



# AMENI KHABTHANI

📍 Mahdia, Tunisia    ✉️ [amenikhabthani2@gmail.com](mailto:amenikhabthani2@gmail.com)    ☎️ +216 23 513 019  
🌐 <https://www.linkedin.com/in/amenikhabthani/>    📄 <https://github.com/Amenikhabthani>

---

## SUMMARY

I am a Data Scientist with a Master's in Data Science, AI, and HealthTech, seeking opportunities to apply my expertise to real-world challenges. I am passionate about developing innovative models; I thrive in collaborative environments and excel at solving complex challenges. I aim to deliver impactful, data-driven solutions that make a meaningful difference.

---

## EXPERIENCE

### Data Science Intern

#### Satoripop, Sousse

6 months - 2024

- Built ASR and TTS solutions, enabling accurate speech processing in French, English, and Tunisian dialects.
- Fine-tuned the Whisper model for Speech Recognition, reducing the word error rate (WER) by 45% on the target dataset
- Fine-tuned XTTS v2 model for Speech Synthesis, improving speech naturalness and intelligibility based on user feedback.
- Built APIs using the FastAPI framework, decreasing response time by 30% and ensuring seamless integration with external applications.
- Designed user-friendly interfaces for APIs with the Gradio toolkit, improving accessibility and usability.

### Data Science Intern

#### Stunas, Sousse

2 months - 2023

- Developed an OCR model for invoice label recognition using DocTr for text detection and NER techniques for label extraction, reducing manual processing time by 50%.

---

## ACADEMIC PROJECTS

### ECG Segmentation

- Heartbeat classification using CNN architecture.

### Brain MRI Tumor Detection

- Improved the accuracy and efficiency of brain tumor detection by 85% using CNN.

### Sentiment Analysis and Rating Prediction for Medicines

- Created an ML model that forecasts medicine ratings based on various features, leveraging sentiment analysis on user reviews.

### Length of Stay Prediction for Hospital Management

- Established a machine-learning model to predict patient length of stay in healthcare facilities, leveraging historical patient data and clinical features to enhance resource allocation and optimize hospital management.
-

## FREELANCE PROJECTS

---

### Creating a Custom TTS System

- Development of a Text-to-Speech (TTS) model based on XTTS v2 fine-tuned from scratch on Maghrebi voice data.
- Assist and support the client in implementing the chain of the solution: data preparation, training, model evaluation.

### AI-based Astrology API (End-to-End)

- Design and deployment of an intelligent astrology API combining traditional and modern sources.
- Integration of an LLM with prompt engineering, intent classification, and generation of personalized responses.
- Database management with Supabase for storing user profiles, query histories, and custom configurations.

### Seamless Dubbing Automation

- Replaced the existing ElevenLabs-based dubbing service in an API with a custom multilingual dubbing pipeline.
- The new module includes speaker diarization and leverages Whisper Large v3 for ASR, GPT-4 for translation, and hybrid methodology XTTS + Orpheus TTS for TTS synthesis.
- Integrated into a Flask API hosted on Azure App Service
- Handled speech-to-text, speaker segmentation, translation, synthesis, and audio merging to produce aligned, natural-sounding dubbed content.

### [In Progress] Medical ASR & Text-to-Translation System

- Currently developing an ASR system for Modern Standard Arabic (MSA) that preserves domain-specific medical terminology.
- Parallel development of a text-to-translation (T2TT) pipeline for Arabic-to-English translation, emphasizing clinical and semantic accuracy.
- Designed for future integration into healthcare transcription and multilingual medical communication tools.

---

## EDUCATION

---

### MASTER'S DEGREE

Pristini School of AI Oct-2022 / Oct-2024

- Data science, AI, and HealthTech

### BACHELOR'S DEGREE

Iset Mahdia Sep-2019 / Jun-2022

- Electrical Engineering Specializing in Automation and Industrial Computing.

---

## CERTIFICATIONS

---

- Natural Language Processing Specialization ([Link](#))
- Generative AI with Large Language Models ([Link](#))
- Python data structures ([Link](#))
- DeepLearning.AI Specialization ([Link](#))
- Generative AI Engineering with LLM Specialization ([Link](#))
- AWS Academy Cloud Foundations ([Link](#))
- Data Analysis with Python ([Link](#))

---

## SKILLS

---

- Machine Learning • Deep Learning • Scikit-learn • Tensorflow • Pytorch • Transformer • LLM
- Generative AI • TTS • STT • Python • DAX • Gradio • FastAPI • Matplotlib • Seaborn • SQL • NOSQL
- Hadoop • Hive • Communication • Problem-Solving

---

## LANGUAGES

Arabic (native)  
English (Professional)  
French (Professional)  
Italian (Notions)

## INVOLVEMENT

Board team member in Pristini Robotics club  
Board team member in Arsii Pristini club

## PASSIONS

Swimming  
ART activities