

JIUK KIM, ML Engineer

"A developer who adapts well, continuously learns, enjoys solving problems, and persistently delves into challenges."

Contacts

010-9590-2997 · alex.jiuk.kim@gmail.com · <https://www.linkedin.com/in/ziweek> · <https://hire-me-please-gpt.streamlit.app>

Target Role

ML Engineer | Software Engineer

Summary of Skills

Machine Learning Engineering

- Developed and Deployed a full-stack web application based on LLM fine-tuned in PEFT with QLoRA [\[Award Factory\]](#)
- Utilized Quantization techniques with llama-cpp reducing LLM model size by 63% and inference time by 84%
- Converted fine-tuned models to on-device LLMs for Apple Silicon using Apple Core ML and Xcode
- Implemented and monitored RAG architecture for multi-agent network design using LangChain and LangGraph [\[TA-CO\]](#)
- Improved performance of small object detection models by 11% leveraging preprocessing and ensemble [\[desirable-sea\]](#)

Software Development

- Lead development of multiple individual and collaborative projects.
- Capable of developing simple frontend interfaces using React.js, Next.js, and Flutter.
- Skilled in building basic backend API servers with Spring, Nest.js, and FastAPI.
- Experience managing images via Docker Registry and deploying servers using Docker and Docker-Compose.
- Familiar with deploying services on AWS (EC2, Route53, S3, Load Balancer, etc.) and GCP (Gemini API, etc.)

Languages & Frameworks & Certifications

- Python (expert), Java (proficient), JavaScript (proficient) / Korean (native), English (proficient), Japanese (basic)
- Transformers, Hugging Face, LangChain, LangGraph / Next.js, Spring, Nest.js, FastAPI / Docker, Jenkins / AWS
- Engineer Big Data Analysis (2024.07), Test of Practical Artificial Intelligence (2023.11)

Experiences

[Award-Factory: Awards lovingly crafted for you by a hilariously talented generative AI \(2024.07 ~ 2024.12\)](#)

Showcasing Project, in 2024 Google Machine Learning Bootcamp Program

- Built an LLM-based full-stack web application from design to deployment, to create personalized awards with Korean humor for loved ones, using Google/Gemma-2b-it, Transformers, HuggingFace, Docker-compose, Next.js, and FastAPI.
- Performed prompt engineering and fine-tuning in PEFT with QLoRA quantization on the Google/Gemma-2b-it model, to generate titles, citations, and issuers with Korean-style humor in JSON format.
- Applied Quantization with the Q5_K_M option of llama-cpp, to minimize computational costs in the deployment environment, reducing the model size by 63.3% and an inference time by 83.4% without significant performance degradation.
- Converted the fine-tuned model into an on-device LLM for Apple Silicon using Apple Core ML and Xcode, to use the local environment of a 2020 Mac Mini with M1 chip as a server to run inference.

[TA-CO: Two Army rules, Chat Once \(2024.10 ~ Present\)](#)

KATUSA Program, 8th Army (United States)

- Developed a bilingual RAG chatbot delivering US Army and ROK Army regulations, to address potential language barriers and reduce confusion caused by differing regulations, utilizing LangChain, LangGraph, Nomic-Embed, and FAISS.
- Leveraged a quantized PEFT fine-tuned Google/gemma-2-2b-it model as part of a bidirectional translation pipeline, to address issues where other languages were included in the Korean output, resolving the language-specific challenge.

- Conducted preliminary research to implement a Multi-Agent Network, to improve RAG performance by integrating goal-oriented pre-retrieval and post-retrieval processes in a multilingual document, with refactoring agents and tools

[Desirable-Sea: Intelligent Platform for Long-Term Abandoned Vehicle Detection Utilizing Satellite Image \(2023.12 ~ 2024.01\)](#)

1st Prize Winner in 2023 Jeju Island Satellite Data Utilization Competition

- Developed a full-stack computer-vision web application utilizing satellite image data, to address the parking space issue caused by long-term abandoned vehicles in Jeju Island, utilizing MMRotate, WandB, Nest.js, and Next.js.
- Enhanced Small Object Detection model accuracy by 11% in mAP (mean Average Precision), by integrating an image resolution enhancing model in a preprocessing pipeline, by implementing a Weighted Boxes Fusion ensemble.

Careers

8th Army (United States) (2023.07 ~ 2025.01)

Sergeant (E-5), KATUSA (Korean Augmentation to The United States Army) Program, full-time

- Streamlined KATUSA human resource operations by managing a database and dashboard, automating repetitive tasks with macro-based procedures, and delivering efficient data visualization for monitoring and analysis.
- Army Commendation Medal, signed by 8th Army Deputy Commanding General, Brigadier General Sean Crockett

Team OneClean (2023.01 ~ 2023.06)

Full-stack Developer, full-time

- Developed a full-stack web application for customers to book restroom cleaning services using Next.js, Nest.js, and MySQL

KU3DS, Korea University Makerspace (2022.07 ~ 2022.12)

Work Scholarship Student, part-time

- Built a web page to provide information about Korea University's metaverse event using Next.js, Tailwind CSS, and Vercel

Education

GEORGIA INSTITUTE OF TECHNOLOGY (United States) (2024.08 ~ 2027.02)

Online Master of Science in Computer Science

- Specialization Program in Machine Learning

KOREA UNIVERSITY (South Korea) (2017.03 ~ 2024.02)

Bachelor in Business Administration

- Concentration Track in Business Analytics
- Founded the Business Analytics Research Student Club, instructing undergraduate students in artificial intelligence
- Backend Team of the 11th Korea University's Software Startup Society NEXT (LikeLion)

ETC

HONORS AND AWARDS

- Chief of Staff of the Republic of Korea Army Award, in the 11th Republic of Korea Army Entrepreneurship Competition (2024)
- 1st Place Award, presented by the Governor of Jeju Province, in 2023 Jeju Island Satellite Data Utilization Competition (2023)
- Major General Award, in the 11th Republic of Korea Army Entrepreneurship Competition (2023)
- Encouragement Prize, in the 3rd Korea University Smart Campus Data Hackathon (2022)
- Excellence Prize, Samsung x KU future eco-friendly home appliances idea contest (2022)