

Mohamed Atwan

DATA SCIENTIST | NLP SPECIALIST

Egypt, Cairo | +20 155 120 9560 | Mo7amed3twan@gmail.com / Mohamed.Atwan@studentambassadors.com

- **Kaggle** : kaggle.com/mo7amed3twan
- **LinkedIn** : linkedin.com/in/mo7amed3twan
- **GITHUP** : github.com/MO7AMED3TWAN

Brief About Me

Results-driven Data Scientist with 2 years of experience in **Machine Learning, NLP, and Data Science**, skilled in extracting insights and building data-driven solutions. Proven track record of leading a team to a **top 5** finish in the **MTC-AIC2 AI competition** by leveraging data analysis and advanced modeling techniques. Actively engaged in **community knowledge-sharing** as an **Alpha Ambassador**, with a focus on fostering collaboration and innovation in data science.

Skills & abilities

Version Control & Deployment

- Proficient in **Git**, **GitHub**, and **Docker** for version control, code management, and containerization, ensuring seamless deployment and collaboration.

Data Management & Programming

- **Databases**: Skilled in **SQL Server**, **PostgreSQL**, and **MySQL** for efficient data management and retrieval.
- **SQL**: Advanced proficiency in **T-SQL**, **DQL**, **DDL**, and **DML** for complex querying and database manipulation.
- **Python**: Expertise in **Exploratory Data Analysis (EDA)**, **NLTK**, **Spacy**, **PyTorch**, **TensorFlow**, and **Transformers** for **data analysis, manipulation, machine learning**, and **deep learning** tasks.

Data Pipeline Development

- Skilled in **ETL** and **ELT** processes to efficiently **extract, transform, and load** data for analysis and reporting.

Data Analysis & Visualization

- Proficient in **Power BI**, **Tableau**, and **Excel** to create clear, impactful visualizations that drive data-driven decision-making.

Machine Learning & Statistical Modeling

- Experienced in **supervised** and **unsupervised** learning techniques for predictive modeling.
- Knowledgeable in **Deep Learning** models, including **CNN**, **RNN**, and **Transformer** architectures for complex data modeling.

Natural Language Processing (NLP)

- Skilled in fine-tuning **NLP models** for **LLMs** like **BERT**, **GPT-2**, and **T5** to solve language-based tasks.

Big Data & Distributed Computing

- Expertise in **Hadoop** and **Apache Spark** for large-scale data processing and distributed computing.

Cloud Analytics

- Proficient in **Azure Data Factory**, **Azure Synapse Analytics**, and **Azure ML** for orchestrating large datasets, transforming data, and deploying machine learning models.
- Experienced in leveraging **cloud-based solutions** to automate workflows and enable scalable, efficient analytics and machine learning processes.

Experience

Information Technology Institute (ITI)

Business Intelligence Development Trainee

July 2024 – September 2024

- Gained experience in data warehousing, advanced SQL, Power BI, and Tableau.
- Designed and optimized data warehouses for efficient management.
- Created interactive dashboards to provide actionable insights.
- Applied business intelligence tools to solve business challenges.

Digital Egypt Pioneers Initiative (DEPI)

Trainee Data Engineer

July 2024 – October 2024

- Mastered **SQL** for data management and optimization.
- Designed scalable **data warehouses** for efficient storage and retrieval.
- Gained hands-on experience in **Python** for data processing and automation.
- Applied **machine learning** concepts and **MLOps** tools like **MLflow** and **Hugging Face**.
- Developed skills in **Azure cloud services** and platform integration.

Projects

Diabetes Prediction and Management System

Developed an AI-powered solution to improve diabetes diagnosis and management

- **Computer Vision for Early Detection:** Applied a transformer-based model to analyze retinal images and detect early signs of diabetic retinopathy.
- **Risk Stratification Model:** Implemented a machine learning model to assess diabetes risk based on health metrics, offering personalized recommendations.
- **NLP-Based Medical Chatbot:** Designed an AI-driven chatbot to provide patients with timely health information, answer questions, and offer support.

Additionally, there is more work available on my GitHub, where I have shared several projects related to:

- **Data Science:** A range of projects utilizing machine learning, data preprocessing, and visualization techniques to extract insights and build predictive models.
- **Natural Language Processing (NLP):** Projects focusing on text classification, sentiment analysis, and language models such as BERT, GPT-2, and T5 for various applications.
- **Machine Learning & Deep Learning:** Various models and pipelines using libraries like TensorFlow, PyTorch, and Scikit-learn for tasks such as image recognition, time series forecasting, and NLP tasks.

Volunteering

Microsoft Learn Student Ambassadors, KFS Branch

Database Administrator

April 2023 - September 2024

- Managed the club’s database infrastructure, ensuring optimal performance and data integrity.
- Designed and optimized schemas for scalability and efficiency.
- Implemented monitoring solutions to ensure continuous system availability.
- Collaborated with stakeholders to deliver tailored database solutions and troubleshooting support.
- Shared best practices with the community to foster learning and development.

Microsoft Learn Student Ambassadors, KFS Branch

Head of Data Engineering

September 2024 - Present

- Developed and executed the strategic roadmap for the Data Engineering track.
- Led and mentored a team of student mentors, ensuring effective execution of learning objectives.
- Fostered a collaborative, educational environment and provided ongoing guidance and support.

IEEE Kafr Elsheikh Branch

Head of Natural Language Processing

October 2024 - Present

- Led the design and development of advanced NLP systems focused on Arabic ASR, language models, and diarization.
- Defined project roadmaps, ensuring team expertise in RNNs, LSTMs, GRUs, Transformers, and LLMs.
- Integrated state-of-the-art NLP models (e.g., BERT, AraBERT, DeepSpeech) into real-world applications.
- Fine-tuned pre-trained models for speech recognition, text correction, and diacritization tasks.
- Mentored team members, designed training materials, and assessed skills in NLP and machine learning.
- Collaborated with engineers and data scientists to build scalable pipelines for large-scale text and audio processing.
- Optimized models for accuracy, efficiency, and scalability, applying noise reduction and voice diarization techniques.

Education

Kafr Elsheikh University

Bachelor of Science in Artificial Intelligence

Expected Graduation: August 2025

<< Specialized coursework in Machine Learning, Deep Learning, Data Science, and Data Engineering >>

Languages

- **Arabic:** Native/Fluent
- **English:** B1 (Intermediate)
- **German:** A1 (Beginner)