SANGHYUK CHOI

+82 1045708606

sanghyuk-choi@hotmail.com

RESEARCH INTERESTS

Natural Language Processing
LLM(Instruction fine-tuning, Alignment tuning), Natural Language Generation(Data-to-text, Summarization)

Machine Learning
Deep Neural Networks

WORK EXPERIENCE

NCSOFT	*Seoul, Korea*
Application LM Team Lead	*Sep 2023 - now*
Generation Model Team Lead	*Jun 2022 - Jan 2024*
Research Engineer	*Jan 2018 - now*
Stayes Inc.	*Seoul, Korea*
CTO, Co-founder	*Feb 2014 - Feb 2015*
Visionaries Inc.	*Seoul, Korea*
CTO, Co-founder	*Mar 2013 - Feb 2014*

RESEARCH EXPERIENCE

VARCO LLM 2.0: NCSOFT's Large Language Model   *2024 - now*

- Alignment tuning (RLHF, RLAIF)

VARCO Text: NCSOFT's AI Writing Assistant   *2023 - 2024*

- Instruction fine-tuning and alignment tuning from the foundation LLM

AI Weather-Forecast News Article Generation (with Yonhap News Agency)   *2019 - now*

- Generative AI model which specialized on writing weather-forecasting news
- Algorithm for converting raw weather data into a frame-based meaning representation

Developing In-house Generative AI-model-training Framework   *2018 - now*

- Developing an automated (data preprocess - train - serve) ML training framework
  which targets no coding but just preparing the config-file for LLM Training (fine-tuning, alignment)

Baseball Game Summary Generation from the game log   *2018 - 2022*

- Developing AI model generates baseball game summary from the game log
- Deployment system of ML model for serving to a real-time application

Deep Research on Pointer Generator (Copy) Networks   *2020 - 2021*

- Performance enhancement of Pointer Generator network reducing hallucinations

Developing a Korean Word2vec Model   *2016 - 2017*

- Developing and publishing of word2vec model trained with large Korean corpus
- Research on syllable-level model leveraging Korean morphological characteristics

A Quote Recommender System   *2015 - 2016*

- Algorithm for evaluating recommended quote of AI dialog system

Input and Management System for Korean Minsokak Fieldwork Material Database *2015 - 2016*

- System for integrating Korean Minsokak(Folk Music) materials into a single database with in a database management system(DBMS)

EDUCATION

Seoul National University *Seoul, Korea*
M.S. in Computer Science and Engineering *Mar 2015 - Feb 2017*
Supervised by Prof. Sang-goo Lee
Thesis: The Modeling and Training Methods for Syllable-based Korean Word Embeddings

Korea University *Seoul, Korea*
B.S. in Department of Computer Science and Engineering (1st Major) *Mar 2011 - Feb 2015*
Software Technology and Enterprise Program for Unlimited Potential (Relation Major)

PUBLICATIONS

Sanghyuk Choi, Jeong-in Hwang, Hyungjong Noh, Yeonsoo Lee. May the Force Be with Your Copy Mechanism: Enhanced Supervised-Copy Method for Natural Language Generation. arXiv preprint, 2021.

Sanghyuk Choi, Taeuk Kim, Jinseok Seol, Sang-goo Lee. A Syllable-based Technique for Word Embeddings of Korean Words. SCLeM2017(The 1st Workshop on Subword and Character level models in NLP, EMNLP2017).

Sanghyuk Choi, Jinseok Seol, Sang-goo Lee. On Word Embedding Models and Parameters Optimized for Korean. The 28th Annual Conference on Human & Cognitive Language Technology.

Sanghyuk Choi, Sang-goo Lee. On the Implementation of an Input and Management System for Korean Minsokak(Folk Music) Fieldwork Material Database. Studies in Korean Music(Vol.59).

AWARDS

Software Olympiad, Korea University **2nd Prize** *2012*

ACM-ICPC Asia Daejeon Regional **4th Prize** *2012*

ACM-ICPC Asia Daejeon Regional **9th Prize** *2013*

OPEN SOURCE

Contributor of OpenNMT-py: Open-Source Neural Machine Translation and (Large) Language Models

Contributor of PEFT: State-of-the-art Parameter-Efficient Fine-Tuning (PEFT) methods, Huggingface

TEACHING EXPERIENCE

TA of Digital Computer Concept and Practice @ Seoul National University *Spring, 2015*

TA of Digital Computer Concept and Practice @ Seoul National University *Spring, 2016*

LANGUAGES

English(Fluent), and Korean(Native)

TECHNICAL SKILLS

Advanced
Python, Pytorch, Pytorch-lightning, Deepspeed, Huggingface, Tensorflow, C, C++, Java, JavaScript, SQL, and Git

Intermediate
C#, Hadoop, and MATLAB