

Haeri Kim

[✉ haeri.kim1108@gmail.com](mailto:haeri.kim1108@gmail.com) | [🌐 haeri-kim1108.github.io](https://haeri-kim1108.github.io) | [in Haeri Kim](#) | [📱 +82 10-4820-9080](tel:+821048209080)

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST)

2019 - 2021

M.S. in Electrical Engineering

Thesis: Diagnosing Data Samples for Robust GAN Training

Advisor: Prof. Hye Won Chung

Korea Advanced Institute of Science and Technology (KAIST)

2014 - 2019

B.S. in Electrical Engineering

WORK EXPERIENCE

Samsung Research

Jul 2021 - present

- AI Solution Team (AI Research Engineer)
- Research topics: On-device Automatic Speech Recognition (ASR), Streaming ASR, Context-aware ASR, Multilingual ASR

LG CNS

Jun 2017 - Jul 2017

- New Technology Development Team (Internship)
- Research topics: Object tracking

PUBLICATIONS

*: Equal contribution

Conference Papers

A More Accurate Internal Language Model Score Estimation for the Hybrid Autoregressive Transducer

Kyungmin Lee*, Haeri Kim*, Sichen Jin, Jinhwan Park, Youngho Han

Interspeech, 2023

Self-Diagnosing GAN: Diagnosing Underrepresented Samples in Generative Adversarial Networks

Jinhee Lee*, Haeri Kim*, Youngkyu Hong*, Hye Won Chung

Conference on Neural Information Processing Systems (NeurIPS), 2021

Under Review

BBPE16: UTF-16-Based Byte-Level Byte-Pair Encoding for Improved Multilingual Speech Recognition

Hyunsik Kim*, Haeri Kim*, Munhak Lee, Kyungmin Lee

Submitted to *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2026

FUNDING & FELLOWSHIPS

Samsung Electronics Overseas PhD Fellowship

Jul 2026 - Jun 2031

Full tuition and living stipend for 5 years; awarded through a competitive internal selection process

PROJECTS

Speech Large Language Model (SpeechLLM)	May 2025 - present
• Research and development of methods to integrate a speech encoder with a large language model (LLM) for end-to-end training and multi-modal understanding	
Multilingual ASR Model	Jan 2024 - present
• Research and development of a multilingual ASR model that unifies multiple monolingual ASR systems into a single model	
• Reduce total model footprint by replacing 23+ language-specific ASR models with a unified model	
Galaxy S24 Speech Translation System	Jan 2023 - Dec 2023
• Development of streaming on-device ASR models for Galaxy S24 speech translation system	
• Training and optimization of streaming ASR models for multiple languages	
• Related article	
Samsung Bespoke AI	Mar 2022 - Dec 2022
• Development of on-device ASR models for home appliances	
• Model adaptation for home appliance domains (washing machines, air conditioners, refrigerators)	
• Related article	
On-Device ASR Model for Smart TV	Mar 2022 - Dec 2022
• Development of on-device ASR models for Smart TV	
• Research and development of a language model (LM) fusion algorithm to improve recognition of TV program names and other domain-specific entities	
• Development of a spell correction algorithm to enhance named-entity recognition accuracy	
Data Valuation for Generative Models	Mar 2020 - Jun 2021
• Research on devising data valuation measures for generative models	
• Focused on Generative Adversarial Network (GAN) and utilized training dynamics to diagnose training data	
Human Face Tracking Using Multi-Domain Network	Jun 2017 - Jul 2017
• Researched state-of-the-art object tracking technology and implemented it for human face tracking	

PATENT

Multilingual Speech Recognition Method and Device

Kyungmin Lee, Haeri Kim, Jinhwan Park, Hyeongmin Oh

PCT/KR2024/016938, 2024

Electronic device for performing speech recognition and control method thereof

Kyungmin Lee, Jaeyoung Roh, Jonguk Yoo, Youngho Han, Haeri Kim

PCT/KR2023/017891, 2023

Self-Diagnosing Generative Adversarial Networks to Diagnose and Emphasize Underrepresented Samples

Hye Won Chung, Jinhee Lee, Haeri Kim, Youngkyu Hong

KR Patent 10-2022-0032061, 2022

TEACHING & TUTORING

Teaching Assistant

- Information Theory (EE623) Sep 2020 - Dec 2020
- Probability and Introductory Random Processes (EE210) Mar 2020 - Jun 2020
- Signals and Systems (EE202) Sep 2019 - Dec 2019

Tutor

- Programming Structure for Electrical Engineering (EE209) Sep 2019 - Feb 2020
- Introduction to Programming (CS101) Mar 2018 - Jun 2018

SKILLS

Programming Languages Python, C, C++, MATLAB
Libraries PyTorch, TensorFlow, NumPy, pandas, scikit-learn, Matplotlib