



Ismail said *Data scientist*

 ismail.said@insat.ucar.tn  +21655944716  [linkedin.com/in/ismail-said-17b08a1b7](https://www.linkedin.com/in/ismail-said-17b08a1b7)
 <https://github.com/ismailsaid2001>

Profile

As a Data Science student at INSAT, I am actively seeking a final-year internship as a Data Scientist. My expertise includes data exploration, DL/ML, NLP, generative AI, development technologies, and Big Data tools. My professional dedication and effective communication are among my strengths, along with my constant desire to learn

Professional experience

- Data scientist, Tellem.ai** 05/2024 – present | Paris, France(remote)
• Implementation of RAG chat with documents.
• Implementation of the workflow part (translate-summarize, content repurposing, content...).
• Deployment of the AI backend (FastAPI) on Azure App Services / Building the CI/CD pipeline for the AI backend (Azure DevOps).
- AI developer, Wecraft-uvey** 11/2023 – 05/2024 | Tunis, Tunisia
• Implementation of data extraction features within a FastAPI framework.
• Utilized AWS Textract for text extraction from documents and leveraged PaddleOCR for data extraction tasks on invoices and bank statements.
- AI intern, LegalPlace** Internship Report 06/2023 – 08/2023 | Paris, France(remote)
• R&D project (creating a podcast from a blog article).
• Scraping blog content, generating a podcast, cloning Neal deGrasse Tyson's voice ..

Education

- Engineering Degree in Computer Networks and Telecommunications,** 09/2020 – 06/2025 | Tunis, Tunisia
National Institute of Applied Sciences and Technology (INSAT)
- Integrated Preparatory Cycle in Mathematics, Physics, and Computer Science.,** 2020 – 2022 | Tunis, Tunisia
National Institute of Applied Sciences and Technology (INSAT).
- Baccalaureate Degree, Pioneer High School Bizerte** 2016 – 2020 | Bizerte, Tunisie
Honors: 17.73/20.00

AI HACHATHONS

- 3 rd place, Start it up 2.0**
- 4 th place, GeoAI**
- 5 th place, DeepFlow**
- 5 th place, Dataquest**
My Notebooks

Projects

1- End to End data engineering project [↗](#)

Setup of a real-time data streaming pipeline, covering every stage from data ingestion to processing and storage. I used a stack of tools and technologies including **Apache Airflow**, **Python**, **Apache Kafka**, **Apache Zookeeper**, **Apache Spark**, and **Cassandra**, all perfectly containerized using Docker. For more details [↗](#)

2. Interpretability and Multi-Class Segmentation of Human Embryo Images Using Deep Learning

As part of the final project, we delved into the UNet architecture, leveraging its powerful capabilities for precise image **segmentation**. Our approach combined ad hoc methods like **Grad-CAM** applied to the **UNet** model with self-explanatory networks such as **CANET**, utilizing the power of attention modules. For more details [↗](#)

3. Applying DCGANS to generate images based on MNIST and a Landscapes Dataset

As part of our project, my colleagues and I successfully trained a generative model based on Deep Convolutional Generative Adversarial Networks (**DCGANs**) for 300 epochs. It is now capable of generating new images from a dataset published on Kaggle called '**Best Artworks of All Time**'. For more details [↗](#)

4. Predictive marketing campaign

Use unsupervised machine learning (**clustering**) to categorize customers based on certain characteristics of their purchasing behavior: For more details [↗](#)

5. Mini RAG using Neo4j

Development of an augmented generation system using **Neo4j** to retrieve contextually relevant information from a knowledge graph. This project combines a **graph database** with generative models to produce accurate and coherent responses.

6. Personal Blog Website [↗](#)

Creation of a personal blog, Sam3oul's AI Odyssey, where I share articles about AI .

7. Advertising Campaign Management Platform

Website developed as part of a web development workshop, using NestJS for the backend and Angular for the frontend.

Technical skills

Data science

DL/ML, sélection de modèles, manipulation des données, PCA, EDA, modèles d'ensemble
images segmentation-Opencv-Yolov8

NLP

Self Attention Models-Bidirectional LSTM RNNs-Advanced Transformer Models

MLOps

MLflow-ML pipelines-ETL pipelines-Monitoring
CI/CD-Docker-Azure devops

Backend

FastAPI-Spring boot /Quarkus-Nest js

Computer science

Semaphores and synchronisation-Problem solving-Algorithms and Data Structures-Graph theory-UML-Design patterns-Linux-Bash scripting

frameworks IA

TensorFlow-PyTorch-Langchain-Langfuse

Big Data

YARN-Hadoop Map Reduce-Spark

Data engineering

Apache Airflow - Apache Kafka - Zookeeper - Control Center - Schema Registry

Base de données

MySQL,cassandra,postgres,pinecone,chromadb, MongoDB, Neo4j

Autres

Git/Github,Méthodologies Scrum,Trello, Figma

Volunteering Experience

Machine learning instructor, *Dataoverflow*

02/2024 – 05/2024

Vice president, *junior entreprise INSAT*

07/2023 – 08/2024

Treasurer and sponsoring manager, *IEEE RAS INSAT*

06/2022 – 08/2023

Languages

Arabe
native

French
fluent

English
fluent

Spanish
Elementary