

Omar Ali Bin Talib

AI/ML Engineer · Data Scientist

omar.a.bintalib@gmail.com | +966 56 512 3153 | Jeddah, Saudi Arabia · Open to relocate
linkedin.com/in/omar-bintalib | github.com/omarbintalib | omarbintalib.com

SUMMARY

Computer Science student who takes AI work from an open question to a running system, handling the data, the models, and the APIs that put results in front of users. Hands-on with object detection, agentic AI, LLM-based validation tools, and survey data modeling, all built to be used by real people, not just demoed.

EDUCATION

King AbdulAziz University

B.S. in Computer Science · GPA 4.9 / 5.0

Expected 06/2027

King Abdullah University of Science and Technology (KAUST)

AI Specialization Track · KAUST Academy (NTDP) · ML, deep learning, computer vision, AI research

10/2024 – 08/2025

- Selected as 1 of 10 from 17,000+ applicants nationwide; track included a 440-hour research placement in Toronto.

EXPERIENCE

AI Engineer Intern · King AbdulAziz University, Data Managing Center

07/2026 – 09/2026

- Built Survey Pilot, which pilots questionnaires on persona-conditioned LLM respondents (5 response styles), computes Cronbach's alpha, bootstrap CIs, and item-total correlations, and flags leading or double-barreled items.
- Benchmarked synthetic responses against human baseline data, matching real-survey reliability within 0.05 Cronbach's alpha, letting researchers catch weak scales before recruiting participants.
- Engineered a FastAPI backend orchestrating local Ollama runs and vLLM batch jobs on Modal, with a 9-step pipeline that extracts validated scales from OpenAlex into PostgreSQL/pgvector.

AI Researcher · University of Toronto, Data Sciences Institute

06/2025 – 08/2025

- Engineered Python/Pandas pipelines to harmonize 55 years of PSID/CDS longitudinal survey data (50,000+ household records), resolving cross-wave schema changes, attrition, and sampling-weight shifts.
- Built Random Forest and XGBoost models linking childhood math scores to adult net worth (R^2 of ~0.80), quantifying how numerical literacy persists across generations.
- Presented findings at the SUDS 2025 Research Showcase to 30+ faculty and researchers.

PROJECTS

Bidaya: AI-Powered Map of the Sirah

Source-Grounded RAG Assistant · bidaya-sirah.vercel.app

- Built a RAG assistant that answers Sirah and Quran questions in Arabic and English only from approved sources; an LLM router refuses fatwa and personal questions, and uncited answers are rejected.
- Engineered hybrid retrieval (bge-m3 dense + BM25, fused with RRF) over 2,905 source passages, re-ranked by bge-reranker-v2-m3 and served by FastAPI on a Modal T4 GPU that scales to zero.
- Reached 97.9% correctly sourced answers and 100% correct refusals, with no false statements on 50 adversarial questions; in a 15-user pilot, quiz scores rose from 2.7 to 5.2 out of 6.

DAL Shemagh Detection Challenge (Kaggle) [3rd / 100+ teams]

Two-Stage Object Detection Pipeline · Team Sekiro

- Engineered a YOLO11m + ConvNeXt-Tiny pipeline (detector proposals, classifier filtering false positives) scoring 0.863 on the private leaderboard, within 0.03 of 1st place.
- Expanded training data from only 5 labeled positives with iterative pseudo-labeling, at zero annotation cost.
- Applied HSV augmentation and Weighted Boxes Fusion; published the 3rd-place solution write-up on Kaggle.

Haris: Semantic Census Data Validation Engine [3rd / 130+ teams]

GASTAT Data Innovation Hackathon (GASTAT & SDAIA) · Team of 3

- Built an LLM validation service that detects cross-field contradictions (household relationships, employment, demographics) in 400+ column census tables, automating 160+ manual business rules with GPT-4o and Pydantic.
- Architected a stateless FastAPI backend with an asyncio batch pipeline, cutting real-time validation latency 12x (60s to 5s) and auditing a full 400+ column batch in ~30s.

SKILLS & CERTIFICATIONS

Programming: Python, SQL, Java

ML & AI: PyTorch, Scikit-learn, Computer Vision, LLMs, RAG, NLP, Agentic AI, Prompt Engineering

Data Science: Pandas, NumPy, Feature Engineering, EDA, Matplotlib, Seaborn, Data Visualization, Data Pipelines

Tools: FastAPI, Pydantic, LangChain, Ollama, Modal, PostgreSQL, Jupyter, Git & GitHub

Languages: Arabic (native), English (fluent)

Certifications: TCS Sustainathon 2026 Top Finalist (Top 5 / ~500) · DeepLearning.AI Math for ML & Data Science