

Md. Rafi

+880 1779-188806 | hello@rafiautomation.systems | linkedin.com/in/md-rafi1 | github.com/MdRaf1 | rafiautomation.systems
| Dhaka, Bangladesh | relocating to Kuala Lumpur (Q4 2026) | fully remote | available immediately, full-time hours now

PROFESSIONAL SUMMARY

AI engineer building autonomous LLM systems in Python and TypeScript. Three projects are publicly shipped and installable: a Git safety CLI on PyPI, a CI/CD compliance action on the GitHub Marketplace, and a multimodal debugging agent deployed on Hugging Face Spaces. Focus areas are retrieval-augmented generation, agent tool-calling, and evaluating whether generated code is actually correct. Currently a software engineering undergraduate, available for fully remote contract work, internships, and junior engineering roles.

CORE COMPETENCIES

Retrieval-Augmented Generation (RAG)

LLM Agent Architecture

Tool Calling

LLM Output Evaluation

Prompt Engineering

Python & TypeScript

CI/CD & GitHub Actions

Elasticsearch & ES|QL

EXPERIENCE

Independent AI Engineering Projects - self-directed

Sept 2025 - Present

AI Engineer

Remote (Dhaka, Bangladesh)

- Shipped a Python CLI to PyPI, a TypeScript CI/CD action to the GitHub Marketplace, and a multimodal agent to Hugging Face Spaces.
- Two fixes merged and shipped upstream in an open-source Node.js CLI: a version probe that read a nonzero exit as "not installed", silently disabling a working extractor; and a page-break rule across seven document templates, submitted with a regression test.
- Hackathon entrant: Gemini 3, Chrome Baseline Tooling, Elastic Search AI.
- Prize winner, Code with Kiro Hackathon (Kiro AI, Devpost).

SHIPPED WORK

Git Sensei

Published on PyPI

Natural-language Git assistant that translates developer intent into shell commands, with a guard layer that blocks destructive repository operations before they execute. Packaged and distributed through PyPI so it installs with a single pip command.

Python | OpenRouter API | github.com/MdRaf1/Git-sensei

Baseline Guardian

GitHub Marketplace

CI/CD action that scans CSS against MDN Baseline compatibility data and posts an actionable policy report on every pull request, blocking non-compliant merges. Built for the Chrome Baseline Tooling Hackathon.

Node.js | TypeScript | GitHub Actions | github.com/MdRaf1/baseline-guardian

SessionSolve

Hugging Face Spaces

Multimodal debugging agent. Ingests a screen recording of a bug alongside a full repository, uses Gemini vision and a long context window to localise the fault, proposes a patch, and generates a Playwright regression test to verify the fix visually. Built for the Gemini 3 Hackathon.

Python | Streamlit | Gemini API | Playwright | github.com/MdRaf1/SessionSolve

The Data Guardian

Elasticsearch | RAG

Privacy agent that detects and redacts personally identifiable information from live support tickets, using retrieval over written data-governance policy to judge sensitivity rather than regex matching alone. An ES|QL dashboard logs every redaction for audit. Built for the Elastic Search AI hackathon.

Elasticsearch | ES|QL | RAG | github.com/MdRaf1/the-data-guardian

retrieval-eval

Offline evaluation harness

An offline evaluation harness written to test the retrieval claims The Data Guardian's design had asserted without a measurement behind them. A 300-document corpus built from the public-domain NIST SP 800-53 Rev 5 OSCAL catalog, 36 graded queries across keyword, paraphrase and multi-doc shapes, frozen before any retriever ran, and seven retrievers scored on nDCG@10, Recall@10 and MRR@10. The

useful output is the three things that did not work, each with its mechanism: title boosting came out bit-identical to the baseline on paraphrase, which doubles as evidence that the subset's anti-leakage rule holds; RRF hybrid fusion landed strictly between its two legs on every subset, because rank-only fusion imports a starved lexical candidate list at full weight; and cross-encoder re-ranking was net flat over dense retrieval (0.622 against 0.623 nDCG@10) and cost recall, while clearly helping the weaker lexical stage - it scores the best-judged control negative on 24 of 35 queries, domain mis-calibration rather than wiring. Runs locally on ONNX Runtime; nothing deployed, no users.

Python | ONNX Runtime | NIST SP 800-53 (OSCAL) | [github.com/MdRaf1/the-data-guardian \(retrieval-eval/\)](#)

guardian-analytics [PostgreSQL | measured query plans](#)

The relational counterpart to The Data Guardian, built because a flat audit index answers searches well and analytical questions badly. A normalised five-table PostgreSQL schema over the same compliance-audit domain, a dependency-free deterministic generator that fills it with 201,619 synthetic violations across 18 months, and six queries the flat index cannot answer: repeat-offender intervals, month-over-month policy trends, top-N per department, p95 scan duration, a flagged-but-never-actioned anti-join, and cumulative load. Index choices are measured rather than assumed - EXPLAIN (ANALYZE, BUFFERS) before and after, medians of five runs, and three candidate indexes built, measured and dropped with the numbers that disqualified them. Synthetic data; nothing deployed, no users.

PostgreSQL | SQL | [github.com/MdRaf1/guardian-analytics](#)

Teletext Zero [Live demo](#)

A "slow web" browser that renders live news and weather inside a strict 40x24 Teletext grid with CRT shaders. As much a constraint-engine problem as a UI one.

TypeScript | [github.com/MdRaf1/Teletext-Zero](#)

EDUCATION

B.Sc. in Software Engineering [Daffodil International University, Dhaka, Bangladesh](#) Currently enrolled; first year completed (3 semesters)

Continuing the degree at UniRazak (Kuala Lumpur) from the first year - Q4 2026 intake, pending student visa. Enrolment certificate available on request.

Higher Secondary Certificate (HSC), Science [Uttara High School and College, Dhaka, Bangladesh](#) October 2024

GPA 4.92 / 5.00

CERTIFICATIONS

Cambridge Linguaskill (English) - overall 186, CEFR C1	Cambridge University Press & Assessment	2026
Learn C Skill Path & Python 3 Course	Codecademy	2025

HONORS & AWARDS

Prize Winner - Code with Kiro Hackathon, for a CLI assistant that translates natural language into safe, context-aware Git syntax with interception of destructive commands	Kiro AI Devpost	
Competitor - CodeTrap Programming Contest, 5th overall, 4 algorithmic problems solved	DIUSEC, Daffodil International University	2025

TECHNICAL SKILLS

Programming Languages: Python, TypeScript, JavaScript, C, SQL

AI & LLM Integration: Retrieval-Augmented Generation (RAG), agent architecture, tool calling, prompt engineering, LLM output evaluation, Gemini API, OpenRouter, ONNX Runtime inference

Search & Data Ecosystems: PostgreSQL, Elasticsearch, ES|QL, Kibana, Plotly

Infrastructure: GitHub Actions, CI/CD, Playwright, Streamlit, Node.js, React.js, Next.js ([rafiautomation.systems](#)), Git

Spoken Languages: Bengali (native), English (Cambridge Linguaskill C1, 186)