AYMEN EMAD ELDIN

Senior Agentic AI Engineer · Multi-Agent Systems Architect · LLM Specialist

Bangalore, India (Open to GCC / US Remote) · +91 99025 62266 · aymen60003@gmail.com

github.com/Ayman-xmax · linkedin.com/in/ayman-emadeldin-71497a222

Languages: English · Arabic · Hindi — builds and evaluates multilingual AI systems across all three

PROFESSIONAL SUMMARY

Agentic AI engineer who designs, builds, and operates production multi-agent systems — hierarchical orchestration, autonomous task execution, MCP servers and agent-to-agent (A2A) delegation, and the harness engineering around them: instruction contracts, runtime tool governance, human-in-the-loop approval gates, and deterministic guardrails that keep autonomous agents reliable in production. Four years of progressive experience across regulated data operations, AI-driven automation, and agentic systems development on LangGraph, LangChain, CrewAI, and Google ADK, with multi-provider LLM orchestration spanning Anthropic Claude, OpenAI, and Gemini.

Deep RAG practitioner beyond basic chatbot patterns: contextual and cited retrieval, header-aware chunking strategy, provider-aware embeddings, and retrieval evaluation — applied to multilingual corpora (Arabic and English) and multimodal sources spanning text, images, audio, and office documents. Makes architecture decisions with explicit trade-offs across scalability, security, observability, governance, and cost. Owns ambiguous problems end-to-end; open to full US business hours and direct collaboration with US-based teams.

CORE COMPETENCIES

Agentic AI & Orchestration: Google ADK · LangGraph · LangChain · CrewAI · AutoGen · MCP (custom FastMCP servers) · A2A / hierarchical delegation · Sequential, Parallel & Loop workflow agents · planning, tool use, memory

Harness Engineering & Governance: Instruction contracts · runtime tool governance · human-gated state machines & approval gates · structural anti-hallucination guardrails · deterministic pure-code scoring · AI output evaluation

LLM Platforms: Anthropic Claude (API + MCP) · OpenAI GPT-4o · Google Gemini · LiteLLM multi-provider routing · Hugging Face Transformers · Whisper STT · OpenAI TTS

Advanced RAG & Vector Search: Contextual & cited retrieval · chunking strategy · provider-aware embeddings · ChromaDB · Pinecone · pgvector · Weaviate · retrieval optimisation & evaluation

Backend & Data: Python (advanced) · FastAPI · REST microservices · PostgreSQL · Pandas / NumPy · prompt engineering & optimisation

Observability & Cost: Prompt monitoring · cost & latency tracking dashboards · per-turn telemetry · structured logging · LangSmith

Frontend & Integration: React · TypeScript · git PR publishing · Laravel API integration · Slack API · n8n · Make.com · HubSpot

Infrastructure & DevOps: Docker · Kubernetes · AWS · GCP (Vertex AI agent ecosystem) · CI/CD · GitHub Actions

PROFESSIONAL EXPERIENCE

AI Solutions Partner & Strategic Advisor | **SuperAGI** *Feb 2026 – May 2026*

- Guided enterprise clients through agentic AI adoption, advising on deployment configurations by quantifying trade-offs between model quality, latency, and infrastructure cost.
- Conducted beta stress-testing of upcoming SuperAGI features; technical feedback directly shaped product reliability work and roadmap prioritisation.
- Acted as ecosystem bridge between core development and the user community, surfacing emerging use cases for autonomous agents and MCP/A2A integrations.

AI-Driven Operations & Automation Lead | **GameBee** *May 2025 – Feb 2026*

- Architected and deployed end-to-end AI automation using FastAPI microservices and LangGraph orchestration, partnering with product and design to define AI-powered capabilities across departments.
- Led integration of multi-agent systems for complex task orchestration, cutting manual intervention in repetitive processes by roughly 60%.
- Built the observability layer — prompt monitoring, cost tracking, latency dashboards, and AI output evaluation — to keep agentic workflows reliable in production.
- Shipped an internal RAG-backed support assistant (knowledge-base retrieval, conversation logging, smart escalation) that deflected routine queries from the operations team.

Senior Relationship Manager (AI & Data Focus) | **Gameberry Labs** *Jun 2023 – May 2025*

- Built LLM-powered personalisation and engagement models for VIP clients, using behavioural data to lift retention among high-value cohorts.

- Analysed large datasets in PostgreSQL with Google Data Analytics tooling, delivering insights that fed directly into client-satisfaction and revenue decisions.
- Designed CRM automation across n8n and HubSpot — automated communication, reporting, and lead-routing workflows with real-time Slack alerts.
- Mentored junior team members on AI tooling adoption and best practices, raising technical capability across the team.

Operations & Data Specialist | **Standard Chartered Bank** *2021 – 2023*

- Managed complex operations in a high-compliance banking environment, maintaining 100% accuracy across audit systems and regulatory reporting.
- Developed automated Python reporting tools for audit data extraction and verification; built foundations in structured data, PostgreSQL query optimisation, and CI/CD-driven deployment under strict governance standards.

KEY AI PROJECTS

AI/SEO Multi-Agent System — Autonomous Analysis & Publishing Platform *Feb 2026 – Present · Independent engagement*

Python · Google ADK (Vertex AI agent ecosystem) · FastMCP · LiteLLM · RAG · PageRank · Git/PR automation

- Architected a **20+ agent system** on Google ADK using hierarchical A2A delegation and Sequential / Parallel / Loop workflow agents, organised as a **three-phase, human-gated state machine** with explicit approval gates between phases.
- Built a **custom MCP server** (FastMCP, 16 tools) consumed through ADK McpToolset, cleanly decoupling the data layer from agent logic and making tools independently testable and governable.
- Implemented the RAG pipeline end-to-end: header-aware 1,800-character chunking, provider-aware embeddings (gemini-embedding-001 / text-embedding-3-small), and a NumPy cosine vector store with **cited retrieval** so every agent claim is traceable to source.
- Reverse-engineered commercial SEO authority metrics by implementing **PageRank** (damping 0.85, log-normalised) over a self-crawled link graph plus Common Crawl ingest — **replacing a $130/month SaaS dependency** with owned, auditable infrastructure.
- Engineered the harness: **instruction contracts, runtime tool governance, approval gates, and deterministic pure-code scoring** as structural anti-hallucination guardrails, so correctness never depends on prompt compliance alone.
- Delivered **multi-provider LLM support** (Gemini ↔ OpenAI via LiteLLM) and framework-agnostic publishing via a git PR workflow with a Laravel API integration.

MultiModal RAG Pipeline — Production Document Intelligence System

Python · LangChain · ChromaDB · Google Vision OCR · Hugging Face Transformers · PyTorch

- Built a production multimodal RAG pipeline covering ingestion, OCR, vector search, and LLM-powered question answering across text, images, and office documents (PDF, DOCX, PPTX, XLSX).
- Designed the retrieval architecture on ChromaDB with LangChain and OpenAI orchestration for context-aware answer generation over mixed-format knowledge bases.
- Integrated Google Vision OCR, PDFMiner, PDF2Image, and Unstructured for content extraction from scanned documents; Hugging Face Transformers and PyTorch for embedding generation and multimodal understanding.

Foody — Real-Time Voice-to-Voice AI Ordering Agent (Egyptian Arabic)

Python · Whisper API · Anthropic Claude · OpenAI TTS · Silero VAD · LangSmith

- Architected a low-latency voice pipeline localised for Egyptian Arabic — Whisper STT → Claude → OpenAI TTS — achieving sub-300 ms end-to-end response.
- Implemented Silero VAD for utterance segmentation and per-turn telemetry for latency monitoring and production bottleneck identification.
- Designed a stateless, protocol-based architecture allowing STT/TTS providers (e.g., Azure Speech) to be swapped without touching core logic; automated order parsing with structured-output validation.

AgentX — Multi-Agent AI Development System

Python · LangGraph · CrewAI · React/TypeScript scaffolding · LangSmith

- Built a production multi-agent system on LangGraph and CrewAI that generates full-stack frontend applications end-to-end.
- Designed specialised agents (Planner, Frontend, CodeMaster, Visualizer) with a stateless architecture for scalability and fault isolation.
- Implemented automated validation pipelines with AI output evaluation, raising output accuracy by ~40% and cutting manual development effort by ~60%; React/TypeScript scaffolding with function calling and structured outputs for deterministic code production.

EDUCATION & CERTIFICATIONS

Bachelor of Computer Applications (BCA) | **Bangalore City University** *Score: 90%*

Certifications: DeepLearning.AI — Multi-AI Agent Systems (CrewAI) · LangChain — LLM App Development · Hugging Face — Open-Source LLMs & Transformers · IBM — Python for Data Science & AI · Google — Data Analytics Professional · UPenn — Operational Performance · IELTS Band 7.5