AHMAD ABDELAAL

+201101112014 | Giza, Egypt
AhmadAbdelaal7810@outlook.com | linkedin.com/in/ahmad-abdelaal-88a004227
github.com/ahmadsameh8 | orcid.org/0009-0009-9959-3048

EXPERIENCE

Data Scientist Feb 2025 - Present
Central Bank of Egypt *Cairo, Egypt*

- Researching and developing a chatbot focused on banking regulations using large language models and AI agents.

Junior Research Assistant Aug 2024 - Dec 2024
C.I.S Research Center Nile University *Giza, Egypt*

- **Led a team** to conduct and validate methodologies for cross-language plagiarism detection.
- Trained **Siamese Neural Networks** with **Longformer/Arat5** due to data limitations.
- Utilized **Weighted Cross-Entropy and Dice Loss functions**, addressing **dataset imbalances** effectively.
- Achieved an **F1-score of 0.9058 with AraT5 using Weighted Cross-Entropy Loss**, demonstrating superior performance compared to existing methodologies by more than 4%.

Backend Engineer July 2024 - Sept 2024
Paymob *Cairo, Egypt*

- Built and Optimized **160 Endpoint** using Django and Django REST Framework.
- Implemented **242 unit tests** using Pytest, Tested API endpoints using Postman.

AI Engineer Mar 2024 - Aug 2024
Mirando solutions *Cairo, Egypt*

- **Processed and ensured the accuracy** and quality of live stream audio and textual data from **32 channels** in parallel by developing efficient data pre-processing pipelines.
- Leveraged Large Language Models (LLMs) for **language detection, transcription, and translation**.
- Improved models inference rates by **20%**.

EDUCATION

Bachelor of Science in Computer Science: Artificial Intelligence, Nile University
CGPA: 3.66

PROJECTS

Graduation project: Exam generation for academic content - Link
Developed a system to generate academic exams with model answers using LLMs.

- Applied advanced chunking methods using LangChain's **NLTK Text Splitter** to ensure contextual integrity in the exam generation process.
- Crafted **Prompt templates** for 3+ types of questions, then fine-tuned llama 3.1 using **4-bit Quantization** to reduce memory requirements on several Q&A datasets.
- Evaluated the generated content using different syntax, semantics, and **LLM powered evaluation metrics** such as BLEU, BERT Score, Tiger Score, and Prometheus.

Automatic Text Correction for Modern Standard Arabic - Link
Fine-tuned a large language model to correct Arabic syntax and semantics text errors.

- Performed **exploratory data analysis** to examine data properties, distributions, and word associations to understand its structure, identify patterns, and extract insights.
- **Fine-tuned ARAT5** model on text correction and **prompted Llama-3 using zero-shot Chain-of-thought**.

- Evaluated the two models results on Gleu, word rate error metrics, and MaxMatch (M2).

Android Malware Classification using Machine Learning - Link
Developed a classification model to classify malware's then deployed it onto google cloud utilizing DevOps pipelines.

- Loaded the dataset efficiently handling it's large size.
- Utilized the **Recursive Feature Elimination (RFE)** algorithm to select top performing features.
- Trained an **XGBoost classifier** on selected features from the dataset achieving **89% accuracy**.
- Created a **docker** Image on **Google cloud**, then created a cluster on **Kubernetes** engine to make the model accessible to external requests.
- Monitored the cluster configurations and performance using **Prometheus**.

Arabic Song Lyrics Generation - Link
Fine-tuned LLMs to generate coherent Arabic song lyrics based on a single prompt.

- Performed **Web scrapping** and pre-processed Arabic song lyrics from the web to meet project requirements
- **Fine-tuned AraGPT and Acegpt** models for coherent Arabic lyrics generation.
- Deployed the model on **Google Cloud with a Flask API**, then tested the API using **Postman**.

PrivateGpt - Link
Developed a secure, offline chatbot capable of processing and interacting with documents directly on local devices.

- Used **Bloom** LLM to enable the **chatbot** to comprehend and generate responses based on documents content.
- Designed custom functions to pre-process, clean, and extract meaningful information from documents.
- Utilized **LangChain and Streamlit** frameworks to create a friendly user interface, allowing seamless document uploads and real-time interaction without relying on external servers or cloud services.

Course-Specific Chatbot Using RAG - Link
Designed a chatbot, to improve accuracy in responses based on video transcripts from a Stanford CNNs course on youtube.

- **Scrapped and processed video transcript** data using chunking and embedding methods, generating over **3,000 chunks** for efficient model training.
- Developed a comparative **chatbot using Tiny Llama and FLAN-T5**, leveraging **Retrieval-Augmented Generation (RAG)** to enhance the factual accuracy of responses.
- Evaluated chatbot performance through **human review and GPT-4** scoring, achieving a response quality score of **6.732/10** for Tiny Llama, highlighting its domain-specific accuracy over FLAN-T5.

AWARDS

- **1st Place at NU Quantum-AI Hackathon**
  Leveraged dimensionality reduction along with a variational quantum classifier and achieved the highest test data score of 77%.
- **Multiple Places in Undergraduate Research forum, Nile University - Link**
- **3xDean's Honor List**
- **Valeo Graduation Project Support**
  Received mentorship and funding from Valeo for my graduation project.

TECHNICAL SKILLS

- **Languages:** Python, C/C++, SQL
- **Frameworks:** PyTorch, Pandas, scikit-learn, Matplotlib, Seaborn, NumPy, Docker, Kubernetes, LangChain.
- **Software Skills:** Git, GitHub, Jira, Agile.