

Md. Mahinuzzaman Shaan

Dhaka, Bangladesh | +880 177 5780238 | mahinuzzaman.shaan@gmail.com | shaanzeeee.tech | linkedin.com/in/mahinuzzaman-shaan | github.com/shaanzeeee

Professional Summary

AI systems architect with 1+ year production experience shipping LLM-powered workflows, RAG platforms, and voice agents that reduced manual overhead by 40% and improved retrieval accuracy by 25%. Proven track record scaling enterprise automation across 15+ integrated systems and delivering measurable business impact through technical leadership.

Technical Skills

AI/ML:

Retrieval-Augmented Generation (RAG), Large Language Models (LLM), LangChain, LangGraph, NLP, Pinecone, ChromaDB, TensorFlow, PyTorch, Computer Vision

Languages:

Python, JavaScript, TypeScript, SQL, C++, Bash

Backend & DevOps:

FastAPI, Node.js, Express, Docker, Inngest, Selenium, PostgreSQL, MongoDB, REST APIs, API orchestration, Kubernetes, AWS

Frameworks & Tools:

React, Pandas, NumPy, Scikit-learn, Tableau, Streamlit, Plotly, Git, Linux, n8n

Professional Experience

AI Engineer

Softvence Agency

Mar. 2026 – Present

Dhaka, Bangladesh

- Architected production RAG systems increasing retrieval accuracy by 25% through Pinecone hybrid search optimization, enabling 3x faster document processing for enterprise clients while reducing infrastructure costs by 18%.
- Orchestrated 15+ automated enterprise workflows using n8n and Python, reducing manual overhead by 40% (20+ FTE hours/month) via API orchestration and custom LLM agents, improving workflow completion times by 35%.
- Engineered low-latency AI voice agents leveraging Whisper and ElevenLabs, achieving sub-second response times, processing 500+ concurrent conversations with 98% uptime and 92% user satisfaction across production infrastructure.

Data Scientist Trainee

Mastercourse Bangladesh

Nov. 2025 – Mar. 2026

Dhaka, Bangladesh

- Launched end-to-end image recognition system using fastai and PyTorch, achieving 94% accuracy on custom datasets and processing 50,000+ images through a transfer learning and augmentation pipeline.
- Accelerated machine learning life cycles by implementing experiment tracking infrastructure, reducing model iteration cycles by 30% (12 days → 8.4 days) through systematic hyperparameter tuning.
- Optimized collaborative filtering engines, improving tabular model baselines by 15% using Random Forests and deep learning ensembles across 100K+ user interactions.

Projects

LeaseLens | *FastAPI, LangGraph, Multi-Agent AI*

[GitHub](#) | [Live](#)

- Engineered a multi-agent AI pipeline using FastAPI and LangGraph to automate commercial real estate due diligence, enabling rapid extraction and underwriting of financial metrics from complex document repositories.

Twine | *RAG, Google Drive Integration, Persona Intelligence*

[GitHub](#) | [Live](#)

- Developed a sophisticated persona intelligence engine utilizing hybrid RAG architecture and Google Drive integration to deliver high-fidelity, context-aware responses with real-time knowledge curation and administrative oversight.

PC Parts Classifier | *PyTorch, FastAI, ViT, Hugging Face*

[GitHub](#) | [Live](#)

- Architected an end-to-end computer vision pipeline using PyTorch and FastAI to classify hardware across 11 categories, deploying a ViT-Base model that achieved 73.89% accuracy on held-out test data.

RatePunk | *Next.js, Prisma, RAWG API*

[GitHub](#) | [Live](#)

- Created a cyberpunk-themed video game rating platform using Next.js and Prisma, integrating the RAWG API to crowdsource performance benchmarks and monetization metrics for over 500,000 titles.

Education

Bachelor of Science in Computer Science

BRAC University

2025

CGPA: 3.56/4.00 (Distinction)