SOOHWAN KIM

18F, Seocho-daero 74 The Asset Tower, Seocho-District, Seoul, South Korea

📞 (+82)10-4564-4668 ✉ sh951011@gmail.com 🔗 linkedin.com/in/Soo-hwan/ 🔗 github.com/sooftware

OBJECTIVE

Ambitious A.I. research engineer with 2+ years experience in speech and natural language processing fields. Currently, co-founder of a tech startup named TUNiB, research interests includes speech and natural language processing, and MLOps. In particular, recently been very interested in large-scale language model training and big-data processing.

EDUCATION

Kwangwoon University   Mar. 2014 – Feb 2021
Bachelor of Engineering in Electronics and Communication   *Seoul, South Korea*

RESEARCH EXPERIENCE

Spoken Language Lab (Sogang Univ.)   Mar 2020 – Aug 2020
Undergraduate Researcher   *Mapo, Seoul*
- Advisor: Prof. Ji-Hwan Kim
- Research and develop latest end-to-end speech recognition technology.

PROFESSIONAL EXPERIENCE

TUNiB Inc.   Mar 2021 – Present
Co-founder/A.I. Engineer   *Gangnam, Seoul*
- Co-founded with former Kakao Brain team members.
- Developing large-scale language model and open-domain dialogue system.
- Lead the development of the TUNiB Electra model, data construction, modeling, and experimental design.
- 1st ranked in the 2021 AI Online Competition, attracting 300 million won worth of subsidies.

Kakao Brain Corp.   Aug 2020 – Feb 2021
A.I. Engineer   *Pangyo, Gyunggi*
- Team leader: Kyubyong Park
- Research and develop latest speech and natural language processing technologies.
- Developed a open-source natural language processing library named pororo
- Developed multilingual TTS technology. Demo page: https://pororo-tts.github.io/
- Developed English, Chinese, and Korean speech recognition using Wav2vec 2.0 technology.

PROJECT EXPERIENCE

Open-domain Dialogue System | *NLP, DeepSpeed, Transformers, Megatron-LM*   Apr 2021 - Present
- Research and develop various versions of chatbots.
- Participate in the development of all overall systems from data construction to big model training.

TUNiB Electra | *NLP, Transformers, Tokenizers, TPU, Google Cloud Platform*   Aug 2021 - Sep 2021
- 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.

Pororo | *NLP, ASR, TTS, Fairseq, PyTorch*   Aug 2020 - Feb 2021
- 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)

OpenSpeech/KoSpeech | *ASR, PyTorch-Lightning, Hydra*   Jan 2020 - Jun 2021
- Framework that enables easy creation of speech recognition models of various models and various languages.
- Dozens of models including Transformers, ContextNets, Transformers, and Listen Attend Spell.

Awards & Honors

- 2021.08 NIPA President's Award (1st ranked), 2021 AI Online Competition
- 2020.11 President's Award (1st ranked) & People's Choice Award, Kwangwoon Engineering Festival
- 2020.10 3rd ranked, 2020 AI Grand Challenge - Speech Recognition Track
- 2019.10 Finalist (12th Ranked), Naver AI Hackathon - Speech
- 2019.04 People's Choice Award, Kwangwoon University Software Start-up Idea Contest
- 2018.08 Outstanding Award, Samsung Multi-Campus