

KAAMIL MIRZA

Toronto, ON, Canada | +1 905-213-3021 | mirzakaamil@gmail.com
linkedin.com/in/kaamil-mirza | github.com/kaamilmirza | kaamilmirza.com

SUMMARY

Software Engineer with 3 years building production AI and backend systems across Python, TypeScript, Azure, and AWS. Shipped hybrid-retrieval RAG, multi-agent LLM orchestration, voice, payment and reconciliation platforms, and high-throughput ETL for enterprise, healthcare, real-estate and e-commerce clients. Currently completing an Masters in IT Security (AI in Security) at Ontario Tech University.

EXPERIENCE

AFM Industries

May 2026 – Present

AI Systems Developer

Whitby, ON

- Lifted Recall@10 from 71% to 94% with a **hybrid-retrieval RAG** (vector + BM25, RRF fusion) over 1.2M chunks
- Cut p95 query time 2.1s → 340ms by redesigning **PostgreSQL** schemas across product, vector, and **ETL** layers
- Routed tool calls across **Azure AI Search** and **PostgreSQL** nodes via multi-agent orchestration in **LangGraph**
- Cut new-module setup to under a day with a **FastAPI** scaffold wired to **Azure AI Search** and **GitHub** PR checks
- Unified reporting on a normalized **SQL** warehouse via **Power BI** dataflows and **Azure Functions** pipelines
- Chose **Python/FastAPI** over **Azure Logic Apps**, **Copilot Studio**, and **Power Automate** by benchmarking cost and vendor lock-in with engineering stakeholders

Independent Contractor

Nov 2025 – May 2026

Software Engineer (Python, Full-Stack) – Contract

Remote

- Eliminated dev/prod drift for a client platform team by migrating services to permanent infra via **GitLab CI/CD**
- Unblocked nightly runs by refactoring **Airflow** DAGs and **AWS** task groups, fixing **Gunicorn** auth failures
- Blocked breaking changes by automating **API** contract checks via **Swagger** and **Git** hooks across cross-functional teams
- Surfaced silent write failures by extending storage observability across **AWS EFS**, **rclone**, and **HDFS**
- Upgraded **Django 5.2** and **Node.js 22** and enforced **ruff/Pylint** monorepo-wide

Xelpmoc Design & Tech.

Jan 2025 – Aug 2025

Software Engineer

Hyderabad, India

- Reduced pipeline runtime 75% by engineering **Python/Spark ETL** processing 10M+ records daily
- Reconciled 10,000+ financial files daily at 92%+ accuracy by building a **Python/AWS** reconciliation system with automated exception handling
- Selected the production retrieval stack by architecting a configurable **RAG** platform in **LangChain** and benchmarking 25+ pipelines on latency and precision
- Extracted 5,000+ documents daily at 92% accuracy by building **OCR** pipelines with **PyTorch** and **LayoutLM**; sustained 10K+ messages/sec on **Kafka/Redis** microservices

Lumara Jewels

Aug 2023 – Dec 2024

SDE-1, Full-Stack Developer

Hyderabad, India

- Automated captures, refunds, settlements, and reconciliation by integrating **payment gateway APIs**
- Secured payment callbacks by designing **REST APIs** with **S2S** verification and signed **webhooks**
- Handled concurrent payment failures safely with event-driven services and **idempotent retries**
- Served faceted search over 5,000+ SKUs at sub-20ms by engineering **Elasticsearch** indices and query tuning
- Kept search in sync by deploying **MongoDB** change-stream pipelines with bulk indexing and retries
- Cut UI iteration cycles 50% by architecting **server-driven UI (SDUI)** with a real-time template engine
- Replaced manual discount entry with a promotion engine supporting stacked rules and timed activations
- Reduced cold-start latency 60% by optimizing **Redis** caching for product filters

SDE Intern, Full-Stack Developer

Aug 2023 – Mar 2024

- Priced 5,000+ SKUs at sub-100ms with a **TypeScript/Node.js** engine over **MongoDB**
- Removed hardcoded pricing by designing modular adapters for material rates, weights, and tax rules
- Delivered 30+ invoice and transactional mail templates by wiring **AWS S3** storage to **Resend**
- Lifted organic CTR 35% and **Lighthouse** SEO to 100 with **SSR**, canonical URLs, and **JSON-LD** in **Next.js**
- Cut escaped bugs 65% by reaching 90%+ test coverage across 50+ components with **Jest**, **RTL**, and **Cypress**

EDUCATION

Ontario Tech University
Masters of Information Technology Security – Artificial Intelligence in Security (MITS-AIS) — GPA: 4.0/4.3
• **Coursework:** Machine Learning, Network Data Analysis, Secure Software Systems, Cryptography and Secure Communications, Attack and Defense, Security Policies and Risk Management, Law and Ethics of IT Security

Sep 2025 – Apr 2027
Oshawa, ON

MLR Institute of Technology (Affiliated to JNTU Hyderabad)
B.Tech. Computer Science and Information Technology — CGPA: 7.66/10
• **Coursework:** Data Structures, Design and Analysis of Algorithms, Database Management Systems, Operating Systems, Computer Networks, Object-Oriented Programming (Java), Artificial Intelligence, Machine Learning, Data Warehousing and Data Mining, Cloud Computing, Software Engineering, Cryptography and Network Security, Probability and Statistics

Aug 2020 – Apr 2024
Hyderabad, India

PROJECTS

Razorpay-MCP Server – NPM Package [Open Source]
Node.js, Express.js, TypeScript, Razorpay
• Authored an open-source NPM package implementing an **MCP** server for AI assistants in **TypeScript/Node.js**
• Exposed payments, orders, and settlements via **REST APIs** with pagination and structured error handling
• Supported flexible deployment by implementing stdio and SSE transports with **Docker**

Apr 2025

Companion – Note-Sharing Mobile Application | *Flutter, Node.js, MongoDB, AWS, Firebase*
• Reached 2,000+ downloads and 200 DAUs by launching a campus app on **Flutter, Node.js**, and **MongoDB**
• Enabled 500+ note shares and real-time chat with **Firebase** and **WebSockets**

Aug 2023

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Go, Java, Dart, C, Bash, HTML/CSS

AI & ML: LangGraph, LangChain, PyTorch, TensorFlow, Hugging Face, Scikit-learn, RAG, Vector Search, LLM Orchestration, Prompt Engineering, Fine-Tuning (SFT, GRPO), NLP, Computer Vision, Generative AI

Backend & Frameworks: FastAPI, Django, Node.js, Express.js, React, Next.js, Pandas, NumPy, Flutter, REST APIs, GraphQL, WebSockets, Microservices, Event-Driven Architecture

Cloud & DevOps: Azure (AI Search, Functions, Logic Apps, OpenAI), AWS (S3, Lambda, EC2, EFS, CloudFront), Docker, Kubernetes, CI/CD, GitHub Actions, GitLab CI, Git, Linux/Unix

Data & Storage: PostgreSQL, MongoDB, Redis, Elasticsearch, MySQL, Apache Kafka, Apache Spark, Airflow, Power BI, ETL/ELT, Pinecone, FAISS, Qdrant, Weaviate, Milvus

Payments & Integrations: Stripe, PayPal, Razorpay, Payment Gateway APIs, Webhooks, Idempotency, Reconciliation, Invoicing, Swagger/OpenAPI

Testing & Observability: Pytest, Jest, Cypress, React Testing Library, Grafana

ACHIEVEMENTS

- **Hack the 6ix 2026 – Winner** | GitHub | Demo (Jul 2026, University of Toronto – 1st of 133 teams): Built **Tabi**, a **LangGraph** multi-agent travel agent with a supervisor node self-scoring 4 candidate replies per turn; post-trained **Qwen3.5-4B** via SFT → distillation → **GRPO** to 100% schema validity. Stack: Gemini, MongoDB Atlas, FastAPI, Docker.
- **Google Hack4Change – Winner** (Aug 2023, T-Hub / Google India / ISB): Built a **Python/TensorFlow** crop monitoring model at 92% accuracy using satellite imagery.
- **Major League Hacking – 2nd Place** (2023): Built a full-stack solution with **React, Node.js**, and **MongoDB** for cognitive retraining of children with disabilities.

PUBLICATIONS

- **A Mobile Application for Crop Optimization using Satellite Imagery and Disease Detection using AI** – IJRASET, ISSN 2321-9653, Vol. 12 Issue IV, Apr 2024 [Link]