Young-Ju Lee Young-Ju Lee

Data & AI Engineer

iyju990312@naver.com · github.com/0Ju12 · geoverselab.com/ko/research-history/13/ · LinkedIn

SUMMARY

M.S. in Environment and Energy (Jeonbuk National University, 2026) with a B.S. in Mineral Resources and Energy Engineering. As an undergraduate researcher I wrote a first-author paper interpreting seismic images of large overseas CO2 storage sites in light of plate-tectonic history, and in the IAB convergence-major capstone I designed and trained a spatiotemporal deep learning model (U-LSTM) that predicts wildfire spread. In my master's I separated overlapping gravity and magnetic signals with wavenumber-domain correlation analysis and published a first-author paper in SPE Journal (SCIE); across both degrees: 3 papers, 6 conference presentations and a best student paper award. At the POSCO Youth AI & Big Data Academy (with POSTECH) I built the training data and ran model experiments for an English expression correction system (FLAN-T5 + RAG) that won an Excellence Award, and at the Samsung Software & AI Academy for Youth (SSAFY) I am building LLM data pipelines and multimodal model ensembles. Two research stays at the University of Calgary; ADsP (Advanced Data Analytics Semi-Professional) certified.

EXPERIENCE

M.S. — Separating gravity & magnetic signals with wavenumber correlation to resolve subsurface structure (SPE Journal, first author)

2024.03 – 2026.08

Research · Jeonbuk National University · Geophysical Computing Lab · Graduate researcher · 6-month stay at University of Calgary

Wavenumber correlation of gravity & magnetic data for salt/basement structure (M.S. · SPE Journal first author) 2024.03 – 2026.08 · Graduate researcher details ↗

- Built a wavenumber correlation filtering (WCF) pipeline — upward continuation and normalization, then the correlation coefficient CC(k) between the two signals at each wavenumber
- Designed two indices — SLFI from positively correlated components (basement) and DLFI from weak/negative ones (salt)
- Chose thresholds from energy–correlation trade-off curves instead of by eye (upward continuation 1,500 m at the knee point; SLFI band 0.5–1.0 keeps ~half the energy at correlation 0.85)
- Validated against three independent published models with map, profile and 3-D density comparisons
- Presented Abu Dhabi processing and analysis results at 10+ regular lab research meetings (2024.05–2026.01) and at the 2025.11 Basic Research Lab workshop
- Researcher on national R&D projects, including the IITP global talent program
- **SLFI basement structure agrees with three independent models at r ≥ 0.94 (profile correlation 0.95–0.98 after shift alignment)**
- **DLFI reproduces the salt distribution and suggests an additional salt uplift west of Ghasha**
- **First-author paper in SPE Journal (SCIE, 2026, with Univ. of Calgary and Khalifa Univ.) · 3 conference talks (2024)**
- **Best Student Paper (poster), KSEG Fall Conference 2024 · M.S. degree conferred 2026.08.21**

Research stay at the University of Calgary — joint interpretation of gravity & magnetic data

2024.06 – 2024.12 · Visiting graduate researcher details ↗

- Funded by the MSIT/IITP Global Talent program in digital fields (AI & digital-twin smart CCS)
- Worked with the Department of Geomatics Engineering on wavenumber correlation of airborne gravity and magnetic data
- Extended the work into a 3-institution collaboration with Khalifa University (UAE)
- Joined a field geophysical survey in Kananaskis (Canadian Rockies) and co-wrote the English Field Exploration Report (2024.08)
- **Led to the first-author SPE Journal paper (SCIE, 2026) and the M.S. thesis**

Signal processing (wavenumber domain) · Correlation analysis · Index design · Gravity & magnetic data · International collaboration

Undergraduate — Seismic interpretation of overseas CO2 storage with plate-tectonic history (first author) · spatiotemporal deep learning capstone

2022 – 2024.02

Research · AI project · Jeonbuk National University · IAB convergence major · Geophysical Computing Lab · Undergraduate researcher · capstor

Seismic interpretation of large overseas CO2 storage sites considering plate-tectonic history (first-author paper) 2022 – 2024.02 · Undergraduate researcher · first author details ↗

- Reconstructed the plate-tectonic history of four regions with GPlates (PALEOMAP PaleoAtlas)
- Interpreted seismic images of two representative CCS projects per region — Quest & Weyburn-Midale, Northern Lights & PORTHOS, Gorgon & Moomba, Uthmaniyah & Qatar LNG — focusing on reservoir and cap rock
- Short stay at Univ. of Calgary (2023.12–2024.02): processed gravity data from the Utah FORGE geothermal site
- **First-author paper: Analysis on Seismic Interpretation for Overseas Large-scale CO2 Storage Considering Geological History Related to Plate Tectonics, Geophysics and Geophysical Exploration 27(1), 1–22 (2024)**
- **Showed that stratigraphy in each seismic image reflects the region's tectonic history → a basis for rejecting geologically unrealistic velocity models when processing domestic data**
- **1 second-author paper (CarbonNet CCS, Gippsland Basin, 2023) · 3 conference talks (2022–2023)**

IAB capstone — Wildfire spread prediction with spatiotemporal deep learning

2023.09 – 2023.12 · Model design & training details ↗

- Rasterized fire fronts from Korea Forest Service maps plus wind, NDVI and elevation in ArcGIS for four large fires (3 for training, 1 held out)
- Designed a shallow U-LSTM (2,736 parameters) to fit the small dataset (4 fires, 35 time steps)
- Expanded the sequences with sliding windows (length 1–19) into 151 training batches; evaluated with Overall Accuracy and F1
- **Unseen fire (Gangneung): Accuracy 0.90 · F1 0.85 (validation 0.95 · 0.81)**
- **Refactored the team notebooks into a tested Python package and documented the limitations on GitHub**

Seismic interpretation · GPlates · CCS · Python · TensorFlow · U-LSTM · ArcGIS

Bootcamps — SSAFY 16th cohort · POSCO Youth AI & Big Data Academy 30th cohort 2025.06 – present

AI project · Samsung SSAFY · POSCO HRD Center · POSTECH · LLM fine-tuning · data pipelines · backend

SSAFY — ThRead LLM comment-analysis pipeline · AI Challenge VQA ensemble

2026.07 – present · Backend & data pipeline · model training (AI Challenge) details ↗

- Comment pipeline: Selenium collects live comments → the LLM returns only the numbers of inappropriate comments and code removes them → pandas/regex cleaning → IQR length thresholds → LLM augmentation (retry on count mismatch) → summary, each step stored
- Streamed progress of 30-second jobs with StreamingHttpResponse (NDJSON) and distinguished failure cases (no company, no comments, LLM failure)
- AI book service: custom Django auth (open-redirect protection), AI recommendations restricted to registered book IDs; 45 mock-based tests without external APIs
- AI core project 2 · National-treasure RAG guide (2026.10, solo): collected treasures and official descriptions from the Korea Heritage Service open API (XML, EUC-KR) and auto-linked each to Korean Wikipedia; two ChromaDB collections with fixed chunk IDs (idempotent re-indexing); per-question LLM query extraction and document selection with caching and offline fallback; LangGraph retrieve→generate where [evidence N] tags are checked in code and unknown numbers removed; Streamlit + FastAPI UIs; 50 tests with a fake LLM and Streamlit AppTest
- AI Challenge (2026.09, image + question 4-choice VQA, model training & refinement): Qwen 27B/4B QLoRA, Gemma, OCR and multi-view probability ensembling; accepted changes by fixes − breaks and per-type regressions; SHA-256 manifests reproduce the submission without a GPU
- **AI Challenge public score 0.94608 (4B baseline) → 0.96961**
- **86 automated tests passing; all project-5 requirements implemented, plus progress streaming and analysis history**

POSPEAK (KILL KONG) — Konglish → natural English correction AI (FLAN-T5 · Qwen + RAG)

2025.06 – 2025.08 · Dataset building & model experiments details ↗

- [Own] Built and cleaned word-level and context-level correction pairs
- [Own] Fine-tuned FLAN-T5 with LoRA and compared three strategies (word→context, context→word, joint) with BERTScore
- [Final service KILL KONG] Qwen-2.5 (1.5B) fine-tuning + a 630-entry Konglish RAG DB + per-user memory, combined with LangChain in a React Native · FastAPI app
- **Found catastrophic forgetting in sequential training (BERTScore F1 0.47 → 0.32) and switched to joint training**
- **Joint training reached validation BERTScore F1 ≈ 0.74 · team Excellence Award**

Python · LLM fine-tuning · QLoRA · RAG · FAISS · Django · Selenium · Pandas

PUBLICATIONS

[Paper] Wavenumber Correlation of Gravity and Magnetic Anomalies for Constraining Salt/Basement Structure in Offshore Abu Dhabi, *SPE Journal (SCIE) · first author* 2026

[Paper] Analysis on Seismic Interpretation for Overseas Large-scale CO2 Storage Considering Geological History Related to Plate Tectonics, *Geophysics and Geophysical Exploration 27(1), 1–22 · first author (in Korean)* 2024

[Paper] Geological Analysis of the CarbonNet CCS Project in the Gippsland Basin, Australia, *Geophysics and Geophysical Exploration · second author (in Korean)* 2023

[Talk] A global talent program for smart CO2 storage operation through AI and digital-twin convergence 2024

[Talk] Status of overseas CCS demonstration projects and strategy for international collaboration 2024

[Talk] A global talent program for smart CO2 geological storage site operation 2024

[Talk] Geophysical monitoring case analysis of Sleipner, Norway, and international collaboration plan 2023

[Talk] Deep-learning seismic imaging research in Saudi Arabia and international collaboration plan 2023

[Talk] Applicability of GPU-based 4D seismic monitoring 2022

SKILLS

Languages & frameworks Python, TensorFlow, PyTorch, Hugging Face Transformers, Pandas, Django, Selenium, SQL

AI & data analysis Spatiotemporal DL (U-LSTM), LLM fine-tuning (LoRA/QLoRA), Model ensembles, RAG (LangGraph · ChromaDB · FAISS), F1 · BERTScore

Spatial & geophysical data ArcGIS, Rasters, Gravity & magnetics, Wavenumber correlation, Seismic interpretation, GPlates

Domain CCS, Energy & resources, Wildfire prediction

EDUCATION

Samsung SW & AI Academy (SSAFY) 2026.07 – present
16th cohort Training program

Jeonbuk National Univ. (Graduate) 2024.03 – 2026.08
Environment and Energy M.S.

POSCO Youth AI & Big Data Academy 2025.06 – 2025.08
30th cohort Completed

Jeonbuk National Univ. ~ 2024.02
Mineral Resources and Energy Engineering B.S.

AWARDS

Project Excellence Award, POSCO Youth AI & Big Data Academy (30th) 2025.08

Best Student Paper (poster), KSEG Fall Conference 2024 2024.11

CERTIFICATIONS

ADsP (Advanced Data Analytics Semi-Professional) 2022.09