AYARI Mohamed

Final Year Computer Science Engineering Student

+216 98 130 015 | mohamed.ayari@supcom.tn | Mohamed9Ayari | med-ayari/ | EL Ghazella, 2083 Ariana, Tunisia

Professional Experiences

Jun. 2024 - Sep. 2024

Oratio (Machine Learning Engineer)

- Led the development of an on-premises generative AI-based system, utilizing LLMs and RAG to deliver personalized feedback and real-time analysis.
- Planned and executed a scalable cloud solution that repurposed many components from the on-premises development, resulting in a 3-month reduction in overall development time.
- Designed and implemented a streaming data pipeline with a CDC pattern, achieving a 90% reduction in latency compared to previous solutions.
- Implemented advanced model monitoring and experiment tracking with Comet ML, optimizing system performance and maintaining high accuracy through continuous model improvements.
- Mentored interns in Gen AI and software engineering best practices, fostering their development and contributing to team growth.

- Utilized: Azure, Kubernetes, IaC, Data Engineering, MLOps, LLMs, RAG, NoSQL

Jan. 2024 - May. 2024

CoaChess (Machine Learning Engineer Intern)

- Led the end-to-end development of a generative AI-driven chess system, architecting scalable data pipelines and optimizing retrieval processes with a custom RAG system for higher accuracy.
- Fine-tuned and deployed a quantized LLM, enhancing real-time AI feedback.
- Spearheaded backend development, integrating machine learning components into the system while ensuring performance and seamless API deployments.
- Partnered with cross-functional teams to deliver a user-friendly interface, successfully bridging AI functionality with front-end applications for a cohesive user experience.

- Utilized: LangChain, AWS SageMaker, LLMs, Vector Databases, MLOps

Jun. 2023 - Sep. 2023

ACTIA (Machine Learning Engineer Intern)

- Applied extensive data augmentation techniques to enhance the model's performance.
- Researched the application of HydraNets for active vision tasks, leveraging ResNet (50, 100) encoder backbones and multiple prediction heads (including classification, object detection, and orientation). The resulting research report detailed further utilization for model development, training, and deployment on a modest four-wheeled robot.
- Utilized PyTorch and harnessed the computing capabilities of TPUs to fine-tune large pre-trained models.
- Conducted real-time classification using an Nvidia Jetson AGX board.

- Utilized: CUDA, PyTorch, Computer Vision, Research Paper, C++, Deep learning

May. 2022 - Aug. 2022

SIRYOS (Data Science Intern)

- Developed and maintained the analytics dashboards for SIRYOS's engineering team and SIRYOS's clients.
- Developed components of the internal ETL tool in Python and SQL.
- Collaborated with the backend engineer to unify the frontend and backend of the analytics stack inside Amazon Redshift.

- Utilized: Python, Flask, SQL, Git, Power BI, Pandas, PostgreSQL, Amazon Redshift, Scripting

Technical Skills

Programming Languages:	Python, SQL, OOP
Frameworks/Libraries	PyTorch, LangChain, Sklearn, Pandas, NumPy, FastAPI, Airflow
Databases:	MongoDB, Qdrant, Postgres, S3
Ops:	Docker, Kubernetes, Terraform, Pulumi, Git