

# FAHEEM KHAN

Senior Solution Architect • GenAI & Agentic AI Systems

Tustin, CA 92782 | 949.677.1997 | fahkhan172@gmail.com | faheemkhan.link | linkedin.com/in/faheem-profile

## PROFILE

Solution and enterprise architect with 12+ years of progressive architecture experience and 3+ years designing and delivering production LLM, RAG, and agentic AI systems. Owns GenAI architecture end to end: LLM selection, retrieval design, agent orchestration with MCP-based tool integration, guardrails and AI governance, observability, and reliability, on Microsoft Foundry / Azure AI Search and AWS (Bedrock, serverless). Deep financial-services background covering title and escrow, Fedwire wire transfers, reconciliation, and regulatory compliance. Sets reference architectures, coding standards, and secure integration patterns that align engineering teams, and turns complex AI designs into clear diagrams and documentation for business stakeholders.

## CORE SKILLS

**GenAI / LLMs:** Claude, GPT (incl. Realtime API), Gemini, DeepSeek; LLM selection & prompt design

**RAG & Retrieval:** Azure AI Search, Pinecone (vector DB), Neo4j (graph)

**Agentic AI:** MCP servers, tool/function calling, Microsoft AI Framework, Copilot Studio, GitHub Copilot

**AI Platforms:** Microsoft Foundry, Amazon Bedrock, Ollama, Foundry Local

**AI Governance:** Responsible AI policies, guardrails, secure integration patterns, conversation logging & audit

**Observability:** Prometheus, Grafana, Splunk, AppDynamics, CloudWatch, X-Ray

**Real-time Comm.:** LiveKit, Amazon Connect, Dialogic IVR

**Languages:** Python, C#, TypeScript, JavaScript, SQL, C++

**Backends & APIs:** .NET Core, ASP.NET, Node.js/Express, REST, GraphQL, gRPC, SOAP; React

**Cloud:** AWS (Lambda, ECS, EKS, EC2, Step Functions, S3, API Gateway), Azure (App Service, Functions, APIM), Docker

**Eventing:** Kinesis, SNS/SQS, Azure Service Bus, Event Grid, Tibco EMS

**Data:** Snowflake, Oracle, MS SQL, Aurora PostgreSQL, DynamoDB, Domo

**CI/CD & DevSecOps:** GitHub Actions, Azure DevOps, CloudFormation, SonarQube, Veracode, Secrets Manager

**Security:** Cognito, Azure AD, IAM, OAuth 2.0, SAML, OIDC, KMS, VPC

**Architecture:** Microservices, Event-Driven, Pub/Sub, API-First, CQRS; TOGAF, Well-Architected; DDD, SDD, TDD, BDD

**Gateways:** MuleSoft, AWS API Gateway, Istio, Akana

## WORK EXPERIENCE

### Veros Real Estate Solutions, Santa Ana, CA

08/2025 – Present

Senior Solution Architect

- **Production RAG:** Architected and delivered a production RAG chatbot inside an external appraisal and valuation application that helps users analyze comparables (comps) and apply USPAP guidelines for improved accuracy and compliance. Retrieval runs on Azure AI Search, with models hosted in Microsoft Foundry.
- **GenAI framework:** Built a custom GenAI framework on the Microsoft AI Framework with governance controls, monitoring, and user-conversation logging built in, so AI applications ship with consistent guardrails and auditability.
- **Agentic AI:** Led the architecture of an agentic email-response system powered by Claude that automates the investigation of customer-reported appraisal discrepancies. The agent calls tools through MCP integrations with Snowflake, Oracle databases, and external web sources.
- **AI governance:** Led enterprise AI governance and responsible-AI adoption, defining policies, guardrails, and secure integration patterns. Rolled out GitHub Copilot and MCP servers to give AI tools secure access to internal systems such as Azure DevOps.
- **Reference architecture:** Authored the company-wide Microservices Strategy that modernized legacy systems into a cloud-native architecture. It set standards for serverless compute, observability, and CI/CD automation, aligned all engineering teams on one architectural vision, raised deployment velocity, and reduced system coupling.

### Allama-AI, Irvine, CA

12/2024 – 07/2025

Consultant – Cloud and AI

- Architected an AI-powered customer-service platform (ai.365Sols.com) that provides 24/7 voice support. It uses LiveKit for real-time communication, AWS serverless microservices (Lambda, API Gateway, Cognito, DynamoDB), and Stripe for payments and subscriptions.

- Built a real-time AI interview platform (CallforInterview.com) for a recruiting firm with LiveKit, OpenAI, Python, React, and AWS (Lambda, API Gateway, Cognito). The platform generates questions dynamically during voice interviews and produces candidate scorecards that help recruiters filter and evaluate candidates. Wrote C# REST APIs to manage interview sessions and candidate data.

## **First American Title, Santa Ana, CA**

**11/2019 – 11/2024**

### *Senior Solution Architect*

- Delivered an AI solution using OpenAI Studio, ChatGPT, and Azure Cognitive Services that improved document management and operational efficiency.
- Designed an enterprise-wide reconciliation solution that prevented more than \$22M in write-offs, built on AWS API Gateway, DynamoDB, a React UI, and a rule engine on AWS EKS. It improved financial accuracy, compliance, and efficiency.
- Re-architected on-premises middleware into a serverless AWS architecture (S3, Lambda, Step Functions, Secrets Manager), reaching 99.9% availability and faster speed to market.
- Improved scalability and maintainability of the Core API by 15% by leading a team to modularize the monolith with Domain-Driven Design, modernize the framework, and centralize business logic in a unified service layer.
- Raised availability of the monolithic Settlement and Title system to 99.5% with an in-place deployment strategy and an optimized CI/CD pipeline that enabled zero-downtime releases.
- Led design of modular platforms using event-driven and microservices architectures for rapid scaling and smooth integration across systems.

## **First American Title, Santa Ana, CA**

**07/2013 – 10/2019**

### *Enterprise Architect*

- Designed a mission-critical outgoing wire-transfer application to the Fedwire system using Azure App Service, Functions, Service Bus, and Azure SQL. Secured requests and responses between trusted applications with Gemalto Luna HSM-based signing and verification.
- Enabled incoming wire processing with a BAI2-standards integration from Fedwire and an automated SFTP integration with Comerica Bank for timely deposit of received funds.
- Designed an Azure-based overdraft workspace (App Service, ExpressRoute, Azure AD) that lets escrow staff approve overdraft transactions remotely and securely.
- Led a multi-year TRID compliance program that overhauled the core system with a UX redesign, precise Closing Disclosure rendering, MISMO transformation, and internal and external system integrations.
- Implemented automated screening of flagged entities during data entry, with a review workflow, to comply with Executive Order 13224.

## **EDUCATION**

---

**Bachelor's Degree in Software Engineering** — SZABIST University