Mainak Mallick

Atlanta, GA | mainakmallick.github.io | mainakmallick2018@gmail.com | linkedin.com/in/mainakmallick | (470) 830-4196

Software Engineer and MS student at Georgia Tech specializing in building scalable, full-stack solutions and intelligent systems. Proven track record of delivering production-ready features in cloud-native environments using React, TypeScript, Java, RESTful APIs, and Kubernetes. Passionate about transforming cutting-edge ML research into production-ready tools that empower users.

EDUCATION

Georgia Institute of Technology, College of Computing — *Atlanta, GA*
 Major: Master of Science, Computational Science and Engineering, **GPA**: 4.0 — *Dec 2026*
 Relevant Courses – Deep Learning (CS 7643), Computer Architecture (CS 6756), Applied NLP (CSE 6470)

Indian Institute of Engineering Science and Technology — *Shibpur, India*
 Major: Bachelor of Technology, Computer Science, **GPA**: 3.7 — *May 2022*

PROFESSIONAL EXPERIENCE

Graduate Research Assistant, GeorgiaTech — *Dec 2024-Present*
- Developed a retrieval-augmented generation (**RAG**) pipeline using **LangChain** and **Google Gemini APIs**, improving contextual retrieval accuracy for academic question answering by **29%**.
- Fine-tuned **LLaMA 7B** models on domain-specific datasets using prompt engineering on **AWS SageMaker**, boosting factual grounding by **34%**.

Graduate Research Assistant, GeorgiaTech — *Aug 2024-Dec 2024*
- Built a full-stack real-time dashboard using **ReactJS, TypeScript** and **cloud-native APIs** to visualize robotic sensor metrics and fault states.
- Automated scalable data ingestion pipelines using **Bash** and **Python scripts** for **orchestration and parallel processing** across distributed simulation nodes and cloud storage backends.
- Integrated system observability features (logs, metrics, alerts), reducing debug latency and deployment overhead by **23%**, and minimizing manual errors by **29%**.

Software Engineer (Analyst), HSBC Bank — *July 2022-Aug 2024*
- Developed real-time risk analytics tools using **AngularJS**, **NodeJS**, and **PostgreSQL**, improving insights delivery for credit ops.
- Designed and deployed **RESTful Kafka** streaming APIs in **Java/Spring Boot** with multi-threaded data handling and persistent storage integration.
- Led microservice migration to **Kubernetes** with **CI/CD** integration, improving scalability by **40%** and reducing release cycle time by **30%**.

RELEVANT PROJECTS

Fine-Tuning LLaMA 3B for Code Completion: ReFT vs IA3 Comparison — *May 2025-Present*
Python, Pytorch, HuggingFace, CUDA, HPC, LoRA, LLM Fine Tuning
- Compared ReFT vs IA3 method for fine-tuning of LLaMA 3B model on HumanEval benchmark for code generation.
- Implemented custom PyTorch C++/CUDA extensions to optimize model inference throughput on HPC systems.
- Benchmarked ReFT's efficiency gain (10.6%) over IA3 on LiveCodeBench against IA3's lightweight adaptation.

AI Agent based Latex Editor Web App — *Apr 2025-May 2025*
Python, ReactJS, JavaScript, MongoDB, FastAPI, LangChain, Agentic AI, Monaco Editor
- Developed an AI-Agent integrated Latex Editor using React.js, FastAPI, MongoDB, OpenAI API and LangChain to automatically edit the Latex Script based on user input and search queries.

Aether - Research Assistant App — *Sep 2024-Dec 2024*
Electron.js, Rust, Node.js, Firebase (Auth & Firestore), Google Gemini API
- Developed a cross-platform floating assistant in Rust/Electron for real-time screen analysis and interactive research workflows with Google Gemini API.

PUBLICATIONS

- Ensemble-Based Model-Agnostic Meta-Learning with Operational Grouping for Intelligent Systems, *MDPI, 2024*.

TECHNICAL SKILLS

Languages/Databases: Python, C++, Java, Rust, JavaScript, Typescript, SQL, MongoDB, Ruby, PostgreSQL, MySQL

Web/App Development: HTML/CSS, SpringBoot, Angular, React, Node.js, Flask, Ruby on Rails, Flutter, GraphQL, RESTful API

Infrastructure/DevOps: Docker, Kafka, Kubernetes, Terraform, Git, Jenkins, Linux, CI/CD pipeline

AI/Cloud Technologies: Scikit-Learn, PyTorch, Deep Learning, LLM, RAG, NLP, CNN, AWS(SageMaker), GCP(Vertex AI)