

MOSTAFA ABDELHAMED MAHMOUD

AI & Data Science Engineer

Cairo, Egypt | abdelhamedmostafa190@gmail.com | +20 105 056 9668

[LinkedIn](#) | [GitHub](#) | [Hugging Face](#) | [Kaggle](#)

PROFESSIONAL SUMMARY

AI & Data Science Engineer with hands-on experience designing and deploying production-grade ML systems, semantic search engines, and generative AI solutions. Proficient in Python, TensorFlow, and PyTorch, with a track record of building scalable RAG pipelines, diffusion models, and NLP applications that serve real users. Published a Hugging Face dataset accumulating 19,100+ downloads. Completed three industry internships across data analysis, AI application development, and cloud-based data pipelines. Committed to ethical AI innovation.

TECHNICAL SKILLS

Languages: Python, R, Java, Javascript, SQL, C++

ML / DL Frameworks: TensorFlow, PyTorch, Scikit-learn, Keras, LangChain

NLP & Generative AI: Hugging Face Transformers, BERT, model fine-tuning, Diffusion Models, RAG

Data & Visualization: NumPy, Pandas, Matplotlib, Seaborn, Power BI, Excel, Databricks

Backend & APIs: FastAPI, Flask, Pydantic

Databases & Vectors: PostgreSQL, MySQL, MongoDB, ChromaDB

MLOps & Cloud: GCP, Docker, GitHub, Jupyter, VS Code, Google Colab, Kaggle, HuggingFace

Computer Vision: OpenCV, CNN architectures, image generation pipelines

PROFESSIONAL EXPERIENCE

Think for Advanced Technologies

Jul 2025 - Sep 2025

AI Trainee Intern | On-site, Cairo, Egypt

- Architected and deployed an AI-powered Text Embedding System using FastAPI, MongoDB, and ChromaDB, cutting average query latency by 35% through async operations and optimized vector indexing.
- Delivered an end-to-end Retrieval-Augmented Generation (RAG) pipeline - text chunking, embedding, vector search, achieving 89%+ semantic retrieval accuracy on benchmark queries.
- Applied layered CSR architecture and OOP best practices, reducing codebase coupling and cutting onboarding time for new contributors by approximately 40%.

British University in Egypt (BUE)

Feb 2024 - Mar 2024

AI & Applications Intern | On-site, Cairo, Egypt

- Preprocessed and normalized multi-source datasets for ML pipelines, reducing missing-value rate to under 2% and improving downstream model accuracy.
- Implemented supervised, unsupervised, and reinforcement learning algorithms, conducting comparative evaluations across 5+ model architectures.
- Built and evaluated generative AI prototypes using GANs and LLMs (GPT, BERT) for text summarization, translation, and question-answering tasks.

Zewail City of Science and Technology (ZC)

Aug 2023 - Sep 2023

AI & AI Tools Intern | On-site, Giza, Egypt

- Gained foundational AI knowledge covering data ethics, confidentiality, and generative AI tools for image and text generation.
- Delivered a group project applying generative AI tools, receiving recognition from supervisors for creative problem-solving.

PROJECTS

Semantic Search Engine

Project Link: github.com/MostafaAI10/TeXt-Embedding-Model

- Built a scalable semantic search engine supporting PDF upload, auto text extraction, vector embedding, and NL queries with tag-based filtering; designed for multilingual embedding models.
- Implemented CSR architecture with FAISS/ChromaDB backend; reduced keyword-search false positives by enabling vector-based similarity retrieval.

Generative AI - DDPM Diffusion Model

Project Link: github.com/MostafaAI10/NVIDIA-Diffusion-Models-Training

- Implemented a Denoising Diffusion Probabilistic Model (DDPM) from scratch in PyTorch, successfully generating high-quality handwritten digits with a custom noise schedule.
- Completed as part of NVIDIA's Deep Learning Institute training; code starred by the open-source community on GitHub.

TwinnyAI Personas Dataset

Dataset Link: huggingface.co/datasets/Mostafa190/TwinnyAI-Personas-Dataset

- Curated and published a structured AI personas dataset on Hugging Face, Kaggle, amassing 19,100+ downloads, supporting research in conversational AI and persona-driven language models.

Reddit Stock Sentiment Analysis

Project Link: github.com/MostafaAI10/Reddit-Sentiment-Analysis

- Built an NLP pipeline that scrapes Reddit (r/WallStreetBets and similar subreddits) to identify the most-mentioned stock tickers, processing 14,200+ comments across 8 posts in a single run.
- Applied VADER SentimentIntensityAnalyzer to compute bearish/neutral/bullish compound scores per ticker, delivering ranked sentiment reports (GME, TSLA) with configurable upvote and author-filtering parameters.

PUBLICATIONS

A 3D Deep Learning Framework for Dental CBCT Structural Analysis for Oral Health Risk Prediction

IEEE 2026 ICEENG International Conference for Innovations in Intelligent Computing and Cybersecurity (IICC)

Mostafa Abdel Hamed et al. | Conference: Cairo, Egypt, May 2026

DOI: [10.1109/IICC69623.2026.11582452](https://doi.org/10.1109/IICC69623.2026.11582452)

- Achieved Dice = 0.97, ROC-AUC = 0.960, PR-AUC = 0.958, F1-score = 0.927 on dental CBCT structural segmentation and oral health risk prediction using a 3D deep learning pipeline.

CERTIFICATIONS

- | | |
|--|----------|
| • AI Engineer for Data Scientists Associate - DataCamp
Credential ID: AEDS0019338093398 | Feb 2026 |
| • Generative AI with Diffusion Models - NVIDIA Deep Learning Institute
Credential ID: 6YSchyAvQGWRCf1CJoKOIw | Jan 2026 |
| • International AI Accreditation - International Accreditation Organization (IAO)
Serial No: ZIN0342 | Nov 2023 |

EDUCATION

Elsewedy University of Technology (SUT) & Amity University

Dual Degree Program - Faculty of Engineering Technology

B.Sc. Artificial Intelligence & Data Science (SUT) · BTech. Data Science & AI (Amity)

Oct 2024 - Aug 2028

German-Arab Chamber of Industry and Commerce (AHK)

Diploma. Electrical and Photovoltaic Systems

Oct 2020 - Jul 2023