

Shoumik Majumdar

Website: [Shoumik Majumdar](#) / LinkedIn: [Shoumik Majumdar](#) / Email: shoumikmajumdar@gmail.com

Skills

- **Data Languages:** Python, C, Java, R, Shell Script, SQL.
- **Databases:** SQL Database (MySQL, Google Cloud SQL), NoSQL database (Cloud Firestore), Vector Databases (Vertex AI Vector Search, FAISS, chromaDB, pgvector), Data warehouse (BigQuery).
- **Tools/Platforms:** Deep Learning (PyTorch, TensorFlow), Machine Learning (Keras, Pandas, Numpy, OpenCV, Scikit-learn, MLFlow), DevOps (Git, Docker, Cloud Build), Generative AI (LLMs, LangChain, Prompt Engineering, RAG frameworks, NL2SQL, Agents, PEFT), Google Cloud Platform.

Professional Experience

Senior Machine Learning Engineer (*Quantiphi Inc*) Boston, MA 2024 - present

- Architected and implemented a robust **document classification** and **entity extraction** pipeline for a **public sector insurance fund**, enhancing document processing efficiency for **300+ analysts**.
- Orchestrated highly scalable pipelines on Google Cloud Platform to process and serve **300K** documents per month resulting in **90% reduction** in manual effort. Stack: **Python, Document AI, Vertex AI, Cloud Run, Cloud Functions**.
- Engineered a **Retrieval Augmented Generation (RAG)** solution for a **software company's** business communication platform. **Enhanced user experience** by enabling swift **video search** and navigation using relevant timestamps.
- Conceived a data processing pipeline to **transcribe videos** and apply advanced chunking techniques, **enhancing relevance** of timestamps and **reducing hallucination** through grounding. Stack: **Python, LangChain, Google Video Intelligence, Vertex AI Vector Search**.
- Created a **Multi-modal RAG**-based chat prototype for a **global construction company**, facilitating rapid access to equipment details to O&M workers from **10K pages** of multi-modal content.
- Enhanced operational efficiency by **decreasing equipment downtime** and improving employee onboarding procedures and garnered **positive feedback** from O&M workers for the prototype. Stack: **Python, LangChain, Unstructured, Vertex AI Vector Search, Cloud Run**.
- Engineered a **NL2SQL** prototype, empowering **manufacturing** category managers with actionable data insights, diminishing reliance on data scientists, fostering **faster decision-making processes**. Stack: **Python, BigQuery, Vertex AI Vector Search**.

Machine Learning Engineer (*Quantiphi Inc*) Boston, MA 2021 - 2023

- Developed and deployed **end-to-end ML framework** on Google Cloud Platform for a semiconductor manufacturing company, reducing average workflow runtime from **96 hours to 8 hours**.
- Led development of a **model monitoring framework** aimed at detecting and addressing drift, automating model retraining and deployment processes to ensure model freshness in production.
- Orchestrated the **migration** of client's on-prem services to fully managed, scalable and cost-efficient frameworks on GCP **ensuring future readiness** and improved **resource utilization**. Stack: **Python, Docker, Kubeflow, Cloud Composer, Vertex AI**.
- Created an **image segmentation** model using GCP's **Visual Inspection AI** platform for real-time detection of knots in yarn and automatic shutdown of tufting machines **reducing downtime** by **76%**.
- Developed and deployed ingestion and inference pipelines on the edge using **Kubernetes** to ensure **µs latency**. Stack: **Python, Visual Inspection AI, Kubernetes**.
- Designed research studies, documenting insights, leading to a **publication**, a **patent application** and AI accelerator development within the Applied Research team.

Education

Master of Science, Computer Science *Boston University* Boston, USA 2019-2021
Bachelor of Engineering, Computer Science *University of Mumbai* Mumbai, India 2014-2018

Online Courses & Certifications

- Databricks Certified Machine Learning Professional. 2023
- Databricks Certified Machine Learning Associate. 2023
- Google Cloud Certified - Professional Machine Learning Engineer. 2022
- Google Cloud Certified - Associate Cloud Engineer. 2021

Publications

- [Ani-GIFS](#) : A benchmark video dataset for domain generalization on action recognition task in GIFs.
- [Physics-informed neural networks for modeling astrophysical shocks](#).