

Md Arian Rahman

AI Engineer | Agentic Systems, Applied ML & Solution Engineering

Dhaka, Bangladesh | Open to remote | +8801797919113 | arianrahman431@gmail.com
[linkedin.com/in/arianrahman](https://www.linkedin.com/in/arianrahman) | github.com/Arian-Rahman | huggingface.co/spaces/soothsayerzzz

PROFESSIONAL SUMMARY

AI Engineer with **3+ years** delivering agentic systems, ML forecasting pipelines and backend products across **retail analytics and legal technology**. Works directly with clients from requirements and solution design through delivery.

EXPERIENCE

AI Engineer — Astha IT

Jul 2025 – Present

Own end-to-end delivery of agentic workflows and ML forecasting for a B2B retail-planning platform, partnering directly with the client's owner and coordinating changes across **agent, .NET and data services**.

The platform turns sales, pricing, inventory and promotion data into **demand forecasts and promotion plans**. Its chat interface lets retail teams explore results and **configure or launch forecasting workflows**.

Agentic workflow service | *LangGraph, FastAPI, WebSockets, Claude and Gemini on Vertex AI, BigQuery, Vertex Agent Engine*

- **Conversational forecast analysis:** built a LangGraph supervisor that routes planner questions across BigQuery SQL, **Vertex AI Agent Engine Sandbox for autonomous programmatic multi-step analysis** and grounded product knowledge; reused session state to **accelerate iterative forecast analysis**.
- **Client-led workflow automation:** translated client requirements into a **guided chat workflow for project setup and forecast launch**, with live validation, explicit write confirmation, safe retries and duplicate-launch protection.
- **Secure project-context switching:** enabled **secure project-context switching within a single chat thread** using isolated per-project state and tenant/user-scoped access; validated context attribution and cross-project leakage controls with a **19-case live evaluation harness**.
- **Reliable, resumable chat:** built resumable, non-blocking chat execution across server instances, **keeping concurrent conversations responsive** and preserving recoverable history through cancellation, disconnects, retries and dead-letter handling.
- **Grounded product knowledge:** designed a manifest-driven knowledge architecture for a bounded product domain, providing **predictable step-by-step guidance, citations and clarification** while controlling retrieval cost.

ML forecasting and promotion engine | *BigQuery ML ARIMA_PLUS_XREG, Python, Polars, Cloud Run, Pub/Sub*

- **End-to-end retail forecasting:** built an **11-phase BigQuery ML ARIMA_PLUS_XREG forecasting pipeline** that generates ad hoc SQL for **autonomous, self-guided data transformation**, spanning feature preparation, back-testing, training, promotion forecasting and result enrichment across product and geography dimensions.
- **Forecast reliability:** engineered **repeatable forecast runs** across series mapping, parallel-job isolation, future-feature preparation, regressor ordering and standardization. Deterministic series mapping keeps product series consistently assigned across parallel runs, **preserving forecast comparability and reproducibility**.
- **SKU-level evaluation:** added **SKU-level holdout scorecards** against naive and trend baselines to **expose lifecycle-driven failures** and prevent blended accuracy metrics from masking poorly performing product series.
- **Observable model runs:** made model runs **observable and operationally safe** through phase checkpoints, execution limits, BigQuery job polling, status updates, cost tracking and bounded forecast outputs.
- **Promotion optimization:** built a per-product optimizer that searches a 2.5-percentage-point discount ladder across **four objectives under margin and sell-through constraints**. Implemented constant-elasticity demand response and **verified monotonic uplift behavior** across discount levels with synthetic scenarios.

Software Developer / AI Developer — Beraten Software (Remote, USA)

Jul 2023 – Jun 2025

- **Legal-document assistant:** built an Azure OpenAI assistant over **10+ legal and court documents**, using structure-aware retrieval, reranking, **cite-or-decline answers and clarification** to provide reliable step-aware procedural guidance.
- **Backend engineering:** built **.NET Core backend services, APIs and application modules**; integrated AI capabilities and **improved application responsiveness and manual workflows** through caching and automation.

SKILLS

Languages and engineering: Python (asyncio, Pydantic), SQL, C#/.NET Core, FastAPI, WebSockets, REST APIs

Agentic AI and LLMs: LangGraph, multi-agent orchestration, tool calling, structured outputs, human-in-the-loop controls, prompt and context engineering, guardrails; Vertex AI Agent Engine

Retrieval and evaluation: RAG, structure-aware chunking, reranking, deterministic evaluation harnesses;

ML and data: BigQuery ML, time-series forecasting, feature engineering; BigQuery, PostgreSQL, Redis, Pandas, Polars

Cloud and reliability: Google Cloud Platform (Cloud Run, Pub/Sub, Cloud Storage), Docker

Agentic engineering workflow: six-stage workflow covering planning, implementation, code review, refactoring, unit testing and release preparation, using reusable agent skills and verification gates for consistent delivery

Solution delivery: client discovery, requirements analysis, solution and system design, PoC-to-production delivery, technical documentation and cross-service integration

EDUCATION

North South University — B.Sc. Computer Science & Engineering · GPA 3.73/4.00 · Magna Cum Laude · 50% academic scholarship

RECOGNITION

Judge, Infinity AI BuildFest 2026 — BRAC University