

Dongin Jung

Seoul, Republic of Korea
dongin1009@gmail.com
<https://github.com/dongin1009>

Experience QRAFT Technologies, Seoul, Republic of Korea
NLP/LLM Researcher on AI Tech Lab Apr. 2024 - Present
NLP/LLM Researcher Intern on AI Tech Lab Jan. 2024 - Mar. 2024

- LLM domain adaptation to know financial terminologies
- Summarization & translation quality evaluation by open-source LLM
- Finance chatbot with Korean, financial, legal knowledge understanding

SK Planet, Seongnam, Republic of Korea
AI Engineer Intern on IoT Development Team Apr. 2023 - Jul. 2023

- Detecting defective parts and unifying multiple types in manufacturing industry
- Real-time road object detection & tracking in real-world environment
- Denoise weather status on video for weather removal and clear video

Education Chung-Ang University, Seoul, Republic of Korea
M.S. in Artificial intelligence Mar. 2021 - Feb. 2023

- Advisor: Prof. Yoon-Sik Cho (*in* Data Science Lab)
- Thesis: Data Relation Modeling Methods for Multiple Fake News Detection
- GPA: 4.34 / 4.5

Wonkwang University, Iksan, Republic of Korea
B.S. in Computer Software Engineering Mar. 2017 - Feb. 2021
B.S. in Digital Forensic (Double Major) Mar. 2018 - Feb. 2021

- Advisor: Prof. Jongmin Lee (*in* Computing System Lab)
- GPA: 3.91 / 4.5

Skills **Research Areas:** Natural Language Understanding & Natural Language Generation,
Graph Neural Network, Contrastive Learning, LLM, PEFT
Environment: Linux (CAU DS Lab. server admin), Git, Anaconda, Slurm, DeepSpeed
Framework: PyTorch, torch geometric, Tensorflow, Huggingface, PySpark, LangChain

Publications

Journal

- **Dongin Jung**, EungYeop Kim, and Yoon-Sik Cho, “Topological and Sequential Neural Network Model for Detecting Fake News”, IEEE Access (Vol. 11), 2023
 - Paper: <https://ieeexplore.ieee.org/document/10363175>
 - Code: <https://github.com/dongin1009/TSNN-DFN>

- **Dongin Jung**, Misuk Kim, and Yoon-Sik Cho, “Detecting Documents With Inconsistent Context”, IEEE Access (Vol. 10), 2022
 - Paper: <https://ieeexplore.ieee.org/document/9875276>
 - Code: <https://github.com/dongin1009/GraDID>

Conference

- Seongmin Kang, **Dongin Jung**, and Yoon-Sik Cho, “Video Retrieval with Tree based Video Segmentation”, International Conference on Database Systems for Advanced Applications (DASFAA 2023)
 - Paper: https://link.springer.com/chapter/10.1007/978-3-031-30675-4_29
- **Dongin Jung**, and Yoon-Sik Cho, “Automatic Conversation Turn-Taking Segmentation in Semantic Facet”, 2023 International Conference on Electronics, Information, and Communication (ICEIC 2023)
 - Paper: <https://ieeexplore.ieee.org/document/10049858> [PDF]
 - Code: <https://github.com/dongin1009/semantic-turn-taking-prediction>
- **정동인**, 김응엽, 강성민, 조운식, “이미지 검색을 위한 대조 학습 모델의 한국어 학습 방법”, 대한전자공학회 2022년도 하계종합학술대회
 - Paper: <https://www.eiric.or.kr/KeyDocs/tmp/2813-2816.pdf>
 - Code: <https://github.com/dongin1009/Ko-CLIP>

Submitted & Preparing

- Dae-Sung Wang, **Dongin Jung**, Hyun-Taek Hong, Yoon-Sik Cho, and Chan-Gun Lee, “Exploiting Component Information with a Context-based Language Model for Effective Bug Triage”, (Under review)
- **Dongin Jung**, Jaehyeok Ahn, and Yoon-Sik Cho, “Joint Contrastive and Supervised Learning in Human Activity Recognition”, (in preparation)
 - Code: https://github.com/dongin1009/joint_supcon_har

Projects

- Text Fake News Detection (sponsored by KOCCA) Mar. 2021 - Dec. 2022
 - Defined fake news into 2 types: inconsistent content news and non-fact news.
 - Proposed concept of ‘supernode’ and detected fake news by using language models and graph neural networks.
- HW/SW platform for interactive self-monitoring in real-time addressing mental health problems based on brain-cognitive behavioral data (sponsored by IITP) Mar. 2021 - Dec. 2021
 - Recognized depression using time-series smartphone sensor values of users.
 - Analyze sensor feature of user by considering feature characteristics.

- Polyglot: Multilingual Large Language Models Oct. 2022 - Present
 - Collect and preprocess large-scale [Romance language text data](#).
 - The goal is to develop large-scale text generation models, named ‘Polyglot-v2’, based on GPT architecture.
 - Work as a member of the ‘polyglot-Korean Pre-training’ group belonging to EleutherAI (AI open-source community).
- [KR3: Korean Restaurant Review with Ratings dataset](#) Apr. 2021 - Dec. 2021
 - Built a Korean positive & negative restaurant text review dataset.
 - Performed task-adaptive pretraining and qualified the additional pretraining effect on cross-task review datasets.
 - Worked as a member of the ‘time-series group’ belonging to DIYA (AI club).
- [Development of an Online Learning Management System](#) (sponsored by SW Central Center at Wonkwang Univ.) Jun. 2020 - Dec. 2020
 - Developed an online LMS Platform for contactless learning.

Patents 정동인, “시각장애인을 위한 음성 챗봇 시스템 및 방법”, 출원번호: 10-2021-0020237, 2021
 정동인, 강현지, 김수빈, 윤지환, 장형국, 조영진, “스마트 화분 장치”, 등록번호: 10-2170452 (출원번호: 10-2020-0029729), 2020

Activities [EleutherAI](#), Polyglot team contributor Oct. 2022 - Present
[DIYA](#), time-series team member Apr. 2021 - Dec. 2021
[Artineer](#), department programming club president Mar. 2019 - Feb. 2020
[LIKELION](#) at Wonkwang University, staff Mar. 2019 - Feb. 2020

Awards and Honors Dacon [물류 유통량 예측 경진대회](#), 29th/237 (top 12.2%) 2021
 Dacon [동서발전 태양광 발전량 예측 AI 경진대회](#), 11th/234 (top 4.7%) 2021
 Creative Competence Award (Greatest in Creative Score), Wonkwang Univ. 2021
 Best Award on Major SW Competition, Wonkwang Univ. 2020
 Excellence Award on LINC+ Startup Idea Competition, Wonkwang Univ. 2020
 Bronze Award on Student Portfolio Competition, Wonkwang Univ. 2019
 Bronze Award on Creative Software Competition, Wonkwang Univ. 2019
 Bronze Award on Corporate Linked Capstone Design, Wonkwang Univ. 2019