee**, Hyunkyung Bae, and Kyomin Jung

- ACL Workshop on Document-grounded Dialogue and Conversational Question Answering (**ACL-W 2022**)

Learning to Select Question-Relevant Relations for Visual Question Answering

Jaewoong Lee, Heejoon Lee, **Hwanhee Lee**, and Kyomin Jung

- NAACL Workshop on Multimodal Artificial Intelligence Workshop (**NAACL-W 2021**)

Relevance Similarity Scorer and Entity Guided Reranking for Knowledge Grounded Dialog System

Hyunkyung Bae, Minwoo Lee, Ahhyeon Kim, and **Hwanhee Lee**, Cheongjae Lee, Cheonum Park, Donghyeon Kim, and Kyomin Jung

- AAAI Workshop on Dialog System Technology Challenge (**AAAI-W 2021**)

ViLBERTScore: Evaluating Image Caption Using Vision-and-Language BERT

Hwanhee Lee, Seunghyun Yoon, Frank Dernoncourt, Doo Soon Kim, Trung Bui, and Kyomin Jung

- EMNLP Workshop on Evaluation and Comparison of NLP Systems (**EMNLP-W 2020**)

Multimodal Semantic Transformer Network with Retrieval Style Word Generator

Hwanhee Lee, Seunghyun Yoon, Frank Dernoncourt, Doo Soon Kim, Trung Bui, and Kyomin Jung

- AAAI Workshop on Dialog System Technology Challenge (**AAAI-W 2020**)

Attention-based Convolutional Neural Networks for Multi-label Emotion Classification

Yanghoon Kim, **Hwanhee Lee**, and Kyomin Jung

- NAACL Workshop on Semantic Evaluation 2018 (**NAACL-W 2018**)

Research Projects

Plug-in External-Memory Models for Augmenting Knowledge of Large Language Models

Funded by the National Research Foundation of Korea, PI *Sep. 2025 - Aug. 2026*

- Research on Next-generation RAG via Plug-in External Memory

Next-Generation O2O Fashion Retail System: Enhancing Sales Through Integration of Personal Shoppers and Immersive Virtual Fitting

Funded by the Ministry of Trade, Industry and Energy, Co-PI *July. 2025 - Dec. 2028*

- Developing Conversational Fashion AI Agent

Development of an LLM System for Financial Transactions

Funded by Seoul Business Agency, Co-PI *Sep. 2025 - Aug. 2026*

- Developing Transaction Classification System with RAG

LLM based Claim Report Generation

Funded by Doosan Enerbility, PI *Jun. 2025 - May. 2026*

- Claim Report Generation with Agentic RAG System

Development of a LLM System for Financial Transaction Categorization to Enhance Risk and Customer Management in Financial Institutions

Funded by the Ministry of SMEs and Startups, Co-PI *May. 2025 - Dec. 2025*

- Developing Transaction Classification System with RAG

Risk Assessment of Construction Site based on LLMs

Funded by Doosan Enerbility, PI *Jun. 2024 - May. 2025*

- Risk Assessment and Prediction with Large Language Models

Improvement of Speech Recognition Rate Using LLMs

Funded by Seoul 120 Dasan Call Foundation, PI *April. 2024 - Dec. 2024*

- Detecting Errors of ASR System for Dialogue Data with Large Language Models

RAG based Conversational Question Answering System

Funded by URP, Co-PI *Mar. 2024 - Dec. 2024*

- Developing ConvQA System for Real Estate Domain

Anomaly Detection in Fuel Cell

Funded by Doosan Fuel Cell, PI *June. 2023 - May. 2024*

- Developing Anomaly Detection System for Full Cell

Factual Error Correction on Grounded Text Generation Systems

Funded by Adobe Research (Research Gifts, 2023-2024) *May. 2023 - Current*

- Developed Factual Error Detection and Correction System for Abstractive Summarization Systems
- Published a Paper at EMNLP 2024

Academic Services

Organizing Committee

IJCNLP-AACL 2023 (Diversity & Inclusion Chair)

Area Chair

ACL 2024-2026, EMNLP 2024-2025, NAACL 2024-2025, COLING 2025

Reviewer

ACL 2021-2023, EMNLP 2020-2023, COLING 2022-2024, AAAI 2021-2026, NeurIPS 2023-2025, ICLR 2026, ICML 2026

Advisor

Software Policy & Research Institute   *Aug. 2022 - Mar. 2023*

- Technical Advising on Hyperscale Language Models

Selected Invited Talks

Sungkyunkwan University   *Suwon, Korea*

Building Secure and Faithful LLM Agents   *Jan. 2026*

Ulsan National Institute of Science and Technology   *Ulsan, Korea*

Building Faithful and Secure Language Models   *Oct. 2025*

Korea Computer Congress 2025   *Jeju, Korea*

Faithfulness in Language Models   *July. 2025*

Korea Advanced Institute of Science & Technology   *Daejeon, Korea*

Cognitive and Reasoning Capabilities of Hyperscale AI (NextGenAI-X Workshop)   *Mar. 2025*

Adobe Research   *San Jose, CA, USA*

Toward Truthful and Safe Language Models   *Jan. 2025*

The 36th Annual Conference on Human & Cognitive Language Technology   *Seongnam, Korea*

Toward Truthful and Safe Language Models   *Oct. 2024*

Hanyang University   *Seoul, Korea*

Hallucination of Large Language Models   *March. 2024*

Korea Institute of Science and Technology Information   *Daejeon, Korea*

Recent Research on Factual Consistency and Applications to the Scientific Domain   *January. 2024*

Korea Advanced Institute of Science & Technology   *Daejeon, Korea*

Research on Text Generation Systems   *Oct. 2023*

Korea Electronics Technology Institute   *Seongnam, Korea*

Large Language Models ChatGPT: How it Works and How to Use   *July. 2023*

Eötvös Loránd University   *Budapest, Hungary*

Research on Text Generation Systems   *July. 2023*

Chung-Ang University   *Seoul, Korea*

Large Language Models ChatGPT: How it Works and How to Use   *April. 2023*

LG AI Research   *Seoul, Korea*

Factual Consistency of Text Generation Systems   *Jan. 2023*

BrainLink 2022   *Incheon, Korea*

Factual Consistency of Text Generation Systems   *Oct. 2022*

Korea Electronics Technology Institute   *Seongnam, Korea*

Factual Consistency of Text Generation Systems   *Aug. 2022*

Teaching Experiences

Natural Language Processing   *Chung-Ang University*

Lecturer   *Spring 2026*

Advanced Natural Language Processing: Large Language Models   *Chung-Ang University*

Lecturer   *Spring 2023-2026*

Objected Oriented Programming   *Chung-Ang University*

Lecturer   *Spring 2024-2025*

Embedded Artificial Intelligence   *Chung-Ang University*

Lecturer   *Fall 2023-2025*

Artificial Intelligence Design   *Chung-Ang University*

Lecturer   *Fall 2023-2025*

Probability and Random Process   *Chung-Ang University*

Lecturer   *Spring 2023*