WIEM ABBASSI

AI Engineer | Information Security Master's Student

Tunis, Tunisia (open to relocation - EU / Canada) | wiemabbassi14@gmail.com | +216 20 09 60 70

linkedin.com/in/wiem-abbassi | github.com/wiemabbassi | wiemabbassi.tech

PROFESSIONAL SUMMARY

AI Engineer specializing in LLM security and Retrieval-Augmented Generation (RAG) for cybersecurity applications, seeking an AI-for-Security internship (or related AI/ML security opportunity) in the EU or Canada. Built and red-teamed a prompt-injection detection pipeline (Garak, PyRIT) and deployed production-oriented AI systems in Python with FastAPI, Docker, PyTorch, Hugging Face Transformers, and vector databases (Pinecone, ChromaDB, FAISS). Computer Science graduate (ISI, 2025), currently pursuing a Master's in Information Security.

TECHNICAL SKILLS

Programming: Python, Bash, SQL, JSON

Machine Learning: Supervised & Unsupervised Learning, Anomaly Detection, Reinforcement Learning

Deep Learning & NLP: NLP, Computer Vision, Generative AI

LLM / Generative AI: LLMs, Prompt Engineering, RAG, Fine-Tuning, LangChain, LlamaIndex, AI Agents, Multi-Agent Systems

AI Security & Safety: LLM Security, Prompt Injection & Jailbreak Defense, PII Masking, Red-Teaming (Garak, PyRIT), Responsible AI, Bias Detection

Frameworks & Libraries: PyTorch, Hugging Face Transformers, scikit-learn, spaCy, Pandas, NumPy

Backend & Deployment: FastAPI, REST APIs, Docker, Git/GitHub, Model Monitoring (Prometheus, Grafana)

Databases & Vector Stores: PostgreSQL, Pinecone, ChromaDB, FAISS

PROFESSIONAL EXPERIENCE

AI Security Engineer Intern | *RFC* **Jul 2026 – Aug 2026**

- Engineered an asynchronous **FastAPI reverse-proxy gateway** inspecting prompts and responses between client applications and two LLM backends (Ollama, OpenAI) in real time, enforcing data-privacy and safety policies on every inference call.
- Cut response time on blocked malicious requests by **~96% (~2.5s to under 100ms)** by architecting a 3-tier detection pipeline - regex/heuristic rules, fine-tuned DeBERTa-v3 and RoBERTa classifiers, and Meta Llama Guard 3 - intercepting prompt injection and jailbreak attempts pre-inference.
- Implemented bi-directional PII tokenization with Microsoft Presidio and spaCy NER, masking **40+ sensitive entity types** (SSNs, API keys, credit card numbers) before model submission and restoring original values on the outbound response.
- Developed a behavioral anomaly-detection layer (Isolation Forest over session-level usage signals), paired with a Redis-backed token-bucket rate limiter and a YAML policy engine mapping risk scores to allow/flag/block decisions.
- Benchmarked the pipeline against Garak and PyRIT red-teaming suites, establishing a **47% baseline** prompt-injection resistance for the underlying llama3.2 model and identifying Long Prompt and DAN Jailbreak attacks at **100% pre-mitigation success** in Giskard scans - findings that directly shaped detection-rule thresholds.

Artificial Intelligence Engineer Intern | *RFC* **Feb 2025 – May 2025**

- Designed and deployed a **RAG pipeline** integrated with Microsoft Sentinel to automate incident investigation, enriching security alerts with threat intelligence from 2 external sources (OTX, Maltiverse).
- Benchmarked candidate LLMs against a cybersecurity query dataset and standardized on the best-performing model for generating threat analysis and remediation recommendations.
- Automated a SOAR playbook orchestrating incident enrichment, analysis, and response end-to-end, replacing a previously manual, multi-step triage process with a single automated flow.

Business Intelligence Analyst Intern | *MAVISION* **Aug 2024 – Aug 2024**

- Cleaned and preprocessed raw datasets, resolving data-quality and consistency issues to produce analysis-ready data for recurring stakeholder reporting.
- Built interactive Power BI dashboards that aggregated key business metrics into stakeholder-facing insights.

AI / ML PROJECTS

Tunisian Darija Agricultural AI Assistant - *RAG, CNN, Multimodal*

- Built a RAG pipeline grounded in 3 verified agricultural sources (FAO, EPPO, CIHEAM) with anti-hallucination safeguards, powering voice-based Q&A in Tunisian Arabic (Darija) for olive growers.
- Developed a CNN-based image classifier for olive leaf disease diagnosis, integrated with contextual crop-management guidance to support in-field decisions.

Threat Hunting Query Generator via AI - *Ollama, Streamlit, MITRE ATT&CK*

- Prototyped a threat-hunting query generator using local LLMs (Ollama) to translate natural-language threat descriptions into validated queries across 3 platforms - Splunk SPL, Microsoft Sentinel KQL, and Elasticsearch DSL.
- Developed a deterministic parsing and post-processing engine that extracts search parameters, corrects time-window mismatches, and maps threat scenarios to MITRE ATT&CK TTPs, surfaced through an interactive Streamlit dashboard.

AI-Based Data Poisoning Detection - *Isolation Forest, LLM-Based Validation, MISP*

- Developed an anomaly-detection module for an open-source threat-intelligence pipeline, combining an Isolation Forest model with LLM-based semantic validation to flag poisoned or contradictory indicators of compromise (IOCs) before they reach analyst queues.
- Implemented rule-based checks for malware/APT-sector mismatches and implausible IOC combinations, strengthening validation coverage prior to MISP ingestion.

OWASP Top 10 for LLM Applications - *LLMGoat, Mistral 7B*

- Evaluated the OWASP Top 10 risks for LLM applications through hands-on adversarial testing with LLMGoat and a locally-hosted Mistral 7B model, running prompt-injection attacks against a live target.
- Mitigated identified vulnerabilities using input filters, validation rules, and safer prompt-handling logic.

EDUCATION

Master's Degree in Information Security - Higher Institute of Computer Science (ISI) *Sep 2025 – Present*
Bachelor's Degree in Computer Science - Higher Institute of Computer Science (ISI) *Sep 2022 – Jun 2025*

CERTIFICATIONS

- Building LLM Applications With Prompt Engineering - NVIDIA, Apr 2026
- Claude Code in Action - Anthropic, Mar 2026
- Azure AI Fundamentals - Microsoft, Nov 2024
- Python - Kaggle, Sep 2024

LANGUAGES

English: Fluent | French: Fluent | Arabic: Native