

Md Rahinul Islam Bhuiyan

rahin577@gmail.com | github.com/cyberjaya101 | cyberjaya101.github.io

SUMMARY

Second-year Computer Science undergraduate at the University of Malaya, specializing in Artificial Intelligence. Builds and ships production-style AI systems end to end: security tooling for AI agents, LLM infrastructure, and applied NLP, backed by published research in mechanistic interpretability. Comfortable across the stack, from model training and evaluation to APIs, dashboards, and deployment.

EDUCATION

University of Malaya

Bachelor of Computer Science (Artificial Intelligence)

Kuala Lumpur, Malaysia

March 2028 (Expected)

- Concentrations: Machine Learning, Natural Language Processing, LLM Engineering | Coursework: Data Structures & Algorithms, Database Systems, Software Engineering

RESEARCH

Honeypot Detector Interpretability — Trustworthy AI for Security

June 2026

Sole Author | DOI: [10.5281/zenodo.20972391](https://doi.org/10.5281/zenodo.20972391)

- Trained a two-layer transformer on EVM opcodes from 543 real Ethereum contracts (F1 = 0.98, reproducible across 5 seeds); used per-head ablation, attention attribution, and activation patching to test whether a security detector's decisions are trustworthy and reproducible, not an artifact of a single training run.

EXPERIENCE

Independent AI Developer

Kuala Lumpur, Malaysia

Freelance AI Engineer

Jan 2024 - Present

- Architected an autonomous AI assistant (LLMs + n8n) automating multi-step workflows across integrated tools and APIs; prototyped an AI-powered wellness conversational agent.

PROJECTS

warden — Provenance Firewall for AI Agents

Jun 2026

Solo Developer

Private repository

- Built a fail-secure, default-deny governor that blocks AI-agent tool calls driven by untrusted data instead of the user; added a provenance-validated GRC agent, hash-chained and signed audit logs, REST/MCP API integrations, and a Prometheus-instrumented dashboard. 95+ tests, CI-verified.

guard — Security Scanner for AI Applications [GitHub](#)

May 2026

Solo Developer

- Built a pre-commit Python security scanner detecting leaked API keys and prompt-injection risks (10+ secret patterns, 20+ SAST signatures) as a zero-config CLI for CI/CD pipelines.

measure — Production LLM Evaluation Dashboard [GitHub](#)

May 2026

Solo Developer

- Built an LLM evaluation dashboard scoring faithfulness, answer relevancy, contextual precision, and hallucination rate on-device, enabling data-driven model comparison and regression testing at zero API cost.

router — Multi-LLM Router with Cost Optimization [GitHub](#)

Mar-Apr 2026

Solo Developer

- Engineered a FastAPI REST service that routes queries by complexity across OpenAI GPT-4, Claude, and Gemini APIs, cutting LLM inference cost 60% with real-time cost tracking and fallback logic.

CommentSense — L'Oréal x University Datathon

Aug-Sep 2025

Team Leader & NLP Engineer

- Led an end-to-end NLP pipeline (sentiment analysis, spam detection, category tagging) over 10,000+ social-media comments at 85% accuracy for L'Oréal marketing teams, with a Streamlit reporting dashboard.

LEADERSHIP

UM International Students Association (UMISA)

Kuala Lumpur, Malaysia

Event Vice Director

Sep 2025 - Present

- Built and maintain a task-management dashboard improving planning efficiency across 15+ cross-functional members; led end-to-end execution of cultural events for 300+ attendees.

SKILLS

Programming: Python (clean, testable code, pytest), SQL, Java, R

AI / ML: PyTorch, LangChain, RAG, Prompt Engineering, Agentic AI, Model Evaluation, NLP, LLMs (GPT-4, Claude, Gemini)

Security & Governance: AI Agent Security, Prompt-Injection & Secret Scanning, GRC Workflows, Risk Assessment, Audit & Compliance Tooling

Data & Analytics: Statistical Analysis, Data Pipelines, Dashboards & Storytelling (Streamlit, Plotly)

Tools: FastAPI, REST APIs, Git/GitHub, Docker, CI/CD, AWS, Linux, MongoDB