Oussama Zorrig

contact@oussamazorrig.dev

50 336 307

Sousse, Tunisia

https://www.linkedin.com/in/oussamazorrig/

Profil

I'm an AI Engineer and team leader with a strong foundation in computer science and a deep specialization in Artificial Intelligence. With extensive hands-on experience in **Generative AI**, **Large Language Models (LLMs)**, and **AI Agents**, I design and implement intelligent systems that autonomously reason, retrieve, and act. My expertise also includes building **Retrieval-Augmented Generation (RAG)** architectures that integrate external knowledge bases for accurate and context-aware responses.
Beyond technical proficiency, I combine **creative problem-solving**, **machine learning expertise**, and **strong project management skills** to lead end-to-end AI initiatives from research and development to deployment. I thrive in collaborative environments, communicate effectively with diverse stakeholders, and continually push the boundaries of what AI systems can achieve in real-world applications.

Skills

Programming Languages
Python, Java, C, PHP, SQL, Matlab.

IT environment
Windows, Linux, Git, AWS.

frameworks and libraries
Tenserflow, Keras, Numpy, Pandas,
OpenCV, Flask, Llama Index, LLamaParse,Langchain, LangGraph, Pytorch.

Professional Experiences

September 2024 – present
Ireland (remote)

Smylor, AI Engineer

January 2025 – December 2025
Sousse, Tunisia

Nexilan, AI Engineer and Team Leader

- Lead the design and development of AI-driven solutions, with a strong focus on Generative AI and Large Language Models (LLMs) to deliver real-world applications for clients across industries.
- Manage cross-functional teams through the full project lifecycle from ideation to deployment, ensuring timely delivery, technical quality, and alignment with business goals.
- Drive innovation by integrating state-of-the-art ML techniques into production environments, leveraging a solid foundation in Machine Learning and AI architecture.
- Coordinate with stakeholders to define project requirements, prioritize tasks, and translate business challenges into scalable AI solutions.

- Foster a collaborative team culture focused on continuous learning, creative problem-solving, and staying at the forefront of AI advancements.

July 2024 – October 2025
Berlin, Germany

Pharma.AI, AI Engineer, Freelance

Contributed to innovative projects in the pharmaceutical industry and pharmacovigilance by applying cutting-edge Generative AI and Machine Learning techniques.

February 2024 – June 2024
Texas, Dallas, USA

AxeGen AI, AI Engineer Intern

- Developed an AWS concierge bot designed to guide users through various AWS services and resources using Retrieval-Augmented Generation (RAG).
- Collected data from the AWS website to train the model.
- Applied RAG techniques on GPT-4 to enhance chatbot capabilities.
- Created APIs using Flask to interface with the bot.
- Deployed the APIs on an AWS EC2 instance for production use.
- Improved user interaction and automation for AWS-related queries.

July 2023 – August 2023
Tunis, Tunisia

Data Science Intern at MASS Analytics (On-site)

During this internship, I worked with genetic algorithm and attempted to solve computationally expensive problems through the implementation of surrogate model and parallelization techniques

February 2022 – May 2022
Sousse, Tunisia

Data Science Intern at Sastec-Group (On-site)

Artificial Intelligence System for Fruit and Vegetable Quality:
I had developed an AI model using the CNN algorithm architecture. The tasks I had accomplished are as follows:

- Preparation of a large dataset.
- Optimization of the model's hyper-parameters.
- Training the model using the CNN architecture.
- Testing the model, which achieved an accuracy of 0.97.
- Deployment of the model on a web page.

Projects

Arabic chatbot

Backend for a Chatbot Capable of Understanding and Interpreting Dreams.

Technologies used:

- Retrieval-Augmented Generation (RAG)
- Flask (APIs)
- Docker
- Git

Development of Drug Adverse Reaction Prototype

Developed a prototype utilizing Retrieval-Augmented Generation (RAG) to determine whether the mechanism of action of a drug or substance could potentially be the cause of a specific adverse reaction.

Technologies used:

- Utilized LLAMA Index, LLAMA Parsing, and a Vector Database for efficient data retrieval and analysis.

- Prompt optimization techniques for improved accuracy and response generation.
- Git

E-Learning platform

Led the AI integration for an E-Learning platform, enabling users to interact with a chatbot that explains course material and assists in solving exercises across various subjects. Implemented advanced Retrieval-Augmented Generation (RAG) techniques to provide accurate, context-aware educational support tailored to each user's input.

Emotion Detection

- Developed an emotion detection system using Convolutional Neural Networks (CNN) to classify facial expressions.
- Implemented the model using Python and TensorFlow/Keras for training and testing.
- Integrated the model with real-time computer vision applications using OpenCV.

AI Agents System using LangGraph

Designed and developed a multi-agent system using LangGraph to orchestrate intelligent tool selection for dynamic inputs. Built two specialized agents capable of autonomously choosing the appropriate processing pipeline based on the input type:

- For image inputs, the agents leverage OpenAI Vision for visual understanding and extraction.
- For document inputs, they utilize LlamaParse for structured document parsing and content retrieval. This architecture enables seamless handling of multimodal data and efficient reasoning through adaptive tool management, demonstrating practical integration of AI Agents, LLMs, and RAG techniques.

Languages

English — C1 **French** — C1

Education

2023 – June 2025 **Computer Engineering,** EPI - International Multidisciplinary School

2019 – 2022 **Bachelor's degree in Computer Engineering with a specialization in IoT and Embedded Systems,** ISITCom