

Mohamed Khaled Shams

AI Engineer | LLM, RAG & Agentic AI Systems

Cairo, Egypt · +20 100 922 2392 · imohamedshamss@gmail.com · [linkedin.com/in/imohamed-shams](https://www.linkedin.com/in/imohamed-shams) · mshams.cloud · github.com/IShamss · credly.com/users/mshams

SUMMARY

AI Engineer at IBM Consulting building production LLM, RAG and multi-agent systems for enterprise and government clients, including regulated pharma. Works across agentic AI, document intelligence, Model Context Protocol (MCP) and human-in-the-loop workflows on AWS, Red Hat OpenShift and Microsoft Copilot Studio, and takes client work from discovery and demos to production.

EXPERIENCE

AI Engineer · IBM Consulting · Cairo, Egypt · Mar 2025 – Present

- Awarded IBM Employee of the Quarter (Q3 2026) for exceptional delivery.
- Delivered AI engineering across two client engagements (pharma and government) and two IBM AI assets, from technical evaluation and client demos through production deployment.
- Competed in and taught at Bobathon 2026 and the 2026 IBMer watsonx Challenge, earning both event badges.
- Instructed the IBM Enterprise Advantage course and the DPSE class course for IBM colleagues.
- Built Salesforce AI agents for the Concentrix–Pearson bid and contributed to the proposal (presales).
- Mentored long-term interns on the Nestlé program and undergraduate summer interns, and served as a mentor in an HCD and QA upskilling program, earning the IBM Mentor badge.

Military Service · Egyptian Armed Forces · Egypt · Jan 2024 – Feb 2025

PMO Engineer · IT Systems Corporation · Cairo, Egypt · Jul 2023 – Dec 2023

ERP consulting.

- Supported a portfolio of ~10 concurrent ERP implementation projects, each with a 4–5 person delivery team, for clients including EGA (Mercedes-Benz), MUST University, DoorTech and Royal Star.
- Built MS Project schedules; wrote SOWs, RACI charts, stakeholder analyses and risk registers; ran client kickoffs.
- Created standard project templates for the PMO team, speeding up status reporting and document management.

Freelance AI Engineer & Full-Stack Developer · Fiverr, Upwork and direct clients · Remote · Aug 2021 – Dec 2023

- Reached Level 2 on Fiverr and Top Rated on Upwork with a 5-star rating, delivering for clients worldwide.
- Built AI features for Talk4Impact, an education platform, using OpenAI models: generated performance reports for students and tutors, and AI-run job interviews in which candidates answer written questions by voice.
- Built and deployed full-stack web apps with React, Node.js/Express, Laravel, MongoDB and PostgreSQL.
- Tutored mathematics and programming for international students, the most successful service across platforms.

Technical Mentor · Udacity · Remote · 2022 – 2023

- Mentored 100+ students in the Full Stack Web Developer Nanodegree and hosted 10+ live technical sessions.

Computer Vision Intern · IBM · Cairo, Egypt · Jul 2022 – Sep 2022

- Built an end-to-end Arabic license-plate recognition pipeline reaching ~85% accuracy on Egyptian plates, using YOLOv5 for detection, OpenCV image processing for character segmentation and KNN for classification; demoed it to IBM Egypt leadership.
- Collected and manually labeled a dataset of local car images, and built a self-labeling tool that runs the pipeline on an image and lets the user correct plate characters before saving.

PROJECTS

ADCA: Enterprise AI Agents (Microsoft Copilot Studio) · IBM Consulting · Abu Dhabi Customs Authority (government) · Sep 2026 – Present

- Build AI agents in Microsoft Copilot Studio with skills and tools that send and reply to emails, schedule events and send notifications, operating with admin-level Microsoft 365 permissions.
- Ground the agents in a large client-specific knowledge base for retrieval-augmented (RAG) answers.
- Deploy the agents to Microsoft Teams and Microsoft 365 as enterprise copilots.

GiniDocs: Agentic Regulatory Drafting · IBM Consulting · Boehringer Ingelheim (pharma) · Jul 2025 – Aug 2026

- Built, on a team of eight, an agentic application that turns ~5,000 pages of source documents into a ~500-page regulatory submission draft, producing 90% of required content correctly and cutting effort per document by 70% versus the manual baseline.
- Owned the document ingestion pipeline end to end, from OCR and layout parsing through chunking to indexing in Amazon OpenSearch, processing ~500-page document sets in ~30 minutes.
- Led the evaluation of IBM Docling against Azure Document Intelligence and recommended Docling for self-hosted deployment in a secure pharma environment and accurate parsing of complex nested tables; presented the evaluation and ingestion design to client stakeholders.
- Built multi-page table reconstruction that combines deterministic rules (consecutive pages, matching table dimensions) with LLM-based context matching, correctly merging 95% of tables split across pages, plus a vision-language model (VLM) step that reapplies source table styling.

- Deployed Docling as a REST service on Red Hat OpenShift (client) and AWS ECS Fargate/EC2 with load balancing and autoscaling.
- Contributed to the Azure-to-AWS migration, replacing Azure AI Search with Amazon OpenSearch and Blob Storage with S3, and helped split the monolith into containerized microservices (ingestion, generation, API routing) behind Amazon API Gateway.
- Designed the storage lifecycle: an autoscaling OpenSearch index treated as a rebuildable search cache, S3 as system of record, and older documents archived to S3 Glacier Deep Archive.

IBM Product Workbench (SDLC-AI): Multi-Agent SDLC Automation · IBM Consulting (IBM Enterprise Advantage) · Feb 2026 – Jun 2026

- Built core components of IBM Product Workbench, a multi-agent SDLC automation platform on LangChain DeepAgents, now deployed on IBM Consulting Advantage (ICA) for IBM teams globally; helped build its orchestrator, planner and executor agent framework.
- Implemented concurrent agent execution and performance optimizations, cutting end-to-end run time roughly in half.
- Built Model Context Protocol (MCP) connectors to Jira and IBM Context Studio for project tickets and team knowledge.
- Implemented lifecycle hooks that add human-in-the-loop review and approval to agent workflows.
- Validated the platform against previously delivered IBM projects: given only the original teams' documentation, it produced close reproductions of the delivered systems, assessed through human review.
- Demoed the platform to IBM leadership and clients, and delivered an internal ICA masterclass to IBM practitioners.

DocuFlow / DocumentsAI: Document Intelligence Asset · IBM Consulting (CatalystAI) · Mar 2025 – Sep 2025

- Built, on a team of eight, a document-intelligence asset tailored to client requirements and demoed before signing; after the demo, a Fortune 500 logistics company purchased the solution.
- Delivered LLM-based structured extraction with OpenAI GPT models for invoices, tax forms and reports, cutting manual extraction time by 80%; users define fields in natural language, and the system detects the document type and proposes a schema of common fields.
- Developed a multi-agent chat-with-documents system with AutoGen that writes Python code and runs it in an isolated container to aggregate values across documents and tables, answering questions beyond the extraction schema.
- Built the document ingestion pipeline and FastAPI backend (REST endpoints, routing) serving ~10 concurrent users.
- Implemented validation and confidence scoring, combining LLM confidence with human feedback to measure accuracy.
- Built Model Context Protocol (MCP) integrations: a SQL Server MCP server for writing extracted values, a mail MCP server for completion and failure notifications, and support for custom MCP servers.

EDUCATION

B.Sc. Computer and Communication Engineering · Cairo University · Giza, Egypt · Sep 2018 – Sep 2023

SKILLS

GenAI & LLMs: LLM application development, RAG, embeddings, chunking, hybrid search, prompt engineering, structured output, function/tool calling, Model Context Protocol (MCP), vision-language models (VLM), OpenAI GPT, Claude Code, evaluation, human-in-the-loop

Agentic AI: multi-agent orchestration, LangChain, LangChain DeepAgents, AutoGen, Salesforce agents, planner/executor architectures, agent code execution, lifecycle hooks

Document Intelligence: OCR, IBM Docling, Azure Document Intelligence, multi-page table reconstruction, schema-driven field extraction, confidence scoring

Computer Vision & ML: YOLOv5, OpenCV, KNN, dataset labeling

Languages & Backend: Python, JavaScript, SQL, FastAPI, Node.js, Express, Laravel, React, REST APIs, microservices

Data & Search: Amazon OpenSearch, Azure AI Search, PostgreSQL, MongoDB, SQL Server, S3 lifecycle and Glacier Deep Archive

Cloud & DevOps: AWS (ECS Fargate, EC2, Lambda, API Gateway, S3, IAM, CloudWatch), Red Hat OpenShift, Microsoft Copilot Studio, Microsoft 365, Microsoft Teams, Docker, load balancing, autoscaling, Git

Delivery & Enablement: client discovery, solution demos, presales and proposals, stakeholder communication, statements of work, RACI, risk registers, MS Project, Jira, technical training, mentoring, public speaking

Languages: Arabic (native) · English (fluent) · German (basic) · French (basic)

CERTIFICATES

Claude Certified Architect – Foundations (Anthropic, Aug 2026) · **Claude Partner Badge – Claude Code** (Anthropic, Sep 2026) · **AWS Certified AI Practitioner** (AWS, Dec 2025) · **IBM Generative & Agentic AI Data Engineer** (IBM, Apr 2026) · **IBM Generative & Agentic AI Foundation** (IBM, Mar 2026) · **AWS Partner: Generative AI Technical** (AWS, Apr 2026) · **AWS Partner: Agentic AI Essentials** (AWS, Mar 2026) · **AWS Partner: Generative AI Essentials** (AWS, Mar 2026) · **IBM Delivery Central Platform Foundations** (IBM, Apr 2026) · **IBM Garage Essentials** (IBM, Mar 2026)

COMMUNITY & SPEAKING

Speaker, Build with AI · Google Developer Groups (GDG) · [Event post](#)

Speaker, Google I/O Extended · Google Developer Groups (GDG) · [Event post](#)

Youth Representative, Egypt · Center for International Policy and Sustainability (CIPS)