



# Soohwan Kim

TUNiB CO-FOUNDER · AI ENGINEER

4F, 20 Pangyo-ro 289beon-gil, Seongnam District, Republic of Korea

☎ (+82) 10-4564-4668 | ✉ sh951011@gmail.com | 🏠 sooftware.io | 📱 sooftware | 🌐 Soo-hwan

“AI / Startup / Open-source / Competition”

## Summary

I am an A.I. research engineer with several years of experience in the fields of speech and natural language processing. Prior to this, I worked at Kakao Brain. Currently, I am one of the co-founders of a tech startup called TUNiB, alongside my former teammates from Kakao Brain.

Our goal is to develop advanced A.I. technology, striving to create chatbots with conversational abilities akin to Samantha from the movie “Her.” While our current chatbot service, Dearmate, may not yet reach Samantha’s level of conversation, we believe that in the near future, we will be able to achieve such a milestone.

## Professional Experience

### TUNiB

CO-FOUNDER/AI LEAD

Pangyo, Gyunggi

Mar. 2021 - Present

- Co-founded with former Kakao Brain team members.
- Raised approximately 3 billion won in seed investment from Pearl Abyss Capital, DSC Investment, and Naver D2. (2021.10)
- [NLP] Lead development the Dearmate project: <https://demo.dearmate.ai>
- [NLP] Development of own sLLM named Cheetah
- [NLP] Responsible for modeling Blenderbot dataset projects (national project)
- [NLP] Responsible for modeling KommonGen dataset projects (national project)
- [NLP] Development of the TUNiB demo NLP technology: <https://tunibridge.ai>
- [NLP] Lead the Safety Check project: <https://tunibridge.ai/safety>
- [NLP] Developing an open-domain dialogue system.
- [NLP] Development of the framework for few-shot evaluation of autoregressive language models.
- [NLP] Led the text data collection (1.2TB) (used in the Polyglot project)
- [NLP] TUNiB-ELECTRA: <https://github.com/tunib-ai/tunib-electra>
- [NLP] NLP-Tasks: natural language processing tasks and examples
- [NLP] Implementation of Luna-Transformer: Linear Unified Nested Attention
- [Competition] 3rd place in 2022 National AI Grand Challenge - NLP (2 billion won in subsidies for commercialization)
- [Competition] 1st place in 2022 National AI Online Competition - NLP (200 million won in subsidies for commercialization)
- [Competition] 3rd place in 2021 National AI Grand Challenge - Speech Understanding
- [Competition] 1st place in 2021 National AI Online Competition - NLP (300 million won in subsidies for commercialization)

### Kakao Brain

AI ENGINEER (FULL-TIME)

Pangyo, Gyunggi

Nov. 2020 - Feb. 2021

- Team Leader: Kyubyong Park
- Speech and natural language processing team.
- [NLP] PORORO: Platform Of neuRal mOdelS for natuRal language prOcessing: <https://github.com/kakaobrain/pororo>
- [SPEECH] Implementation of Conformer: Convolution-augmented Transformer for Speech Recognition
- [SPEECH] Implementation of Speech Transformer
- [SPEECH] Implementation of the “Listen, Attend and Spell” model based on the Fairseq framework.

## Internship

### Kakao Brain

AI ENGINEER

Mapo, Seoul

Aug. 2020 - Oct. 2020

- Team Leader: Kyubyong Park
- Speech and natural language processing team.
- [SPEECH] Multilingual Text-To-Speech
- [SPEECH] English, Chinese, and Korean Speech Recognition using Wav2vec 2
- [SPEECH] Korean Wav2vec 2 pre-training

### Spoken Language Lab (Sogang Univ.)

UNDERGRADUATE RESEARCHER

Mapo, Seoul

Mar. 2020 - Aug. 2020

- Department of Computer Science & Engineering Sogang University
- Advisor: Prof. Ji-Hwan Kim
- Homepage: <https://speech.sogang.ac.kr>

## Service

### Dearmate

AI LEAD

- AI chatbot platform with a variety of characters with unique personas.
- Running the service with our own sLLM(named Cheetah).
- Quickly adding new characters and improving dialogue performance.
- [Web Demo] [Play Store] [App Store]

TUNiB

2022 - Present

### TUNiBridge

AI LEAD

- NLP Cloud API Services
- API Services: Text Safety Check, Emotion Prediction, Dialect Translation etc.
- Portfolio: NIA, wrtn, Nexon, Krafton, UpStage, Mirae Asset etc.
- Link: <https://tunibridge.ai>

TUNiB

2021 - Present

## OpenSource Project

### Pororo

KAKAO BRAIN

- Library that can solve more than 30 natural language processing tasks in just 3-4 lines with pre-trained models.
- Developed speech related modules. (Automatic Speech Recognition, Text To Speech)
- Link: <https://github.com/kakaobrain/pororo>

1.2k+ Stars

Aug 2020 - Feb 2021

### OpenSpeech / KoSpeech

PERSONAL

- Framework that enables easy creation of speech recognition models of various models and various languages.
- Dozens of models including Transformers, Conformer, Transformers, and Listen Attend Spell.
- Link(OpenSpeech): <https://github.com/openspeech-team/openspeech>
- Link(KoSpeech): <https://github.com/sooftware/kospeech>

500+ Stars

Jan 2020 - Jun 2021

### TUNiB Electra

TUNiB

- Released two version of ELECTRA models. (Korean-English bilingual model, Korean-only model).
- Conducted performance comparison experiments between several tokenizers made using the Tokenizers library.
- Link: <https://github.com/tunib-ai/tunib-electra>

100+ Stars

Aug 2021 - Sep 2021

### Conformer

PERSONAL

- PyTorch implementation of "Conformer: Convolution-augmented Transformer for Speech Recognition" (INTERSPEECH 2020)
- Link: <https://github.com/sooftware/conformer>

700+ Stars

Jan 2021 - Jan 2021

## Education

### Kwangwoon University

BACHELOR OF ENGINEERING IN ELECTRONICS AND COMMUNICATION

- **Major:** Electronics & Communication Engineering, **Minor:** Data Science
- **Major GPA :** 3.90 / 4.5, **Total GPA :** 3.67 / 4.5

Seoul, South Korea

Mar. 2014 - Feb. 2021

#### • Relevant Coursework

- |                             |                   |                         |                       |
|-----------------------------|-------------------|-------------------------|-----------------------|
| • Machine Learning          | • Database        | • Operating System      | • Computer Network    |
| • AI & Speech Processing    | • Data Structures | • Computer Architecture | • Data Communication  |
| • Digital Signal Processing | • Data Mining     | • System Programming    | • Network Programming |

#### • Awards

- 2020.11 **President's Award**, Multi-Y Capstone Design Competition
- 2020.11 **People's Choice Award**, Multi-Y Capstone Design Competition
- 2019.10 **People's Choice Award**, Software start-up idea contest - Fall
- 2019.04 **People's Choice Award**, Software start-up idea contest - Spring
- 2018.08 **Outstanding Award**, Samsung Multi-Campus

#### • Honors

- 2020.03 **Honors Scholarship**, Samsung Scholarship
- 2019.03 **Honors Scholarship**, Samsung Scholarship
- 2019.03 **Academic Scholarship**, Kwangwoon University
- 2018.09 **Academic Scholarship**, Kwangwoon University
- 2018.03 **Academic Scholarship**, Kwangwoon University

## Military Service

### Republic of Korea Marine Corps

SERGEANT

- Served and discharged from Republic of Korea Marine Corps as Signaller.

*The 2nd Marine Division*

Sep. 2015 - Jun. 2017

## Talks

### The development process of the TUNiB Electra model

2021 SSDC (SAMSUNG SOFTWARE DEVELOPMENT CONFERENCE)

- Presentation Session: [YouTube](#)

*Seoul, S.Korea*

Nov. 2021

### Korean Speech Recognition

2021 LANGCON

- Presentation Video: [YouTube](#)

*Seoul, S.Korea*

Aug. 2021

### Know-how to win the 2021 AI online competition

2021 AI ONLINE COMPETITION, PANGYO AI CAMP

- Presentation Video: [YouTube](#)

*Seoul, S.Korea*

Aug. 2021

## Writing

### Technical Blog

QUICK PEEK

- Beam search: <https://sooftware.io/beamsearch>
- Attention Mechanism: <https://sooftware.io/attention>
- Seq2seq: <https://sooftware.io/seq2seq>

*sooftware.io*

Aug. 2018 - Present

## Skills

**Language** Python, C, Java, Shell Script, SQL

**Deep Learning** PyTorch, PyTorch-Lightning, Transformers, Fairseq, DeepSpeed

**MLOps** Triton Inference Server, Faster Transformer, Docker, FastAPI

## Publication

[1] Chaewon Park, [Soohwan Kim](#), Kyubyong Park, and Kunwoo Park, "A Korean Hate Speech Dataset with Toxicity Scales: Cognitive Reflection Test for Improving Model Fairness", pre-print, 2023.

[2] [Soohwan Kim](#), Seyoung Bae, and Cheolhwang Won, "Open-Source Toolkit for End-to-End Korean Speech Recognition", ELSEVIER, SIMPAC (Software Impacts), Vol.7, 100054, 2021.

[3] [Soohwan Kim](#), Seyoung Bae, Cheolhwang Won, and Suwon Park, "KoSpeech: Open-Source Toolkit for End-to-End Korean Speech Recognition", arXiv, 2009.03092, 2020.

## Competitions

2022 **3rd Ranked**, National AI Grand Challenge - Policy-assisted AI

*IITP*

2022 **1st Ranked**, National AI Online Competition - Machine Reading Comprehension Track

*MSIT*

2021 **3rd Ranked**, National AI Grand Challenge - Speech Understanding Track

*IITP*

2021 **1st Ranked**, National AI Online Competition - Conversation Sentiment Classification Track

*MSIT*

2020 **3rd Ranked**, National AI Grand Challenge - Speech Understanding Track

*IITP*

2019 **Finalist (12th Ranked)**, Naver AI Hackathon - Speech

*Naver*

## Honors & Awards

2022 **Minister's Award - Ministry of Science and ICT (MSIT)**, 2022 National AI Grand Challenge

*MSIT*

2022 **Minister's Award - Ministry of Science and ICT (MSIT)**, 2022 National AI Online Competition

*MSIT*

2021 **President's Award - IITP**, 2021 National AI Grand Challenge

*IITP*

2021 **President's Award - NIPA**, 2021 National AI Online Competition

*NIPA*