

MOHAMMAD HAYAT

AI / ML Engineer | Generative AI & LLM Solutions Architect

Dubai, United Arab Emirates | +971 50 508 3587 | hayat28mir@gmail.com | [Linkedin](#) | [Github](#)

PROFESSIONAL SUMMARY

AI/ML Engineer and Generative AI Solutions Architect with 3 years of hands-on experience designing, building, and deploying production-grade LLM applications, agentic AI systems, and enterprise AI platforms on Microsoft Azure and Google Cloud Platform (GCP). Specialized in Retrieval-Augmented Generation (RAG), Agentic RAG, multi-agent orchestration, and tool/function calling using Azure AI Foundry, Azure OpenAI, Azure AI Search, Semantic Kernel, LangChain, LangGraph, and CrewAI. Experienced in defining AI architecture and technical strategy, leading solution design sessions with business stakeholders, evaluating LLMs, frameworks, and vendors, and implementing LLMOps practices for evaluation, observability, guardrails, and Responsible AI. Proven record of delivering scalable, explainable, domain-specific AI for ESG and sustainability reporting, finance, marketing, and business process automation.

CORE COMPETENCIES & TECHNICAL SKILLS

Generative & Agentic AI: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Agentic RAG, Hybrid Search, Re-ranking, AI Agents, Multi-Agent Systems, Agent Orchestration, Model Context Protocol (MCP), Tool / Function Calling, Structured Outputs, Prompt Engineering, Fine-tuning (LoRA / PEFT)

Microsoft Azure AI: Azure AI Foundry, Azure OpenAI Service (GPT-4o, GPT-4.1, o-series reasoning models), Azure AI Foundry Agent Service, Azure AI Search (vector, hybrid & semantic ranking), Azure AI Document Intelligence, Azure AI Content Safety, Prompt Flow, Azure Machine Learning

Azure Cloud & Infrastructure: Azure Functions, Azure App Service, Azure Container Apps, Azure Kubernetes Service (AKS), Azure Cosmos DB, Azure Blob Storage, Azure Key Vault, Azure API Management, Azure Monitor, Application Insights

Google Cloud AI: Vertex AI, Google Gemini (Gemini 2.5 Pro / Flash), Vertex AI Agent Builder, Cloud Run, Cloud Storage

Frameworks & Libraries: LangChain, LangGraph, LlamaIndex, Semantic Kernel, CrewAI, AutoGen, Hugging Face Transformers, FastAPI, Pydantic, PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy

LLMOps & Evaluation: LLM Evaluation, RAGAS, Azure AI Evaluation, LangSmith, Prompt Versioning, Guardrails, Observability & Tracing, Cost & Latency Optimization, Responsible AI

Vector Databases & Data: Azure AI Search, FAISS, ChromaDB, pgvector, Embeddings, Chunking Strategies, Semantic Search, Document Parsing, OCR, Natural Language Processing (NLP), Feature Engineering, SQL Server, PostgreSQL

Programming Languages: Python, C#, .NET, SQL

MLOps & DevOps: Git, GitHub Actions, Azure DevOps, Docker, Kubernetes, CI/CD Pipelines, REST APIs, Microservices, MLOps, Model Deployment & Monitoring

PROFESSIONAL EXPERIENCE

AI / ML Engineer

Jan 2026 – Present

Nextoria Info Tech | Dubai, UAE

- Built and launched PersonaXpand end to end, an AI social media management platform live on web, iOS, and Android, featuring multi-network publishing, an LLM-based Digital Twin, and AI auto-replies in the user's voice.
- Engineered AI capabilities across Nextoria's product suite, including markXpand (AI marketing and multi-channel ad campaigns) and ReachXpand (AI social media automation with grounded customer auto-replies for businesses).
- Defined AI architecture and technical strategy for enterprise Generative AI applications built on Azure AI Foundry and Azure OpenAI (GPT-4o, GPT-4.1), automating marketing content generation and business workflows.
- Architected Agentic RAG solutions combining Azure AI Search (hybrid vector + keyword search with semantic ranking), Azure AI Document Intelligence for document ingestion, and enterprise data sources to deliver grounded, context-aware responses.
- Built AI agents and multi-agent workflows using Semantic Kernel and Azure AI Foundry Agent Service with tool/function calling and structured outputs to integrate with internal systems via REST APIs.
- Delivered end-to-end AI applications on Microsoft Azure, from solution design and model integration through API development (Python, C#, .NET), containerized deployment on Azure Container Apps and Azure Functions, and CI/CD with Azure DevOps.
- Implemented LLMOps practices including prompt versioning, LLM evaluation for groundedness and relevance, Azure AI Content Safety guardrails, and monitoring with Application Insights to keep AI services reliable, secure, and observable in production.
- Led solution design sessions with business stakeholders, translating requirements into AI architectures and evaluating LLMs, frameworks, cloud services, and vendors for cost, latency, and accuracy trade-offs.

Tech Stack: Python, C#, .NET, Azure AI Foundry, Azure OpenAI, Azure AI Search, Azure AI Document Intelligence, Semantic Kernel, Azure Functions, Azure Container Apps, Azure DevOps, Application Insights, SQL, Git, Docker

AI Developer

Jan 2024 – Jan 2026

BriskFlow

- Designed and built an AI-powered ESG XBRL tagging engine using NLP and RAG pipelines, automating taxonomy tagging across multiple ESG reporting standards (IFRS, ESRS, SASB).
- Developed Brisky, an LLM-driven ESG chatbot built with LangChain agents and Google Gemini on Vertex AI that maps sustainability concepts to taxonomy tags and delivers concise, explainable reporting guidance.
- Engineered the retrieval layer with optimized chunking strategies, embeddings, and re-ranking over FAISS and ChromaDB vector stores, improving the precision of taxonomy tag recommendations.
- Implemented document parsing and auto-recommendation systems that highlight taggable ESG text and suggest XBRL tags, reducing manual annotation time by approximately 60%.

- Deployed a FastAPI backend containerized with Docker for scalable LLM inference and RESTful integration, using structured outputs and validation logic to ensure consistent, auditable tagging decisions.

Tech Stack: Python, LangChain, Google Vertex AI (Gemini), FastAPI, Pydantic, FAISS, ChromaDB, Docker, ESG Taxonomies (IFRS, ESRS, SASB), XBRL

KEY PROJECTS

PersonaXpand – AI Social Media Manager personaxpand.com

Live on Web, iOS App Store & Google Play

- Built PersonaXpand end to end, a production AI SaaS platform that lets users write a post once and publish it to 8 social networks (Instagram, Facebook, X, LinkedIn, YouTube, TikTok, Threads, Bluesky), with each version automatically adapted to the platform's character limits, media formats, and hashtags.
- Developed an LLM-powered "Digital Twin" that learns a user's tone, topics, and writing style from their existing posts and drives personalized captions, hashtags, AI-generated images, and replies.
- Engineered an AI Auto-Reply system that detects intent and responds to comments, DMs, and mentions in the user's voice, with human-in-the-loop approval mode, per-platform automation controls, and guardrails that prevent reply loops.
- Delivered AI analytics and a Reputation Audit that generates brand reports, best-time-to-post recommendations, and trending topic suggestions.
- Integrated official platform APIs through OAuth, with AES-256 encrypted tokens, quota-aware publishing, circuit breakers for rate-limit errors, and GDPR/CCPA-compliant data controls.

markXpand – AI Marketing Platform markxpand.com

Live in Production

- Engineered AI capabilities for an end-to-end marketing platform that turns a business's products and objectives into live ad campaigns across Google Ads, Meta, YouTube, TikTok, LinkedIn, Snapchat, Email, and WhatsApp.
- Built AI workflows for campaign strategy, competitor intelligence from live web research (pricing, USPs, SWOT, market analysis), platform recommendations, and budget allocation across channels.
- Developed generative AI pipelines that extract a Brand Kit from a company website and produce on-brand ad copy, email subject lines, images, and short-form videos in every platform size and aspect ratio.
- The platform has delivered 14M+ ad impressions and reached 3.7M+ people for 15+ brands, per publicly reported figures.

ReachXpand – AI Social Media Automation for Businesses reachxpand.com

Live in Production

- Engineered AI features for a multi-tenant social media automation platform for businesses, publishing to 8 networks with a separate workspace, Brand Kit, and AI voice per brand.
- Built a grounded customer auto-reply system that answers only from approved brand facts (FAQs, offers, links), refuses to invent answers, and hands unknown questions to a human, with per-platform approval modes.
- Developed an autoposting engine that plans weekly content, generates images and captions, and creates brand videos from a brief through strategy, storyboard, and render for Reels, TikTok, Shorts, and YouTube.
- Implemented AI inbox classification that separates leads from complaints, plus a reputation audit that analyzes Google reviews and customer conversations into a weekly brand health score.

AI-Powered ESG Tagging Assistant (Brisky)

- Built an intelligent ESG chatbot with a FastAPI backend, a LangChain agent, and Google Gemini (Vertex AI) integration for taxonomy mapping and sustainability queries.
- Automated ESG report parsing to highlight taggable text and recommend XBRL tags with explainable reasoning.

AI Recommendation System for ESG XBRL Reports

- Developed a RAG-based engine grounded in multiple ESG taxonomy standards for XBRL tag prediction.
- Built an AI agent that suggests, validates, and automates XBRL tagging decisions.

EDUCATION

Master of Computer Applications (MCA) Islamic University of Science & Technology, Kashmir	2021 – 2023
Bachelor of Computer Applications (BCA) Amar Singh College, Srinagar	2017 – 2020