

Seongmin Park

seongminpark.com | github.com/seongminp | llov0708@gmail.com | Google Scholar

EDUCATION

- **University of Pittsburgh** Pittsburgh, PA, USA
M.S. in Information Science Aug. 2018 – Dec. 2019
- **Carnegie Mellon University** Pittsburgh, PA, USA
B.S. in Electrical and Computer Engineering Aug. 2013 – Dec. 2017

WORK EXPERIENCE

- **Google** Seoul, South Korea
Forward Deployed Engineer, GenAI, Google Cloud Starting Jul. 2026
 - Joining the Google Cloud generative-AI team to design, build, and deploy GenAI solutions with enterprise customers.
- **Goodnotes** Remote
Machine Learning Engineer Jun. 2025 – Jul. 2026
 - Built and shipped LLM-powered meeting-assistant features (live summaries, agenda notes, and context-aware chat) in Goodnotes, the #1 global note-taking app with tens of millions of users.
 - Deployed on-device speech-to-text and language models across iOS, macOS, Windows, Android, and Web using ONNX, OpenVINO, and QNN for fully on-device production inference.
 - Fine-tuned and prompt-tuned models for real-time speech summarization, improving output quality and responsiveness for live meetings.
 - Built and scaled ML evaluation pipelines that raised accuracy and lowered latency, enabling faster and more reliable releases.
 - Partnered across product, design, QA, and engineering to take research prototypes to production at scale.
- **ActionPower** Seoul, South Korea
Language Team Lead, Machine Learning Researcher Jun. 2020 – Jun. 2025
 - Led a three-person team building the language-AI stack behind “Daglo,” a speech-to-text service; joined from post-seed through post-Series B as monthly active users grew 125×.
 - Shipped transcript summarization, retrieval-augmented generation (RAG), Model Context Protocol, instruction and alignment tuning, keyword extraction, error correction, text normalization, and knowledge distillation in Python with large language models.
 - Partnered with Seoul National University Hospital to develop a model for suicide-risk assessment.
 - Delivered pre-sales PoCs for 20+ B2B and B2G engagements (React, FastAPI, Gradio, Streamlit) and contributed to Daglo’s web frontend.
- **OkeyDokey** Seoul, South Korea
Smart Contract Developer May 2017 – Aug. 2018
 - Built open-source Ethereum smart contracts in Solidity to control IoT devices, with Android (Java) and iOS (Swift) clients; ported the go-ethereum (Go) client to run on the Android NDK.
 - Company acquired by Colong E&C after tenure.
- **SK Techx** Seoul, South Korea
Mobile Developer Intern, Volo Team May 2016 – Aug. 2016
 - Maintained an iOS (Objective-C) travel-journal social app with ~80,000 users and helped port it to Android.

TECHNICAL SKILLS

Languages: Python, JavaScript, Solidity, Java, Swift, Objective-C, Go, Zig

ML & NLP: LLMs; instruction & alignment tuning (DPO, RLHF); RAG; summarization; knowledge distillation; speech-to-text / ASR; on-device & edge inference (ONNX, OpenVINO, QNN)

Web & Tools: React, FastAPI, Gradio, Streamlit

HONORS & AWARDS

- **1st place**, Korean AI Competition 2023 (1st of 233 teams); Minister of Culture, Sports and Tourism Award.
- **1st place**, Korean AI Competition 2024 (1st of 280 teams); Minister of Culture, Sports and Tourism Award.
- Competition focus: LLM instruction tuning, DPO, RLHF, ensembling, summarization, and classification.

PUBLICATIONS

- [JMIR Med Inform 2025] Automated Speech Analysis for Screening and Monitoring Bipolar Depression: Machine Learning Model Development and Interpretation Study.
- [ICASSP 2025] Prompt Crossing: Evaluating Whether LLM Response Stems from Jailbreak or Normal Prompt.
- [COLING 2024] RT-VQ2A2: Real Time Vector Quantized Question Answering with ASR.
- [COLING 2024] Non-Essential is NEcessary: Order-agnostic Multi-hop Question Generation.
- [ICASSP 2024] Unsupervised Extractive Dialogue Summarization in Hyperdimensional Space.
- [Interspeech 2023] Unsupervised Dialogue Topic Segmentation in Hyperdimensional Space.
- [ACL Findings 2023] Cross-task Knowledge Transfer for Extremely Weakly Supervised Text Classification.
- [COLING 2022] LIME: Weakly Supervised Text Classification Without Seeds.
- [COLING 2022 Workshop] Leveraging Non-dialogue Summaries for Dialogue Summarization (Transcript Understanding Workshop).
- [ACL 2022 Workshop] Unsupervised Abstractive Dialogue Summarization with Word Graphs and POV Conversion.
- [EMNLP 2021 Workshop] Finetuning Pretrained Transformers into Variational Autoencoders.
- [Preprint 2021] Improving Distinction between ASR Errors and Speech Disfluencies with Feature Space Interpolation.
- [TPRC 2019] Technology Adoption in Spectrum Sharing: Estimating the Impact on Incumbents in the 3.5 GHz Band.

PATENTS

Inventor on **16 patents and applications** across the United States (6), South Korea (9), and Japan (1). Selected U.S. grants:

- **US 11,749,260** (speech recognition using grapheme information).
- **US 11,887,002** (data generation via an encoder–decoder neural-network model).
- **US 11,721,325** (generating training data for speech-recognition error detection).
- **US 12,197,862** (identifying the word corresponding to a target word in text).
- **US 11,640,493** (dialogue summarization with word graphs).

PROJECTS

- **Hama NLP** (open-source Zig, Python, and JavaScript library for client-side, sub-megabyte natural language processing): grapheme-to-phoneme conversion, part-of-speech tagging, and on-device automatic speech recognition.

ENGLISH PROFICIENCY

PTE Academic (Sep. 2025): **90 / 90** (Listening 90, Reading 90, Speaking 90, Writing 90).