Devam Mondal

908-449-7970 | devammondal123@gmail.com | linkedin.com/in/dmmo | github.com/Dodesimo | U.S. Citizen

EDUCATION

Georgia Institute of Technology — May 2027
B.S. in Computer Science | Concentrations: Intelligence + Modeling and Simulation | GPA: 3.93 — Atlanta, GA
- **Involvement:** Grand Challenges LLC, Machine Learning Club, Big Data Big Impact, Karting Club
- **Coursework:** Data Structures & Algorithms, Objects & Design, Artificial Intelligence, Machine Learning, Computer Vision, Computer Organization, Computer Systems & Networks

TECHNICAL SKILLS

Languages & Developer Tools: Python, Java, Go, C++, C, JavaScript, Git, Docker, Linux, Vercel, Terraform
Frameworks & Libraries: Flask, FastAPI, Django, React, Node.js, Express.js, OpenCV, LangChain, HuggingFace
Databases & Cloud Platforms: MongoDB, SQL, Firebase, AWS EC2 + SageMaker + S3 + SQS, Google Cloud
Concepts: Agile, REST APIs, Databases, LLMs + RAG, Distributed Systems, Information Retrieval, ML (PyTorch)

EXPERIENCE

Millennium Management — December 2025 – Present
Incoming Software Engineer Intern — Miami, FL
- Incoming C++ engineer on a latency-critical trading team building real-time market-data ingestion systems

Cisco — May 2025 – Present
Software Engineer Intern — Atlanta, GA
- Prototyped and deployed **4** Go API endpoints with Terraform and Jenkins to explore Gmail message remediation for the Secure Email Threat Defense team, expanding product reach
- Implemented a bulk remediation and reclassification pipeline using two AWS SQS queues and AWS Lambda workers, deploying to beta production and empowering administrators to process **5,000+** emails concurrently
- Engineered a time-throttle mechanism for bulk remediation requests using a AWS DynamoDB audit table, increasing backend reliability and reducing system load spikes by **15%**
- Integrated **5** LLM engines to increase spam classification accuracy, improving remediation efficiency

InPharmD *(YC W21)* — Feb 2025 – Apr 2025
Artificial Intelligence Engineer Intern — Atlanta, GA
- Refactored two web-scraping utility APIs using FastAPI and SERP, reducing execution time by **30%** to **19 seconds** while considering **50+** online vetted sources
- Designed a novel deep-research agentic workflow with TF-IDF document search using OpenRouter, MongoDB, and AWS EC2, generating **20+** page drug reports for pharmacists through LLMs and cutting diagnosis times by **80%**
- Developed a custom reverse-search algorithm mapping patient conditions to appropriate drug dosages, improving deep research result validation

Hackensack Meridian Health — May 2024 – Feb 2025
Software Engineer Intern — Remote
- Streamlined single-cell RNA sequencing workflow time by **85%** with a React web application hosted on an Google Compute Engine instance, boosting research productivity
- Implemented a PostgreSQL database schema for storing single-cell RNA sequencing metadata and results, enabling efficient querying of **10,000+** cell annotations and streamlining data retrieval by **60%**

PROJECTS

C++ Transformer w/ Inference Endpoint | *C++, Eigen, Crow*
- Implemented a Transformer language model from scratch in C++, including single-head attention, cross-entropy loss, and layer normalization, using Eigen matrix operations
- Built a custom word-level tokenizer that constructs vocabularies of **5,000+** tokens from **10,000** lines of raw text using `std::vector` and `std::unordered_map`
- Designed a binary model serialization pipeline and engineered a REST inference endpoint with Crow, enabling real-time text generation from saved model weights

Toffee | *React, FastAPI, Firebase, Firestore, OpenRouter, Vercel*
- Developed a full-stack AI study tool with React and FastAPI that converts documents into quizzes and flashcards
- Designed five API endpoints to convert PDFs, plaintext, and YouTube transcripts into JSON-formatted quiz questions using Google Gemini Flash and Regex, deployed on Vercel
- Used Firestore to save user sessions and **100+** generated flashcards across MCQ, T/F, FITB quiz formats