GEONSIK MOON

(917) 257-7860 | geonsik.moon@columbia.edu | linkedin.com/in/gsmoon97 | github.com/gsmoon97

Education

Columbia University Expected Dec 2026
Master of Science, Computer Science (Machine Learning Track) *New York, NY*

- Cumulative GPA: 3.92/4.00
- Teaching Assistant: High Performance Machine Learning (COMS 6998) with Prof. K. El Maghraoui
- Research Collaboration: Optimizing Granite Speech Model architectures with IBM Research

National University of Singapore Jun 2022
Bachelor of Computing, Computer Science (Honors) *Singapore*

- Specialization Awards: Distinction in Artificial Intelligence, Merit in Database Systems

Experience

LLM Training Operation Specialist May 2024 – Aug 2025
ByteDance *Singapore*

- Orchestrated data pipelines for *CodeContests+* benchmark, supervising ∼70 annotators and implementing calibration protocols for 17K+ Reinforcement Learning (RL) training data points to align multi-agent systems
- Led collaboration with subject matter experts for *AetherCode* benchmark, curating high-difficulty problems (IOI, ICPC) and comprehensive test suites to evaluate code reasoning, where SOTA models achieve only 35.5% Pass@1
- Engineered the feedback loop for software engineering agents and identified root causes of failure by analyzing 3K+ evaluation outputs, improving model performance by 15%
- Executed instruction-tuning alignment for foundational models, curating 2K+ expert-validated Supervised Fine-Tuning (SFT) samples to enhance multi-lingual coding capabilities
- Built Python-based automation tools and utilized advanced prompt engineering (CoT, Few-Shot) to streamline data format conversion and trajectory analysis workflows

NLP Research Assistant Sep 2022 – Feb 2024
National University of Singapore *Singapore*

- Published two *ACL 2024* papers, proposing a novel framework for timeline summarization and benchmarking encoder vs. decoder models on word semantic understanding tasks
- Fine-tuned open-source LLMs (Mistral, Llama 2, FLAN-T5) via QLoRA (4-bit quantization), optimizing training stability and convergence by tracking experiments with Weights & Biases
- Architected an incremental clustering pipeline using LangChain and ChromaDB, deploying LLM-based pairwise classification to automate large-scale data organization
- Engineered scalable GEC web applications using Docker and Flask, resulting in system demonstration papers at *EACL 2023* and *IJCNLP-AACL 2023*

Machine Learning Engineer May 2022 – Sep 2022
Apple (via TransPerfect) *Singapore*

- Optimized Siri's Natural Language Understanding (NLU) performance by refining model inputs based on large-scale error analysis of production logs
- Explored synthetic data augmentation using Transformer-based models (BERT, T5) to expand linguistic coverage for low-resource queries
- Automated the dialogue optimization workflow via Python and Regex, reducing dataset noise by ∼30% and significantly improving downstream model inference relevance

Technical Skills

Programming Languages: Python, Java, JavaScript, C++, Go, SQL
ML Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn
LLM Training & Evaluation: LoRA/PEFT, LM Eval Harness, Weights & Biases
LLM Infrastructure: LangChain/LangGraph, vLLM, Ollama, llama.cpp, Chroma, Pinecone
Cloud & DevOps: AWS (Bedrock, EC2, S3), GCP, Docker, CUDA, Git