Ahmed Ayman Al-Hofy

Cairo, Egypt | +201027557259 | ahmedalhofy42@gmail.com

https://github.com/AhmedAyman4 | https://www.linkedin.com/in/ahmed-alhofy/ | View my Portfolio!

Summary

Aspiring Data & Machine learning Scientist with skills in Python, SQL, ML, NLP, and data analysis. Proficient in Pandas, Scikit-learn, Tableau, and Power BI. Information Systems student at the Faculty of Computers & AI, University of Sadat City, Passionate about using data to solve everyday challenges and applying AI to create practical, impactful solutions.

Education

Bachelor's Degree, Faculty of Computers & Artificial Intelligence – University of Sadat City   *Graduating June 2026*

CGPA: 3.3, Major: Information Systems.

Experience

AI Engineer | Konecta · Internship   *New Cairo, Egypt | Jul 2025 - Present*

- Collected, cleaned, and structured data using Python and web scraping tools to support reliable analysis and modeling.
- Built and optimized machine learning models (classification, clustering, image recognition) and implemented RAG-based chatbots with embeddings, document indexing, and multi-turn conversations.
- Automated data and AI workflows using n8n, improving efficiency and reducing manual processing.

Data scientist | Digital Egypt Pioneers Initiative - DEPI · Internship   *Alexandria, Egypt | Oct 2024 – May 2025*

- Learned to analyze data using Python, SQL, and analytics tools to uncover trends, opportunities, and actionable business insights.
- Learned to build, deploy, and track ML models using Python, Scikit-learn, and MLflow, enhanced with AI and prompt engineering for seamless integration.

Intern | CIB Egypt · Internship   *Cairo, Egypt | Jul 2024 - Jul 2024*

- Developed data, financial, and entrepreneurial skills to strengthen problem-solving and decision-making.

Skills

- **Technical:** Python (ML/DS), SQL, Bash, HTML/CSS, LangChain, LangGraph, LLMs (Hugging Face), Prompt Engineering, RAG, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Web Scraping (Playwright, BeautifulSoup).
- **Tools & Skills:** AWS, Google Cloud Platform (GCP), Vector Databases, Stats, Tableau, Power BI, Excel, Docker, Git, OOP, Problem-Solving, Communication.

Projects

Retail Analytics Copilot (Hybrid RAG + SQL Agent)

- Built a Dockerized Retail Analytics Copilot in Streamlit featuring a stateful LangGraph agent that routes user queries to SQL, RAG, or HYBRID paths using chain-of-thought reasoning and keyword heuristics; implemented a self-correcting, schema-constrained SQL pipeline with DSPy, regex sanitization, and error-driven retries; and fused BM25 document retrieval with natural language constraint extraction and a synthesizer to generate cited, contextual answers.

Semantic Book Recommender:

- **Accomplished a 40% improvement in recommendation relevance** (compared to keyword-based search) by engineering a semantic search engine using **Sentence-Transformers** for **vector embeddings** and **Zero-Shot classification** to map books to specific emotional and categorical tags.
- **Developed and deployed a production-ready full-stack application** with a **Next.js** frontend and **FastAPI** backend, achieving seamless multi-environment scalability through **Docker** containerization and automated **GitHub Actions** workflows.
- **Optimized data discovery and user insight** by preprocessing a dataset of 7,000+ books through comprehensive **EDA**, **NLP-driven sentiment analysis**, and the creation of custom-tagged datasets in **Jupyter Notebooks**, enabling users to find titles based on mood and complex semantic queries.

Customer Segmentation:

- Performed customer segmentation on wholesale spending data using K-Means (selected via Elbow and Silhouette analysis), Hierarchical, DBSCAN, and MeanShift clustering supported by log-transformation, PCA for dimensionality reduction, and rigorous evaluation metrics; derived six actionable customer profiles (e.g., Fresh-First, Bulk Buyers, Household Essentials) to inform targeted marketing, inventory planning, and high-value client prioritization.

Data Scraper (Real Estate & E-Commerce):

- Built robust Python scrapers using Playwright and BeautifulSoup to extract 1,400+ real estate listings from Nawy.com and 250+ gaming laptop products from Noon.com; automated infinite scrolling and pagination with error-resilient parsing, structured data into pandas DataFrames, and saved versioned CSVs, reducing manual data collection effort by over 90%.

Automated Document Processing:

- Built an AI-powered document processing system using OCR (PyTesseract, pdfplumber) and LangChain with Google's Gemini LLM to classify and extract structured data from invoices, purchase orders, and approvals, reducing manual processing by 80%; deployed a Streamlit/Flask app with real-time extraction and MongoDB-backed semantic search for interactive querying and duplicate detection.

Courses And Certifications

SAP Certified Associate - SAP Generative AI Developer Issued by SAP	Sep 2025
AWS Certified AI Practitioner Issued by Amazon Web Services Training and Certification	Apr 2025
Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate Issued by Oracle	Jun 2025
Data Scientist Associate Issued by Data Camp	Aug 2024
IBM SkillsBuild Data Analytics Certificate Issued by IBM SkillsBuild	Jun 2025
AWS Academy Graduate - AWS Academy Machine Learning for Natural Language Processing	Aug 2024
AWS Academy Graduate - AWS Academy Cloud Foundations	Jun 2024
AWS Academy Graduate - AWS Academy Data Engineering	Jun 2024
Applied Data Science with Python - Level 2 Issued By IBM	May 2024
Google Data Analytics Professional Certificate Issued by Coursera	Jul 2024
The Machine Learning Process A-Z – Issued By 365 Data Science	Nov 2024
McKinsey Forward Program - Issued by McKinsey & Company	Dec 2024