

Hassan M. Khalil

hassan4khalil13@gmail.com · hmk60@mail.aub.edu · +961 71 266 776 · linkedin.com/in/hassankhalil · github.com/hassankhalil

Profile

AI/ML Engineer and Data Scientist who ships production AI systems end-to-end — and the experiments and evaluation pipelines that keep them reliable. Builds and optimizes production ML models (gradient boosting and deep learning) on MongoDB- and PostgreSQL-backed data pipelines; runs A/B tests, statistical analysis, and exploratory data analysis; and ships LLM/RAG applications and multi-agent systems (LangGraph / LangChain). Designs high-signal datasets, benchmarks, and evaluation protocols to surface failure modes and quantify improvement, then communicates results through clear written reports and verbal presentations for technical and non-technical stakeholders. Production work has cut recruitment screening time by **92%**, reduced support handling time by **60%**, and processed **20M+** records. Remote-first, with full overlap across US and European business hours.

Technical Skills

Programming: Python (Pandas, NumPy, scikit-learn, PyTorch, TensorFlow, Keras, XGBoost, FastAPI, Flask), SQL, R, TypeScript, JavaScript, C++; REST API design and SDK integration (boto3, OpenAI, Anthropic SDKs).

Statistics, Mathematics & Machine Learning: Hypothesis testing, experimentation and A/B testing, exploratory data analysis (EDA), statistical inference; applied mathematics (linear algebra, calculus, optimization); supervised/unsupervised learning, gradient boosting (XGBoost, CatBoost, LightGBM), deep neural networks, CNNs, sequence models, autoencoders and anomaly detection; feature engineering, class-imbalance handling (SMOTE / SMOTE-ENN), PCA; predictive modeling — building and validating models with cross-validation.

Data Engineering & Databases: Data collection, handling, cleaning, preprocessing, and quality validation; ETL pipelines; SQL and NoSQL — MongoDB (schema design, aggregation pipelines, compound/multikey indexing, query optimization), PostgreSQL (pgvector), MySQL, Neo4j, Supabase, Redis.

Data Visualization & BI: Data storytelling and stakeholder reporting — Tableau, Power BI, matplotlib, seaborn; KPI dashboards; workflow automation with n8n (225+ production workflows, also Zapier / Make).

Evaluation & Data Design: Evaluation harnesses, benchmarks, and eval protocols; high-signal dataset curation, failure-mode analysis and error taxonomies, human-in-the-loop labeling; agentic-eval metrics (task success, tool-call accuracy, groundedness); knowledge-graph, ontology, and metadata-catalog design (Neo4j).

LLMs & Generative AI: RAG pipelines, prompt engineering, embeddings, retrieval, and inference optimization; cloud APIs (OpenAI, Anthropic Claude, Google Gemini, Azure OpenAI, AWS Bedrock) and local open-weight deployment (GPU/CPU).

AI Agents & Orchestration: Multi-agent systems, LangChain, LangGraph, Pydantic AI, tool-using agents, Model Context Protocol (MCP) servers; agent evaluation, tracing, and decision logic.

Cloud & MLOps: AWS (S3, Lambda, SageMaker; Bedrock — familiar), Azure OpenAI, GCP Vertex AI (familiar); Docker, CI/CD (GitHub Actions), Git version control, automated testing, model monitoring & observability; secure coding and documentation.

Experience

Monty Mobile — Beirut [LinkedIn](#)

May 2025 – Present

AI Engineer & Data Scientist

- Built the ML core of an **AI Recruitment Platform** — automated CV parsing, embedding-based **semantic scoring**, and candidate ranking — cutting screening time **92%** (12 min → under 1 min) and saving 200+ hours/month.
- Designed and managed the platform's **MongoDB** data layer — document schemas, compound/multikey indexing, aggregation pipelines for scoring/ranking, and batched ingestion/ETL with schema validation — reducing ranking-query latency via index and pipeline optimization and serving both the recruiter app and the RAG assistant from one documented operational store.
- Surfaced shortlist-quality **failure modes** in the live pipeline, built a labeled ranking benchmark with precision@k and recruiter-agreement metrics, and ran **champion–challenger A/B evaluation** that lifted shortlist precision before rollout.
- Built an autonomous **eSIM support voice agent** with multi-step tool use, real-time call routing, escalation and fallback logic (Twilio + Ultravox + OpenAI), cutting handling time **60%** and enabling 24/7 coverage.
- Architected a stacked **XGBoost + CatBoost + LightGBM** churn-prediction ensemble — exploratory data analysis (EDA), SMOTE-ENN resampling, and threshold tuning — surfacing **actionable insights** on churn drivers and delivering at-risk dashboards and retention triggers to Sales as a key retention/revenue lever.
- Built a **SIM-box fraud-detection** pipeline over **20M+ call records** — large-scale data collection, cleaning, and feature engineering feeding an autoencoder-based anomaly-detection model; adopted by R&D for continued research.
- Containerized and deployed scoring/ranking services (Docker, FastAPI) on **AWS** via **GitHub Actions CI/CD** with automated tests, secret scanning, and versioned model artifacts; added **continuous monitoring** with p95-latency and health dashboards plus prediction-drift checks.
- **Tech:** Python, PyTorch, scikit-learn, LangChain, OpenAI / Azure, MongoDB, PostgreSQL, n8n, AWS, Docker, Power BI.

Mustashari AI (Consulting Automation Startup) — Beirut [LinkedIn](#)

Jan 2025 – Present

Founder & Lead AI Automation Engineer (part-time)

- Founded and led the technical direction of an AI consulting platform delivering **multi-agent** workflow automation for enterprise clients across Sales, Finance, and Professional Services.
- Designed a **benchmark-research module** running 5 specialized agents across frontier models (OpenAI, Claude, Gemini, Perplexity); defined the scoring rubric, comparison protocol, and source-grounding checks, and analyzed cross-model disagreement to select highest-signal outputs — producing standardized reports in under 30 minutes vs. days.
- Built an **evaluation harness** measuring task success rate, tool-call accuracy, groundedness, and robustness across agent workflows, and authored the internal benchmark report adopted as the team's model-selection reference.
- Shipped a suite of consulting agents (Benchmark, Gap Analysis, Deck Reviewer, PowerPoint Generator) and 35+ automation pipelines (RFP/RFI drafting, PM reporting, Git → LLM → docs) with **REST API** and webhook integrations.
- **Tech:** Python, FastAPI, LangChain, LangGraph, OpenAI Deep Research, n8n, PostgreSQL, Supabase, Docker.

Multiple Clients (MK Energies, OTS)

Oct 2025 – Present

Contract AI & Automation Engineer (Project-Based)

- Delivered end-to-end AI/automation for production use — LLM/RAG pipelines, ML models, and **REST API** integrations on Python/FastAPI with MongoDB and PostgreSQL.

American University of Beirut

Nov 2024 – Mar 2025

Research Assistant — ML & Data Engineering

- Built reproducible ETL and feature pipelines and trained/validated ML models for wind-turbine control analysis over high-frequency multi-sensor time-series, delivered to the PI as a structured research deliverable; optimized PostgreSQL query performance with composite indexing.

Amazon ML Mentorship Program

May 2024 – Sep 2024

Machine Learning Intern

- Selected for a competitive 5-month mentorship under Amazon engineers; built and validated production-grade ML models on automated AWS (S3 → SageMaker) data pipelines.

Selected Projects

AgentForge — Multi-Agent Engineering Orchestration

2025 – Present

- Orchestration layer coordinating a fleet of specialized AI coding agents in parallel via isolated git worktrees, with 8+ custom subagents (architect, reviewer, test generator, security auditor) and inter-agent handoff protocols.
- Built an **MCP server layer** integrating Supabase memory and a Neo4j knowledge graph with a domain **ontology** (typed entities/relations) driving retrieval and entity resolution, plus a custom RAG retrieval service for persistent cross-session memory.
- Added governance hooks — secret scanning, audit logging, linting gates, and human-approval checkpoints — for safer, auditable automation. **Tech:** Claude Agent SDK, MCP, LangGraph, Supabase, Neo4j, Python, TypeScript, GitHub Actions.

ByChat — Agentic RAG with Knowledge Graph

Sep 2024 – Nov 2024

- Agentic RAG system where the agent autonomously selects vector search, graph traversal, or hybrid retrieval; streaming FastAPI backend serving real-time answers over 20+ technical documents. **Tech:** Python, Pydantic AI, FastAPI, PostgreSQL (pgvector), Neo4j, Graphiti, OpenAI, Docker.

Matchify — Agentic Job-Matching Platform

Nov 2024 – Feb 2025

- Coordinator agent orchestrating CV/JD parsing, semantic matching, and explainable shortlists via multi-step decision logic. **Tech:** Python, FastAPI, PostgreSQL, OpenAI, Pydantic AI.

CancerScope — Deep-Learning Classification

2024

- Deep-learning classifier reaching **89% accuracy** with balanced per-class precision/recall on a held-out split across 4 breast-cancer subtypes (AUBMC clinical data), using SMOTE and PCA. **Tech:** Python, Keras, scikit-learn.

Education

American University of Beirut (AUB)

2022 – 2026

B.E. in Computer Science and Engineering — Graduated May 2026

Georgia Institute of Technology (OMSCS)

2026

M.S. in Computer Science — Admitted; matriculating 2026 (part-time, alongside full-time work)

Lycée Franco-Libanais Verdun

2022

French Baccalaureate, Science Track

Certifications, Coursework & Languages

Certifications: Deep Learning Specialization — Andrew Ng (Coursera: NNs, CNNs, Sequence Models, ML System Design) · Local RAG — Zero To Mastery (2025) · Prompt Engineering — Vanderbilt (Coursera) · AWS Certified Machine Learning (in progress, expected Aug 2026) · Docker & Kubernetes (Udemy).

Relevant Coursework: Agentic Systems (EECE 503P), Machine Learning, Artificial Intelligence in Industry, Software Engineering, Databases, Data Structures & Algorithms, Statistics, Data-Centric Engineering.

Languages: English (Fluent) · French (Fluent) · Arabic (Native).