

Hamza Khaled Mahmoud Ahmed

Machine Learning Engineer — Data Scientist — AI Engineer

hamzakhaledklk@gmail.com

github.com/h19overflow — [Hamza Khaled LinkedIn](#)

Cyberjaya, Malaysia — +60 122938594

Summary

Highly motivated Computer Science student (expected graduation **May 2026**) specializing in **Data Science**, **Machine Learning**, and **Artificial Intelligence**, with 1.5+ years of hands-on project experience in **predictive analytics**, **deep learning**, **computer vision**, **natural language processing**, and **AI automation systems**. Possesses a robust foundation in **statistics**, **advanced mathematics**, **data analysis**, **algorithm development**, and **predictive modeling**. Proven ability to manage the **end-to-end machine learning lifecycle**:

- **Data engineering**, preprocessing, **feature engineering**, and **data pipeline** development
- **Exploratory data analysis (EDA)**, **statistical inference**, and **hypothesis testing**
- **Model development** (classical ML, deep learning, neural networks, LLMs, RAG, AI agents), **training**, and **hyperparameter optimization**
- **Model evaluation**, **performance tuning**, **cross-validation**, and **deployment** strategies

Passionate about applying AI/ML to build intelligent, high-impact solutions. Demonstrated project success in **Advanced RAG Systems**, **Full-Stack AI Development**, **Computer Vision** (95% accuracy), **Natural Language Processing**, **Fraud Detection** (95.67% accuracy), and **AI Automation**. Seeking an entry-level **Machine Learning Engineer**, **AI Engineer**, or **Data Scientist** position to contribute to data-driven projects in **FinTech**, **Healthcare**, or **AI/Tech**.

Education

BSc in Computer Science (Specialization in Data Science)

March 2023 – Present

Multimedia University, Malaysia

Expected Graduation: May 2026

CGPA: 3.63 / 4.0

Achievements: 4-time Dean's List Award Winner

Relevant Coursework: Statistics, Calculus, Discrete Mathematics, **Machine Learning Algorithms**, Data Analysis, **Deep Learning**, Object-Oriented Programming, Object-Oriented Analysis & Design, Database Management, **Artificial Intelligence**.

Selected Projects

Graph-Powered Agentic RAG System — Advanced AI Research & Development

Technologies: Python, LightRAG, LangGraph, Google Gemini API (Vision Pro & Flash), PostgreSQL, Pydantic, MinerU

- **Architected and built an end-to-end Retrieval-Augmented Generation (LightRAG) system** to solve the "fragmented context" problem inherent in traditional vector search-based RAG.
- **Engineered a multi-stage, multi-modal ETL pipeline** that intelligently processes complex PDFs, using a Vision Language Model (VLM) to analyze images and diagrams, and prepares the data for ingestion into a knowledge graph.
- **Implemented a sophisticated multi-agent system using LangGraph**, where specialized AI agents collaborate to analyze user queries, determine optimal hybrid retrieval strategies (vector + graph), and synthesize fragmented answers into coherent, actionable narratives.
- **Demonstrated a significant leap in AI reasoning capabilities** by enabling the system to understand and traverse the relationships between concepts, moving beyond simple keyword matching to structured, human-like understanding.

LangGraph Agentic Auditing System — AI-Powered Financial Automation

Technologies: LangGraph, Python, LLMs, Pandas, Multi-Agent Systems

- Designed a multi-agent auditing system using **LangGraph** to automate complex financial workflows, including End-of-Service (EOS) calculations and interactive payroll analysis.

- Architected stateful, graph-based workflows to manage data flow and control, enabling both fully automated processing and complex **human-in-the-loop (HITL)** interactions.
- Deployed specialized AI agents for tasks such as data classification and dynamic transformation based on **natural language instructions**, functioning as an auditor's "co-pilot".
- Implemented parallel execution within the graph to concurrently run independent calculations, significantly improving the efficiency of the final EOS benefit calculation.

Agentic Workbench — Full-Stack AI Document Processing & Analytics Platform

Technologies: LangGraph, LangChain, FastAPI, Google Gemini (LLM), Google Vision (OCR), SQLite, React, TypeScript, Docker, Python

- Engineered a full-stack, AI-powered workbench to automate **structured data extraction** from documents and enable **natural language-based analytics** and visualizations.
- Orchestrated complex, multi-agent workflows using **LangGraph** for an end-to-end document processing pipeline, incorporating OCR, LLM-based extraction, and a **human-in-the-loop (HITL)** review stage.
- Developed a robust backend using **FastAPI** to serve **RESTful APIs** and **WebSockets** for real-time status updates, interfacing with Google Gemini for data structuring and Google Vision for OCR.
- Implemented a data persistence layer using **SQLite** and designed an agentic querying system that translates natural language questions into executable **SQL queries** for data analysis.

Intelligent Customer Service Assistant with Hybrid ML/LLM Architecture

Technologies: Python, LangChain (LangGraph), Scikit-learn (Random Forest), ONNX, MongoDB Atlas, Redis

- Engineered a **cost-effective hybrid architecture** by developing a custom Random Forest intent classifier (**99% accuracy**), optimized with **ONNX Runtime** for sub-millisecond inference, reducing reliance on expensive LLM calls.
- Architected an advanced agentic system using **LangGraph**, featuring custom subgraphs for reliable structured output and a novel "pre-hook context fetching" mechanism to minimize token consumption.
- Implemented an **autonomous memory management system** using **MongoDB Atlas** to store and retrieve user history, enabling the agent to personalize conversations and adapt its communication style.
- Optimized system latency by integrating **Redis** as a high-speed caching layer for user memory and vector embeddings, ensuring rapid data retrieval for real-time interactions in a production-ready environment.

Pneumonia Detection using Hypertuned ResNet50V2 and Simulated Federated Learning

Technologies: TensorFlow, Keras (ResNet50V2), Flower (flwr), Deep Learning

- Engineered a **deep learning model** for pneumonia detection from X-ray images, leveraging a fine-tuned **ResNet50V2** network to achieve **95% accuracy** and **0.90 F1-score**.
- Optimized model performance through comprehensive **hyperparameter tuning** and robust **data augmentation** strategies for improved **generalization**.
- Designed a simulated **Federated Learning** environment using the Flower framework, demonstrating **privacy-preserving model training** principles.

AI-Powered Flashcard Generator with RAG and Web Interface

Technologies: Google Gemini API, RAG, Flask, React, FAISS, LangChain, Python, Pandas, NumPy, Vector Databases

- Built an automated content generation system to create study materials from PDF documents, enhancing learning efficiency through automated **document understanding**.
- Implemented a robust **Retrieval-Augmented Generation (RAG)** architecture leveraging **Google's Gemini LLM** and a **FAISS vector database** for intelligent **semantic search** and accurate content generation.
- Designed a full-stack web application using Flask for the **backend API** and React for the frontend to facilitate seamless PDF uploads and interactive content review.

Fraud Detection Model (Blockchain Transactions) — Machine Learning Classification

Technologies: XGBoost, Random Forest, Scikit-learn, Pandas, SMOTE, Ensemble Methods

- Developed and evaluated multiple models for detecting fraudulent blockchain transactions, achieving **95.67% test accuracy** with an optimized Random Forest model.
- Implemented **SMOTE** to effectively address significant **class imbalance**, enhancing model robustness for **anomaly detection**.

Skills

- **Programming Languages:** **Python** (Expert), **SQL** (Proficient, PostgreSQL, MySQL, **SQLite**), **Java** (Intermediate), **Kotlin** (Intermediate), **TypeScript** (Intermediate), **R** (Basic)
- **ML Deep Learning Frameworks:** **TensorFlow**, **Keras**, **PyTorch**, **Scikit-learn**, **XGBoost**, **Hugging Face Transformers**, **Flower** (Federated Learning), **LangChain**, **LangGraph**, **LightRAG**, **CrewAI**, **ONNX**
- **Data Analysis & Visualization:** **Pandas**, **NumPy**, **SciPy**, **Matplotlib**, **Seaborn**, **Tableau**, **EDA**
- **Cloud Platforms & MLOps:** **Google Cloud Platform** (GCP), **Vertex AI** (Gemini Models), **Docker**, **Git**, **GitHub**, **Jupyter Notebooks**, **CI/CD** (conceptual)
- **Databases Data Management:** **SQL** (PostgreSQL, MySQL), **NoSQL** (MongoDB Atlas), **Vector Databases** (FAISS), **Knowledge Graphs**, **Redis**, **Data Modeling**, **Data Warehousing** (Conceptual)
- **AI/ML Domains & Techniques:**
 - **Computer Vision:** **OpenCV**, **YOLO**, **MediaPipe**, **Object Detection**, **Image Classification**, **OCR**
 - **Natural Language Processing:** **LLMs** (Gemini), **Advanced RAG** (LightRAG), **Knowledge Graphs**, **Structured Data Extraction**, **Sentiment Analysis**, **Text Classification**, **Text Embeddings**, **Prompt Engineering**, **AI Agent Systems**
 - **Core ML:** **Regression**, **Classification**, **Clustering**, **Ensemble Methods**, **Feature Engineering**, **Hyperparameter Optimization**, **Model Evaluation**, **Anomaly Detection**, **Predictive Analytics**
- **Web Development (AI Integration):** **React**, **TypeScript**, **Flask**, **FastAPI**, **RESTful APIs**
- **Statistical & Mathematical Foundations:** **Statistical Modeling**, **Hypothesis Testing**, **A/B Testing**, **Linear Algebra**, **Calculus**, **Probability & Statistics**, **Optimization**

Soft Skills: **Analytical Problem-Solving**, **Critical Thinking**, **Data-driven Decision Making**, **Technical Communication**, **Team Collaboration**, **Project Leadership**, **Adaptability**, **Continuous Learner**.

Languages

English: Fluent

Arabic: Native