

Adil Shamim

SENIOR AI ENGINEER

adilshamim696@gmail.com | [+880 1321073452](tel:+8801321073452) | adilshamim.me | github.com/AdilShamim8 | linkedin.com/in/adilshamim8

3 years shipping production LLM systems — agentic RAG, multi-agent orchestration, and the evaluation and guardrail infrastructure that makes them trustworthy. Currently CTO of a UK fintech AI product.

EXPERIENCE

CTO & AI Engineer | FinTech Startup, UK (Remote)

Jan 2026 – Present

AI Intern → AI Engineer → CTO

- Architected and shipped the production Finance AI Agent: tool-calling over live financial data, retrieval pipelines returning grounded and auditable answers, and citation validation against retrieved evidence.
- Own the AI stack end to end — prompt and tool design, evaluation harness, guardrails, deployment — and set technical direction as CTO; hardened the agent against prompt injection and retrieval poisoning, with role-scoped access filtered before inference.

Founder & AI Engineer | ReWoo (rewoo.tech)

Jul 2026 – Present

- Building an AI-workforce platform of 10 role-scoped AI executives spanning sales, support, finance, marketing, operations, and analytics; landed the first paying UK client 10 days after founding.
- Automated 3 business functions for early customers behind an agent quality gate that vets 100% of customer-facing output before it reaches a user.

Student Researcher | ELITE Research Lab LLC (Queens, NY — Remote)

Sep 2026 – Present

- Empirical research on foundation model evaluation and cognitive task modeling: benchmark design, measurement methodology, reproducible pipelines.

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

Jun 2025 – Mar 2026

- Solo-built and operated an AI-tools discovery platform: 400+ tools across 15 categories, automated moderation pipeline, Learn AI education hub.
- Shipped **Toolly Studio**, a zero-backend workspace of text, encoding, data-conversion, and calculation utilities with built-in AI assistance, running entirely client-side.

SELECTED ENGINEERING WORK

Agentic AI Eval Harness github.com/AdilShamim8/Agentic-AI-Eval-Harness

A CI merge gate for AI agents — pytest, where the unit under test is an entire agent. Framework-agnostic adapters (LangGraph, OpenAI Agents SDK, CrewAI, plain Python) run each agent against 280 hash-pinned golden cases, scoring final outcomes and trajectory behavior, then failing the merge when quality drops below committed thresholds. Deterministic, κ -calibrated LLM-judge, and behavioral evaluators localize failures by layer; judge calibration reached Cohen's $\kappa = 1.0$.

Agentic RAG Platform github.com/AdilShamim8/Agentic-RAG-Platform

Production agentic RAG over changing organizational knowledge: hybrid dense + lexical retrieval fused with RRF over pgvector + Postgres FTS, cross-encoder reranking (BGE-reranker-v2-m3), and an agent state machine doing planning, query decomposition, and multi-step retrieval. Citation validation against retrieved evidence, RBAC enforced inside the retrieval SQL, and hardening against prompt injection and retrieval poisoning; golden-dataset evals gate CI with OpenTelemetry + Langfuse tracing.

TECHNICAL SKILLS

Agentic AI LangGraph, LangChain, OpenAI Agents SDK, CrewAI, ReAct / ReWOO / Reflexion / supervisor / swarm patterns, tool & function calling, MCP & A2A, memory and state management, agent evals & guardrails

LLM / GenAI RAG architecture (hybrid retrieval, RRF, cross-encoder reranking, citation validation), multi-provider orchestration (OpenAI, Anthropic, Gemini, Ollama), fine-tuning (LoRA/PEFT, SFT, RLHF), LLM-as-judge calibration, RAGAS

ML / DL & Speech PyTorch, HuggingFace Transformers, TensorFlow, speaker diarization (pyannote.audio, WeSpeaker), Whisper, wav2vec2, knowledge distillation, XGBoost, scikit-learn

Platform & MLOps Docker, Kubernetes, vLLM, FastAPI, Qdrant / Chroma / FAISS, Postgres + pgvector, Redis, Prometheus & Grafana, OpenTelemetry, Langfuse / LangSmith, MLflow, ZenML, GitHub Actions CI/CD, AWS, GCP

Programming Python, SQL, JavaScript / TypeScript, Bash

RESEARCH & RECOGNITION

- First author** — *Bengali Hallucination Detection with Retrieval and Programs: A System Study* — inference-only judging architecture: response-conditioned query rewriting, six hybrid retrieval routes, and a router dispatching each case to a 27B textual judge or a 14B agent that reasons directly or executes generated Python. 0.956 private macro-F1 on the Olikbochon challenge.
- First author** — *Bangla Diarizz: Domain-Adapted Speaker Diarization for Bengali Long-Form Audio*, BUET CSE Fest 2026 / DL Sprint 4.0 (107 teams): fine-tuned pyannote segmentation plus WeSpeaker ResNet34-LM embeddings with gap-stitching post-processing; 56% wall-clock speedup (1h22m → 36m), private DER 0.286, rank 19.
- Kaggle Notebooks Master** (167 / 62,365), top 1% finish (29 / 4,082) | startup funding, ITBI Pitch Fest 2026 | 610+ GitHub stars.

EDUCATION & CERTIFICATIONS

B.Sc. Computer Science & Engineering | BNIST

Feb 2023 – Present

Certifications AI Engineer for Developers Associate, DataCamp (2026) | Machine Learning, Stanford / Coursera | CS50, Harvard

Languages English (fluent) | Bengali (native) | Hindi (conversational)