# Awab Zemzemi

**AI Engineering | LLM Applications, AI Agents & MLOps**

Tunis, Tunisia | +216 53 532 055 | awab.zemzemi@insat.ucar.tn
LinkedIn | GitHub | **Final-year internship: February 2027 (4–6 months)**

## Education

**National Institute of Applied Sciences and Technology (INSAT)** Expected 2027
*Engineering Degree in Computer Networks and Telecommunications*
Master's-equivalent qualification Tunis, Tunisia

**Relevant coursework**: Machine Learning, Deep Learning, Software Architecture, Microservices, Cloud Computing.

## Experience

**Bit&Books Inc.** Jul 2026 – Aug 2026
*AI Intern* Remote – Ottawa, Canada

- Built an **OCR-to-LLM document extraction pipeline** to convert invoices and receipts into structured bookkeeping transactions.
- Added evidence-based recovery of missing line items and total checks to validate extracted document data.
- Used **BGE-M3 embeddings** and **Qdrant** to retrieve Chart of Accounts candidates; an LLM selected accounts and flagged uncertain matches for review.
- Served **vLLM** inference on RunPod GPUs with a task queue and parallel workers, processing **1,880 items**.

**KIAN Technology** Jul 2025 – Aug 2025
*DevOps Intern* Tunis, Tunisia

- Deployed **Wazuh** on DigitalOcean Kubernetes with **Falco** DaemonSets and webhooks for runtime security monitoring and alerts.
- Used **Terraform** to automate cloud infrastructure provisioning.

**ACTIA** Jul 2024 – Aug 2024
*Mobile & Web Development Intern* Tunis, Tunisia

- Built an e-bike battery monitoring application with **Flutter**, **Flask**, and **Firebase**.
- Built a health dashboard and automated builds using **Docker** and **Jenkins CI/CD**.

## Projects

**Autonomous AIOps – Kubernetes Diagnostic Agent (Ongoing)**
Python, LangGraph, FastAPI, Kubernetes API, Kind

- Developing a **LangGraph-based AI agent** for Kubernetes cluster diagnostics, using dynamic tool selection and evidence-based reasoning to investigate infrastructure issues.
- Implemented **read-only diagnostic tools** to collect Pod health, container logs, events, restarts, and resource metrics, supporting root-cause investigation.
- Integrated the agent with **FastAPI** and validated diagnostic workflows against healthy and failing workloads in a local **Kubernetes test lab**.

**Second Brain – AI Knowledge Assistant** [GitHub]
React, FastAPI, LangGraph, Neo4j, ChromaDB, MCP, LangSmith

- Built a **LangGraph-based knowledge assistant** combining **GraphRAG** over Neo4j relationships with ChromaDB semantic search and MCP tools for grounded question answering.
- Developed an **LLM-powered note ingestion pipeline** for classification, entity extraction, and indexing across PostgreSQL, ChromaDB, and Neo4j.
- Evaluated answers on a **31-question dataset** using **LLM-as-a-judge** in LangSmith to assess correctness, relevance, hallucinations, and conciseness.

**SentinelML – Bank Transaction Anomaly Detection** [GitHub]
Kafka, FastAPI, MLflow, Docker, Jenkins

- Built an **online inference pipeline** streaming bank transactions through Kafka to FastAPI for real-time anomaly detection.
- Developed a **retraining pipeline** with Jenkins automation and MLflow experiment tracking.
- Evaluated anomaly detection on **1.5M transactions** (**ROC-AUC 0.875, PR-AUC 0.678, 72% recall**).

## Technical Skills

**Programming & Web**: Python, Java, SQL, Bash, FastAPI, Flask, NestJS, Angular, REST APIs
**LLMs & Agents**: LangGraph, LangChain, RAG, GraphRAG, Tool Calling, MCP, LLM Evaluation
**ML & Retrieval**: Scikit-learn, PyTorch, Pandas, NumPy, BGE-M3, Qdrant, ChromaDB, Neo4j, MLflow
**Infrastructure & MLOps**: vLLM, Docker, Kubernetes, Kafka, Terraform, RunPod, Jenkins, GitHub Actions, Git, pytest

## Activities & Languages

**Activities**: Securinets INSAT (Securiday organizing team, CTF); ACM INSAT (Codequest SysAdmin staff).
**Languages**: Arabic (native), English (fluent), French (fluent). **Training**: Cisco Networking Academy (Networking; Switching & Routing).