

Md. Shohanur Islam Sobuj

Machine Learning Engineer · MLOps · LLMs · Multimodal AI

📍 Köln, Germany | ✉ contact@shohanursobuj.dev | 🌐 shohanursobuj.dev
📄 shohanursobuj | 🔄 shohanursobuj | 🎓 Google Scholar

SUMMARY

Machine Learning Engineer with 6+ years delivering production ML systems across MLOps, LLMs, RAG, multimodal AI, NLP, and computer vision. Currently at Anymate Me GmbH in Köln, building an agentic slide-to-video pipeline (PDF/PPTX → RAG agent → TTS → avatar lip-sync) with p95 latency under 3 minutes and 99.4% uptime. Published researcher with 12 peer-reviewed papers across Nature, IEEE, ACL, and NeurIPS venues — 225+ Google Scholar citations, h-index 8.

WORK EXPERIENCE

Machine Learning Engineer

Anymate Me GmbH

Köln, Germany

Dec 2024 – Present

- Architected end-to-end agentic slide-to-video pipeline: PDF/PPTX → RAG content agent → TTS → avatar lip-sync, achieving p95 latency < 3 min for 10-slide decks.
- Built PPTX content rewriting agent using LLM + RAG, enabling users to restructure and localise presentation content without manual editing.
- Implemented groundedness validation (DeBERTa-v3 NLI) on generated scripts, maintaining 96.2% auto-approval rate in production.
- Designed MLOps pipelines on GCP — model versioning in MLflow, canary deploys, automated rollback on metric regression.
- Reduced pipeline error rate to < 4% through stage-level observability (Grafana + Loki) and per-job retry logic.

Machine Learning Engineer

Business Automation Ltd.

Dhaka, Bangladesh

Nov 2023 – Oct 2024

- Implemented Change Data Capture (CDC) pipeline with MySQL, Debezium, Apache Kafka, and Zookeeper — enabling real-time data sync across microservices.
- Developed SmartRemarks NLP system for automated content analysis and sentiment classification, handling 10K+ requests/day in production.
- Built OCR-based TIN certificate validation system using computer vision, achieving 95%+ extraction accuracy and reducing manual processing time by 80%.
- Designed event-driven microservices architecture for real-time ML inference, standardising deployment workflows across environments.
- Led migration of ML-backed services to containerised deployments (Docker + Kubernetes).

Machine Learning Engineer

Anchorblock Technology LLC

Dhaka, Bangladesh

May 2022 – Oct 2023

- Led design, development, and delivery of 13 AI/ML products across conversational AI, NLP, computer vision, and FinTech.
- Architected LLM-powered enterprise knowledge bots using advanced RAG techniques; shipped **Finix** (AI banking assistant) to Google Play.
- Built distributed ML infrastructure handling millions of data points daily with auto-scaling on AWS (EC2, ECS, S3).
- Implemented CI/CD pipelines for ML model deployment using GitHub Actions — reduced deployment time by 60%.
- Led technical architecture decisions for distributed computing systems across 3 concurrent client projects.

Freelance ML Engineer

Self-Employed

Remote

Jan 2019 – Apr 2022

- Delivered 30+ ML projects for international clients across Europe, North America, and Asia.
- Specialised in NLP (text classification, sentiment analysis, NER) and computer vision (object detection, image classification).
- Built end-to-end data pipelines and predictive models for e-commerce, healthcare, and finance clients.

EDUCATION

Hajee Mohammad Danesh Science and Technology University (HSTU) Dinajpur, Bangladesh
B.Sc. (Engineering) in Electrical and Electronic Engineering

- **Thesis:** Diagnosis of Diabetic Retinopathy Based on Convolutional Neural Networks Using Optical Coherence Tomographic Images.
- Research focus: Natural Language Processing, Deep Learning, Bangla language models. Published 4 papers during undergraduate studies.

TECHNICAL SKILLS

LLMs & Generative AI: RAG, PEFT, LoRA, QLoRA, Prompt Engineering, LangChain, Hugging Face Transformers, Fine-Tuning, Embeddings, Vector Search


MLOps & Infrastructure: MLflow, Docker, Kubernetes, GCP, AWS, GitHub Actions, CI/CD, Apache Kafka, Debezium, Apache Airflow

ML & Deep Learning: PyTorch, TensorFlow, Scikit-learn, Transformers, Computer Vision, NLP, OCR, Time-Series Forecasting, TTS

Programming & APIs: Python, SQL, FastAPI, Flask, REST APIs, JavaScript

Databases & Storage: PostgreSQL, MySQL, MongoDB, Redis, Qdrant, Elasticsearch

PUBLICATIONS

 Google Scholar: 225+ citations · h-index 8 | 12 peer-reviewed papers

LLM-Mixer: Multiscale Mixing in LLMs for Time Series Forecasting. M. Kowsher, **M.S.I. Sobuj**, N.J. Prottasha, et al. *NeurIPS Workshop*, 2025. [15 citations]

PEFT A2Z: Parameter-Efficient Fine-Tuning Survey for Large Language and Vision Models. N.J. Prottasha, ..., **M.S.I. Sobuj**, ..., M. Kowsher, et al. *arXiv*, 2025. [14 citations]

OCR-Enhanced Digital Signatures for Tamper-Proof Document Integrity Verification. M.M. Uddin, **M.S.I. Sobuj**. *QPAIN 2025 (IEEE)*, 2025.

Securing Electric Vehicle Performance: Machine Learning-Driven Fault Detection and Classification. **M.S.I. Sobuj**, et al. *IEEE Access*, 2024. [60 citations]

Parameter-Efficient Fine-Tuning of Large Language Models Using Semantic Knowledge Tuning. N.J. Prottasha, A. Mahmud, **M.S.I. Sobuj**, et al. *Scientific Reports (Nature Publishing Group)*, 2024. [36 citations]

Leveraging Pre-trained CNNs for Efficient Feature Extraction in Rice Leaf Disease Classification. **M.S.I. Sobuj**, M.I. Hossen, M.F. Mahmud, M.U.I. Khan. *iCACCESS 2024 (IEEE)*, 2024. [11 citations]

L-TUNING: Synchronized Label Tuning for Prompt and Prefix Tuning in LLMs. M. Kowsher, **M.S.I. Sobuj**, A. Mahmud, N.J. Prottasha, P. Bhat. *Tiny Papers @ ICLR 2024*. [8 citations]

Contrastive Learning for Universal Zero-Shot NLI with Cross-Lingual Sentence Embeddings. M. Kowsher, **M.S.I. Sobuj**, N.J. Prottasha, M.S. Arefin, Y. Morimoto. *Findings of EMNLP 2023*. [3 citations]

An Enhanced Neural Word Embedding Model for Transfer Learning. M. Kowsher, **M.S.I. Sobuj**, M.F. Shahriar, N.J. Prottasha, M.S. Arefin, P.K. Dhar, T. Koshiba. *Applied Sciences (MDPI)*, 2022. [43 citations]

A Classical Approach to Handcrafted Feature Extraction for Bangla Handwritten Digit Recognition. M.F. Wahid, M.F. Shahriar, **M.S.I. Sobuj**. *ICECIT 2021 (IEEE)*. [23 citations]

BanglaLM: Data Mining Based Bangla Corpus for Language Model Research. **M.S.I. Sobuj**, et al. *ICIRCA 2021 (IEEE)*. [9 citations]

An Efficient Approach on Sentiment Analysis of Bangla Social Media Data Using FastText. **M.S.I. Sobuj**, et al. *Research on Computational Language*, 2021. [2 citations]

LANGUAGES

English: Professional Working (C1) | **German:** Basic (A1) | **Bengali:** Native