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 Models (LLM). He is currently working as a Principal Research Engineer & Technical Leader at the Upstage LLM Team. His current interests include forming an LLM-based ecosystem, and he has initiated projects such as SOLAR, Open Ko-LLM Leaderboard, Dataverse, Evalverse and Up 1 Trillion Token Club. In 2023, he received his Ph.D. from Korea University, supervised by Professor Heuiseok Lim, for his work on "Data-Centric Neural Machine Translation". From 2018 to 2019, he worked at SYSTRAN as a Research Engineer. Chanjun is the founder and chief scientist 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 and Publication Chair for the DMLR Workshop. He has published more than 190 papers in the field of NLP. Finally, he is selected for Forbes 30 Under 30 Korea in the SCIENCE / SW field.

My research philosophy is anchored in service-driven research, focusing on bridging the divide between foundational natural language processing (NLP) theories and their practical, real-world applications. I am committed to developing straightforward yet effective research methodologies that deliver immediate and tangible benefits to the industry, while also advancing our fundamental understanding of NLP. By emphasizing service-oriented research, I aim to address the industry pressing needs with innovative solutions that are both accessible and impactful. My recent primary interests lie in the development of efficient, purpose-trained Large Language Models (LLMs), with particular focus on: (1) fundamental capabilities of LLMs; (2) evaluation of LLMs; (3) foundational NLP research addressing recency and truthfulness in LLMs; (4) cross-lingual NLP; and (5) multidisciplinary NLP research that bridges disciplines to advance NLP applications. Through these avenues, I strive to contribute to both the academic and practical advancements in the field of NLP.

- Large Language Model (LLM)

- Data-centric AI

- Machine Translation

- Automatic Speech Recognition

Education

Korea University *Seoul, South Korea*

PH.D IN COMPUTER SCIENCE AND ENGINEERING, GPA: 4.44 / 4.5 *Sep. 2019 - Aug.2023*
- **Advisor: Heuiseok Lim**
- **Natural Language Processing & Artificial Intelligence Lab:** `http://nlp.korea.ac.kr/`
- **Founder and Leader of KU-NMT Group:** `https://kunmt.org/`

Busan University of Foreign Studies (BUFS) *Busan, South Korea*

B.S. IN NATURAL LANGUAGE PROCESSING & CREATIVE CONVERGENCE, GPA: 4.12 / 4.5 *Mar. 2015 - February. 2019*
- **Advisor: Pummo Ryu**
- **Full Scholarship**

Professional Experience

Upstage *Seoul, South Korea*

PRINCIPAL RESEARCH ENGINEER & TECHNICAL LEADER *March. 2022 - Present*
- About Upstage: `https://www.upstage.ai/`
- Principal Research Engineer & Technical Leader at Large Language Model (LLM) Team
- Head of LLM related Data and Evaluation (Leading Data-Centric LLM Team)
- Director and Founder of the Open Ko-LLM Leaderboard: `https://huggingface.co/spaces/upstage/open-ko-llm-leaderboard`
- Director and Founder of the Up 1 Trillion Token Club: `https://en.upstage.ai/1tt`
- Director and Founder of the Dataverse: `https://github.com/UpstageAI/dataverse/`
- Director and Founder of the Evalverse: `https://github.com/UpstageAI/evalverse`
- Extensive experience in leading multiple private LLM projects across various fields such as healthcare, Law, Finance, e-commerce, mathematics, and journalism
- Establishment of evaluation system to develop LLM Application (AskUp Biz)
- Participate in the Product Search Pack, Document AI Pack, and RecSys Pack
- Design and Development of DMOps (Data Management Operation and Recipes). Data-Centric AI Research
- Semantic search project with AMOREPACIFIC & LG U+

SYSTRAN *Seoul, South Korea*

RESEARCH ENGINEER *Sep. 2018 - July. 2019*

- 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, CJ Logistics, Samsung SDS speech Recognition Solution 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)
- Reason for leaving: Graduate School Admission

SYSTRAN *Seoul, South Korea*

RESEARCH INTERN *June. 2018 - August. 2018*

- 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)

Academic Services

Data-centric Machine Learning Research (DMLR) *Anywhere on Earth*

PUBLICATION CHAIR *September.2023-Present*

- DMLR: `https://dmlr.ai/`
- Hosted at ICLR 2024

Widening NLP (WiNLP) Workshop *Anywhere on Earth*

PROGRAM CHAIR *March.2023-Present*

- WiNLP: `https://www.winlp.org/`
- Hosted at EMNLP 2023, EMNLP 2024

COLING 2022 *South Korea*

VIRTUAL SOCIAL CHAIR *Sep.2021-Present*

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

Program Committee *South Korea*

PROGRAM COMMITTEE *January.2021-Present*

- Reviewer of ACL Rolling Review (ACL, EMNLP, NAACL, EACL etc.) (2022 Current), NeurIPS 2024 Track Datasets and Benchmarks, ACL 2024-ML4AL, NAACL 2024-NLP+CSS, NAACL 2024-Insights, NAACL 2024-BEA, ICML 2023-Data-centric Machine Learning Research (DMLR), ACL 2023-Industry Track, ACL 2023-BEA, ACL 2022-Insights, NAACL 2022-Industry Track

Reviewer *South Korea*

REVIEWER *January.2021-Present*

- ICLR 2024 Workshop Proposal Reviewing, ETRI Journal, IEEE Access, International Journal of Human-Computer Interaction, Cybernetics and Systems, Applied Sciences, Electronics, Entropy, Connection Science

External Activities

AI Ethics and Trustworthiness Forum (AI 윤리·신뢰성 포럼 위원) - AI 전략최고위협의회 윤리·안전분과 *South Korea*

COMMITTEE MEMBER *April.2024-Present*

- News: `https://www.edaily.co.kr/news/read?newsId=02758486638857432&mediaCodeNo=257&OutLnkChk=Y`

National Institute of Korean Language (국립국어원 AI 말평) *South Korea*

ADVISOR *May.2024-Present*

- Consulting related to the evaluation of Korean LLM

Information Technology Committee of the Republic of Korea Army (육군발전자문위 정보화분과 자문위원) *South Korea*

ADVISOR *May.2024-April.2025*

- News: `https://www.edaily.co.kr/news/read?newsId=02758486638857432&mediaCodeNo=257&OutLnkChk=Y`

AI Privacy Policy Council (AI 프라이버시 민·관 정책협의회) *South Korea*

Committee Member *October.2023-Present*

- News: `https://www.youtube.com/watch?v=Z1GEsgLKzc4`

Ministry of Food and Drug Safety *South Korea*

Advisor *December.2023-Present*

- Homepage: `https://www.mfds.go.kr/index.do`

Yonsei University-Institute of Language and Information Studies *South Korea*

Advisor *December.2023*

- Homepage: `https://ilis2.yonsei.ac.kr/ilis/index.do`
- Invited by Professor Hansaem Kim

Software Policy & Research Institute *South Korea*

Advisor *August.2023*

- Homepage: `https://spri.kr/`

Selected Publications

Open Ko-LLM Leaderboard: Evaluating Large Language Models in Korean with Ko-H5 Benchmark
Chanjun Park, Hyeonwoo Kim, Dahyun Kim, SeongHwan Cho, Sanghoon Kim, Sukyung Lee, Yungi Kim, Hwalsuk Lee
ACL 2024

Length-aware Byte Pair Encoding for Mitigating Over-segmentation in Korean Machine Translation
Jungseob Lee, Hyeonseok Moon, Seungjun Lee, Sugyeong Eo, **Chanjun Park**, Hyunwoong Ko, Jaehyung Seo, Seungyoon Lee, Heuiseok Lim, **(Corrsponding Author)**
ACL 2024-Findings

KoCommonGEN v2: A Benchmark for Navigating Korean Commonsense Reasoning Challenges in Large Language Models
Jaehyung Seo, Jaewook Lee, **Chanjun Park**, SeongTae Hong, Seungjun Lee, Heuiseok Lim
ACL 2024-Findings

Hyper-BTS Dataset: Scalability and Enhanced Analysis of Back TranScription (BTS) for ASR Post-Processing
Chanjun Park, Jaehyung Seo, Seolhwa Lee, Junyoung Son, Hyeonseok Moon, Sugyeong Eo, Chanhee Lee, Heui-Seok Lim
EACL 2024-Findings, 2024

Generative Interpretation: Toward Human-Like Evaluation for Educational Question-Answer Pair Generation
Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heui-Seok Lim
EACL 2024-Findings, 2024

Solar 10.7 b: Scaling large language models with simple yet effective depth up-scaling
Sanghoon, Lee, Dahyun Kim, **Chanjun Park**, Wonsung Kim, Wonho Song, Yunsu Kim, Hyeonwoo Kim, Yungi Kim, Hyeonju Lee, Jihoo Kim, Changbae Ahn, Seonghoon Yang, Sukyung Lee, Hyunbyung Park, Gyoungjin Gim, Mikyoung Cha, Hwalsuk Lee, Sunghun Kim, **(Equal Contribution (First Co-Author), (Corrsponding Author))**
NAACL 2024 - Industry Track (2024). 2024

Informative Evidence-guided Prompt-based Fine-tuning for English-Korean Critical Error Detection
Dahyun Jung, Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
IJCNLP-AACL 2023, 2023

KEBAP: Korean Error Explainable Benchmark Dataset for ASR and Post-processing
Seonmin Koo, **Chanjun Park**, Jinsung Kim, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim, **(Equal Contribution (First Co-Author))**
EMNLP 2023, 2023

PEEP-Talk: A Situational Dialogue-based Chatbot for English Education
Seugnjun Lee, Yoonna Jang, **Chanjun Park**, Jungseob Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Seounghoon Lee, Bernardo Yahya, Heui-Seok Lim
ACL 2023-System Demonstrations, 2023

Improving Formality-Sensitive Machine Translation using Data-Centric Approaches and Prompt Engineering
Seugnjun Lee, Hyeonseok Moon, **Chanjun Park**, Heui-Seok Lim
ACL 2023 - IWSLT 2023, 2023

Doubts on the reliability of parallel corpus filtering
Hyeonseok Moon, **Chanjun Park**, Seonmin Koo, Jungseob Lee, Seungjun Lee, Jaehyung Seo, Sugyeong Eo, Yoonna Jang, Hyunjoong Kim, Hyoung-gyu Lee
Expert Systems with Applications (2023) p. 120962. Elsevier, 2023

CHEF in the Language Kitchen: A Generative Data Augmentation Leveraging Korean Morpheme Ingredients
Jaehyung Seo, Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, **Chanjun Park**, Heui-Seok Lim
EMNLP 2023, 2023

FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue
Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuiseok Lim
LREC 2022, 2022, Marseille, France

PicTalky: Augmentative and Alternative Communication for Language Developmental Disabilities

Chanjun Park, Yoonna Jang, Seolhwa Lee, Jaehyung Seo, Kisu Yang, Heuiseok Lim
AACL-IJCNLP 2022: System Demonstrations, 2022, Taipei, Taiwan

Priming Ancient Korean Neural Machine Translation
Chanjun Park, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim
LREC 2022, 2022, Marseille, France

QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Gyeongmin Kim, Jungseob Lee, Heui-Seok Lim
COLING 2022, 2022

KU X Upstage's Submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
WMT 2022 - EMNLP 2022 - Proceedings of the Seventh Conference on Machine Translation, 2022, Abu Dhabi

Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing
Hyeonseok Moon, **Chanjun Park**, Seolhwa Lee, Jaehyung Seo, Jungseob Lee, Sugyeong Eo, Heuiseok Lim
LREC 2022, 2022, Marseille, France

A Dog Is Passing Over The Jet? A Text-Generation Dataset for Korean Commonsense Reasoning and Evaluation
Jaehyung Seo, Seounghoon Lee, **Chanjun Park**, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuiseok Lim
NAACL 2022-Findings, 2022, Seattle, United States

PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge
Jaehyung Seo, Dongsuk Oh, Sugyeong Eo, **Chanjun Park**, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuiseok Lim
Knowledge-Based Systems (2022) p. 109861. Elsevier, 2022

Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification
Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim
NAACL 2021: Industry Papers, 2021

Selected Research Experience

Solar
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
June. 2023. - Present

- Role of Principal Research Engineer & Tech Leader in LLM Team
- We introduce the first 10.7 billion (B) parameter model, SOLAR-10.7B (Specialized and Optimized Llm and Applications with Reliability). It's compact, yet remarkably powerful, and demonstrates unparalleled state-of-the-art performance in models with parameters under 30B.
- Achieving Global No. 1 on the Open LLM Leaderboard
- URL: `https://huggingface.co/upstage/SOLAR-10.7B-v1.0`
- Paper: `https://arxiv.org/abs/2312.15166`
- Press: `https://www.yna.co.kr/view/AKR20231214066100017?input=1195m`
- Blog: `https://www.upstage.ai/feed/product/solarmini-performance-report`
- API (SaaS Console): `https://console.upstage.ai/`
- Poe: `https://poe.com/Solar-Mini`
- AWS JumpStart: `https://bitly.ws/3gyjc`
- LangChain: `https://python.langchain.com/docs/integrations/chat/upstage/`
- **Product Lineup**: Solar Chat, Solar Chat Document, Solar Embedding, Solar Translation (Ko<->En), Solar Summarization, Solar, Groundedness Check, Solar Function Call

(Project) Development of technologies for sustaining LLMs and incorporating recency for up-to-date relevance
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
April. 2024. - Present

- Conducting fundamental research and developing benchmarks related to the recency of LLMs.
- Conducting joint research with ETRI, NC Soft, Seoul National University, and KAIST
- Skills: LLM, NLP, Data

(Project) Jasmine Technology Solution Private Thai LLM Project
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
April. 2024. - Present

- Data and evaluation responsibility roles related to Private Thai LLM
- Deployment of Enterprise-Exclusive Thai LLM
- Skills: LLM, NLP, Data

(Project) Law&Company Private LLM Project
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
April. 2024. - Present

- Data and evaluation responsibility roles related to Private Law LLM
- Deployment of Enterprise-Exclusive Law LLM
- Press: `https://www.epnc.co.kr/news/articleView.html?idxno=241298`
- Skills: LLM, NLP, Data

(Project) DeepLearning.AI
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
January. 2024. - June.2024

- Development and planning of lecture content related to pretraining large language models: `https://www.deeplearning.ai/short-courses/pretraining-llms/`
- Hosted by Andrew Ng
- Skills: LLM, NLP, Data

Evalverse
Seoul, South Korea
Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
March. 2024. - Present

- Open-Source Project for LLM Evaluation
- URL: `https://github.com/UpstageAI/evalverse/tree/main`
- Director and Founder of the Evalverse
- Skills: LLM, NLP, Data

Dataverse
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
March. 2024. - Present

- Open-Source Project for Data Extract, Transform, Load (ETL) Related to LLM
- URL: `https://github.com/UpstageAI/dataverse/`
- Press: `https://www.aitimes.com/news/articleView.html?idxno=157776`
- Director and Founder of the Dataverse
- Skills: LLM, NLP, Data

Open Ko-LLM LeaderBoard
Seoul, South Korea
Pᴿᴵɴᴄᴵᴾᴬᴸ Rᴇsᴇᴬʀᴄʜᴇʀ ɪɴ Uᴘsᴛᴀɢᴇ
August. 2023. - Present

- The Open Ko-LLM Leaderboard objectively evaluates the performance of Korean Large Language Model (LLM)
- URL: `https://huggingface.co/spaces/upstage/open-ko-llm-leaderboard`
- Press: `https://www.aitimes.kr/news/articleView.html?idxno=28851`
- HuggingFace Blog: `https://huggingface.co/blog/leaderboard-upstage`
- Director and Founder of the Open Ko-LLM Leaderboard
- Skills: LLM, NLP, Data

Up 1T Token Club

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

August. 2023. - Present

- Role of overall responsibility for the 'Up 1T Token Club'. Through this, I aimed to share models developed using Korean data, thereby expanding accessibility to AI technology. In other words, I aspire to spearhead a national initiative that moves beyond just sharing models to fostering a culture of sharing data.
- URL: `https://en.upstage.ai/1tt`
- Press: `https://www.seoul.co.kr/news/newsView.php?id=20230814500068`
- Director and Founder of the Up 1 Trillion Token Club
- Skills: LLM, NLP, Data

(Project) Connect Wave Private LLM Project

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

September. 2023. - April. 2024

- Data and evaluation responsibility roles related to Private E-commerce LLM
- Deployment of Enterprise-Exclusive E-commerce LLM
- Press: `https://newsis.com/view/?id=NISX20230911_0002445129&cID=10406&pID=13100`
- Skills: LLM, NLP, Data

(Project) Math LLM Project

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

September. 2023. - December. 2023.

- Technical PM roles
- Achieving State of the Art (SOTA) in GSM8K and Math Benchmark
- Collaboration Underway Among QANDA, KT, and Upstage
- Press: `https://n.news.naver.com/mnews/article/001/0014430469?sid=105`
- Mata Official Blog: `https://ai.meta.com/blog/llama-2-mathgpt-mathpresso-qanda-upstage-open-source-llm/`
- Paper: `https://arxiv.org/abs/2404.03887`
- Skills: LLM, NLP, Data

(Project) BIG KINDS LLM Project

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

September. 2023. - December. 2023.

- Data and evaluation responsibility roles related to Private LLM
- Deployment of Enterprise-Exclusive LLM for Press Field
- Skills: LLM, NLP, Data

(Project) Medical LLM Project

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

July. 2023. - December. 2023.

- Data and evaluation responsibility roles related to Private LLM
- Exploration of Medical Use Cases using LLM
- Skills: LLM, NLP, Data

(Project) Hanwha Life Insurance LLM Application Project

Seoul, South Korea

PRINCIPAL RESEARCHER IN UPSTAGE

July. 2023. - November. 2023.

- Development of an Insurance-Specialized Conversational QA System based on LLM
- Responsible for designing LLM evaluation
- Skills: LLM, NLP, Data

Large Language Model (LLM) TF

Seoul, South Korea

RESEARCHER IN UPSTAGE

June. 2023. - July. 2023.

- Research for Large Language Model (LLM)
- Participate Hugging Face Open LLM Leaderboard
- As of July 20, 2023, achieved first place worldwide
- Press: `https://www.aitimes.com/news/articleView.html?idxno=152958`
- Skills: LLM, NLP, Data

AskUp Biz

Seoul, South Korea

RESEARCHER IN UPSTAGE

June. 2023. - Present

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

DMOps for Document AI

Seoul, South Korea

RESEARCHER IN UPSTAGE

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 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

RESEARCHER IN UPSTAGE

Seoul, South Korea

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)

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

RESEARCHER IN ELEUTHER AI

Seoul, South Korea

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) Hyundai Motor - Korean-Prompt-based Engineering for Few-shot Research

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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) NAVER PAPAGO - Parallel Corpus Filtering and Mining

RESEARCHER IN KOREA UNIV

Seoul, South Korea

December. 2021. - June. 2022.

- Research for Static and Dynamic Parallel Corpus Filtering and Mining
- Human-inspired curriculum learning based NMT
- Publication of research paper in Expert Systems With Applications (SCIE): `bit.ly/3K8VCM2`
- Supported by NAVER PAPAGO
- Skills: Neural Machine Translation, Parallel corpus filtering, Pytorch, Fairseq

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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

(Project) National Security Research Institute - Domain Specialized Neural Machine Translation (Security)

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

April. 2020. - July. 2023

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

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

September. 2019. - August. 2022

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

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

RESEARCHER IN SYSTRAN

Seoul, South Korea

June. 2018. - July.2019

- 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 Groud Test (FGT) progress and software maintenance
- Skills: JAVA, JAVA FX, Kaldi

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

RESEARCHER IN SYSTRAN

Seoul, South Korea

April. 2019. - April.2019

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

LCloud-API Platform Service Development

RESEARCHER IN SYSTRAN

Seoul, South Korea

January. 2019. - June.2019

- Planning and development of natural language processing API Platform Service
- Gravitee 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: Gravitee IO,Microservices architecture(MSA), IT Infrastructure, Mockup, Marketing, Strategic Planning, Technology management, Amazon AWS

Speech recognition transcription tool development

RESEARCHER IN SYSTRAN

Seoul, South Korea

January. 2019. - July. 2019

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

(Project) CJ Logistics Speech Recognition Agent

RESEARCHER IN SYSTRAN

Seoul, South Korea

Aug. 2018. - July. 2019

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

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

RESEARCHER IN BUFS

Seoul, South Korea

Sep. 2016. - June. 2018

- 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)

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

RESEARCHER IN BUFS

Seoul, South Korea

Sep. 2017. - June. 2018

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

(Project) Local industry development project in connection with public institutions   *Seoul, South Korea*

RESEARCHER IN BUFS   *Sep. 2016. - Aug. 2017*

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

(Project) Basic Research on Semantic Frame Generation for Short Questions (ExoBrain)   *Seoul, South Korea*

RESEARCHER IN BUFS   *Oct. 2015. - Feb. 2016*

- Exobrain Project

(Project) Local industry development project in connection with public institutions   *Seoul, South Korea*

RESEARCHER IN BUFS   *Sep. 2016. - Aug. 2017*

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

(Project) Basic Research on Semantic Frame Generation for Short Questions (ExoBrain)   *Seoul, South Korea*

RESEARCHER IN BUFS   *Oct. 2015. - Feb. 2016*

- Exobrain Project

Honors & Awards

2023.10	**Forbes 30 Under 30 Korea**, Forbes	*South Korea*
2023.10	**Best Paper Award (Language Model 2 Section), 2023**, The 35th Annual Conference on Human & Cognitive Language Technology (HCLT2023)	*South Korea*
2023.10	**Best Paper Award (Data Bias & Ethics Section), 2023**, The 35th Annual Conference on Human & Cognitive Language Technology (HCLT2023)	*South Korea*
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*

Skills

Programming Skills	Python, JAVA, C / C++
Machine Learning & Deep Learning	Pytorch, Pytorch lightning, Keras, Tensorflow, Huggingface, OpenNMT, Fairseq, Kaldi, FAISS
Natural Language Processing	Large Language Model, Data-Centric AI, Machine Translation, Speech Recognition, Information Retrieval
Product & Project Manager	MS Office, Notion, Slack, Pitch, Linear, Whimsical, Recruitee
Software Skills	Linux, AWS, Docker, 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	**METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION,** Granted Patent (10-2409667)	*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 TRANSLATING ANCIENT KOREAN USING MACHINE TRANSLATION,** Granted Patent (10-2425922)	*South Korea*
2020.09	**APPARATUS AND METHOD FOR OUTPUTTING IMAGE CORRESPONDING TO LANGUAGE,** Granted Patent (10-2476497)	*South Korea*
2021.05	**Method For Generating Training Data And Method For Post-Processing Of Speech Recognition Using The Same,** Granted Patent (10-2557810), PCT application	*South Korea*
2021.11	**DEVICE AND METHOD FOR GENERATING OF TRAINING DATA FOR QUALITY ESTIMATION IN MACHINE TRANSLATION,** Granted Patent (10-2593447)	*South Korea*
2021.09	**APPRATUS FOR CORPUS PROCESSING, APPARATUS AND METHOD AND MATHINE TRANSLATION,** Granted Patent (10-2574167)	*South Korea*
2021.03	**Improving speech recognition performance using TTS in domain-specific environment,** 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	**DEVICE AND METHOD FOR GENERATING OPTIMAL TRANSLATION SUBTITLE USING QUALITY ESTIMATION,** Apply for a patent	*South Korea*
2021.09	**METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION,** PCT International Application	*South Korea*

Teaching Experience

Fast Campus   *South Korea*
INSTRUCTOR   *August. 2023-Present*
- Language Model to Large Language Model (LM to LLM)
- Invited by Fast Campus
- URL: `https://fastcampus.co.kr/b2g_kdigitaltraining_ai`

Fast Campus   *South Korea*
INSTRUCTOR   *August. 2023-Present*
- Natural Language Processing (NLP) Basic
- Invited by Fast Campus
- URL: `https://fastcampus.co.kr/b2g_kdigitaltraining_ai`

Fast Campus   *South Korea*
INSTRUCTOR   *August. 2023-Present*
- Learning ChatGPT Utilization and Service Construction with AskUP
- Invited by Fast Campus
- URL: `https://fastcampus.co.kr/data_online_askup`

Upstage   *South Korea*
INSTRUCTOR   *July. 2023-Present*
- Course Name: Finance Specialized Large Language Model for Everyone
- Upstage Online Course

BoostCamp - NAVER Connect Foundation   *South Korea*
MASTER (INSTRUCTOR)   *February. 2023-Present*
- Course Name: Data-centric NLP
- HomePage: `https://boostcamp.connect.or.kr/program_ai.html`

Korea University   *South Korea*
TEACHING ASSISTANT   *Autumn 2021*
- (BDC101) Introduction to Natural Language Processing in Big Data

Korea University   *South Korea*
HEAD TEACHING ASSISTANT   *Autumn 2020*
- (BDC101) Introduction to Natural Language Processing in Big Data
- (DFE610) Natural Language Processing for Digital Finance Engineering

Korea University *South Korea*

Teaching Assistant *Spring 2020*

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

Invited Talk

KCC 2024 *South Korea*

Presenter *June. 2024*

- SOLAR: The Next Frontier in Large Language Models by Upstage and its Ecosystem
- Invited by KCC 2024 Organizer

Kyungpook National University *South Korea*

Presenter *June. 2024*

- The Ecosystem of LLMs from a Real-World Perspective
- Invited by Professor Dongsuk Oh

Applied ML: LLMs and Knowledge Graphs - Tokyo *South Korea*

Presenter *May. 2024*

- The Ecosystem of LLMs from a Real-World Perspective
- Invited by Kaitlan Melone

Chungnam National University *South Korea*

Presenter *May. 2024*

- Data and Evaluation Methods for Trustworthy AI
- Invited by Professor Sangkeun Jung

AI Safety Compass Conference 2024 *South Korea*

Presenter *May. 2024*

- Data and Evaluation Methods for Trustworthy AI
- Invited by Team Cookie
- URL: `https://2024asc.oopy.io/`

Chonbuk National University *South Korea*

Presenter *May. 2024*

- SOLAR: The Next Frontier in Large Language Models by Upstage and its Ecosystem
- Invited by Professor Hyun-Je Song

TTA *South Korea*

Presenter *May. 2024*

- LLM Evaluation and its Ecosystem
- Invited by TTA

ETRI *South Korea*

Presenter *May. 2024*

- SOLAR: The Next Frontier in Large Language Models by Upstage and its Ecosystem
- Invited by Dr. Kyoungman Bae

AI Tech 2024 - AI Frontier for AI Era *South Korea*

Presenter *May. 2024*

- SOLAR: The Next Frontier in Large Language Models by Upstage and its Ecosystem
- Invited by Professor AI Expo Korea
- URL: `https://dubiz.co.kr/Event/249`

No code · Low code Hyper-automation Conference 2024 *South Korea*

Presenter *May. 2024*

- The Ecosystem of LLMs from a Real-World Perspective
- Invited by Professor ZDNET Korea
- URL: `https://jco.zdnet.co.kr/conference/hyper2024`

Hallym University *South Korea*

Presenter *April. 2024*

- SOLAR: The Next Frontier in Large Language Models by Upstage and its Ecosystem
- Invited by Professor Yu-Seop Kim

Korean Society of Artificial Intelligence in Medicine

PRESENTER

South Korea

April. 2024

- Current state of Large-Scale Generative AI Development
- Invited by Korean Society of Artificial Intelligence in Medicine

NIA

PRESENTER

South Korea

April. 2024

- Open Ko-LLM Leaderboard and ites Ecosystem
- Invited by NIA

Korea University College of Medicine - Intelligent Medical Data Lab

PRESENTER

South Korea

February. 2024

- SOLAR: The Next Frontier in Large Language Models by Upstage
- Invited by Korea University College of Medicine

Fast Campus

PRESENTER

South Korea

February. 2024

- Job Employment Special Lecture
- Invited by Fast Campus

FuriosaAI

PRESENTER

South Korea

February. 2024

- SOLAR: The Next Frontier in Large Language Models by Upstage
- Invited by FuriosaAI

The University of Tokyo - Center for Data-Driven Discovery

PRESENTER

South Korea

January. 2024

- SOLAR: The Next Frontier in Large Language Models by Upstage
- Invited by Professor Kyunghyun Cho (`https://kyunghyuncho.me/`)

Defense Agency for Technology and Quality (DTaQ)

PRESENTER

South Korea

January. 2024

- Current State of Artificial Intelligence (AI) and Its Application Strategies in the Defense Sector
- Invited by Defense Agency for Technology and Quality (DTaQ)

National Library of Korea

PRESENTER

South Korea

January. 2024

- Upstage LLM One Pager
- Invited by National Library of Korea

Ministry of Science and ICT - Data Utilization Council

PRESENTER

South Korea

December. 2023

- Upstage LLM One Pager
- Invited by Ministry of Science and ICT

Korean-English Joint Seminar on Artificial Intelligence (AI) Safety and Reliability

PRESENTER

South Korea

December. 2023

- Upstage Vision for Ethical and Trustworthy Large Language Models
- Invited by TTA
- URL: `https://www.tta.or.kr/tta/selectBbsNttView.do?key=74&bbsNo=105&nttNo=13121&fbclid=IwAR2XW8KEGB6sBCmKk2OHvIuZ9SP9QjiyfQ7dB1XhMy5k1PI31tHOa-6Kcno`

Digital Innovation Forum (Ministry of Culture, Sports and Tourism)

PRESENTER

South Korea

November. 2023

- Generative AI era: Copyright issues and countermeasures
- Invited by Ministry of Culture, Sports and Tourism
- URL: `https://www.culture.go.kr/digicon/sub02/sub04.php`

Kyungpook National University

PRESENTER

South Korea

November. 2023

- Upstage LLM One Pager
- Invited by Kyungpook National University

Future Technology Exchange Conference in Southeast Region

PRESENTER

South Korea

November. 2023

- Upstage LLM One Pager
- Invited by Pukyong National University

Jeju National University
SOUTH KOREA

PRESENTER
November. 2023

- Upstage LLM One Pager
- Invited by Jeju National University

Busan University of Foreign Studies
South Korea

PRESENTER
November. 2023

- From Language Model to Large Language Model
- Invited by Busan University of Foreign Studies

KB Kookmin Bank
South Korea

PRESENTER
November. 2023

- Areas of Generative AI Application in Business
- Invited by KB Kookmin Bank

Human-Inspired AI Research
South Korea

PRESENTER
November. 2023

- Upstage and Private LLM
- Invited by Human-Inspired AI Research

NIA
South Korea

PRESENTER
October. 2023

- Upstage LLM and latest B2B LLM trends
- Invited by NIA (National Information Society Agency)

Woongjin Thinkbig
South Korea

PRESENTER
September. 2023

- Upstage and Private LLM
- Invited by Woongjin Thinkbig

Kyobo Life Insurance
South Korea

PRESENTER
September. 2023

- Upstage and Private LLM
- Invited by Kyobo Life Insurance

HD Korea Shipbuilding & Offshore Engineering
South Korea

PRESENTER
September. 2023

- Upstage and Private LLM
- Invited by HD Korea Shipbuilding & Offshore Engineering

Korea University
South Korea

INSTRUCTOR
September. 2023

- From Language Model to Large Language Model
- Invited by Professor Changhee joo, Heuiseok Lim

HanYang University
South Korea

INSTRUCTOR
September. 2023

- From Language Model to Large Language Model
- URL: `https://gsai.hanyang.ac.kr/front/community/notice/view?id=164286`
- Invited by Professor Taeuk Kim

Korea Institute of Patent Information
South Korea

PRESENTER
September. 2023

- Upstage LLM and latest B2B LLM trends
- Invited by Korea Institute of Patent Information

AGI Town in Seoul
South Korea

PRESENTER
September. 2023

- The use case of solving customer problems using LLM
- URL: `https://bitly.ws/UVNt`
- Invited by AGI Town in Seoul

Fast Campus
South Korea

PRESENTER
September. 2023

- Job Employment Special Lecture
- Invited by Fast Campus

HD Korea Shipbuilding & Offshore Engineering
PRESENTER

South Korea

August. 2023

- Upstage LLM and latest B2B LLM trends
- Invited by HD Korea Shipbuilding & Offshore Engineering

Upstage Webinar
PRESENTER

South Korea

August. 2023

- The Future of Finance/Insurance Transformed by Generative AI
- Invited by Upstage
- URL: `bit.ly/3qomK2U`

Google I/O Extended 2023 Incheon
PRESENTER

South Korea

August. 2023

- From Language Model to Large Language Model
- Invited by Google I/O Extended 2023 Incheon
- URL: `https://festa.io/events/3780`

POSCO RIST
PRESENTER

South Korea

August. 2023

- Data-Centric AI in Real-World
- Invited by POSCO RIST

Google I/O Extended 2023 Seoul
PRESENTER

South Korea

July. 2023

- From Language Model to Large Language Model
- Invited by Google I/O Extended 2023 Seoul
- URL: `https://festa.io/events/3683`

TensorFlow Korea LLM Day
PRESENTER

South Korea

July. 2023

- Language Model to Large Language Model
- Invited by TensorFlow Korea (TFKR)
- URL: `https://festa.io/events/3717`

Jeju National University
INSTRUCTOR

South Korea

July. 2023

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

AI EDucation Alliance Policy lab (AIEDAP)
INSTRUCTOR

South Korea

June. 2023

- Deep Learning Understanding and Practice
- Invited by AIEDAP

KIDA (Korea Institute for Defense Analysis)
PRESENTER

South Korea

May. 2023

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

Upstage
PRESENTER

South Korea

April. 2023

- NLP based Large Language Model for All

AI · DATA SUMMIT 2023
PRESENTER

South Korea

February. 2023

- Real-World Centric AI
- Invited by AI · DATA SUMMIT 2023
- URL: `https://www.youtube.com/watch?v=2ONXHKGV3uI&t=2s`

Sunmoon University
PRESENTER

South Korea

December. 2022

- Real-World Centric AI
- Invited by Sunmoon University

Kyungsung University
PRESENTER

South Korea

August. 2022

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

Hankuk University of Foreign Studies   *South Korea*

Presenter   *July. 2022*

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

Dongguk University   *South Korea*

Presenter   *January. 2022*

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

Busan Social Welfare Development Institute   *South Korea*

Invited Advisor   *July. 2021*

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

LLsoLLu   *South Korea*

Presenter   *March. 2020*

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

NC SOFT   *South Korea*

Presenter   *February. 2020*

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

Dongguk University   *South Korea*

Presenter   *January. 2020*

- 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`

SKC   *South Korea*

Instructor   *Oct. 2019 - Nov.2019*

- 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`

SK T Academy   *South Korea*

Instructor   *Aug. 2019*

- 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`

Naver Tech Talk   *South Korea*

Presenter   *Aug. 2019*

- 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`

All Publications

Preprints

sDPO: Don't Use Your Data All at Once
  Dahyun Kim, Yungi Kim, Wonho Song, Hyeonwoo Kim, Yunsu Kim, Sanghoon Kim, **Chanjun Park**, **(Corrsponding Author)**
  arXiv preprint arXiv:2403.19270 (2024). 2024

SAAS: Solving Ability Amplification Strategy for Enhanced Mathematical Reasoning in Large Language Models
  Hyeonwoo Kim, Gyoungjin Gim, Yungi Kim, Jihoo Kim, Byungju Kim, Wonseok Lee, **Chanjun Park**, **(Corrsponding Author)**
  arXiv preprint arXiv:2404.03887 (2024). 2024

Evalverse: Unified and Accessible Library for Large Language Model Evaluation
  Jihoo Kim, Wonho Song, Dahyun Kim, Yunsu Kim, Yungi Kim, **Chanjun Park**, **(Corrsponding Author)**
  arXiv preprint arXiv:2404.00943 (2024). 2024

Translation of Multifaceted Data without Re-Training of Machine Translation Systems
Hyeonseok Moon, Seungyoon Lee, Seongtae Hong, Seungjun Lee, **Chanjun Park**, Heuiseok Lim
arXiv preprint arXiv:2404.16257 (2024). 2024

Dataverse: Open-Source ETL (Extract, Transform, Load) Pipeline for Large Language Models
Hyunbyung Park, Sukyung Lee, Gyoungjin Gim, Yungi Kim, Dahyun Kim, **Chanjun Park**, **(Corrsponding Author)**
arXiv preprint arXiv:2403.19340 (2024). 2024

ChatLang-8: An LLM-Based Synthetic Data Generation Framework for Grammatical Error Correction
Jeiyoon Park, **Chanjun Park**, Heuiseok Lim, **(Corrsponding Author)**
arXiv e-prints (2024) arXiv–2406. 2024

Enhancing Consistency and Role-Specific Knowledge Capturing by Rebuilding Fictional Character's Persona
Jeiyoon Park, **Chanjun Park**, Heuiseok Lim, **(Corrsponding Author)**
arXiv preprint arXiv:2405.19778 (2024). 2024

Self-Improving-Leaderboard (SIL): A Call for Real-World Centric Natural Language Processing Leaderboards
Chanjun Park, Hyeonseok Moon, Seolhwa Lee, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim
arXiv preprint arXiv:2303.10888 (2023). 2023

Empirical study on BlenderBot 2.0 Errors Analysis in terms of Model, Data and User-Centric Approach
Jungseob Lee, Midan Shim, Suhyune Son, Yujin Kim, **Chanjun Park**, Heuiseok Lim, **(Equal Contribution (First Co-Author))**
arXiv preprint arXiv:2201.03239 (2022). 2022

Language Chameleon: Transformation analysis between languages using Cross-lingual Post-training based on Pre-trained language models
Suhyune Son, **Chanjun Park**, Jungseob Lee, Midan Shim, Chanhee Lee, Yoonna Jang, Jaehyung Seo, Heuiseok Lim, **(Equal Contribution (First Co-Author))**
arXiv preprint arXiv:2209.06422 (2022). 2022

TOP CONFERENCE

Open Ko-LLM Leaderboard: Evaluating Large Language Models in Korean with Ko-H5 Benchmark
Chanjun Park, Hyeonwoo Kim, Dahyun Kim, SeongHwan Cho, Sanghoon Kim, Sukyung Lee, Yungi Kim, Hwalsuk Lee
ACL 2024

Explainable CED: A Dataset for Explainable Critical Error Detection in Machine Translation
Dahyun Jung, Sugyeong Eo, **Chanjun Park**, Heuiseok Lim, **(Corrsponding Author)**
NAACL 2024 - Student Research Workshop

Length-aware Byte Pair Encoding for Mitigating Over-segmentation in Korean Machine Translation
Jungseob Lee, Hyeonseok Moon, Seungjun Lee, Sugyeong Eo, **Chanjun Park**, Hyunwoong Ko, Jaehyung Seo, Seungyoon Lee, Heuiseok Lim, **(Corrsponding Author)**
ACL 2024-Findings

Exploring Inherent Biases in LLMs within Korean Social Context: A Comparative Analysis of ChatGPT and GPT-4
Seungyoon Lee, Dongjun Kim, Dahyun Jung, **Chanjun Park**, Heuiseok Lim, **(Corrsponding Author)**
NAACL 2024 - Student Research Workshop

KoCommonGEN v2: A Benchmark for Navigating Korean Commonsense Reasoning Challenges in Large Language Models
Jaehyung Seo, Jaewook Lee, **Chanjun Park**, SeongTae Hong, Seungjun Lee, Heuiseok Lim
ACL 2024-Findings

Model-Based Data-Centric AI: Bridging the Divide Between Academic Ideals and Industrial Pragmatism
Chanjun Park, Minsoo Khang, Dahyun Kim
ICLR 2024 - Data-centric Machine Learning Research (DMLR) Workshop (2024). 2024

Hyper-BTS Dataset: Scalability and Enhanced Analysis of Back TranScription (BTS) for ASR Post-Processing
Chanjun Park, Jaehyung Seo, Seolhwa Lee, Junyoung Son, Hyeonseok Moon, Sugyeong Eo, Chanhee Lee, Heui-Seok Lim
EACL 2024-Findings, 2024

Detecting Critical Errors Considering Cross-Cultural Factors in English-Korean Translation
Sugyeong Eo, Jungwoo Lim, **Chanjun Park**, DaHyun Jung, Seonmin Koo, Hyeonseok Moon, Jaehyung Seo, Heui-Seok Lim
LREC-COLING 2024, 2024

Leveraging Pre-existing Resources for Data-Efficient Counter-Narrative Generation in Korean
Seungyoon Lee, **Chanjun Park**, DaHyun Jung, Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, Heui-Seok Lim, **(Corrsponding Author)**
LREC-COLING 2024, 2024

Generative Interpretation: Toward Human-Like Evaluation for Educational Question-Answer Pair Generation
Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heui-Seok Lim
EACL 2024-Findings, 2024

Solar 10.7 b: Scaling large language models with simple yet effective depth up-scaling
Sanghoon, Lee, Dahyun Kim, **Chanjun Park**, Wonsung Kim, Wonho Song, Yunsu Kim, Hyeonwoo Kim, Yungi Kim, Hyeonju Lee, Jihoo Kim, Changbae Ahn, Seonghoon Yang, Sukyung Lee, Hyunbyung Park, Gyoungjin Gim, Mikyoung Cha, Hwalsuk Lee, Sunghun Kim, **(Equal Contribution (First Co-Author), (Corrsponding Author))**
NAACL 2024 - Industry Track (2024). 2024

Synthetic Alone: Exploring the Dark Side of Synthetic Data for Grammatical Error Correction
Chanjun Park, Seonmin Koo, Seolhwa Lee, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

DMOps: Data Management Operation and Recipes
Eujeong Choi, **Chanjun Park**, **(Corresponding Author)**
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

Proceedings of the Seventh Widening NLP Workshop (WiNLP 2023)
Bonaventure FP Dossou, Isidora Tourni, Hatem Haddad, Shaily Bhatt, Fatemehsadat Mireshghallah, Sunipa Dev, Tanvi Anand, Weijia Xu, Atnafu Lambebo Tonja, Alfredo Gomez, **Chanjun Park**
EMNLP 2023- Widening NLP Workshop (WiNLP 2023), 2023

Informative Evidence-guided Prompt-based Fine-tuning for English-Korean Critical Error Detection
Dahyun Jung, Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
IJCNLP-AACL 2023, 2023

Knowledge Graph-Augmented Korean Generative Commonsense Reasoning
Dahyun Jung, Jaehyung Seo, Jaewook Lee, **Chanjun Park**, Heuiseok Lim
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

Transcending Traditional Boundaries: Leveraging Inter-Annotator Agreement (IAA) for Enhancing Data Management Operations (DMOps)
Damrin Kim, NamHyeok Kim, **Chanjun Park**, Harksoo Kim, **(Corresponding Author)**
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

Inter-Annotator Agreement in the Wild: Uncovering Its Emerging Roles and Considerations in Real-World Scenarios
NamHyeok Kim, **Chanjun Park**, **(Corresponding Author)**
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

KEBAP: Korean Error Explainable Benchmark Dataset for ASR and Post-processing
Seonmin Koo, **Chanjun Park**, Jinsung Kim, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim, **(Equal Contribution (First Co-Author))**
EMNLP 2023, 2023

Toward Practical Automatic Speech Recognition and Post-Processing: a Call for Explainable Error Benchmark Guideline
Seonmin Koo, **Chanjun Park**, Jinsung Kim, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim, **(Equal Contribution (First Co-Author))**
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

PEEP-Talk: A Situational Dialogue-based Chatbot for English Education
Seugnjun Lee, Yoonna Jang, **Chanjun Park**, Jungseob Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Seounghoon Lee, Bernardo Yahya, Heui-Seok Lim
ACL 2023-System Demonstrations, 2023

Improving Formality-Sensitive Machine Translation using Data-Centric Approaches and Prompt Engineering
Seugnjun Lee, Hyeonseok Moon, **Chanjun Park**, Heui-Seok Lim
ACL 2023 - IWSLT 2023, 2023

Data-Driven Approach for Formality-Sensitive Machine Translation: Language-Specific Handling and Synthetic Data Generation
Seugnjun Lee, Hyeonseok Moon, **Chanjun Park**, Heuiseok Lim
ICML 2023 - Data-centric Machine Learning Research (DMLR) Workshop (2023). 2023

Alternative Speech: Complementary Method to Counter-Narrative for Better Discourse
Seungyoon Lee, Dahyun Jung, **Chanjun Park**, Seolhwa Lee, Heuiseokv Lim
ICDM 2023 - The First Workshop on Data-Centric AI (DCAI) (2023). 2023

CHEF in the Language Kitchen: A Generative Data Augmentation Leveraging Korean Morpheme Ingredients
Jaehyung Seo, Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, **Chanjun Park**, Heui-Seok Lim
EMNLP 2023, 2023

FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue
 Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuiseok Lim
LREC 2022, 2022, Marseille, France

Priming Ancient Korean Neural Machine Translation
 Chanjun Park, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim
LREC 2022, 2022, Marseille, France

QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Gyeongmin Kim, Jungseob Lee, Heui-Seok Lim
COLING 2022, 2022

KU X Upstage's Submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
WMT 2022 - EMNLP 2022 - Proceedings of the Seventh Conference on Machine Translation, 2022, Abu Dhabi

Focus on FoCus: Is FoCus focused on Context, Knowledge and Persona?
SeungYoon Lee, Jungseob Lee, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Jaehyung Seo, Jeongbae Park, Heui-Seok Lim
COLING 2022 - Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge, 2022

A Self-Supervised Automatic Post-Editing Data Generation Tool
Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, SeungJun Lee, Heuiseok Lim
ICML 2022 - DataPerf workshop, 2022 (2022). 2022

Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing
Hyeonseok Moon, **Chanjun Park**, Seolhwa Lee, Jaehyung Seo, Jungseob Lee, Sugyeong Eo, Heuiseok Lim
LREC 2022, 2022, Marseille, France

A Dog Is Passing Over The Jet? A Text-Generation Dataset for Korean Commonsense Reasoning and Evaluation
Jaehyung Seo, Seounghoon Lee, **Chanjun Park**, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuiseok Lim
NAACL 2022-Findings, 2022, Seattle, United States

Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification
Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim
NAACL 2021: Industry Papers, 2021

FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue
Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuiseok Lim
AAAI 2022 - Artificial Intelligence for Education(AI4EDU) (2021). 2021

How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus
Chanjun Park, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim
NeurIPS 2021 - Data-centric AI (DCAI) workshop (2021). 2021

Two Heads are Better than One? Verification of Ensemble Effect in Neural Machine Translation
Chanjun Park, Sungjin Park, Seolhwa Lee, Taesun Whang, Heuiseok Lim
EMNLP 2021 - Proceedings of the Second Workshop on Insights from Negative Results in NLP, 2021, Online and Punta Cana, Dominican Republic

BTS: Back TranScription for Speech-to-Text Post-Processor using Text-to-Speech-to-Text
Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim
ACL 2021 - Proceedings of the 8th Workshop on Asian Translation (WAT2021), 2021, Online

Dealing with the Paradox of Quality Estimation
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim, **(Equal Contribution (First Co-Author))**
MT Summit - Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), 2021, Virtual

A New Tool for Efficiently Generating Quality Estimation Datasets
Sugyeong Eo, Chanjun Park, Jaehyung Seo, Hyeonseok Moon, Heuiseok Lim
NeurIPS 2021 - Data-centric AI (DCAI) workshop, 2021

Syntax-enhanced Dialogue Summarization using Syntax-aware information
Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heui-Seok Lim
NerIPS 2021 16th Women in Machine Learning Workshop, 2021

Towards Syntax-Aware Dialogue Summarization using Multi-task Learning
Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heui-Seok Lim
EMNLP 2021 -Widening NLP (WiNLP2021) workshop, 2021

Automatic Knowledge Augmentation for Generative Commonsense Reasoning
Jaehyung Seo, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim
NeurIPS 2021 - Data-centric AI (DCAI) workshop, 2021

INTERNATIONAL JOURNAL

Towards Harnessing the Most of ChatGPT for Korean Grammatical Error Correction
Chanjun Park, Seonmin Koo, Gyeongmin Kim, Heuiseok Lim
Applied Sciences 14.8 (2024) p. 3195. MDPI, 2024

Analysis of the Effectiveness of Model, Data, and User-Centric Approaches for Chat Application: A Case Study of BlenderBot 2.0
Chanjun Park, Jungseob Lee, Suhyune Son, Kinam Park, Jungsun Jang, Heuiseok Lim
Applied Sciences 14.11 (2024) p. 4821. Multidisciplinary Digital Publishing Institute, 2024

Exploiting Hanja-based Resources in Processing Korean Historic Documents written by Common Literati
Hyeonseok Moon, Myunghoon Kang, Jaehyung Seo, Sugyeong Eo, **Chanjun Park**, Yeongwook Yang, Heuiseok Lim
IEEE Access (2024). IEEE, 2024

Enhancing Machine Translation Quality Estimation via Fine-Grained Error Analysis and Large Language Model
Dahyun Jung, **Chanjun Park**, Sugyeong Eo, Heuiseok Lim
Mathematics 11.19 (2023) p. 4169. Multidisciplinary Digital Publishing Institute, 2023

Uncovering the Risks and Drawbacks Associated with the Use of Synthetic Data for Grammatical Error Correction
Seonmin Koo, **Chanjun Park**, Seolhwa Lee, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim, **(Equal Contribution (First Co-Author))**
IEEE Access (2023). IEEE, 2023

A Survey on Evaluation Metrics for Machine Translation
Seungjun Lee, Jungseob Lee, Hyeonseok Moon, **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, Seonmin Koo, Heuiseok Lim
Mathematics 11.4 (2023) p. 1006. Multidisciplinary Digital Publishing Institute, 2023

Doubts on the reliability of parallel corpus filtering
Hyeonseok Moon, **Chanjun Park**, Seonmin Koo, Jungseob Lee, Seungjun Lee, Jaehyung Seo, Sugyeong Eo, Yoonna Jang, Hyunjoong Kim, Hyoung-gyu Lee
Expert Systems with Applications (2023) p. 120962. Elsevier, 2023

Mimicking Infants' Bilingual Language Acquisition for Domain Specialized Neural Machine Translation
Chanjun Park, Woo-Young Go, Sugyeong Eo, Hyeonseok Moon, Seolhwa Lee, Heuiseok Lim
IEEE Access 10 (2022) pp. 38684–38693. IEEE, 2022

PicTalky: Augmentative and Alternative Communication for Language Developmental Disabilities
Chanjun Park, Yoonna Jang, Seolhwa Lee, Jaehyung Seo, Kisu Yang, Heuiseok Lim

The ASR Post-Processor Performance Challenges of BackTranScription (BTS): Data-Centric and Model-Centric Approaches
 Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Heuiseok Lim
Mathematics 10.19 (2022). 2022

Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC
 Chanjun Park, Midan Shim, Sugyeong Eo, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Heuiseok Lim
Applied Sciences 12.11 (2022) p. 5545. Multidisciplinary Digital Publishing Institute, 2022

Word-level Quality Estimation for Korean-English Neural Machine Translation
Sugyeong Eo,   **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim,   **(Equal Contribution (First Co-Author))**
IEEE Access 10 (2022) pp. 44964–44973. IEEE, 2022

Research on Utilization Strategy of User Engagements in Korean Fake News Detection
Myunghoon Kang, Jaehyung Seo,   **Chanjun Park**, Heuiseok Lim
IEEE Access (2022) pp. 1–1. 2022

AI Student: A Machine Reading Comprehension System for the Korean College Scholastic Ability Test
Gyeongmin Kim, Soomin Lee,   **Chanjun Park**, Jaechoon Jo
Mathematics 10.9 (2022) p. 1486. Multidisciplinary Digital Publishing Institute, 2022

K-NCT: Korean Neural Grammatical Error Correction Gold-Standard Test Set Using Novel Error Type Classification Criteria
Seonmin Koo,   **Chanjun Park**, Jaehyung Seo, Seungjun Lee, Hyeonseok Moon, Jungseob Lee, Heuiseok Lim,   **(Equal Contribution (First Co-Author))**
IEEE Access (2022). IEEE, 2022

BERTOEIC: Solving TOEIC Problems Using Simple and Efficient Data Augmentation Techniques with Pretrained Transformer Encoders
Jeongwoo Lee, Hyeonseok Moon,   **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim
Applied Sciences 12.13 (2022) p. 6686. Multidisciplinary Digital Publishing Institute, 2022

An Automatic Post Editing with Efficient and Simple Data Generation Method
Hyeonseok Moon,   **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim,   **(Equal Contribution (First Co-Author))**
IEEE Access (2022). IEEE, 2022

Return on Advertising Spend Prediction with Task Decomposition-Based LSTM Model
Hyeonseok Moon, Taemin Lee, Jaehyung Seo,   **Chanjun Park**, Sugyeong Eo, Imatitikua D Aiyanyo, Jeongbae Park, Aram So, Kyoungwha Ok, Kinam Park
Mathematics 10.10 (2022) p. 1637. MDPI, 2022

Dense-to-Question and Sparse-to-Answer: Hybrid Retriever System for Industrial Frequently Asked Questions
Jaehyung Seo, Taemin Lee, Hyeonseok Moon,   **Chanjun Park**, Sugyeong Eo, Imatitikua D Aiyanyo, Kinam Park, Aram So, Sungmin Ahn, Jeongbae Park
Mathematics 10.8 (2022) p. 1335. MDPI, 2022

Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners
Jaehyung Seo, Hyeonseok Moon, Chanhee Lee, Sugyeong Eo,   **Chanjun Park**, Jihoon Kim, Changwoo Chun, Heuiseok Lim
IEEE Access 10 (2022) pp. 107587–107597. 2022

PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge
Jaehyung Seo, Dongsuk Oh, Sugyeong Eo,   **Chanjun Park**, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuiseok Lim
Knowledge-Based Systems (2022) p. 109861. Elsevier, 2022

AI for Patents: A Novel yet Effective and Efficient Framework for Patent Analysis
Junyoung Son, Hyeonseok Moon, Jeongwoo Lee, Seolhwa Lee,   **Chanjun Park**, Wonkyung Jung, Heuiseok Lim
IEEE Access 10 (2022) pp. 59205–59218. IEEE, 2022

Comparative Analysis of Current Approaches to Quality Estimation for Neural Machine Translation
Sugyeong Eo,   **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim,   **(Equal Contribution (First Co-Author))**
Applied Sciences 11.14 (2021) p. 6584. Multidisciplinary Digital Publishing Institute, 2021

Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network
Kuekyeng Kim,   **Chanjun Park**, Jaehyung Seo, Heuiseok Lim
IEEE Access (2021). IEEE, 2021

Exploring the Data Efficiency of Cross-Lingual Post-Training in Pretrained Language Models
Chanhee Lee, Kisu Yang, Taesun Whang,   **Chanjun Park**, Andrew Matteson, Heuiseok Lim
Applied Sciences 11.5 (2021) p. 1974. Multidisciplinary Digital Publishing Institute, 2021

Who speaks like a style of Vitamin: Towards Syntax-Aware Dialogue Summarization using Multi-task Learning
Seolhwa Lee, Kisu Yang,   **Chanjun Park**, João Sedoc, Heuiseok Lim
IEEE Access (2021). IEEE, 2021

An Empirical Study on Automatic Post Editing for Neural Machine Translation
Hyeonseok Moon,   **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim,   **(Equal Contribution (First Co-Author))**
IEEE Access 9 (2021) pp. 123754–123763. IEEE, 2021

Variational Reward Estimator Bottleneck: Towards Robust Reward Estimator for Multidomain Task-Oriented Dialogue
Jeiyoon Park, Chanhee Lee,   **Chanjun Park**, Kuekyeng Kim, Heuiseok Lim
Applied Sciences 11.14 (2021) p. 6624. Multidisciplinary Digital Publishing Institute, 2021

Neural Spelling Correction: Translating Incorrect sentences to Correct sentences for Multimedia
 Chanjun Park, Kuekyeng Kim, YeongWook Yang, Minho Kang, HeuiSeok Lim
Multimedia Tools and Applications (2020) pp. 1–18. Springer, 2020

Ancient Korean Neural Machine Translation
　Chanjun Park, Chanhee Lee, YeongWook Yang, Heuiseok Lim
　IEEE Access 8 (2020) pp. 116617–116625. 2020

Comparison of the evaluation metrics for Neural Grammatical Error Correction with Overcorrection
　Chanjun Park, YeongWook Yang, Chanhee Lee, Heuiseok Lim
　IEEE Access 8 (2020) pp. 106264–106272. 2020

Decoding Strategies for Improving Low-Resource Machine Translation
　Chanjun Park, Yeongwook Yang, Kinam Park, Heuiseok Lim
　Electronics 9.10 (2020) p. 1562. Multidisciplinary Digital Publishing Institute, 2020

INTERNATIONAL CONFERENCE

Enhancing Relation Extraction with Triple Representation
　Chanjun Park, Sugyeong Eo, Yoonna Jang, Jungseob Lee, Aram So, HeuiSeok Lim
　International Conference on Information, System and Convergence Applications (ICISCA 2023), 2023

Exploring the Potential of Large Language Model for Korean Grammatical Error Correction
　Seonmin Koo, **Chanjun Park**, Aram So
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2023), 2023

Assessing the Retrieval-based Generation Capabilities of Large Language Models: A Call for a New Benchmark
　Jungseob Lee, Junyoung Son, Taemin Lee, **Chanjun Park**, Myunghoon Kang, Yuna Hur
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2023), 2023

Evaluating the toxicity of ChatGPT for Gender and Race in Korean
　Seungyoon Lee, **Chanjun Park**, Heuiseok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2023), 2023

Hallucination of Large Language Models Derived by the Colloquial Style
　Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, **Chanjun Park**, Heuiseok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2023), 2023

Prompt-based Learning for English-German Critical Translation Error Detection
　Sugyeong Eo, **Chanjun Park**, Heuiseok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

A Dataset for Korean Graph-to-Text Generation
　Dahyun Jung, Lee. Seungyoon, SeungJun Lee, Sugyeong Eo, **Chanjun Park**, Danielle Aiyanyo Imatitikua
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

Masked language modeling-based Korean Data Augmentation Techniques Using Label Correction
　Myunghoon Kang, Jungseob Lee, SeungJun Lee, Hyeonseok Moon, **Chanjun Park**, Aram So
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

Automatic Generation of Learning Data for Korean Speech Recognition Post-Processor
　Seomin Koo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

Fortune Telling via Deep Learning based Generation Model
　Hyeonseok Moon, Jungseob Lee, Jaehyung Seo, Sugyeong Eo, **Chanjun Park**, Jeongbae Park
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

Efficient Way for Constructing Hate Speech-Counter Narrative Dataset
　Lee. Seungyoon, Suhyune Son, Dahyun Jung, **Chanjun Park**, Yuna Hur
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022), 2022

A Study on Jeju Dialect ASR processor using BackTranScription
　Chanjun Park, Jaehyung Seo, Seolhwa Lee, Heonseok Moon, Sugyeong Eo, Yoonna Jang, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Cross-lingual Zero-shot Transfer for Quality Estimation
　Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heonseok Moon, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Proposal of Automatic Essay Scoring Method Using 4-level Natural Language Analysis
　Yujin Kim, **Chanjun Park**, Seolhwa Lee, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Error Type Categorization Criteria for Korean Spelling Correction
　Seonmin Koo, **Chanjun Park**, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

A High-level Machine Reading Comprehension Model
　Jeongwoo Lee, Hyeonseok Moon, **Chanjun Park**, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Comparative study of Korean speech recognition based on SpecAugment and Kaldi
　Seounghoon Lee, **Chanjun Park**, Jaehyung Seo, Gyeongmin Kim, HeuiSeok Lim
　International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Context Detector: Persona Based Natural Conversation using Conversational Agent
　Seungjun Lee, Yoonna Jang, **Chanjun Park**, HeuiSeok Lim

International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Domain Specialized Automatic Post Editing System
Heonseok Moon, **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, HeuiSeok Lim
International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Korean Generative Commonsean Reasoning
Jaehyung Seo, **Chanjun Park**, Heonseok Moon, Sugyeong Eo, HeuiSeok Lim
International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

AI for Construction: Development of communication platform for construction field using Natural Language Processing
Midan Shim, **Chanjun Park**, Yuna Hur, HeuiSeok Lim
International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

Categorization of Common Error Types in the Off-the-shelf ASR systems for Korean
Junyoung Son, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim
International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

A Comparative Study on Korean Relation Extraction with entity position information
Suhyune Son, Yuna Hur, Jungwoo Lim, Midan Shim, **Chanjun Park**, HeuiSeok Lim
International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021

PicTalky: Deep learning-based speech therapy platform for developmental disabilities
 Chanjun Park, Yanghee Kim, Yoonna Jang, Umadevi G.R, HeuiSeok Lim
The International Conference on Big data, IoT, and Cloud Computing, 2020

Domain-specialize Neural Machine Translation Methodology
 Chanjun Park, HeuiSeok Lim
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2020

Toward High Quality Parallel Corpus Using Monolingual Corpus
 Chanjun Park, Youngdae Oh, Jongkeun Choi, Dongpil Kim, HeuiSeok Lim
The 10th International Conference on Convergence Technology, 2020

Data Augmentation Method for Korean Neural Machine Translation
 Chanjun Park, Kinam Park, HeuiSeok Lim
INTERNATIONAL CONFERENCE ON INFORMATION, SYSTEM AND CONVERGENCE APPLICATIONS, 2020

Self- Supervised Korean Spelling Correction via Denoising Transformer
 Chanjun Park, Sungjin Park, HeuiSeok Lim
INTERNATIONAL CONFERENCE ON INFORMATION, SYSTEM AND CONVERGENCE APPLICATIONS, 2020

Can artificial intelligence translate the Annals of the Joseon Dynasty ?
 Chanjun Park, Sungjin Park, Kinam Park, Jaechoon Jo, HeuiSeok Lim
The 10th International Conference on Convergence Technology, 2020

Design Neural Machine Translation Model Combining External Symbolic Knowledge in translation task
Sugyeong Eo, **Chanjun Park**, HeuiSeok Lim
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2020

Free Talking NAO Robot for Children
 Chanjun Park, Jaekue Choi, Sungjin Park, Yanghee Kim, HeuiSeok Lim
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019

Improved Machine Translation Performance Using Two Stage Subword Tokenization
 Chanjun Park, Sora Choi, HeuiSeok Lim
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019

Quality before Quantity: Improved Machine Translation performance using parallel corpus filtering
 Chanjun Park, Seunggyu Han, HeuiSeok Lim
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019

Improved Machine Translation Performance Using Relative Ratio of Original and Synthetic Corpus
 Chanjun Park, Jang Mook Kang
International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019

Domestic Journal

Survey on Large Language Models
 Chanjun Park, Lee. Wonsung, Yungi Kim, Kim. Jihoo, Lee. Hwalsuk
Communications of the Korean Institute of Information Scientists and Engineers (Communications of KIISE) 41.11 (2023) pp. 8–24. 2023

Policy-based performance comparison study of Real-time Simultaneous Translation
Jungseob Lee, Hyeonseok Moon, Jaehyung **Chanjun Park** Seo, Sugyeong Eo, Seungjun Lee, Seonmin Koo, Heuiseok Lim
Journal of the Korea Convergence Society 13.3 (2022) pp. 43–54. Korea Convergence Society, 2022

Error Analysis of Recent Conversational Agent-based Commercialization Education Platform
Seungjun Lee, **Chanjun Park**, Jaehyung Seo, Heuiseok Lim
Journal of the Korea Convergence Society 13.3 (2022) pp. 11–22. Korea Convergence Society, 2022

Deep Learning based Korean Dialect Machine Translation Research Considering Linguistics Features and Service
Sangbeom Lim, **Chanjun Park**, Yeongwook Yang
Journal of the Korea Convergence Society 13.2 (2022) pp. 21–29. Korea Convergence Society, 2022

Korean and Multilingual Language Models Study for Cross-Lingual Post-Training (XPT)
Suhyune Son, **Chanjun Park**, Jungseob Lee, Midan Shim, Chanhee Lee, Heuiseok Lim
Journal of the Korea Convergence Society 13.3 (2022) pp. 77–89. Korea Convergence Society, 2022

A study on performance improvement considering the balance between corpus in Neural Machine Translation
 Chanjun Park, Kinam Park, Sugyeong Eo, Hyeonseok Moon, HeuiSeok Lim
Journal of the Korea Convergence Society 12.5 (2021) pp. 23–29. Korea Convergence Society, 2021

A Study on Verification of Back TranScription(BTS)-based Data Construction
 Chanjun Park, Jaehyung Seo, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim
Journal of the Korea Convergence Society 12.11 (2021) pp. 109–117. Korea Convergence Society, 2021

Research on Subword Tokenization of Korean Neural Machine Translation and Proposal for Tokenization Method to Separate Jong-sung from Syllables
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, HeuiSeok Lim
Journal of the Korea Convergence Society 12.3 (2021) pp. 1–7. Korea Convergence Society, 2021

Research on Recent Quality Estimation
Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, HeuiSeok Lim
Journal of the Korea Convergence Society 12.7 (2021) pp. 37–44. Korea Convergence Society, 2021

Study on Zero-shot based Quality Estimation
Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Hyeonseok Moon, Heuiseok Lim
Journal of the Korea Convergence Society 12.11 (2021) pp. 35–43. Korea Convergence Society, 2021

Classification and analysis of error types for deep learning-based Korean spelling correction
Seonmin Koo, **Chanjun Park**, Aram So, Heuiseok Lim
Journal of the Korea Convergence Society 12.12 (2021) pp. 65–74. Korea Convergence Society, 2021

Empirical study on BlenderBot 2.0's errors analysis in terms of model, data and dialogue
Jungseob Lee, Suhyune Son, Midan Shim, Yujin Kim, **Chanjun Park**, Aram So, Jeongbae Park, Heuiseok Lim
Journal of the Korea Convergence Society 12.12 (2021) pp. 93–106. Korea Convergence Society, 2021

Filter-mBART Based Neural Machine Translation Using Parallel Corpus Filtering
Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, JeongBae Park, HeuiSeok Lim
Journal of the Korea Convergence Society 12.5 (2021) pp. 1–7. Korea Convergence Society, 2021

Recent Automatic Post Editing Research
Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, HeuiSeok Lim
Journal of Digital Convergence 19.7 (2021) pp. 199–208. Journal of Digital Convergence, 2021

The Verification of the Transfer Learning-based Automatic Post Editing Model
Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim
Journal of the Korea Convergence Society 12.10 (2021) pp. 27–35. Korea Convergence Society, 2021

Study on Decoding Strategies in Neural Machine Translation
Jaehyung Seo, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim
Journal of the Korea Convergence Society 12.11 (2021) pp. 69–80. Korea Convergence Society, 2021

Cross-Validated Ensemble Methods in Natural Language Inference
Kisu Yang, Taesun Whang, Dongsuk Oh, **Chanjun Park**, HeuiSeok Lim
Journal of KIISE 48.2 (2021) pp. 154–159. Korean Institute of Information Scientists and Engineers (KIISE), 2021

Neural Machine translation specialized for Coronavirus Disease-19 (COVID-19)
 Chanjun Park, Kyeong-Hee Kim, Ki-Nam Park, Heui-Seok Lim
Journal of the Korea Convergence Society 11.9 (2020) pp. 7–13. Korea Convergence Society, 2020

An AI Service to support communication and language learning for people with developmental disability
 Chanjun Park, Yanghee Kim, Yoonna Jang, Umadevi G.R, HeuiSeok Lim
Journal of the Korea Convergence Society 11.6 (2020) pp. 51–57. 2020

Automatic Post Editing Research
 Chanjun Park, HeuiSeok Lim
Journal of the Korea Convergence Society 11.5 (2020) pp. 1–8. 2020

A Study on the Performance Improvement of Machine Translation Using Public Korean-English Parallel Corpus
Chanjun Park, HeuiSeok Lim
Journal of Digital Convergence 18.6 (2020) pp. 271–277. 2020

Domestic Conference

Prompt Based Machine Translation
 Chanjun Park
Korea Computer Congress 2023, 2023

A Study on Data Quality Estimation Using Inter Annotator Agreement(IAA) in Data Management Operations(DMOps)
Damrin Kim, NamHyeok Kim, **Chanjun Park**, Harksoo Kim
Korea Computer Congress 2023, 2023

Are Vector Size and Response Time Proportional?: Empirical Study of Response Time according to Index Type in Semantic Search
 Chanjun Park, Eujeong Choi, Dayone Suh Suh, Minjeong Kim, Suwon Shin, Juhwan Yoon, Seungwon Jeong, Jaekyoung Bae
Korea Software Congress 2022 (2022) pp. 350–352. 2022

Is Modeling Essential for Model Improvement? Encoder performance improvement through a Data-Centric Approach

Seungwon Jeong, Suwan Kim, **Chanjun Park**
Korea Software Congress 2022 (2022) pp. 377–379. 2022

BackTranScription (BTS)-based Jeju Automatic Speech Recognition Post-processor Research
 Chanjun Park, Jaehyung Seo, Seolhwa Lee, Heonseok Moon, Sugyeong Eo, Yoonna Jang, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Word-level Korean-English Quality Estimation
Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heonseok Moon, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Persona-based Korean Conversational Model
Yoonna Jang, Jungwoo Lim, Yuna Hur, Kisu Yang, **Chanjun Park**, Jaehyung Seo, Seungjun Lee, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Proposal of Automated Essay Scoring Method based on Deep-Learning
Yujin Kim, **Chanjun Park**, Seolhwa Lee, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Error Type Categorization for deep learning-based Korean Spelling Correction
Seonmin Koo, **Chanjun Park**, Heuiseok Lim
Korea Software Congress 2018 (2021) pp. 1279–1281. 2021

Model Training and Data Augmentation Schemes For the High-level Machine Reading Comprehension
Jeongwoo Lee, Hyeonseok Moon, **Chanjun Park**, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Comparative study of Korean speech recognition based on SpecAugment and Kaldi
Seounghoon Lee, **Chanjun Park**, Jaehyung Seo, Gyeongmin Kim, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning (Best Paper Award)
Seungjun Lee, Yoonna Jang, **Chanjun Park**, Minwoo Kim, Bernardo N Yahya, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Deep Learning based Korean Dialect Machine Translation Research
Sangbeom Lim, **Chanjun Park**, Jaechoon Jo, Yeongwook Yang
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Verification of the Domain Specialized Automatic Post Editing Model
Heonseok Moon, **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation (Best Paper Award)
Jaehyung Seo, **Chanjun Park**, Heonseok Moon, Sugyeong Eo, Myunghoon Kang, Seounghoon Lee, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Con-Talky: Information Extraction and Visualization Platform for Communication of Construction Industry
Midan Shim, **Chanjun Park**, Yuna Hur, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Categorization and Analysis of Error Types in the Korean Speech Recognition System (Best Paper Award)
Junyoung Son, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

A Comparative Study on Korean Relation Extraction with entity position information
Suhyune Son, Yuna Hur, Jungwoo Lim, Midan Shim, **Chanjun Park**, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Attention Patterns and Semantics of Korean Language Models
Kisu Yang, Yoonna Jang, Jingwoo Lim, **Chanjun Park**, Hwanseok Jang, HeuiSeok Lim
The 33rd Annual Conference on Human Cognitive Language Technology, 2021

Quality, not Quantity? : Effect of parallel corpus quantity and quality on Neural Machine Translation
 Chanjun Park, Yeonsu Lee, Chanhee Lee, HeuiSeok Lim
The 32st Annual Conference on Human Cognitive Language Technology, 2020

Alleviation of Overcorrection Problem in Neural Korean Spelling Correction
 Chanjun Park, Yeonsu Lee, Kisu Yang, HeuiSeok Lim
The 32st Annual Conference on Human Cognitive Language Technology, 2020

Design Neural Machine Translation Model Combining External Symbolic Knowledge
Sugyeong Eo, **Chanjun Park**, HeuiSeok Lim
The 32st Annual Conference on Human Cognitive Language Technology, 2020

Cross-Lingual Transfer of Pretrained Transformers to Resource-Scarce Languages (Best Paper Award)
Chanhee Lee, **Chanjun Park**, Gyeongmin Kim, Dongsuk Oh, HeuiSeok Lim
The 32st Annual Conference on Human Cognitive Language Technology, 2020

MMA: Multi-modal Message Aggregation for Korean VQA
Sungjin Park, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim
The 32st Annual Conference on Human Cognitive Language Technology, 2020

Korean Spell Correction based on Denoising Transformer
 Chanjun Park, Sol Jeong, kisu Yang, Sumi Lee, Jaechoon Joe, HeuiSeok Lim

The 31st Annual Conference on Human Cognitive Language Technology, 2019

Parallel Corpus Filtering and Korean-Optimized Subword Tokenization for Machine Translation
  Chanjun Park, Gyeongmin Kim, HeuiSeok Lim
The 31st Annual Conference on Human Cognitive Language Technology, 2019

Optimization of Data Augmentation Techniques in Neural Machine Translation
  Chanjun Park, Kuekyeng Kim, HeuiSeok Lim
The 31st Annual Conference on Human Cognitive Language Technology, 2019

Comparative Study of Sentiment Analysis Model based on Korean Linguistic Chracteristics
  Gyeongmin Kim, **Chanjun Park**, HeuiSeok Lim
The 31st Annual Conference on Human Cognitive Language Technology, 2019

Cross-Validated Ensemble Methods in Natural Language Inference (Best Paper Award)
  Kisu Yang, Taesun Whang, Dongsuk Oh, **Chanjun Park**, HeuiSeok Lim
The 31st Annual Conference on Human Cognitive Language Technology, 2019

Korean Spacing Module using Eojeol dictionary and Morpheme dictionary
  Chanjun Park, Jin-hyuk Choi, Pummo Ryu
Korea Software Congress 2018, 2018

Knowledge-enhancing Real-time Lecture interpreter service Development
  Chanjun Park, Minho Kang, Byunggun Park, Suntae Kim, Myeongjin Hwang, Dongpil Kim, Young-kil Kim
Korea Software Congress 2018, 2018

Voice recognition performance improvement using TTS system in address recognition system
  Chanjun Park, Suntae Kim, Minho Kang, Myeongjin Hwang, Byunggun Park, Yeonsu Kim
Korea Software Congress 2018, 2018

Improved speech recognition performance in business meeting environments using domain-specific method
  Chanjun Park, Suntae Kim, Minho Kang, Byunggun Park, Myeongjin Hwang, Yeonsu Kim
Korea Software Congress 2018, 2018

Spam/Ham Classification using Ensemble Technique of Machine Learning Algorithms
  Chanjun Park, Pummo Ryu
Korea Computer Congress 2018, 2018

Two-Stage Korea Compound Noun Decomposer
  Chanjun Park, Pummo Ryu
The 30st Annual Conference on Human Cognitive Language Technology, 2018

Dictionary based Compound Noun Decomposer using Resource extension
  Chanjun Park, Hyejin Won, Pummo Ryu
Korea Computer Congress 2018, 2018

BOOK CHAPTERS

Data-Centric Neural Machine Translation - A Real-World Approaches
  Chanjun Park
Ph.D. Thesis (2023). 2023

Natural Langugae Processing Bible
  Heuiseok Lim, **Korea University NLPAI Lab**
Human Science, https://www.aladin.co.kr/shop/wproduct.aspx?partner=rss&ISBN=K412637214, 2020

Extracurricular Activities during Undergraduate Studies ──────────

Operate Naver's natural language processing blog   *South Korea*
BLOGGER   *2016.Dec-Present*
- 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

OpenNMT Forum Members   *South Korea*
MEMBERS   *2019.July-2019.July*
- Active as OpenNMT Forum Members
- Develop and share OpenNMT Pytorch, Tensorflow, Library Tutorial

ETRI's machine translation quality evaluator   *South Korea*
ESTIMATOR   *2019.October-2019.December*
- Quality evaluation of machine translation hosted by ETRI

Completed the natural language processing tutorial (2016-2019) *South Korea*

Student *2016 - 2019*

- Special Interest Group of Human and Cognitive Language Technology

Journalist of the Ministry of Science and ICT *South Korea*

Reporter *2016.Feb-2017.Feb*

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

BIT Computer Advanced Course Completion *South Korea*

Student *2016.January-2016.August*

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

BIT Computer Short Course Completion *South Korea*

Student *2015.June-2015.August*

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