

ADIL SHAMIM

AI Engineer

+880 1321073452 · adilshamim696@gmail.com · adilshamim.me
github.com/AdilShamim8 · linkedin.com/in/adilshamim8 · kaggle.com/adilshamim8

SUMMARY

AI Engineer with 3 years building and shipping production LLM systems, RAG pipelines, and multi-agent workflows. Founder of ReWoo (paying UK client 10 days after launch) and Toolly (400+ AI tools, open-sourced); now building a production Finance AI Agent at a UK-based fintech startup. 1st-author paper on Bengali speaker diarization — 56% inference speedup (BUET CSE Fest 2026). Kaggle Master — Top 1% competition finish (29/4,082). Earlier, shipped a recommender system that lifted client sales 10% in 3 months.

SKILLS

Agentic AI: LangChain · LangGraph · ReAct & orchestrator-worker agents · tool / function calling · MCP & A2A protocols · memory & state management · agent evaluation & guardrails

LLM & Generative AI: RAG architecture · multi-provider orchestration (OpenAI, Anthropic, Gemini, Ollama) · fine-tuning (LoRA / PEFT, SFT, RLHF) · knowledge distillation · prompt engineering · RAG evaluation (RAGAS)

ML/DL & Speech AI: PyTorch · TensorFlow · Hugging Face Transformers · speaker diarization · Whisper · wav2vec2 · pyannote.audio · WeSpeaker · XGBoost · LightGBM · scikit-learn

MLOps & Deployment: Docker · Kubernetes · vLLM · FastAPI · REST APIs · Qdrant / Chroma / FAISS · Redis · Prometheus · Grafana · LangSmith / Langfuse · MLflow · GitHub Actions CI/CD · pytest · AWS / GCP

Programming: Python · SQL · JavaScript · Streamlit · Flask · Git

EXPERIENCE

Founder & AI Engineer — ReWoo · rewoo.tech

Jul 2026 – Present

- Founded an AI-workforce startup deploying teams of 10 role-scoped AI executives — sales, support, finance, marketing, operations, analytics — that run business functions autonomously for early-stage founders.
- Landed a paying UK client 10 days after founding; automated 3 business functions — marketing, HR operations, and collections — generating international revenue from day one.
- Engineered an agent quality gate that vets 100% of customer-facing output against a quality benchmark before release.

AI Engineer — UK-based FinTech Startup · Remote

Apr 2026 – Present

- Building a production Finance AI Agent on a small founding team — agent architecture, tool-calling over financial data sources, and retrieval pipelines for grounded, auditable answers.
- Own end-to-end delivery across the agent stack: prompt and tool design, evaluation, guardrails, and deployment.

Founder & AI Engineer — Toolly · toolly.tech (open-sourced)

Jun 2025 – Mar 2026

- Designed, shipped, and ran an AI-tools discovery platform solo — 400+ curated tools across 15 categories with a community submission and moderation pipeline, search and filtering, a Learn AI hub, and usage analytics.
- Built Toolly Studio, a Streamlit + Bria AI image-generation app with batch export and Docker packaging; open-sourced the full platform on GitHub after 11 months of iteration.

PROJECTS

Multimodal RAG Production github.com/AdilShamim8/Multimodal_RAG_Production

FastAPI · Qdrant / Chroma / FAISS · Hugging Face Transformers · Docker · Prometheus · pytest

- Re-architected a 71-cell research notebook into a modular, production-grade multimodal RAG package — swap embedding, reranking, and generation providers or vector stores (Chroma, Qdrant, FAISS) via a single env var.
- Serves text- and image-query retrieval over 10,096 image+text pairs via FastAPI with rate limiting, caching, retries, and Prometheus metrics; 11-module pytest suite, docker-compose stack (API, Qdrant, Redis, Grafana), CPU-only mock mode.

QuantScope — Quantitative Stock Analysis Platform github.com/AdilShamim8/QuantScope

Python · FastAPI · LangChain · yfinance · Docker · pytest

- Compresses 2–4 hour manual stock analyses to ~45 seconds across 35+ global exchanges with exchange-aware annualization; a 6-provider LLM fallback chain (OpenAI, Anthropic, Google, Ollama, Mistral, Cohere) degrades gracefully to template output.
- Enforces strict separation — deterministic indicator and risk math, LLM for explanation only; ATR risk engine (position sizing, stop-loss, 2% per-trade cap) validated by 33 unit tests; Dockerized (non-root, health checks).

PUBLICATION

Bangla Diarizz: Domain-Adapted Speaker Diarization for Bengali Long-Form Audio — 1st Author

BUET CSE Fest 2026 · DL Sprint 4.0 — paper selected for on-site presentation (107 teams) · ResearchGate

- Fine-tuned pyannote segmentation with WeSpeaker ResNet34-LM embeddings and knowledge distillation for Bengali long-form audio — 56% inference speedup (1h 22m to ~36m across 14 test files), 3.4× real-time on CPU, DER 0.19 (dev).

LEADERSHIP & AWARDS

- Winner — AI Hackathon, DUET CSE Carnival 2026:** national festival with 1,200+ students from 100+ universities.
- Kaggle:** Notebooks Master (ranked 167 of 62,365) · Top 1% competition finish (29/4,082, Road Accident Risk).
- Startup Funding — ITBI Pitch Fest 2026, CUET:** pitched ReWoo and secured funding at Bangladesh's first campus incubator.
- Open Source:** 610+ GitHub stars across 58 public AI/ML repositories (github.com/AdilShamim8).

EDUCATION

B.Sc. in Computer Science & Engineering — BNIST

Feb 2023 – Present

Coursework: Linear Algebra · Probability & Statistics · Data Structures & Algorithms · Machine Learning · Data Science

Certifications: AI Engineer for Developers Associate — DataCamp (2026) · Machine Learning — Stanford (Coursera) · CS50 — Harvard

Languages: English (Fluent) · Bengali (Native) · Hindi (Conversational)