Seifeldin Mahmoud
AI Engineer

Abu Dhabi • contact@saifmb.com • saifmb.com • LinkedIn • GitHub

ABOUT

AI Engineer focused on cost-efficient low-latency inference. Experienced in building AI pipelines under explicit hardware constraints and reliability-critical backend architectures. Expected graduation May 2027

EDUCATION

AL AIN UNIVERSITY   **ABU DHABI, UAE**
Bachelor of Science in Software Engineering   **2023-EXPECTED MAY 2027**
- Honors List, 3.81/4 GPA
- Completed 78 credit hours

EXPERIENCE

AL AIN UNIVERSITY   **UAE**
UNDERGRADUATE RESEARCH ASSISTANT   **NOVEMBER 2025 — Present**
- Conducting a systematic literature review on structured token pruning techniques to identify gaps in Vision Transformers' computational efficiency, specifically FLOPs, inference latency and accuracy retention
- Built Trace, a Chrome extension that tracks and summarizes articles, streamlining multi-user team workflows

LuxAI   **UAE**
Data Science Intern   **JULY 2025 — Present**
- Built Reelwise, an SEO analysis pipeline intercepting internal GraphQL traffic via Playwright for high-fidelity metadata extraction. Leverages Semantic Embeddings and Faster-Whisper for deep content analysis
- Architected a triage pipeline that reduced heavy GPU inference by 60% by filtering low-signal inputs via CPU

SMART NAVIGATION SYSTEMS   **UAE**
Software Engineering Intern   **MAY 2025 — November 2025**
- Spearheaded end-to-end development of Himaya71, a critical emergency response platform. Engineered a Django backend aggregating high-frequency IoT/YOLOv8 data, automating stale record cleanup via cron jobs
- Deployed a Next.js command center for live occupancy visualization, providing real-time calls to action

PROJECTS

Oryx Engine — GitHub
- Architected a geo-aware SEO engine based on unsupervised clustering. Routed queries to language-specific embedding models then optimized dynamic K-Means for stable semantic grouping. Built a hybrid pipeline merging TF-IDF extraction with dense vector analysis to drive Generative Engine Optimization (GEO)

Resume Optimizer — GitHub
- Developed a privacy-first resume optimization platform using Next.js and Gemini. Implemented a two-phase prompt chaining architecture with Server-Sent Events (SSE) achieving 72% client-perceived latency reduction
- Engineered a custom input sanitization layer with 40+ regex patterns to detect context-aware prompt injection and XSS attacks, validated via Playwright E2E test suites and Upstash distributed rate limiting

AstroScope (PlanetX 2025) — GitHub
- Engineered a cross-platform fault-tolerant RAG pipeline, implementing a dual-tier workflow strategy that falls back from live data, ensuring availability during API overloads

SKILLS

- AI/ML: TensorFlow, scikit-learn, Hugging Face, NumPy, OpenCV, RAG, Vector DBs
- Languages: Python, Bash, SQL, TypeScript, JavaScript, Java, CSS, HTML
- Frameworks: Django, Node.js, Next.js, React, Tailwind CSS, GraphQL
- DevOps/Deployment: Linux, Docker, Git, Github Actions, Playwright, PostgreSQL

CERTIFICATIONS

HarvardX CS50AI — Introduction to Artificial Intelligence   **July 2025**
DeepLearning.AI & Stanford — Supervised Machine Learning: Regression and Classification   **July 2025**
DeepLearning.AI & Stanford — Advanced Learning Algorithms   **Dec 2025 - Present**
DeepLearning.AI & Stanford — Unsupervised Learning, Recommenders, Reinforcement Learning   **Dec 2025 - Present**