이영주 Young-Ju Lee

데이터 수집부터 정제·검증·색인까지 파이프라인을 만드는 데이터 엔지니어

iyju990312@naver.com
github.com/0Ju12 · 연구 이력 · LinkedIn
포트폴리오 0ju12.github.io

파이프라인 3종 오픈API·크롤링 수집 → 정제 → 색인 | **테스트 136개** ThRead 86 · 국보 RAG 50 (외부 호출 없이)

SCIE 1편 SPE Journal 제1저자 · 분석 코드 공개 | **SSAFY 16기** 삼성 청년 SW·AI 아카데미

PROJECTS · 핵심 경험

부트캠프 — SSAFY 16기 · 포스코 청년 AI·BigData 아카데미 30기 2025.06 – 진행 중

삼성 청년 SW·AI 아카데미(SSAFY) · 포스코인재창조원·POSTECH · LLM 파인튜닝 · 데이터 파이프라인 · 백엔드 · Python, LLM 파인튜닝, QLoRA, RAG

- 국보 입문 가이드(개인): 국가유산청 오픈API(XML·EUC-KR)·위키백과 수집 → 정제 → ChromaDB 색인. 청크 ID 고정으로 재실행해도 중복 없는 색인, 질문별 수집 캐시·위키 장애 시 저장본 대체, 429/5xx Retry-After 재시도
- ThRead: Selenium 댓글 수집 → LLM 판단을 코드로 검증 → pandas 정제·IQR 길이 기준·증강·요약, NDJSON 진행 상황 스트리밍 · 알라딘 API 페이지네이션·중복 제거 수집 · Django 백엔드 · 자동 테스트 86개
- AI 챌린지(VQA): 제출 전 자동 검사(6,714행·ID·확률 행합) + SHA-256 매니페스트로 GPU 없이 재현 가능한 제출 관리 · Public Score 0.96961
- POSPEAK(KILL KONG): 콩글리시 교정 데이터셋 구축, FLAN-T5 LoRA 학습 전략 비교(F1 0.32→0.74) + RAG · 장려상

석사과정 — 중력·자력 데이터 파수 상관 분석으로 지하 구조 규명 (SPE Journal 제1저자) 2024.03 – 2026.08

전북대학교 대학원 환경에너지융합학과 · 지구물리컴퓨팅연구실 · 석사과정 연구원 · 캘거리대 장기 파견 · 신호처리(파수 영역), 상관 분석, 지표 설계, 중력·자력 데이터

- 항공 중력·자력 격자 자료를 파수 영역에서 처리하는 분석 파이프라인을 만들고, 겹친 신호를 분리하는 지표(SLFI·DLFI) 설계 — 독립 모델 3개와 상관계수 0.94 이상 일치
- 분석 코드를 Python 패키지·테스트로 정리해 공개 · 캘거리대 6개월 파견, 3개 기관 공동연구 → SCIE SPE Journal 제1저자 (2026)
- 국가R&D 과제 참여 연구원 · 정기 연구미팅 발표 10회 이상 · 2025 기초연구실 워크숍 발표

학부 — 판구조 기반 해외 CO2 저장소 탄성파 해석 (제1저자) · IAB 융합전공 시공간 딥러닝 2022 – 2024.02

전북대학교 자원에너지공학과 · IAB 융합전공 · 지구물리컴퓨팅연구실 · 학부연구생 · 캡스톤 모델 설계·학습 담당 · 탄성파 해석, GPlates, CCS, 중력 데이터 처리

- IAB 융합전공 캡스톤: ArcGIS로 산불 래스터 데이터를 구축하고 U-LSTM으로 화선 예측 — 미학습 산불 테스트 Accuracy 0.90 · F1 0.85
- 제1저자 논문: GPlates 판구조 복원 + 대규모 CO2 저장소 탄성파 해석 (지구물리와 물리탐사, 2024) · 제2저자 논문 1편 · 학회 발표 3건

EDUCATION · 교육

삼성 청년 SW·AI 아카데미 (SSAFY) 16기 교육 과정 — 2026.07 – 진행 중

전북대학교 대학원 환경에너지융합학과 석사 — 2024.03 – 2026.08

포스코 청년 AI·BigData 아카데미 30기 수료 — 2025.06 – 2025.08

전북대학교 자원에너지공학과 학사 — ~ 2024.02

SKILLS · 기술

수집·처리 Python, Pandas, SQL, Selenium, 오픈API(XML·JSON), 웹 크롤링

저장·검색 ChromaDB, FAISS, 벡터 색인, Django ORM

품질·운영 pytest 자동 테스트, GitHub Actions CI, SHA-256 재현성 매니페스트, 오류 로깅

분석·ML PyTorch, TensorFlow, Hugging Face, LangGraph RAG, 공간 데이터(ArcGIS·래스터)

PUBLICATIONS · 연구 성과

Wavenumber Correlation of Gravity and Magnetic Anomalies for Constraining Salt/Basement Structure in Offshore Abu Dhabi — SPE Journal (SCIE) · 제1저자 · 게재 승인 — 2026

판구조론적 역사를 고려한 해외 대규모 이산화탄소 지중저장소 탄성파 해석 결과 분석 — 지구물리와 물리탐사 27(1), 1–22 · 제1저자 — 2024

호주 깁스랜드 분지 CarbonNet CCS 프로젝트의 지질학적 분석 — 지구물리와 물리탐사 · 제2저자 — 2023

외 학회 발표 6건 (전체 목록은 포트폴리오 참고)

AWARDS · CERTIFICATIONS

포스코 청년 AI·BigData 아카데미 30기 프로젝트 장려상 — 2025.08

2024 한국지구물리·물리탐사학회(KSEG) 추계학술대회 학생 우수논문(포스터) 발표상 — 2024.11

ADsP (데이터분석 준전문가) — 2022.09