

Chanjun Park

UPSTAGE

338, Gwanggyojungang-ro, Suji-gu, Yongin-si, Gyeonggi-do, Republic of Korea

☎ (+82) 10-2580-9066 | ✉ chanjun.park@upstage.ai / bcj1210@naver.com | 📱 [parkchanjun](#) | 📺 [bcj1210](#)

Research Interests

Chanjun Park is a researcher in the field of Natural Language Processing (NLP), with a focus on Data-Centric AI, Machine Translation and Large Language Model (LLM). He is currently working as an AI Research Engineer at Upstage since March 2022. In 2023, he received Ph.D. from Korea University under the supervision of Professor Heuseok Lim for the work on “Data-Centric Neural Machine Translation”. From 2018 to 2019, he worked at SYSTRAN as a Research Engineer. Chanjun is the founder and leader of the KU-NMT Group, and has received the Naver Ph.D. Fellowship in 2021. He served as the Virtual Social Chair at COLING 2022, and is currently serving as the Program Chair for the WinLP Workshop. He has published more than 150 papers in the field of NLP.

- Data-centric AI
- Machine Translation
- Large Language Model (LLM)
- Real-World-centric AI

Education

Korea University

PH.D IN COMPUTER SCIENCE AND ENGINEERING, GPA: 4.44 / 4.5

- **Advisor: Heuseok Lim**
- **Natural Language Processing & Artificial Intelligence Lab**
- **Founder and Leader of KU-NMT Group.**

Seoul, South Korea

Sep. 2019 - Aug. 2023

Busan University of Foreign Studies (BUFS)

B.S. IN NATURAL LANGUAGE PROCESSING & CREATIVE CONVERGENCE, GPA: 4.2 / 4.5

- **Advisor: Pummo Ryu**
- **Full Scholarship**

Busan, South Korea

Mar. 2015 - Aug. 2018

Academic Services

Widening NLP (WinLP) Workshop

PROGRAM CHAIR

- WinLP: <https://www.winlp.org/>

Anywhere on Earth

March.2023-Present

COLING 2022

VIRTUAL SOCIAL CHAIR

- COLING 2022: <http://coling2022.org/>
- Virtual Social Chair

South Korea

Sep.2021-Present

Program committe

PROGRAM COMMITTEE

- ICML 2023-ICML 2023-Data-centric Machine Learning Research (DMLR), ACL 2023-Industry Track, ACL 2023-BEA, EACL 2023, EACL 2023-Insights, ACL 2022-Insights, NAACL 2022-Industry Track, Reviewer of ARR (2022 Current)

South Korea

January.2021-Present

Reviewer

REVIEWER

- IEEE Access, International Journal of Human-Computer Interaction, Cybernetics and Systems, Applied Sciences, Electronics, Entropy, Connection Science

South Korea

January.2021-Present

Selected Publications

Honors & Awards

2023.02	Best Paper Award , Korea University	South Korea
2023.02	Research Encouragement Scholarship , Korea University	South Korea
2022.12	1st place in WMT 2022 QE Task 3, 2022 , WMT 2022 - EMNLP 2022	Abu Dhabi
2022.10	Best Paper Award, 2022 , The 34th Annual Conference on Human & Cognitive Language Technology (HCLT2022)	South Korea
2021.12	Naver Ph.D. Fellowship, 2021 , Naver	South Korea
2021.10	Best Paper Award (NLP Application 2 Section), 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	Best Paper Award (Language Resource Section), 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	Best Paper Award (QA and Speech Section), 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.07	Ranked 4th on the CommonGen 1.1 Leaderboard (Nov. 2022 Ranked 7th, CommonGen 1.1) , CommonGen 1.1	South Korea
2020.11	1st Place in Flitto Hackathon (Team Lead), 2020 , Flitto	South Korea
2020.10	Best Paper Award, 2020 , The 32nd Annual Conference on Human & Cognitive Language Technology (HCLT2020)	South Korea
2020.05	Best practices for using NIA AI training data(Korean-English Neural Machine Translation model) , AI Hub	South Korea
2019.10	Best Paper Award, 2019 , The 31st Annual Conference on Human & Cognitive Language Technology (HCLT2019)	South Korea
2019.10	1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead) , Microsoft	South Korea
2019.03	Graduate School Associate Scholarship , Sungkyunkwan University	South Korea
2018.10	Next Generation Information Processing NLP Competition 2018: Participation Award , Next-generation information computing technology development business	South Korea
2017.06	Bit Computer Excellence Award (President Award) , Bit Computer	South Korea
2015-2019	Full Scholarship , BUFS	South Korea
2016-2017	Scholarship for academic excellence , Sooyoungro Church	South Korea

Selected Research Experience

AskUp Biz

RESEARCHER IN UPSTAGE

Seoul, South Korea

June. 2023. - Present

- Research for Large Language Model (LLM) based 3rd Party Platform
- Responsible for LLM evaluation
- AskUp Seargest and AskUp QA
- Skills: Data-Centric AI, LLM, NLP, Search

Large Language Model (LLM) TF

RESEARCHER IN UPSTAGE

Seoul, South Korea

June. 2023. - Present

- Research for Large Language Model (LLM)
- Skills: LLM, NLP, Data

DMOps for Document AI

RESEARCHER IN UPSTAGE

Seoul, South Korea

October.2022. - June.2023

- Data Management Operations (DMOps) research related to OCR
- Building training and evaluation data for OCR
- Presentation at AI · DATA SUMMIT 2023: https://conference.etnews.com/conf_info.html?uid=237
- Publication of 7 research paper in ICML-Data-centric Machine Learning Research (DMLR): <https://dmlr.ai/>
- Press: <https://www.aitimes.com/news/articleView.html?idxno=151905>
- Skills: Data Management, DataOps, OCR

(Project) LGU + VOD Search POC Project

RESEARCHER IN UPSTAGE

Seoul, South Korea

June. 2022. - October.2022

- Gold Testset and evaluation system design for tag extraction model comparison
- Responsible for designing search system evaluation. Planning and designing blind tests for comparison with commercial systems
- Press: <http://www.asiaa.co.kr/news/articleView.html?idxno=90842>
- Skills: Customer Experience (CX), Quality Assurance (QA)

(Project) Amorepacific Search commercialization Project

Seoul, South Korea

RESEARCHER IN UPSTAGE

March. 2022. - June.2022

- Participated in Amorepacific's AP Mall search commercialization project
- Stability test planning and design
- Response time testing planning and design
- Establishment of systematic QA process based on bug testing
- Skills: Customer Experience (CX), Quality Assurance (QA)

Product Search Pack

Seoul, South Korea

RESEARCHER IN UPSTAGE

March. 2022. - October. 2022

- Develop Product Search Pack
- Establishment of evaluation system to develop Product Search Pack and responsible for data, planning, and evaluation
- Quality Assurance for Search System, Gold Testset Planning
- Press: <https://sedaily.com/NewsView/265XEWO00Z>
- Skills: Data Manager, User Experience (UX), Customer Experience (CX), Quality Assurance (QA)

(Project) Polyglot: Large Language Models of Well-balanced Competence in Multi-languages

Seoul, South Korea

RESEARCHER IN ELEUTHER AI

September. 2022. - May.2023

- Various multilingual models such as mBERT, BLOOM, and XGLM have been released. Therefore, someone might ask, "why do we need to make multilingual models again?" Before answering the question, we would like to ask, "Why do people around the world make monolingual models in their language even though there are already many multilingual models?" We would like to point out there is a dissatisfaction with the non-English language performance of the current multilingual models as one of the most significant reason. So we want to make multilingual models with higher non-English language performance. This is the reason we need to make multilingual models again and why we name them 'Polyglot'.
- Skills: Language Model, Data management

(Project) Korean-Prompt-based Engineering for Few-shot Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

March. 2022. - July. 2022.

- Plain Template Insertion: Korean-Prompt-based Engineering for Few-shot Learners
- Supported by Hyundai Motor
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9913979>
- Skills: Prompt learning, Few-shot learning, Pytorch, Huggingface

(Project) Return on Advertising Spend (ROAS) Prediction Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - March. 2022.

- Return on Advertising Spend (ROAS) refers to the ratio of revenue generated by advertising projects to its expense.
- We present a method for predicting ROAS that does not require controlled experiments in data acquisition and validates its effectiveness through comparative experiments. Specifically, we propose a task deposition method that divides the end-to-end prediction task into the two-stage process; occurrence prediction and occurred ROAS regression.
- Supported by Data voucher
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/10/1637>
- Skills: ROAS, Regression, Pytorch

(Project) Hybrid Retriever System for Industrial Frequently Asked Questions Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - March. 2022.

- We propose a method of returning responses to customer inquiries using only data that can be easily obtained from companies based on the deep learning information retrieval module. In this work, we hybridize dense embedding and sparse embedding to make it robust to professional terms and suggest new functions to scale the results returned by the two modules and adjust the weight ratio.
- Supported by Data voucher
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/8/1335>
- Skills: Information Retrieval, Pytorch

Fake News Detection Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2022. - Present

- Research for Utilization Strategy of User Engagements in Korean Fake News Detection
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9841541>
- Skills: Graph Neural Network, Pytorch, Fairseq

(Project) NAVER PAPAGO - Parallel Corpus Filtering and Mining

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - June. 2022.

- Research for Static and Dynamic Parallel Corpus Filtering and Mining
- Human-inspired curriculum learning based NMT
- Supported by NAVER PAPAGO
- Skills: Neural Machine Translation, Parallel corpus filtering, Pytorch, Fairseq

Data-centric Neural Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

July. 2021. - Present

- Research for Data-centric Neural Machine Translation, Automatic Post Editing, Quality Estimation, Generative Commonsense Reasoning
- Develop Efficient tool for building high quality parallel corpus and A New Tool for Efficiently Generating Quality Estimation Datasets and A Self-Supervised Automatic Post-Editing Data Generation Tool
- Publication of 3 research paper in NeurIPS 2021 - Data-centric AI (DCAI) workshop: <https://eval.how/dcai2021/index.html>
- Publication of research paper in ICML 2022 - DataPerf workshop: <https://sites.google.com/view/dataperf2022>
- Publication of 7 research paper in ICML-Data-centric Machine Learning Research (DMLR): <https://dmlr.ai/>
- Improving Formality-Sensitive Machine Translation using Data-Centric Approaches and Prompt Engineering
- Publication of research paper in IWSLT 2023 : <https://sites.google.com/view/dataperf2022>
- A Survey on Evaluation Metrics for Machine Translation
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/11/4/1006>
- Skills: Neural Machine Translation, Web Application, Fairseq

Korean Commonsense Generation and Reasoning Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - Present

- KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation
- Automatic Knowledge Augmentation for Generative Commonsense Reasoning
- Publication of research paper in NAACL 2022: <https://aclanthology.org/2022.findings-naacl.172/>
- Publication of research paper in Knowledge-Based Systems: <https://www.sciencedirect.com/science/article/abs/pii/S0950705122009546>
- Skills: Commonsense Reasoning, Pytorch, Huggingface

Image Retrieval Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - September. 2021.

- Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9585128>
- Skills: Image Retrieval, Pytorch, Huggingface

(Project) Development of patent document invention sentence extraction and key phrase extraction module

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - December. 2021.

- AI for Patent research
- T5 based Key sentence extraction and key phrase generation
- Supported by "LG Innotek"
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9779775>
- Skills: Automatic Summarization, Sentence classification, Information Extraction Pytorch, Huggingface

Dialogue Summarization Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - December. 2021

- Syntax-Aware Dialogue Summarization using Multi-task Learning
- Skills: Dialogue system, Automatic Summarization, Pytorch, Huggingface
- Publication of research paper in EMNLP 2021 -Widening NLP (WiNLP2021) workshop
- Publication of research paper in NeurIPS 2021 - Women in Machine Learning (WiML 2021) workshop
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/abstract/document/9664379/>

Dialogue System Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2021. - December. 2021

- Robust Reward Estimator for Multidomain Task-Oriented Dialogue
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/14/6624>
- BlenderBot 2.0 Errors Analysis in terms of Model, Data and User-Centric Approach
- Publication of research paper in Arxiv: <https://arxiv.org/abs/2201.03239>
- Focus on FoCuS: Is FoCuS focused on Context, Knowledge and Persona?
- Publication of research paper in COLING 2022 Workshop: <https://aclanthology.org/2022.ccgpk-1.1.pdf>
- Skills: Dialogue system, Pytorch, Huggingface, ParlAI

(Project) Subtitle production and editing technology development for colloquial broadcasting contents

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2021. - December. 2021.

- Development of Colloquial Aware Neural Machine Translation (CANMT) model that is robust to colloquial expression based on broadcast content
- Consortium: Korea Univ, Hanyang Univ, ON THE IT, SBS Contents Hub, CBS
- Supported by IITP
- Korean Colloquial-Sentence-Splitter(Ko-CSS) for simultaneous speech translation
- Neural Architecture Search(NAS) based NMT Research
- Speech recognition post-processor research
- Cross lingual adaptation research
- Skills: Neural Machine Translation, Automatic Speech Recognition, simultaneous speech translation, Pytorch, Huggingface, Fairseq
- Ko-CSS Demo: <http://nlplab.iptime.org:32285/>

BTS: Speech recognition post-processor Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Proposed BackTranScript(BTS). Back Transcription (BTS) has seen limited success without the role of phonetic transcription for building and creating vast datasets automatically.
- Publication of research paper in WAT 2021 (ACL Workshop): <https://aclanthology.org/2021.wat-1.10/>
- Publication of research paper in Mathematics: <https://www.mdpi.com/2227-7390/10/19/3618>
- Skills: Neural Machine Translation, Pytorch, Huggingface, Fairseq

Cross Lingual Pretrained Language Model Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Research for Exploring the Data Efficiency of Cross-Lingual Post-Training in Pretrained Language Models
- Research for combining BART and Parallel Corpus Filtering
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/5/1974>
- Skills: Neural Machine Translation, Pytorch, Huggingface, Fairseq

Automatic Post Editing Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- APE research based on Cross Lingual Pretrained Language Model
- Adapter based APE research
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9528385/>
- Research for Efficient and Simple Data Generation Method - Low Resource APE and Pseudo APE data generation
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9714400>
- Research for Synthetic Data Generation Using Noising Strategies for Automatic Post-editing
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.93.pdf>
- Skills: Neural Machine Translation, Pytorch, Automatic Post Editing, Huggingface, Fairseq

Quality Estimation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Large Scale PLM based Quality Estimation development and research
- Low Resource QE and Pseudo QE data generation
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/14/6584>
- Publication of research paper in MT Summit 2021 - LoResMT Workshop: <https://aclanthology.org/2021.mtsummit-loresmt.1/>
- Publication of research paper in IEEE Access: <https://ieeexplore.ieee.org/document/9761258>
- Publication of research paper in COLING 2022: <https://aclanthology.org/2022.coling-1.460.pdf>
- 1st place in WMT 2022 QE Task 3: KU X Upstage's submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task
- Skills: Neural Machine Translation, Pytorch, Quality Estimation, OpenKiwi, Huggingface, Fairseq

COVID-19 Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

May. 2020. - November.2020

- Neural Machine translation specialized for COVID-19 domain
- Skills: Neural Machine Translation, Pytorch, Python
- Publication of research paper in Journal of the Korea Convergence Society(KCI): <https://bit.ly/2QgQKLK>

(Project) Domain Specialized Neural Machine Translation (Security)

Seoul, South Korea

RESEARCHER IN KOREA UNIV

May. 2020. - November.2020

- Research on Neural Machine Translation specialized for security domain
- Supported by "National Security Research Institute"
- Project Manager, Tech Lead
- Development of the latest translation model based on Transformer/BERT, AI-based Korean-English translation model suitable for the security field, and research trend of the latest AI-based translation technology research
- We propose Cross Communication Method (CCM), a new DS-NMT training methodology, inspired by cognitive scientific interpretation of cross language speech perception. This is the first study to reflect cognitive science theory in DS-NMT training methodology.
- Publication of research paper in IEEE Access: <https://ieeexplore.ieee.org/document/9751075>
- Skills: Neural Machine Translation, Pytorch, Python

(Project) A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2020. - July. 2023

- Neural-Symbolic Model Research and Development
- Skills: Neural Symbolic Computing, Knowledge base, Pytorch, Python
- Supported by IITP
- Demo: <http://nlpplab.iptime.org:32245/>

Ancient Korean Neural Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

March. 2020. - December. 2022

- Ancient Korean Neural Machine Translation Research
- Bilingual Word Embedding for Ancient Korean Neural Machine Translation
- Priming Ancient Korean Neural Machine Translation Research
- Publication of research paper in IEEE Access(SCIE): <https://ieeexplore.ieee.org/document/9125904/>
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.3.pdf>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

Deep Learning based Grammar Error Correction Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - June.2020

- Deep learning based English grammar corrector development
- Publication of research paper in IEEE Access(SCIE): <https://ieeexplore.ieee.org/document/9102992>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

Deep Learning based Korean Spelling Correction Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - Present

- Development of Korean spelling corrector based on deep learning
- We propose a gold test set, K-NCT (Korean Neural grammar Correction Test set), for Korean grammar error correction and a new error type classification guideline.
- Publication of research paper in multimedia tools and applications (SCIE): <https://link.springer.com/article/10.1007/s11042-020-09148-2>
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9938990>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

Neural Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - Present

- Deep learning based Korean-English and English-Korean, Korean-Japanese machine translator development
- Research for Parallel Corpus Filtering, Data Augmentation, Various decoding strategies
- Best practices for using NIA AI training data: <http://www.aihub.or.kr/node/4525>
- Publication of research paper in ELECTRONICS (SCIE): <https://www.mdpi.com/2079-9292/9/10/1562#cite>
- Our paper is accepted in NAACL 2021 Industrial Track: "Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification"
- How to improve the performance of NMT without changing the model through data pre- and post-processing
- Publication of research paper in NAACL 2021: <https://www.aclweb.org/anthology/2021.naacl-industry.13/>
- Verification of Ensemble Effect in Neural Machine Translation
- Publication of research paper in EMNLP 2021 - Workshop on Insights from Negative Results in NLP: <https://insights-workshop.github.io/2021/cfp/>
- Publication of research paper in Applied Sciences (SCIE): <https://www.mdpi.com/2076-3417/12/11/5545>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

AI for Education Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2019. - Present

- FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue
- Publication of research paper in AAAI 2022 - Artificial Intelligence for Education (AI4EDU) workshop: <https://arxiv.org/abs/2112.04126>
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.132.pdf>
- Automatic Essay Scoring Research
- PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning
- Publication of research paper in ACL 2023 - Demo Track: <https://www.mdpi.com/2227-7390/10/9/1486>
- AI student: A Machine Reading Comprehension System for the Korean College Scholastic Ability Test
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/9/1486>
- BERTOEIC: Solving TOEIC Problems Using Simple and Efficient Data Augmentation Techniques with Pretrained Transformer Encoders
- Publication of research paper in Applied Sciences (SCIE) : <https://www.mdpi.com/2076-3417/12/13/6686>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

AI for Accessibility Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - December. 2022

- PicTalky: Design of augmentative alternative communication (AAC) software for children with developmental disabilities
- 1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead)
- Award: <https://bit.ly/2vPaZ91>
- PickTalky Demo: <https://pictalky-311.de.r.appspot.com/>
- Publication of research paper in AACL 2022 Demo Track: <https://preview.aclanthology.org/aac1-22-ingestion/2022.aac1-demo.3/>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

(Project) Research and development of human-inspired multiple intelligence

RESEARCHER IN KOREA UNIV

- Human-inspired language understanding model
- Funded by the Korea Ministry of Science and ICT (MSIT)
- Site: <http://hiaai.co.kr/>

Seoul, South Korea

September. 2019. - August. 2022

(Project) Next-generation information computing technology development business

RESEARCHER IN KOREA UNIV

- Development of Core Application Techniques for Korean Intelligent Systems
- Research and development of Korean natural language processing core technology

Seoul, South Korea

September. 2019. - January. 2021.

(Project) Core technology development of the real-time simultaneous speech translation based on knowledge enhancement

RESEARCHER IN SYSTRAN

- Collaboration with the Korea Electronics and Telecommunications Research Institute (ETRI) Language Intelligence Research Group
- Director: Dr. Young-Gil Kim (ETRI)
- 3rd year: Real-time lecture interpretation solution development
- 4th year: Busan University of Foreign Studies Field Group Test (FGT) progress and software maintenance
- Skills: JAVA, JAVA FX, Kaldi

Seoul, South Korea

June. 2018. - July. 2019

(Project) Neural Machine Translation service for SK Telecom's in-house education program

RESEARCHER IN SYSTRAN

- Development of automation software to provide translation service to SK Telecom's in-house education program
- Skills: Python

Seoul, South Korea

April. 2019. - April. 2019

LCloud-API Platform Service Development

RESEARCHER IN SYSTRAN

- Planning and development of natural language processing API Platform Service
- Gravitree IO-based platform design, planning, marketing materials production, strategy establishment
- REST API management using Swagger
- Transcription API, Translation API, Address Correction API, Speaker Verification API
- Learned Microservices architecture(MSA), IT Infrastructure
- Skills: Gravitree IO, Microservices architecture(MSA), IT Infrastructure, Mockup, Marketing, Strategic Planning, Technology management, Amazon AWS

Seoul, South Korea

January. 2019. - June. 2019

Speech recognition transcription tool development

RESEARCHER IN SYSTRAN

- Development and maintenance of speech recognition transcription tool, modification and addition of various functions
- Skills: Front End, Spring Boot, Flask, AJAX, MySQL, Kaldi,

Seoul, South Korea

January. 2019. - July. 2019

(Project) CJ Logistics Speech Recognition Agent

RESEARCHER IN SYSTRAN

- CJ Logistics artificial intelligence (AI) virtual assistant project
- Skills: Quality Assurance, Quality control, Mockup, Strategic Planning, Kaldi

Seoul, South Korea

Aug. 2018. - July. 2019

Automatic subtitle generation platform service development and planning

RESEARCHER IN SYSTRAN

- Development and maintenance of automatic subtitle generation platform
- Establish platform service strategies such as B2B and B2C
- Various company contacts and SWOT analysis
- Skills: JAVA, JAVA FX, Kaldi, Mockup, Strategic Planning, Technology management, Marketing

Seoul, South Korea

June. 2018. - June. 2019

Planning of English speaking evaluation platform service

RESEARCHER IN SYSTRAN

- Establishing a service strategy using English speaking evaluation technology
- English speaking evaluation technology Q.A & Q.C
- Skills: Quality Assurance, Quality control, Mockup, Strategic Planning, Technology management, Marketing

Seoul, South Korea

June. 2018. - December. 2018

Speech Recognition Research

RESEARCHER IN SYSTRAN

- Research for Automatic Speech Recognition Technology based on Kaldi
- Acoustic Model, Language Model
- Speaker Verification
- English speaking evaluation
- Real-time speech recognition system
- Skills: Kaldi

Seoul, South Korea

Aug. 2018. - July. 2019

Neural Machine Translation Research

RESEARCHER IN SYSTRAN

- Research for Neural Machine Translation
- Skills: OpenNMT Tensorflow, OpenNMT Pytorch

Seoul, South Korea

Aug. 2018. - July. 2019

(Project) Development of multilingual word translation service based on Korean semantics analysis for multicultural family

RESEARCHER IN BUFS

- Han-Tong-E: Multilingual Word Translation Service based on Word Semantic Analysis
- Collaboration with Professor Ok Cheol-young's laboratory at Ulsan University
- Project homepage development
- In-depth understanding of Utagger and seminar on WSD
- Build project-related data and conduct surveys
- Project Page: <https://bufsconv.blogspot.kr/>
- Funded by National Research Foundation of Korea(NRF)

Seoul, South Korea

Sep. 2016. - June. 2018

(Project) Establishment of Gold Standard language resource bank for Korean machine learning and evaluation (Next-generation information computing technology development business)

RESEARCHER IN BUFS

- For the purpose of developing Korean Natural Language Toolkit
- Development of automatic Korean word spacing and Korean compound noun decomposer
- Skills: Python

Seoul, South Korea

Sep. 2017. - June. 2018

(Project) Local industry development project in connection with public institutions

RESEARCHER IN BUFS

- ICT VAEMS: Vertical Aquaculture Energy Management
- Funded by the Korea Institute for Advancement of Technology(KIAT)

Seoul, South Korea

Sep. 2016. - Aug. 2017

(Project) Basic Research on Semantic Frame Generation for Short Questions (ExoBrain)

RESEARCHER IN BUFS

- Exobrain Project

Seoul, South Korea

Oct. 2015. - Feb. 2016

Professional Experience

Upstage

AI RESEARCH ENGINEER

- About Upstage: <https://www.upstage.ai/>
- Interactive AI Team & LLM TF
- Establishment of evaluation system to develop AskUp Biz and responsible for data, planning, and evaluation
- Responsible roles related to LLM evaluation
- Semantic search project with AMOREPACIFIC
- Semantic search project with LG U+
- Semantic Search-related DMOps
- OCR-related DMOps
- Research for Data-centric AI
- Research for Large Language Model

Seoul, South Korea

March. 2022 - Present

SYSTRAN

RESEARCH ENGINEER

- About SYSTRAN: <https://www.systransoft.com/>
- Automatic Speech Recognition(ASR) Team
- Core technology development of the real-time simultaneous speech translation based on knowledge enhancement (ETRI)
- Participated in NTT Docomo speech Recognition Solution Project
- Participated in CJ Logistics speech recognition agent development project
- Participated in Samsung SDS speech recognition project
- Speech recognition service planning and development (Real-time lecture interpretation, Automatic subtitle generation service, English speaking evaluation, Customer-specific recognizer service, Web RTC-based Real-time video interpretation)
- Deep learning based speech recognition research (Acoustic Model, Language Model Training)
- Machine Translation research
- IaaS and PaaS service planning, API Platform service operation planning, B2B B2C planning and development
- Micro Service Architecture Design and Successful API Platform Service Launching Experience (LCloud)
- Governmental project proposal experience
- Reason for leaving: Graduate school

Seoul, South Korea

Sep. 2018 - July. 2019

- Automatic Speech Recognition(ASR) Team
- Participated a project "Core technology development of the real-time simultaneous speech translation based on knowledge enhancement (ETRI)"
- Developed Simultaneous Speech Translation (implemented using JAVA, Python and Kaldi)

Skills

Programming	Python, JAVA, C / C++
Machine Learning & Deep Learning	Pytorch, Pytorch lightning, Keras, Tensorflow
Natural Language Processing	Machine Translation, Speech Recognition, Information Retrieval, Huggingface, OpenNMT, Kaldi, FAISS
Full Stack	Html, CSS, Javascript, Java FX, Oracle, MySQL, Spring Boot, VertX, AWS
Product & Project & Data Manager	MS Office, Notion, Slack, Linear, Whimsical, Recruitee, Redmine
Strategy & Planning	API Platform Service, Platform Service Planning, Gravitee IO, Paas, Iaas, Q.A, Q.C, Mockup, Swagger
Etc	Linux, Docker, Web RTC, Latex
Language Skills	Korean, English, Japanese, Thai, Myanmar

Patents

2019.12	Device and method for correcting Korean spelling , Granted Patent (10-2430918)	South Korea
2020.09	APPARATUS AND METHOD FOR OUTPUTTING IMAGE CORRESPONDING TO LANGUAGE , Granted Patent (10-2476497)	South Korea
2020.09	METHOD OF TRANSLATING ANCIENT KOREAN USING MACHINE TRANSLATION , Granted Patent (10-2425922)	South Korea
2022.04	Correction performance evaluation metrics of neural network machine translation and method of constructing the same , Granted Patent (10-2390154)	South Korea
2020.09	METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION , Granted Patent (10-2409667)	South Korea
2021.03	Improving speech recognition performance using TTS in domain-specific environment , Apply for a patent	South Korea
2021.05	Method For Generating Training Data And Method For Post-Processing Of Speech Recognition Using The Same , Apply for a patent	South Korea
2021.09	DEVICE AND METHOD FOR GENERATING OPTIMAL TRANSLATION SUBTITLE USING QUALITY ESTIMATION , Apply for a patent	South Korea
2021.09	DEVICE AND METHOD FOR GENERATING TRAINING DATA FOR AUTOMATIC POST EDITING , Apply for a patent	South Korea
2021.09	METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION , PCT International Application	South Korea
2021.09	APPARATUS FOR CORPUS PROCESSING, APPARATUS AND METHOD AND MATHINE TRANSLATION , Apply for a patent	South Korea
2021.11	DEVICE AND METHOD FOR GENERATING OF TRAINING DATA FOR QUALITY ESTIMATION IN MACHINE TRANSLATION , Apply for a patent	South Korea

Demo

2019.12	Deep Learning based English Grammar Error Correction , http://nlplab.iptime.org:32292/	South Korea
2019.12	Deep Learning based Korean Spelling Correction , http://nlplab.iptime.org:32288/	South Korea
2020.01	Korea University Neural Machine Translation Platform , http://nlplab.iptime.org:32296/	South Korea
2020.02	Neural Symbolic Inference , http://nlplab.iptime.org:32245/	South Korea
2020.02	English Free Talking and Grammar Error Correction Feed Back System using NAO Robot , https://www.youtube.com/watch?v=Rl6dG_4cnFk	South Korea
2020.03	Korean Ancient Neural Machine Translation Platform , http://nlplab.iptime.org:32242/	South Korea
2020.03	PicTalky: Text to Pictogram Demo for Children with Developmental Disabilities , https://pictalky-311.de.r.appspot.com/	South Korea
2020.05	COVID-19 Machine Translation Platform , http://nlplab.iptime.org:32250/	South Korea
2020.07	Ancient Korean Translation & Machine Reading Comprehension with NAO , https://www.youtube.com/watch?v=nXBZe-t9a-4&t=8s	South Korea
2020.11	Bilingual Education Using NAO , https://www.youtube.com/watch?v=4kCMMa62LDY&t=9s	South Korea
2021.04	Ko-CSS (Korean Colloquial-Sentence-Splitter) for simultaneous speech translation , http://nlplab.iptime.org:32285/	South Korea
2021.06	BTS (BackTranScription) for STT Post-Processor , https://backtrans-317413.de.r.appspot.com/	South Korea
2021.09	Translation platform with AIHUB dataset , http://nlplab.iptime.org:32247/	South Korea
2021.09	Con-Talky , http://nlplab.iptime.org:32245/	South Korea
2021.09	Korea University English Grammar Corrector KakaoTalk Channel , http://pf.kakao.com/_cqkxps	South Korea
2021.09	PEEP-Talk , http://pf.kakao.com/_FlDxgs	South Korea
2021.10	A Tool for generating High-Quality Parallel Corpus considering Human Translation Market , http://nlplab.iptime.org:9090/	South Korea
2021.10	A New Tool for Efficiently Generating Quality Estimation Datasets , http://nlplab.iptime.org:9091/	South Korea
2021.11	A Self-Supervised Automatic Post-Editing Data Generation Tool , http://nlplab.iptime.org:9092/	South Korea

Invited Talk & Teaching Experience

Jeju National University

PRESENTER

- Real-World Artificial Intelligence and Large Language Model for Everyone
- Invited by Jeju National University

South Korea

July. 2023

AI Education Alliance Policy lab (AIEDAP)

PRESENTER

- Deep Learning Understanding and Practice
- Invited by AIEDAP

South Korea

June. 2023

KIDA (Korea Institute for Defense Analysis)

PRESENTER

- Data-Centric AI in the Large Language Model Era
- Invited by KIDA (Korea Institute for Defense Analysis)

South Korea

May. 2023

Upstage

PRESENTER

- NLP based Large Language Model for All

South Korea

April. 2023

BoostCamp

MASTER (INSTRUCTOR)

- Course Name: Data-centric NLP
- HomePage: https://boostcamp.connect.or.kr/program_ai.html

South Korea

February. 2023-Present

AI · DATA SUMMIT 2023

PRESENTER

- Real-World Centric AI
- Invited by AI · DATA SUMMIT 2023

South Korea

February. 2023

Sunmoon University

PRESENTER

- Real-World Centric AI
- Invited by Sunmoon University

South Korea

December. 2022

Kyungsoong University

PRESENTER

- Language and Information Studies and the Future of Artificial Intelligence
- Invited by Kyungsoong University

South Korea

August. 2022

Hankuk University of Foreign Studies

PRESENTER

- Basic practice of natural language processing for everyone
- Invited by Hankuk University of Foreign Studies, Interpreting and Translation Research Institute (ITRI)

South Korea

July. 2022

Dongguk University

PRESENTER

- Artificial intelligence and Machine Translation lectures (3 times)
- Invited by Dongguk University, Department of English Language and Literature

South Korea

January. 2022

Korea University

TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data

South Korea

Autumn 2021

Busan Social Welfare Development Institute

INVITED ADVISOR

- Attending advisory meetings and Focus Group Interview
- ICT-Based Ecosystem Construction Plan to Secure Opportunities to Enjoy Together - A Study on Developmental Disabilities

South Korea

July. 2021

Korea University

HEAD TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data
- (DFE610) Natural Language Processing for Digital Finance Engineering

South Korea

Autumn 2020

Korea University

TEACHING ASSISTANT

- (COSE461) Natural Language Processing
- (DFC615) Artificial Intelligence and Natural Language Processing
- Lecture materials: <https://github.com/Parkchanjun/KU-NLP-2020-1>

South Korea

Spring 2020

LLsoLLu

PRESENTER

- Latest natural language processing trends
- Language Representation, Neural Symbolic Computing
- Invited by LLsoLLu Global Strategy and Planning Team

South Korea

March. 2020

NC SOFT

PRESENTER

- Technology Transfer Seminar
- Korean Spelling Corrector Using Automatic Noise Generation Methodology
- Invited by NC SOFT Language AI Team

South Korea

February. 2020

Dongguk University

PRESENTER

- A.I - NLP - MT for Liberal Arts
- Introduction to Natural Language Processing, RBMT, SMT, NMT, OpenNMT
- Invited by Dongguk University, Department of English Language and Literature professor. Myung-Kwan Park
- Lecture materials: https://github.com/Parkchanjun/Dongguk_AI_NLP_MachineTranslation

South Korea

January. 2020

SKC

INSTRUCTOR

- Deep Learning and Natural Language Processing Course
- Introduction of Text Preprocessing and NLP for Linguistics
- Introduction of Word Embedding
- Introduction of Machine Translation
- Lecture materials(Text Preprocessing): https://github.com/Parkchanjun/SKC_Text_Preprocessing
- Lecture materials(Introduction of Machine Translation): https://github.com/Parkchanjun/SKC_MachineTranslation

South Korea

Oct. 2019 - Nov. 2019

Naver Tech Talk

PRESENTER

- Machine Translation for Everyone
- Invited by Sung Kim who are the Head of Naver Clova A.I
- Lecture videos: <https://tv.naver.com/v/9906991>

South Korea

Aug. 2019

SK T Academy

INSTRUCTOR

- Machine Translation for Everyone
- Invited by SK Planet
- Introduction to Natural Language Processing, RBMT, SMT, NMT, OpenNMT
- Lecture videos: <https://tacademy.skplanet.com/live/player/onlineLectureDetail.action?seq=165>

South Korea

Aug. 2019

Sooyoungro Church

INSTRUCTOR

- Sooyoungro Church Dream Bowl School Instructor (Free Service)
- Basic English Grammar and English Conversation Education

South Korea

Aug. 2017 - Dec. 2017

Korea Press Foundation

INSTRUCTOR

- Reading mentoring

South Korea

Aug. 2015 - Dec. 2015

All Publications

PREPRINTS

TOP CONFERENCE

INTERNATIONAL JOURNAL

INTERNATIONAL CONFERENCE

DOMESTIC JOURNAL

DOMESTIC CONFERENCE

BOOK CHAPTERS

Operate Naver's natural language processing blog

BLOGGER

- URL: <https://blog.naver.com/bcj1210>
- Linguistics for natural language processing
- Natural Language Processing (NLP)
- Machine Translation for Everyone
- Business, strategy, and planning for everyone
- Natural language processing with NLTK
- Software engineering
- Today's CCM

South Korea

2016.Dec-Present

OpenNMT Forum Members

MEMBERS

- Active as OpenNMT Forum Members
- Develop and share OpenNMT Pytorch, Tensorflow, Library Tutorial

South Korea

2019.July-2019.July

ETRI's machine translation quality evaluator

ESTIMATOR

- Quality evaluation of machine translation hosted by ETRI

South Korea

2019.October-2019.December

Completion of basic curriculum for natural language processing and language intelligence

STUDENT

- Korea University Human-Inspired AI & Computing Research Center

South Korea

2019.July-2019.July

Completed the natural language processing tutorial (2016-2019)

STUDENT

- Special Interest Group of Human and Cognitive Language Technology

South Korea

2016 - 2019

Reporter, Korea Electronics and Telecommunications Research Institute (ETRI)

REPORTER

- Served as 3rd and 4th reporters at the Korea Electronics and Telecommunications Research Institute

South Korea

2016.Feb-2018.Feb

Flitto Supporters

SUPPORTERS

- The first plito supporters

South Korea

2017.Dec-2017.Dec

Journalist of the Ministry of Science and ICT

REPORTER

- 8 articles published
- Visit Exobrain Quiz Show
- URL: <https://blog.naver.com/bcj1210/221147196551>

South Korea

2016.Feb-2017.Feb

Planning a video lecture for Natural Language Processing Basic at BUFS

PLANNER

- URL: <https://bit.ly/2NmFJV1>

South Korea

2016.July-2016.July

BIT Computer Advanced Course Completion

STUDENT

- Bet Advanced Course Completion and Bet President Award
- Bit project level 2 license
- Development of intelligent English learning APP

South Korea

2016.January-2016.August

BIT Computer Short Course Completion

STUDENT

- Short-term course completion (C, JAVA, Data Structure)

South Korea

2015.June-2015.August