

# Ammar Mahmoud

## Machine Learning Engineer

 Egypt  01022805974  ammar013783@gmail.com  linkedin.com/in/ammam-mahmoud-b9396921a

### Profile

Machine Learning Engineer focused on LLMs, RAG, and AI agents. Skilled in fine-tuning LLaMA models with QLoRA, deploying scalable systems via vLLM, and building autonomous agents with LangGraph and CrewAI. Developed real-world applications across NLP and medical imaging using CNNs, ViT, and U-Net. Passionate about applying GenAI tools in production environments.

### Education

**Faculty of Computers and Artificial Intelligence, Beni Suf University**

2019/09 – 2023/07

Medical informatics department with GPA 3.15/4

### Professional Experience

#### Machine Learning Intern – Information Technology Institute (ITI) Summer 2023

- Built ML models using Scikit-learn and TensorFlow, boosting forecast accuracy by 15%.
- Optimized pipelines via feature engineering and hyperparameter tuning.
- Collaborated on internal projects and delivered technical presentations.

### Projects

#### Respiratory Disease Detection using Deep Learning and API Deployment, Graduation Project

- Designed and deployed an AI system to detect respiratory conditions (asthma, pneumonia, bronchitis) from sound recordings.
- Achieved 95% accuracy using Convolutional Neural Networks (CNNs) combined with Mel-spectrogram transformations.
  - Reduced diagnosis time by 30% and improved diagnostic accuracy for remote healthcare users.
  - Deployed the solution using FastAPI, Docker, and AWS/GCP, providing seamless access to healthcare professionals across various platforms."

#### Instruction-Based Fine-Tuning with QLoRA & LLaMA-Factory

- Collected articles and used OpenAI API to generate titles and translations.
- Fine-tuned LLaMA base model with QLoRA using LLaMA-Factory.
- Measured base model improvement, estimated costs, and deployed via vLLM.

#### AI Agent using CrewAI

- Simulated real-world collaboration roles using CrewAI, LangChain, and OpenAI.
- Handled document analysis, research, and task delegation through agent teamwork.

#### End-to-End Medical Chatbot using Generative AI

- Built and deployed a medical chatbot powered by LLMs to handle health-related queries with improved accuracy using domain-specific fine-tuning.
- Integrated **LangChain** and medical knowledge bases for retrieval-augmented generation (RAG) to enhance response quality.
- Deployed the chatbot API with **FastAPI** and **Docker**, ensuring scalable and real-time performance on **AWS/GCP**.

#### GPT from Scratch Implementation

- Built a simplified version of the GPT architecture from scratch using PyTorch and NumPy.
- Implemented key components including tokenization, multi-head self-attention, positional encoding, and causal masking.
- Trained the model on a small dataset to showcase its ability to generate coherent text and demonstrate architectural principles.
- Gained deep insights into the internals of transformer-based models and attention mechanisms.

#### Image Classification using ViT, CNN, and Transfer Learning

- Developed an image classification pipeline using both CNN architectures and Vision Transformers (ViT).
- Leveraged transfer learning on pretrained models to enhance performance on medical image datasets.
- Achieved high classification accuracy across multiple categories using PyTorch and TorchVision.
- Applied data augmentation and evaluated models with precision-recall metrics to improve generalization.

## Brain Tumor Segmentation in MRI Images

- Developed a semantic segmentation model to detect and outline brain tumors in MRI scans using a U-Net architecture.
- Trained the model on annotated medical datasets and achieved a Dice Coefficient of 0.85 and IoU of 87%.
- Applied advanced preprocessing techniques including intensity normalization and data augmentation.
- Delivered a solution that supports radiologists by highlighting tumor boundaries with high spatial accuracy.

## Face Mask Detection using Faster R-CNN

- Built a real-time object detection system using Faster R-CNN to monitor face mask compliance in public environments.
- Fine-tuned a pretrained Faster R-CNN model on a custom dataset of masked and unmasked faces, achieving 93% mean Average Precision (mAP) and 90% IoU.
- Enhanced robustness through data augmentation, transfer learning, and inference optimization.

## Image Processing and Feature Extraction using Classical Filters (with GUI)

- Applied 20 classical image filters (e.g., Sobel, Prewitt, Laplacian, Gabor) to extract features and enhance image patterns for CV tasks.
- Built an interactive GUI using Tkinter (or PyQt) to allow users to apply filters and visualize results in real time.
- Implemented the solution with OpenCV and NumPy, and evaluated filter effects on classification models.

## Skin Cancer Classification

- Developed a CNN-based image classification model to distinguish between benign and malignant skin lesions using real-world medical image datasets.
- Achieved 98% accuracy and improved early diagnosis capabilities through robust model training and evaluation.
- Applied data augmentation and image normalization techniques to reduce class imbalance and improve generalization

## Medical Cost Prediction (End-to-End)

- Developed an end-to-end machine learning pipeline to predict individual medical insurance costs based on demographic and health-related attributes.
- Explored multiple regression algorithms including Random Forest, XGBoost, Gradient Boosting, and Linear Regression.
- Performed extensive feature engineering and hyperparameter tuning.
- Achieved an  $R^2$  score of 0.89, with interpretable model outputs to support cost transparency.

## Skills

Languages: Python, C++

Mathematical and Analytical Concepts: Linear Algebra, Statistics, Probability, Calculus, Numerical Optimization

Data Analysis & Preprocessing: Preprocessing, Exploratory Data Analysis (EDA), Data Visualization, Feature Engineering

Machine Learning Concepts: Regression, Classification, Unsupervised Learning, Ensemble Learning

Deep Learning Concepts: Neural Networks, Convolutional Neural Networks (CNNs), Transfer Learning, Object Detection, Segmentation, Autoencoders, Generative Models, Recurrent Neural Networks (RNNs), Attention Mechanism, Transformers, Prompt Engineering, Vector Databases

Tools & Libraries: Hugging Face ,LangChain ,CrewAI,LangGraph, SpaCy ,Matplotlib , Plotly TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, OpenCV, FastAPI, Docker

## Languages

Arabic (native)

English (Conversational)