Oussama CHAABENE

Nationality: Tunisian   **Date of birth:** 06/06/2001   **Gender:** Male

📞 **Phone number:** (+216) 27916977

✉ **Email address:** chaabaneoussama@fss.u-sfax.tn

✉ **Email address:** oussama.chaabene@ieee.org

🟢 **WhatsApp Messenger:** +216 27916977

in **LinkedIn:** https://www.linkedin.com/in/oussama-chaabene/

📍 **Home:** Ibn Charaf Street, Oudhref, 6052 Gabes (Tunisia)

ABOUT ME

I am a recent Data Engineering graduate with specialized expertise in Artificial Intelligence and distributed systems. Through hands on experience developing multi-tenant RAG architectures, optimizing LLM deployments, and building production-ready AI platforms, I have gained deep technical proficiency in modern AI/ML technologies. I am seeking opportunities in AI Engineering, Data Engineering, or ML Infrastructure where I can leverage my skills to drive innovation and deliver impactful solutions.

WORK EXPERIENCE

🏢 *ACCESS International* – Tunis, Tunisia

City: Tunis | **Country:** Tunisia

AI software engineer

[01/06/2025 – 30/06/2025]

- Sport Reservation Platform: Built full-stack booking platform with user interface and back-office dashboard for facility management and real-time scheduling
- WhatsApp Business Integration: Developed conversational booking agent using n8n workflow automation for direct WhatsApp reservations
- SSL Connectivity Setup: Configured secure connections using ngrok for SSL termination and WhatsApp webhook integration

🏢 *Favizone* – Tunis, Tunisia

City: Tunis | **Country:** Tunisia

AI engineer

[01/05/2025 – 31/05/2025]

- Large-Scale LLM Deployment: Deployed and optimized large language models (like Mistral-8x7B, Llama-70B) using vLLM framework with GPU memory management and tensor parallelism
- Performance Benchmarking: Conducted comparative experimentation measuring token throughput, latency, and concurrent user capacity (5-20 simultaneous users) across different hardware configurations
- Hardware Optimization: Performed systematic comparison testing on NVIDIA configurations (like H100, A10, A100, Tesla series on OVHCloud provider), optimizing VRAM utilization and inference speed
- Custom Model Integration: Developed chat templates and tokenizer configurations for non-supported models (like JAIS family), enabling vLLM compatibility
- Production Monitoring: Built automated performance monitoring systems tracking tokens/second, memory usage, and real-time scaling metrics
- Experimenting with enabling Ray cluster on multiple nodes

🏢 *ACCESS International* – Tunis, Tunisia

City: Tunis | **Country:** Tunisia

AI Software engineer Intern

[03/02/2025 – 31/05/2025]

- Developed an innovative end-to-end platform that democratizes AI chatbot creation, transforming a months-long technical process into a streamlined workflow accessible to non-technical users.
- Multi-Tenant RAG Architecture: Implemented isolated vector search with ChromaDB, metadata filtering with intelligent deduplication, semantic chunking for documents and row-based chunking for tabular data
- Multi-Language LLM Integration: Deployed Mistral-Small-24B for English/French interactions with function calling and Cohere Command-R7B for Arabic using vLLM on high-performance infrastructure (160GB VRAM)
- Asynchronous Data Pipeline: Built Celery orchestrated ingestion supporting 7 formats, web scrapping via Selenium Grid, and database connectors
- Automated Prompt Engineering: Integrated COSTAR methodology and context optimization
- Production Infrastructure: Containerized microservices architecture.
- Full-Stack Development: Built React frontend with responsive UI and Django REST backend.
- Real time Analytics: Developed conversation tracking, CSV export handling the data strategy for the client, and performance monitoring dashboard

Freedom Of Dev – Sfax, Tunisia

City: Sfax | **Country:** Tunisia | **Website:** https://www.freedomofdev.com/en/homepage/

AI Engineer Intern

[01/07/2024 – 31/08/2024]

- Developed a tool recommending ML models based on user input.
- Integrated ReactJS frontend and Django backend for file uploads and model generation.
- Used YOLO and AutoKeras for image classification and user model training.
- Provided model evaluation metrics and comparisons in a simple format.

CompiTechnolgy – Sfax, Tunisia

City: Sfax | **Country:** Tunisia

Mobile Developer Intern

[01/07/2023 – 15/09/2023]

- Developed a mobile app prototype for a Canadian startup using Flutter.
- Aligned the prototype with the startup's goals using Figma for UI/UX designs.
- Utilized Android Studio for development and debugging.
- Integrated Firebase for backend functionality and real-time updates

IBeForum – Remote / Bengaluru, India

City: Remote / Bengaluru | **Country:** India

Data Entry Intern

[01/06/2022 – 31/08/2022]

- Actively engaged in data collection, using various techniques to gather information supporting organizational objectives.
- Leveraged Googling and web-based tools to collect pertinent data effectively

EDUCATION AND TRAINING

Engineering degree

Faculty Of Sciences Of Sfax [15/09/2022 – 02/07/2025]

City: Sfax | **Country:** Tunisia | **Website:** https://fss.rnu.tn/useruploads/files/cycleingenieurinfo.pdf | **Field(s) of study:** Data Engineering

Integrated Preparatory (Math, Physics and Computer Science)
Faculty Of Sciences Of Sfax [15/09/2020 – 31/05/2022]

City: Sfax | **Country:** Tunisia | **Website:** https://fss.rnu.tn/useruploads/files/P2I-2021-2022.pdf

Baccalaureate
Ibn Khaldun High School Mettouia [15/09/2016 – 15/06/2020]

City: Gabes | **Country:** Tunisia | **Field(s) of study:** Technology

LANGUAGE SKILLS

Mother tongue(s): Arabic

Other language(s): English | French

SKILLS

Languages and Frameworks

Python / Django / FastAPI / Flask / Java / Spring Boot / Javascript / React / C / SQL / IaC / Git / Linux / Docker / Kubernetes / Hadoop / Spark / Zookeeper / Hive / Kafka / Hbase / Hue / VLLM / Ollama / Redis / Celery / ChromaDB / RAG / LLM

Software

VMware / VirtualBox / PyCharm / PowerBI / Visual Studio / Eclipse / Android Studio / DataSpell

PROJECTS

MLOps Project
Built an MLOps pipeline for predictive modeling with data versioning, experiment tracking, monitoring, scalable deployment, and automated CI/CD, plus a React dashboard for insights.

Data Analytics Project
built a data pipeline using a dockerized Selenium Grid to scrape data from e-commerce sites (Tunisianet, Mytek, Zoom, Wiki), streamed it to Apache Kafka, stored it in Apache Hive, and analyzed it using Apache Hue for business insights.

Blockchain Project
Developed a blockchain-based application to certify and verify document integrity by generating unique fingerprints stored on the blockchain. The system ensures documents remain unaltered by comparing current fingerprints with the original, enhancing security and authenticity by immediately detecting any tampering.

MediReport AI
Developed a healthcare chatbot that acts as a generalist doctor, assisting users with medical inquiries and providing tailored advice. Built using a fine-tuned LLaMA model for accurate responses, a Flask API for seamless backend functionality, and a React-based interface for an intuitive user experience, the chatbot ensures clear, reliable communication to support better healthcare outcomes.

End Of Year Project
Developed a real-time data stream processing system for log analysis using Spark and Kafka, managed with YARN and deployed on Linux environments via VirtualBox, with analysis Using Pyspark and visualization facilitated through Chart.js.

Automated Infrastructure

Configured automation infrastructure with Kubernetes, managed repositories with Nexus, and implemented CI/CD with Jenkins. Used Vagrant, VirtualBox, Linux and GitLab

YouTube Sentiment Classifier

Classified Arabic comments into positive, negative, and neutral categories. Evaluated YouTube videos for pleasantness and ranked them using web scraping, NLP, NLTK, vectorizing, ML, and DL

ELK Log Monitoring System

Deployed and configured an ELK stack, incorporating Filebeat and Redis to enable real-time log collection, improving troubleshooting efficiency by 40% and streamlining incident response times across the operations team .

Pneumonia Detection

Detected pneumonia in X-ray images using deep learning. Built CNNs from scratch and used pretrained models like VGG16 and MobileNetV2 with freezing techniques

HONOURS AND AWARDS

Best IoT Project

Awarded by Technopole Sfax, IEEE Tunisian Section, and University of Sfax

2nd Place PYFAC9 Competition

Focused on AI in the Automotive Field, organized by PYFAC and Primatec

2nd Place Local Solution Challenge

By Google

Best Mentor at MASTER-FASTER

A Data Science Competition

VOLUNTEERING

Member & Former Treasurer GDSC FSS

Former Treasurer & Member SecuriNets Club FSS

Former EfS Manager NATEG ENIS Student Chapter

Former Vice-Chair & Current Chair IEEE Computer Society FSS Student Branch Chapter.