

이동건

(DongGeon Lee)
M.S Student
NLP/AI Researcher

자연어 처리와 대규모 언어모델에 깊은 이해와 관심이 있습니다.

의료, 항공 등 다양한 도메인 특화 언어모델을 직접 개발하고
배포한 경험이 있습니다.

더 나은 세상을 위해 인공지능 기술을 실생활에 적용할 때,
비로소 기술의 가치가 실현된다고 생각합니다.

E-mail: dg.lee@postech.ac.kr

Personal Site: <https://donggeon.github.io>

Etc: [CV](#), [Google Scholar](#), [LinkedIn](#)

Education

포항공과대학교 (POSTECH) / 인공지능대학원 석사과정

2024.02 - PRESENT

인하대학교 / 정보통신공학과 학사

2018.03 - 2024.02

Skills

Programing Language

- Python, Shell Script, C/C++, HTML5/CSS, JavaScript

Frameworks

- PyTorch, transformers, TensorFlow, Keras

Systems and Tools

- Git, Linux, Markdown, LaTeX, MySQL

Language (English)

- OPIc : Intermediate High
- TOEIC : 805

Teaching Experience

엘리스 / Mentor

2024.09 - 2024.11 (3m)

2024년 청소년 SW 동행 프로젝트; 고등학생 대상 SW 및 AI 멘토링

- 기술 프로젝트 및 문제 해결 중점 교육 세션 진행

한국로봇융합연구원 (KIRO) / Instructor (Contract)

2024.03 - 2024.05 (3m)

디지털새싹 프로그램; 포항제철고등학교 SW·AI 위탁교육

- 생성형 AI를 활용한 파이썬 프로그래밍

인하대학교 정보통신공학과 / Teaching Assistant

2023.02 - 2023.06, 2023.08 - 2023.12 (10m)

프로그래밍 실습 지원, 과제 및 프로젝트 평가

- ICE1005: 인공지능 프로그래밍 기초 (2023학년도 1학기)
- ICE3020: 알고리즘 설계 (2023학년도 1학기)
- ICE2004: 자료구조론 (2023학년도 2학기)
- ICE4016: 데이터베이스 설계 (2023학년도 2학기)

(주) 잼코딩 / Instructor (Contract)

2021.10 - 2023.05 (1y 8m)

프로그래밍 정규 강의 (고등학생~초등학생 대상)

- 데이터 분석 및 시각화 (matplotlib, numpy, pandas)
- 파이썬·C언어 프로그래밍 및 알고리즘
- SW 저작권 출원 특강

Research Experience

POSTECH Data Intelligence Lab / Research Assistant

(지도교수: [유헌조](#))

2024.02 - PRESENT

LLM (Large Language Model) 연구

- Cross Domain Adaptation via In-Context Learning

모두의 연구소 LLM Experimental LAB / NLP Team Lead.

2024.02 - PRESENT

RAG (Retrieval Augmented Generation) 연구

- Adaptive Query Rewriting for RAG

인하대학교 Data Intelligence Lab / Research Assistant

(지도교수: [최원익](#))

2022.11 - 2023.11 (1y 2m)

도메인 특화 sLLM (small Large Language Model) 연구

- 항공 안전 도메인 특화 언어모델 [KoAirBERT](#) 개발
- 항공 사고 보고서의 사고 원인 추출을 위한 Fine-tuning 연구

인하대학교 Nursing Informatics Lab / Research Assistant

(지도교수: [조인숙](#))

2021.07 - 2023.05 (1y 10m)

도메인 특화 sLLM (small Large Language Model) 연구

- 병원 내 낙상 사건 진술문 탐지 모델 개발 (F1-score ≥ 0.97)
- 낙상 보고서 자동 생성을 위한 NER 모델 개발

ML production cycle (low level)

- 의료 데이터 수집 및 한국어 특화 전처리

의료 텍스트 데이터 (EMR) 분석

- 의료 데이터 특성에 맞는 전처리 및 익명화 기법 연구
- 약물 처방 패턴에 대한 네트워크 분석

Conference Publications

(Selected)

질문 유형에 따른 검색 증강 생성의 질의응답 성능 분석

이동건*, 박아정*, 이혜리, 남현서, 맹윤희 (*: 공동 1저자)

제 36회 한글 및 한국어 정보처리 학술대회, 2024 [[paper](#)]

Effects of Language Differences on Inpatient Fall Detection Using Deep Learning

Insook Cho, EunJu Lee, **DongGeon Lee**

MedInfo, 2023 [[paper](#)]

Bridging the Reporting Gap of Inpatient Falls to Improve Safety Practices Using Deep-Learning-Based Language Models and Multisite Data

DongGeon Lee, EunJu Lee, Insook Cho

AMIA Clinical Informatics Conference, 2023 [[paper](#)]

Under Review

Enhancing Adverse Event Reporting with Clinical Natural Language Processing Techniques: Using Language Models to Detect Inpatient Falls

etc.

Open Source Contribution

[huggingface/transformers](https://huggingface.co/transformers)

- Contribute to PyTorch usages for better readability ([PR](#))

장학금 및 수상 경력

2024.10 우수 논문상(국립국어원장상) / 2024년도 한글 및 한국어 정보처리 & 한국코퍼스언어학회 공동 학술대회

2024.10 2024 국립국어원 인공지능의 한국어 능력 평가 경진 대회 / 금상(국립국어원장상) / 국립국어원

2024.05 우수 대학원생 입학 장려 장학금 / POSTECH

2024.02 최우수 공학도상 / 인하대학교

2023.08 우수 학생 연구원 연구 장학금 / 인하대학교