

SANGHYUK CHOI

+82 1045708606

sanghyuk-choi@hotmail.com

RESEARCH INTERESTS

Natural Language Processing

LLM(Instruction fine-tuning, Learning from human-feedback), Natural Language Generation(Data-to-text, Abstractive summarization)

Machine Learning

Deep Neural Networks, ML Ops

EDUCATION

Seoul National University

M.S. in Computer Science and Engineering

Seoul, Korea

Mar 2015 - Feb 2017

Supervised by Prof. Sang-goo Lee

Thesis: The Modeling and Training Methods for Syllable-based Korean Word Embeddings

Korea University

B.S. in Department of Computer Science and Engineering (1st Major)

Seoul, Korea

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

WORK EXPERIENCE

NCSoft

Team Leader of Application LM Team

Seoul, Korea

Sep 2023 - now

Team Leader of Generation Model Team

Jun 2022 - Jan 2024

Research Engineer

Jan 2018 - now

Stayes Inc.

CTO, Co-founder

Seoul, Korea

Feb 2014 - Feb 2015

Visionaries Inc.

CTO, Co-founder

Seoul, Korea

Mar 2013 - Feb 2014

RESEARCH EXPERIENCE

VARCO LLM (VARCO Text), NCSoft's large language model *2022 - now*

- Instruction fine-tuning of pretrained GPT model
- Learning from Human-feedback (RLHF, RRHF, SLiC-HF, DPO, PRO) with LLM-as-a-Judge.
- 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)

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

- A machine learning-based weather forecast news generation model
- Algorithm for converting raw weather data to a framed-based meaning representation

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

- Performance enhancement of Pointer Generator network on summarization and data-to-text tasks
- Deep analysis and comparison of prior copy networks

Baseball Game Summary Generation from the game log *2018 - 2020*

- A machine learning-based baseball game summary generation from the game log
- Deployment system of ML model for serving to a real-time application

Implementing a Korean Word2vec Model *2016 - 2017*

- Implementing and Publishing of word2vec model trained with large Korean corpus
- Syllable-level Korean morphological analysis

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)

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*

OPEN SOURCE

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

LANGUAGES

English(Fluent), and Korean(Native)

TECHNICAL SKILLS

Advanced

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

Intermediate

C#, Hadoop, and MATLAB