TAHER NOUIRA

taher.nouira@insat.ucar.tn | linkedin.com/in/taher-nouira | github.com/NouiraTaher | TaherNouira-Portfolio

Profile

5th-year Software Engineering student focused on Natural Language Processing and LLM systems. Experienced in developing retrieval pipelines, prompt safety layers, and document understanding models. Combines algorithmic depth from competitive programming with a systems mindset for reliable AI deployment.

Experience

AI/NLP Engineer, Dydon AI - Herrliberg, Switzerland   April 2025 – October 2025
- Optimized a **Retrieval-Augmented Generation** pipeline for financial document understanding, improving retrieval precision by **15%** and response coherence.
- Integrated **FlashRank** and lightweight rerankers to balance latency and accuracy, achieving a **2.3x** speedup with minimal quality loss.
- Built a lightweight **prompt sanitization system** using regex detection, token escaping, and context rewriting, deployed through an internal LLM security gateway orchestrated with **Docker** and **FastAPI**.
- Developed a **visual document classifier** using **MobileNetV2** for country detection and an **OpenCV** pipeline for EPC identification, reaching **100% accuracy** across France, Germany, and Austria.

Software Developer Intern, BEDI - Monastir, Tunisia   June 2024 – September 2024
- Architected and developed an attendance system with **PyQT**, integrating **ZKTeco** biometric SDK for real-time authentication.
- Designed fault-tolerant microservices using **RabbitMQ** for asynchronous messaging.

Projects

RAG-based Hotel Recommendation System (Academic) : BERT, ChromaDB, Gemini, SHAP
- Built an end-to-end **RAG pipeline** for personalized hotel recommendations, integrating data scraping, fake review filtering, and vector retrieval with **Gemini**-based reasoning.
- Fine-tuned a **BERT** model to detect deceptive reviews and applied **SHAP** analysis for interpretability, improving recommendation reliability and model transparency.

MiniGPT: Transformer Exploration : PyTorch, MLflow
- Implemented a **GPT**-style decoder-only Transformer (2 layers, 4 heads, 0.3M params) from scratch in **PyTorch**, studying attention mechanisms, memory efficiency, and optimization stability on limited hardware.

Serendipity Engine: Cognitive Creativity Agent : Embeddings, LLM Reasoning, Knowledge Graphs
- Designed an intent-driven **creativity agent** combining semantic similarity, LLM reasoning, and knowledge graph linking to generate cross-domain insights.
- Built a modular pipeline for context extraction, embedding-based inspiration retrieval, and idea synthesis for interactive conceptual exploration.

Education

INSAT - National Institute of Applied Science and Technology   2021 – 2026
Software Engineering Diploma

Monastir Pioneer High School   2017 – 2021
Mathematics Baccalaureate — **18.26**/20 (Top **3%** nationally)

Technical Skills

Programming Languages: Python, C++, Java, SQL
Core Systems: Data Structures & Algorithms, Distributed Systems, Cloud Computing
AI & NLP: Deep Learning, Information Retrieval, Transformers, Retrieval-Augmented Generation, Prompt Engineering, Computer Vision
Frameworks & Libraries: PyTorch, TensorFlow, Hugging Face, scikit-learn, spaCy, OpenCV
MLOps & Deployment: MLflow, Docker, DVC, FastAPI, LangChain, Streamlit, Azure ML, REST APIs
Collaboration & Tools: Git, GitHub, Jira, Notion, Slack, Google Workspace

Languages

English: Fluent (C1) | **French:** Fluent (C1) | **Arabic:** Native | **German:** Basic (A2)