Shoumik Majumdar

Website: Shoumik Majumdar • LinkedIn: Shoumik Majumdar • Email: shoumikmajumdar@gmail.com

EDUCATION

Boston University, Boston, USA Sep 2019 - Jan 2021

- Master of Science, Computer Science

University of Mumbai, Mumbai, India Aug 2014 - May 2018

- Bachelor of Engineering, Computer Science

SKILLS

- **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.

WORK EXPERIENCE

Quantiphi Inc: Senior Machine Learning Engineer, Boston Jan 2024 - Present

- Designed and implemented a robust **document classification** and **entity extraction** pipeline on Google Cloud for a **public sector insurance fund**, enhancing document processing efficiency.
- Orchestrated highly scalable pipelines on Google Cloud Platform to efficiently process and serve **300K** documents per month resulting in **90% reduction** in manual effort. Stack: **Python, Document AI, Vertex AI, Cloud Run, Cloud Functions.**
- Designed and developed a **Retrieval Augmented Generation** (RAG) based solution for a **software company** with a business communication platform to **search for key moments** in a videos and guiding users to those moments with relevant timestamps.
- Implemented 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 **Multimodal RAG**-based chat prototype for a **global construction company**, facilitating rapid access to equipment details to O&M workers from **10K pages** of mulitmodal content.
- Received **positive feedback** from O&M workers for the prototype's efficiency, contributing to **decreased equipment downtime** and improved employee onboarding procedures. Stack: **Python, LangChain, Unstructured, Vertex AI Vector Search, Cloud Run**.
- Designed and developed a **NL2SQL** prototype to empower **manufacturing** category managers with data driven insights, reducing dependency on data scientists and **accelerating decision-making processes**. Stack: **Python, BigQuery, Vertex AI Vector Search**.

Quantiphi Inc: Machine Learning Engineer, Boston Jun 2021 - Dec 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 onpremise 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.
- Developed and deployed ingestion and inference pipelines on the edge using **Kubernetes** to enable automatic shutdown of tufting machines **reducing downtime** by **76%**. Stack: **Python, Visual Inspection AI, Kubernetes**.
- Designed research studies, documenting insights, leading to **publications** and AI accelerator development within the Applied Research team.

Boston University: Research Associate (Machine Learning), Boston Sep 2019 - Jun 2021

- Conducted research in computer vision, focusing in domain adaptation and domain generalization.
- Published the first human action recognition video dataset for domain generalization in a peer-reviewed computer science journal. Stack: **Python, Pytorch**.

PUBLICATIONS

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

PROFESSIONAL CERTIFICATIONS

- Databricks Certified Machine Learning Professional. Jun 2023
- Databricks Certified Machine Learning Associate. May 2023
- Google Cloud Certified - Professional Machine Learning Engineer. Oct 2022
- Google Cloud Certified - Associate Cloud Engineer. Dec 2021