YAZEED ABD ALKAREEM MSHAYEKH

Machine Learning Engineer

Amman, Jordan | yzy0202090@ju.edu.jo | +962 799736382

Portfolio | Medium | Github | LinkedIn

EDUCATION

The University of Jordan — **Oct 2020 - July 2024**

Bachelor's program in Data Science (BS in DS)

GPA: 3.6 (Very Good)

EMPLOYMENT

Natural Language Processing Engineer Internship — **Mar 2024 - Present**

SMSM for AI | Internship Updates and Details can be seen via: github repo.

- Learned about the beginnings of search engines, large language models (LLMs), Retrieval Augmented Generation (RAG) systems, data annotation, LLMs dataset creation, Optical Character Recognition (OCR), and PDF mining tools.
- Gained knowledge on building RAG systems from document reading to LLM text generation and fine-tuning LLMs in general, with a specific focus on Arabic.
- Developed team working skills, including how to split tasks among team members and share success.
- Currently working on a dataset research paper, a whole system paper, an Arabic LLM and RAG system, an ASR system for Arabic, and an Arabic annotated dataset.

Co-Founder — **Feb 2024 - Present**

MedFormer a Startup Company | Startup page can be seen via: github repo.

- A small startup company focusing on revolutionizing healthcare with advanced AI solutions.
- Expertise extends beyond standard AI applications, delving into Large Language Models (LLMs) and Multimodal LLMs.
- Developing intelligent systems that process vast amounts of text and seamlessly understand and integrate visual information.
- For an example of our upcoming system, click here.

Machine Learning Engineer Training — **Jan 2023 - May 2023**

ShAI for AI | Training Details can be seen via: github repo.

- Underwent an extensive five-month training program with SHAI, specializing in machine learning.
- Deepened understanding of data preprocessing, handling various data types, implementing model pipelines, and training models.
- Undertook several projects to sharpen skills in data manipulation and model optimization.
- Researched and troubleshooted to resolve debugging issues and errors, enhancing problem-solving capabilities in this domain.

NON-RESEARCH PROJECTS

- **AMUN-RAG**

Under Progress ...

RAG System, specialized in Arabic Language, specifically Egyptian dialect. We're going to build this system by fine-tuning an LLM on a hand-created annotated dataset, by combining the best choices of Large Language Models and Embedding Models that can deal perfectly with the Arabic documents.

- **LoreWeaver**

Details can be seen via github repo.

In Collaboration with my colleague Basel Anaya, we produced "A Novel Generation Multimodal LLM" called LoreWeaver, LoreWeaver utilizes Mistral LLM's powerful text generation capabilities while incorporating additional functionalities to provide users with more diverse and accessible experiences.

- **ScratchyRAG**

Details can be seen via github repo.

In this project, I have built a ScratchyRAG, which is a RAG system built from scratch that allow the user to upload a PDF file and write a query to extract any info from the PDF file.

- **SilentVoice**

Details can be seen via github repo.

Continuous American Sign Language Translation Using Transformer and Conformer Architectures, using Google's Competition Sign Language Dataset, this project tackles the lack of popularity of sign language, creating a solution that automatically translates Sign Language gestures into spoken language.

- **Attendify**

Details can be seen via github repo.

An Automated Attendance System Using Tri-Architecture Fusion for Facial Recognition. Through the utilization of YOLOv9, Siamese Network and RESNET-50, we have successfully achieved a competitive recognition accuracy **99.71**.

VOLUNTEERING AND INITIATIVES

- **Machine Learning Team Leader @ IEEE Geeks Team and Graphic Designer** — **Feb 2024 - Present**

IEEE Computer Society | University of Jordan Branch

Learn about problem solving. I'm now a machine learning and data analysis team leader. Learned about gaming development, One of my gaming project can be seen via github repo.

- **Content Writer and Graphic Designer** **Jan 2023 - Feb 2023**

IEEE Computer Society | University of Jordan Branch

As a member of IEEE, I was involved in writing interesting content to connect with our community and share what IEEE is all about. This work helped me improve my writing and design skills. It also gave me a chance to work alongside professionals and be a part of the active IEEE community.

- **AI Content Writer** **Jan 2023 - Feb 2023**

ShAI for AI

I interned with the ShAI club for AI, where I focused on writing content about the latest developments in AI for about two months.

- **Founder of JUREPO** **Feb 2024**

AI and Data Science Community Discord Server.

Details about the server can be seen via linkedin post.

CERTIFICATIONS

- **Natural Language Processing Specialization - Under Learning Yet**

DeepLearning.AI. Consists of 4 courses. My progress can be seen via github repo.

- **Deep Learning Specialization**

DeepLearning.AI. Consists of 5 courses. Details and Certificate can be seen via github repo.

- **Machine Learning Specialization**

Stanford University and DeepLearning.AI. Consists of 3 courses. Details and Certificate can be seen via github repo.

- **LangChain: Chat with Your Data | Short Course**

Given by Co-Founder and CEO at LangChain. Details and Certificate can be seen via github repo.

- **LangChain for LLM Application Development | Short Course**

Given by Co-Founder and CEO at LangChain. Details and Certificate can be seen via github repo.

RESEARCH PROJECTS

MedFormer: A Biomedical Vision Language Model With RAG System. Mar 2024

Unpublished yet. Draft will be added soon to my github: github repo.

- MedFormer is an advanced Biomedical Vision-Language Model integrated with a RAG System, addressing complex diagnostic challenges in healthcare.
- Utilizes cutting-edge technologies such as Idefics2, LLAMA-3, and Citrinet-512, seamlessly integrating textual, visual, and auditory data.
- Enhances accuracy and efficiency in medical image analysis, automated report generation, and response to complex inquiries.
- Improves diagnostic processes, ensures accessible medical data for all patients, and fosters exploration in biomedical information synthesis, streamlining healthcare delivery and improving patient outcomes.

Attendify: Automated Attendance System Using Tri-Architecture Fusion for Facial Recognition Jan 2024

On hold in Arxiv. Draft can be seen via: github repo.

- Traditional methods of recording attendance are inefficient and prone to human errors, highlighting the need for a robust solution.
- Introducing an automated attendance system utilizing facial recognition techniques to overcome challenges reported in past research.
- Methodology included preprocessing steps and implementing state-of-the-art models: **YOLOv9** for face detection, **ResNet-50** for recognition, and a modified **Siamese network** for similarity measurement.
- Achieved promising results with an accuracy of **99.7%**, showcasing the potential to revolutionize attendance management and integrate with other domains.

ACHIEVEMENTS

First Place

Organized by the STEAM Center in collaboration with Arab Robotics Association.

Me and My teammates secured the first spot In the Computer Vision category at the third Arab Olympiad for Artificial Intelligence. It was a remarkable gathering, with over 350 students from Jordan, Palestine, Qatar, Libya, Yemen, Egypt, Tunisia, Sudan, Algeria, and Lebanon. More details can be seen via linkedin post.

First Place

The 11th National Technology Parade 2024 - The Educational and E-learning Technologies Category

Me and my teammates, with our project "Attendify: Automated Attendance System Using Tri-architecture Fusion For Facial Recognition". secured the first place in the Educational and E-learning Technologies category at the 11th National Technology Parade 2024. More details can be seen via linkedin post.

LANGUAGES

Arabic Native | Mother Tongue | Proficiency

English Very Good | Proficiency