





Malek BEN YOUSSEF

Machine Learning Engineer

✉ malek.benyoussef@insat.ucar.tn  ben-youssef-malek  malekex6  website






EDUCATION

Master of Engineering in Computer Science
National Institute of Applied Sciences and Technology (INSAT) 
2019 – Present | Tunis, Tunisia
Relevant Courses: Linear Algebra, Database Management, Algorithms and Data Structures, Deep Learning, Cloud Computing, Big Data, Optimization Methods, Reinforcement Learning, NLP

LANGUAGES

English: Proficient
French: Proficient
Spanish: Basic

CERTIFICATES

Professional Machine Learning Engineer 
Google Cloud
Deep Learning Specialization 
DeepLearning.AI
Scrum Fundamentals 

SKILLS

Programming Languages
Python, C, C++, Java, SQL
Machine Learning & AI Frameworks
TensorFlow, scikit-learn, LangChain, LlamaIndex, SpaCy, NLTK, Hugging Face
Cloud & MLOps
Google Cloud Platform (GCP), Vertex AI, BigQuery, Looker Studio, Docker, CI/CD, GitHub Actions, Flask
Embedded AI & Robotics
TensorFlow Lite, Edge Computing, Microcontrollers, Tiny ML

PROFESSIONAL EXPERIENCE

Machine Learning Intern
Premier Cloud Inc
July 2024 – present | Victoria, British Columbia, Canada

- Built an NL2SQL system that converts prompts from sales managers into SQL queries and generates natural language responses about customers billing information.
- Worked with Gemini LLM and achieved approximately 95% output accuracy.
- Performed SQL query optimization, improving information retrieval speed by 40%.
- Deployed the application using Google Cloud Run and managed a CI/CD pipeline with Google Cloud Build
- **Skills:** GCP, BigQuery, SQL, Gemini, Vertex AI, Prompt Engineering

Machine Learning Engineer
Silver Brain AI AG
November 2023 – January 2024 | Geneva, Switzerland

- Developed a chatbot using the pre-trained LLaMA 2 model with 7 billion parameters to deliver detailed information about the company's services to clients.
- Employed Retrieval-Augmented Generation (RAG) techniques to optimize LLM text generation, resulting an improvement of approximately 30% in output quality
- **Skills:** Python, LangChain, LlamaIndex, Lama2, Prompt Engineering, FastAPI

Full Stack Developer Intern
Safran
July 2023 – September 2023 | Tunis, Tunisia

- Created a dashboard for 100+ users, that offers quick access to the company's equipment status.
- Integrated data visualization and real-time updates, achieving a 25% improvement in operational efficiency
- **Skills:** Angular, Spring Boot, MySQL

Data Scientist Intern
University of Moncton
July 2022 – September 2022 | New Brunswick, Canada

- Implemented a solution for colon disease detection that reduces 72% of false positives.
- Enhanced diagnostic precision by 85% in colon disease detection through the application of advanced data augmentation and EfficientNet model optimization techniques.
- **Skills:** Python, Keras, TensorFlow, OpenCV, EfficientNet(B0 to B7)

PROJECTS

Autonomous Navigation using Deep Reinforcement Learning
June 2024

- Developed an autonomous robot navigation system employing the TD3 algorithm, resulting in a significant 85% improvement in navigation accuracy.
- Implemented deep reinforcement learning techniques to enable obstacle avoidance, achieving a 92% success rate in various test environments.
- **Skills:** Python, ROS2, Rviz, Gazebo, TensorFlow.

Soccer Chatbot
December 2023

- Built an information retrieval system about soccer based on 1000+ PDF file pages.
- Applied Llama 2 Model and RAG techniques achieving an exceptional 95% accuracy in information retrieval.
- **Skills:** Python, LangChain, Chainlit, FAISS, OpenAI Embeddings, Llama2