Sakib Ahmed

Data Scientist & Full-Stack Developer

sakibahmed2018go@gmail.com
https://sakib.qbitlab.tech/
sakib-ahmed-6339081a7
Sakib323 Sakib323

Education

2018 • **SSC in Science,** The Aided High School, Sylhet **— GPA: 5.00/5.00**

2021 • **HSC in Science,** Blue Bird School & College, Sylhet **— GPA: 5.00/5.00**

2029 • **Bachelor of Computer Science,** Richmond American University London, 566 Chiswick High Road, London, W4 5AN

Professional Experience

Jan 2023 - Nov 2025

Data Scientist, QBitLab, East Zindabazar, Sylhet- 3100, Bangladesh

- Fine-tuned Llama 2/Mistral models using QLoRA for domain-specific tasks.
- Developed a RAG pipeline to allow internal SaaS products to query private documentation, utilizing Pinecone for vector storage.
- Optimized LLM inference using vLLM and quantization techniques (INT8/FP16) to ensure smooth performance on the resource-constrained environments.
- Contributed to the Qbitlab's participation in the Smart Accelerator Bangladesh (SBA) and successfully organized three public product showcasing events at Metropolitan University, Osmani Medical High School, Agragami Girls School, and Bluebird High School.
- Engineered the migration of an internal SaaS product to AWS, cutting infrastructure costs by 22% through optimized resource allocation and improved the API response time by 33% across the platform.
- Mentored junior developers on data-centric application design, clean code practices, and ML model integration, ensuring rapid and robust delivery across agile sprints.
- Deployed lightweight CNN models (MobileNet) within React Native apps for real-time image classification, ensuring on-device processing to maintain privacy and speed.

Feb 2021 - Nov 2022

AI Engineer, Neuron AI, Motijheel, Dhaka - 1000, Bangladesh

- Designed and deployed custom GenAI features for e-commerce clients, including automated product description generators and AI chatbots to improve customer conversion rates.
- Engineered robust ETL pipelines using Apache Airflow to automate data ingestion and transformation, ensuring high-quality data availability for downstream predictive modeling.
- Architected data warehouse solutions on AWS (Redshift/S3) and real-time streaming systems, optimizing data retrieval speeds for high-traffic e-commerce dashboards.
- Developed cross-platform mobile applications (React Native /Firebase) that integrated real-time ML inference, bridging the gap between complex data models and user-friendly mobile interfaces.

2020-2025

AI Engineer & Full stack developer, Fiverr

- Developed several Dense Neural Network models alongside complete frontend and backend architectures for diverse client requirements and specialized tasks. For example, I engineered a Geospatial Intelligence system for building extraction, processing 5GB+ of multi-spectral TIFF files through a high-performance data pipeline. I developed a CNN model for semantic segmentation alongside Random Forest and KNN ensembles for land-use classification. Finally, I architected a self-hostable web platform featuring interactive maps and real-time spatial analytics for data-driven decision-making.

Research Experience

May 2025

A Deep Neural Network Approach for Highly Efficient Lightning Prediction and Warning Systems
The absence of accessible and accurate early warning systems for lightning strikes poses a critical challenge, leading to severe consequences such as loss of life. Addressing this issue requires an innovative, integrated solution. My study investigates the development of a highly accurate lightning prediction model using predictive artificial intelligence, specifically Deep Neural Networks (DNNs), and essential meteorological parameters. This paper also has been published at the prestigious **8th International Conference on Engineering Research, Innovation and Education (ICERIE 2025).**

Jun 2025

Reimagining LLMs: Matrix Multiplication-free Language Model, Metropolitan University, Sylhet, Bangladesh
Constructed a Matrix Multiplication-free Language Model with ternary weights {-1, 0, 1} in order to develop a new LLM architecture that is more cost-effective, lightweight, and scalable, and can run on edge devices, IoT systems, or ultra-low-power applications while also maintaining performance comparable to that of state-of-the-art language models.

Teaching Experience

Jan 2024 - Mar 2025 **Basics of Machine Learning with Hands-On STEM Education Experience - Course Instructor and Mentor, FTP BOOTCAMP**

- Designed and evaluated theoretical and practical assignments on various basic beginner-friendly machine learning models with real-world ML projects like spam detection and house price prediction.
- Mentored sixteen pairs of students on STEM projects, providing hands-on experience with microcontrollers through DIY projects such as smart home automation, line-following robots, and weather monitoring systems.

Jan 2024 - Aug 2025 **Advisor & Mentor teaching mathematical framework and it's application in Machine Learning Models, MUGAS**

- Mentored university students on core mathematical concepts in machine learning, teaching linear algebra applications like linear regression (y = mx + c), explaining terms (m as slope, x as input), the role of chain rule in backpropagation, challenges like gradient vanishing, and mathematical techniques to overcome these issues.

Projects

Apr 2025 **Autonomous Surveillance Drone**

Designed and deployed drone system for real-time disaster situation analysis and automated reporting. At the core is a fine-tuned Vision Transformer (ViT-GPT2), fine tuned FP16 Quantization to captions live video frame. The model is then hosted on Azure Machine Learning (NCv3-series), exposing a REST API for high-speed inference. A Raspberry Pi acts as the edge gateway, encoding video with FFmpeg and transmitting frames to the Azure endpoint. The drone system utilizes Flask and Ngrok for bidirectional communication, allowing for low latency remote motor control over the internet.

Jun 2024 - Jul 2025 **Fine tuned open source LLM**

Fine tuned Open Source LLMs and Vision Transformers, including Llama 2/3 (8B/7B) and IBM Granite, optimizing them for specialized downstream tasks such as Instruction Tuning, Market Analysis, and Automated Code Generation. Utilizing PEFT techniques and QLoRA (4-bit/16-bit), I reduced computational overhead while maintaining high benchmark performance. These models are production-ready and hosted on my Hugging Face repository.

Aug 2024 - Sep 2024 **AI Marketplace**

A platform like an app store for AI models, where users can try my fine-tuned LLMs (like LLaMA 2, LLaMA 3, GRANITE, GPT) and models from Hugging Face. It includes a drag-and-drop playground to build custom AI pipelines like combining text summarization, speech-to-text, and PDF reading. Users can export their workflows as standalone apps like APKs to run on devices like Android.

Oct 2023 - Dec 2023 **Sustainable Disaster Response Alert Mechanism**

Developed for NASA Space Apps Challenge 2023, SDRAM is a disaster communication system that uses LoRa (432 MHz) modules to form a mesh network with up to 12 km range. The network connects to a custom Android app (built with Android Studio) to enable real-time alerts and peer-to-peer communication, even without internet. I also trained a deep neural network to predict lightning events, hosted on Azure, with API integration into the app. The system scales as more LoRa nodes join and was selected as a Global Finalist in the competition.

Aug 2023 - Sep 2023 **Electronic Health Record (EHR) management system**

Built a full-featured EHR system using Android Studio (Java) for managing patient data like medical history, medications, lab results, and more. Integrated OCR for extracting text from prescriptions and used Gemini for intelligent analysis. Trained a custom Vision Transformer (ViT) model for X-ray report analysis, accessible directly from the app. The system includes appointment scheduling, prescription management, and billing, offering a complete and user-friendly solution for healthcare providers.

Apr 2025 **Hear Helper is an AI-powered assistive listening app**

Developed Hear Helper, an AI-powered assistive listening application using Whisper and ChatGPT, transcribing, summarizing, and localizing critical environmental sounds with directional and decibel data for 50k+ users.

Mar 2023 **Vitro- A VR learning platform in Unity(C#)**

Vitro is like a special virtual reality (VR) world (platform) just for learning. It's a place where students and teachers can come together in VR to study and learn collaboratively. In the Vitro store, users can download tools like VR Chemistry Labs, 3D viewers, AI PDF viewers, Virtual Computers, AI voice notes, and more to help them with their studies.

Jun 2024 **Space AR - Solar System Tour**

Space Ar is an augmented reality app that helps users explore the solar system. I had used Unity using C# to develop this app. The app allows users to view planets in 3D and learn about their characteristics. It also includes a feature that lets users take pictures of the planets and share them on social media.

Scholastic Achievements

- **Global Finalist at NASA International Space App Challenge 2023** and developed SDRAM, or Sustainable Disaster Response Alert Mechanism, had revolved around diverse fields from Computer Science and Data Science. I had trained a sequential Deep Neural Network model to predict lightning strikes based on changes in environmental parameters such as temperature, humidity, wind speed etc , along with the development of a half-duplex LoRa communication system tuned at 432 MHz as an alternative communication method during disasters — and lastly, an application (Android Studio and Java) that can access all those features. View Project Here

- **Global Nominee at NASA International Space App Challenge 2024** and built an AR app in Unity (C#) for exoplanet exploration to make learning fun for kids. The app features 3D object interaction, solar system visualization, and exploration of celestial bodies. View Project Here

- **Top 100 global finalist at IBM TechXchange Pre-Conference watsonx Hackathon 2024**. Fine-tuned open-source LLMs like Llama 3, Granite to boost productivity. I built a React Native app hosting these and third-party models, creating a Play Store-like marketplace where users can drag and drop models and tools to build pipelines. View Project

- **In the Microsoft Imagine Cup 2023**, I served as the team lead and lead technologist, securing $10,000 in funding for QbitLab. Our MVP product was Vitro, a VR learning platform (Unity/C#, AI, Azure) where students and teachers collaborate. Its store offers tools like VR Chemistry Labs, 3D viewers, AI PDF readers, virtual computers, and AI voice notes to enhance learning. View Project

Specialized/Research Topics:

- Transformer Architectures — ViT, DiT, LLMs (including matmul-free/ternary transformer variants)
- Convolutional & Hybrid Vision Models — CNN backbones integrated with Transformer heads
- 3D VAE and latent space modeling
- Diffusion Models — DiT-based diffusion pipeline for 3D.
- Multi-modal learning — image-text-shape triad.
- Cross & self-attention Transformers — used in your 3D VAE and DiT.
- Fine-tuned open-source foundation models for domain-specific tasks with FP16/INT8 optimization.

Courses

[Statistical Learning]	Advanced Deep Learning, Deep Learning, Machine Learning, Natural Language Processing, Reinforcement Learning, Multi-Armed Bandits, Probabilistic Graphical Models, Computational Models of Cognition
[Curriculum]	Computer Networks, Database Systems, Operating Systems, Data Structures and Algorithms,Object-Oriented Programming
[Mathematic]	Probability-Statistics-Stochastic Processes, Discrete Mathematics, Linear Algebra, Graph Theory

Skills

Languages	C, C++, C#, Java, Python, HTML, CSS, Javascript, Web Assembly Unity, ARCore, Android Studio, Stanford CoreNLP, Git,
Tools	Bootstrap, jQuery, Emscripten, Blaze, j2Cl
Libraries	NLTK, django, scipy, pandas, sklearn, gensim, keras, tensorflow, pytorch