EESUN MOON

eesun.moon@columbia.edu | LinkedIn | GitHub | NY, United States

EDUCATION

Columbia University New York, NY
MS in Computer Science, GPA: 3.92/4.0 Expected Dec 2025

- Courses: Applied Machine Learning, Natural Language Processing, Spoken Language Processing, Computer Vision, Database

Sejong University Seoul, South Korea
BS in Intelligent Mechatronics Engineering, BE in Data Science, GPA: 4.4/4.5 Feb 2024

- Courses: Artificial Intelligence, Computer Networks, Operating Systems, Image Processing, Data Structures, Web Programming
- Teaching Assistant: Algorithms using C programming, Python fundamentals

TECHNICAL SKILLS

Programming & Databases	Python, C, R, Java | MySQL, PostgreSQL, MongoDB
Machine Learning Frameworks	TensorFlow, Keras, PyTorch, Scikit-Learn, Hugging Face, OpenAI API, LangChain
Development & Model Optimization	Git, Docker, Linux, FastAPI, AWS EC2, GCP | GPU, NPU, ONNX, TFLite

PROFESSIONAL EXPERIENCE

Samsung Research America Mountain View, CA
AI Algorithm/NPU Simulator Research Scientist Jun 2025 – Expected Aug 2025

- Build automation pipeline to profile ViTs and LLMs on Exynos NPU under diverse configurations and clock frequencies, optimizing latency, memory bottlenecks, and MAC Utilization in Galaxy AI's Next-GenAI On-device Platform
- Develop compiler-based analysis framework to extract layer-wise insights, validating and accelerating NPU simulator design
- Explore LLaMA3 inference optimization on NPU, focusing on KV cache quantization and memory-efficient attention

Humaner: Human-centered AI Software Development [GitHub] Seoul, South Korea
Machine Learning Engineer Mar 2024 – May 2024

- Developed and deployed Q&A-based support message generator using GPT-3.5 Turbo and LangChain on AWS EC2, enabling real-time interaction with 500+ live users and boosting **satisfaction by 20%**
- Iteratively refined prompt logic using post-deployment user feedback to improve personalization and message relevance

Sejong University, Mobile Intelligent Embedded System Laboratory [GitHub] Seoul, South Korea
Research Assistant Sep 2021 – Mar 2024

- Led multimodal emotion recognition project for On-device AI using TensorFlow and MongoDB on Linux for government initiatives
- Optimized ONNX-based deep models with score-based fusion of multimodal signals (heart rate, EEG, speech, image), achieving **99.68% classification accuracy** and reducing **power consumption by 3.12x** and **latency by 1.48x** for edge deployment
- Published papers in **IEEE** [1] and **NCAA** [2], and demonstrated live deployment at **KIST**

PROJECTS

CS Advising Assistant Chatbot with LLM, RAG, and Agentic Flow [GitHub] Jan 2025 – May 2025

- Built chatbot with DeepSeek-R1 on Ollama, **eliminating API costs** via local inference; deployed to GCP for production
- Integrated Agentic Flow into LangChain RAG via MCP server, enabling multi-step retrieval and dynamic tool-based reasoning

Sentence Embedding Analysis in LLMs [GitHub] Jan 2025 – May 2025

- Analyzed interpretability of BERT and LLaMA through layer-/domain-specific embeddings, revealing semantic structure shifts
- Validated cluster coherence via Zipf slope steepening (from -0.87 to -1.42) in intermediate/domain-specific embeddings

Ranking-Based Spam Filtering on Social Networking Services [GitHub] Mar 2022 – Jun 2022

- Spearheaded project to prioritize organic user posts over likely ads from social media, earning **1st place** in graduation competition
- Automated data collection with Selenium and implemented unsupervised clustering with cosine similarity-based ranking, achieving **0.8 intra-cluster similarity** as coherence indicator

PUBLICATIONS

[1] **Eesun Moon**, A.S.M Sharifuzzaman Sugar, Hyung Seok Kim, "Multimodal Daily-life Emotional Recognition Using Heart Rate and Speech Data from Wearables," *IEEE Access*, vol. 12, pp. 96635-96648, 2024. DOI

[2] Taein Kim, **Eesun Moon**, Hoyeon Kang, Hyung Seok Kim, "OMER-NPU: On-device Multimodal Emotion Recognition on Neural Processing Unit for Low Latency and Power Consumption," *Neural Computing and Applications*, 2025. DOI.