Ahmed Emad

ahmedyasser200580@gmail.com | +201067197511 | Heliopolis, Cairo, Egypt
LinkedIn | GitHub

Professional Summary

Motivated and skilled Data Science and Engineering student with a strong foundation in machine learning, data engineering, and full-stack data solution development. Demonstrates proficiency in building and deploying end-to-end projects, including AI-powered chatbots, ETL pipelines, and recommendation systems using technologies like Python, SQL, Apache Spark, and Airflow. Combines technical expertise in data manipulation, statistical analysis, and model implementation with soft skills in problem-solving and communication. Eager to apply a passion for data-driven decision-making in a challenging internship or entry-level role in data science or data engineering.

Work Experience

Data Science Intern – Orange Egypt: Smart Village, Egypt   September 2025
- Implemented a classical classification problem using multiple machine learning models, including Logistic Regression, Decision Trees, Random Forest, and XGBoost, to evaluate and compare performance.
- Developed a customer service chatbot capable of answering FAQs, providing information on bundles, and assisting users with tracking their current consumption.

Education

EUI - Egypt University of Informatics   November 2023 - Present
*Bachelor of Science, **Computing and Information Sciences***   New Administrative Capital, Egypt
- **Courses:** OOP (Object-Oriented Programming), Algorithm Design and Analysis, Advanced Programming, Data Structures, Computer Architecture and Organization, Introduction to Probability and Statistics, Computational Linear Algebra, Programming Fundamentals, Introduction to Computing and Programming, Digital Logic Design, Discrete Structures, Entrepreneurship

DEPI - Digital Egypt Pioneers Initiative   July 2025 – Present
Microsoft Data Engineer
- **Coursework:** Python, OOP (Object-Oriented Programming), Data Structures and Algorithms, NumPy, Pandas, Matplotlib, Seaborn, Dash, SQL (MySQL), NoSQL (MongoDB), Data Warehouse and Modeling, Python for Databases and APIs, dbt

Projects

Orange Customer Service Chatbot | [GitHub Repo] | [Demo Video]   September 2025
AI Chatbot, RAG, Vector Databases, LLM Integration, Authentication   Streamlit, ChromaDB, Ollama, LangChain, Python
- Developed an AI-powered customer service assistant for Orange telecom with retrieval-augmented generation (RAG) for accurate responses.
- Built a modern chat interface in Streamlit with Orange branding, supporting FAQs, billing, mobile/internet plans, and troubleshooting.
- Integrated ChromaDB for vector search and Ollama LLMs (Llama 3.2, nomic-embed-text) to enable intelligent, context-aware responses.
- Implemented secure user authentication and personalized assistance using customer profiles (e.g., remaining data, billing details).
- Designed modular, production-ready codebase (app.py, chat_prompts.py, processing.py) with clear separation of concerns.

End-to-End Retail Data Pipeline | [GitHub Repo] | [Demo Video]   September 2025
Data Engineering, ETL, SQL   Apache Airflow, PostgreSQL, Python
- Built a complete ETL process extracting data from PostgreSQL, transforming it with Python/pandas, and generating automated visualizations.
- Implemented Apache Airflow DAGs with proper task dependencies, error handling, and scheduling.
- Created automated daily revenue calculations and time series analysis.
- Generated dynamic charts showing revenue trends over time.

Movie Recommendation System | [GitHub Repo] | [Demo Video]   April 2025 - May 2025
Functional programming, Web Scraping, Web App   Scala, Python
- Developed an AI-driven recommendation system that suggests personalized movies and TV shows using user preferences and content-based filtering.
- Designed a robust data ingestion pipeline using Scala and Apache Spark to process and transform over 20,000+ movie and TV show entries.
- Integrated TMDb API for multilingual metadata (English and Arabic), enhancing recommendation accuracy and user experience.
- Applied functional programming principles (immutable data, ADTs, pattern matching, pure functions) to ensure modularity, readability, and testability.
- Built a responsive user interface with Streamlit to allow real-time querying and recommendation display.
- Implemented filtering and search features (by genre, release year, runtime) and logic to handle edge cases with missing or incomplete data.
- Focused on performance and scalability, enabling the system to handle large datasets efficiently with Spark transformations.

Technical Skills

- **Programming Languages:** Python, C, C++, SQL, Scala, Java
- **Data Visualization:** Matplotlib, Seaborn
- **Data Engineering & Big Data:** Apache Spark, Apache Airflow, ETL Pipelines, SQL (PostgreSQL, MySQL), NoSQL (MongoDB), dbt
- **Data Cleaning & Transformation:** Pandas, NumPy
- **Web Scraping:** Beautiful Soup, Selenium
- **ML Algorithms:** Time Series Analysis, Regression Analysis, Gradient Boosting, K-Nearest Neighbors, Decision Trees
- **ML Frameworks & Tools:** Scikit-learn, TensorFlow, PyTorch, Streamlit, XGBoost, Hugging Face, LangChain
- **Statistical Analysis:** SciPy, A/B Testing, Hypothesis Testing
- **SWE & Automation:** Git, Docker

Soft Skills

- Problem-Solving & Critical Thinking
- Data Storytelling & Communication
- Collaboration & Teamwork
- Business Acumen
- Adaptability & Continuous Learning
- Attention to Detail
- Time Management & Prioritization

Volunteering

Student Ambassador — August 2025
Egypt University of Informatics — EduGate, Egypt
- Represented EUI at EduGate, providing information and guidance about the Computing and Information Sciences program to prospective students, parents, and visitors.

RYLA Program — December 2023
Rotary — Ain El Sokhna, Egypt
- Attended the Multi-Club RYLA leadership program, a transformative experience focused on developing youth leadership, teamwork, and community engagement.

SOL Academy — September 2015 - December 2015
Blooms — American University in Cairo, Egypt
- Participated in the SOL (Sing Out Loud) Academy, which helped refine public speaking skills and overcome challenges related to speaking confidence.

Languages

- **Arabic** - Native
- **English** - Fluent
- **French** – Intermediate

Certifications

- Data Analyst in Python
- SQL Fundamentals
- GitHub Foundations
- AI and Machine Learning Foundations
- Supervised Machine Learning: Regression and Classification
- Data Engineering for Beginners: Python, SQL & Airflow
- Introducing Multimodal Llama 3.2
- Generative AI and LLMs: Architecture and Data Preparation