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
- IBM Research Project Collaboration on Granite Speech Model Optimization

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

- Awarded the Certificate of Distinction in the Artificial Intelligence Focus Area
- Awarded the Certificate of Merit in the Database Systems Focus Area

Experience

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

- Diagnosed systemic failure patterns in software engineering agents through analysis of 3K+ evaluation outputs, producing root-cause reports and actionable patch roadmaps that improved Supervised Fine-Tuning (SFT) performance by 30% to achieve parity with state-of-the-art models
- Directed end-to-end data operations for a multi-agent LLM system by supervising ∼70 annotators, implementing calibration protocols and cross-team coordination, successfully delivering 17K+ high-quality data points for Reinforcement Learning (RL) training that contributed to development of CodeContests+ benchmark
- Led collaboration with subject matter experts for AetherCode benchmark by curating evaluation datasets from top coding competitions (IOI, ICPC) with comprehensive test suites, establishing a rigorous new standard in code reasoning where top models achieve only 35.5% Pass@1
- Managed SFT project for instruction-tuning, curating 2K+ expert-validated data points to align foundational models for complex coding capabilities across JavaScript, Python, SQL, and Go
- Developed optimization tools in Python and implemented prompt engineering techniques (few-shot, Chain-of-Thought (CoT)) to streamline training workflows for data format conversion and trajectory analysis

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

- Co-authored two *ACL 2024* publications, advancing the state-of-the-art of LLM capabilities for downstream tasks of timeline summarization and understanding word semantics
- Implemented end-to-end LLM pipelines utilizing LangChain and Chroma, seamlessly integrating open-source models for incremental clustering algorithm deploying LLM-based pairwise classification
- Fine-tuned LLMs (Mistral, Llama 2, FLAN-T5) using LoRA-based PEFT with 4-bit quantization, employing various hyperparameter tuning based on performance tracking from Weights & Biases
- Published two system demonstration papers on GEC web applications in *EACL 2023* and *IJCNLP-AACL 2023*, detailing innovative design strategies to boost NLP pipeline efficiency and accuracy
- Engineered scalable, production-grade web applications for Grammatical Error Correction (GEC) by leveraging containerized microservices (Docker) and modern web frameworks (Flask, Bootstrap)

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

- Served as a vendor language engineer for the Global Siri Team, performing comprehensive error analyses and and fine-tuning model inputs to enhance the model's natural language understanding
- Explored synthetic data augmentation via transformer-driven methods (BERT, T5), utilizing back-translation and context-aware paraphrasing to expand linguistic coverage while preserving semantic integrity
- Automated a dialogue optimization pipeline using scheduled Python scripts and regex pattern matching, reducing extraneous utterances by ∼30% and enhancing response speed and relevance

Technical Skills

Programming Languages: Python, Java, JavaScript/TypeScript, C/C++, Go, SQL
ML Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn
LLM Training & Evaluation: LoRA/PEFT, RLHF/SFT, 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