Ghulam Mustafa Zafar

AI Architect

Remote (Worldwide) · Lahore, Pakistan · +92 332 7455045 · mustafazafer9999@gmail.com
mustafazafar.dev · linkedin.com/in/musta1337 · github.com/Musta1337

AI and platform architect for multi-tenant SaaS. Owns the target architecture for an enterprise field service platform's AI and data estate: **composable, MACH-aligned** and API-first, built for UK/EU data residency and tenant isolation. Takes systems from build-versus-buy decision to production at scale, and leads the engineers who deliver them.

Experience

Joblogic *(Field Service Management SaaS)* — Lahore, Pakistan

AI Architect, AI and Data Platform — *Feb 2025 – Present*

- **Target architecture.** Defined the composable, MACH-aligned target architecture for the AI platform and the roadmap teams deliver against. Led the build-versus-buy evaluation of Salesforce Agentforce and Microsoft Copilot Studio; neither could model field service scheduling, meet UK/EU residency and tenant isolation, or operate on the live operational data model, so the decision was to build FSM-native.
- **Composable agent platform.** Domain capabilities are exposed as typed, permissioned, API-first services, and agents are composed from them declaratively, so a new agent ships as configuration rather than code. Partner integrations such as Sage and Xero plug in through an extension SDK and run isolated per tenant. **10+ agents** run live for the **100+ tenants** in the AI pilot programme, out of **3,500+** on the platform.
- **Automation as a platform capability.** Now extending the platform to automation: a tenant describes a workflow in plain language, an agent builds it, and it runs on a schedule, on an event or in the background as a serverless Azure Container Apps job. The platform owns provisioning, scaling and isolation, so neither tenants nor product teams manage infrastructure. It is set to replace the Marketplace's earlier automations.
- **Headless, omnichannel delivery.** Designed the agent layer to be headless: in-app chat, real-time voice (Pipecat, Azure Speech, ElevenLabs) and email (Nylas) are channels over the same capability layer, so anything built once is available on every surface, with voice held to a sub-second conversational turn.
- **Security and compliance.** Own the AI security and compliance architecture: prompt injection and jailbreak defence on AWS Bedrock Guardrails, PII redaction before any model call, UK/EU data residency, and tenant isolation enforced at the retrieval layer. Accountable for the AI controls under ISO 27001 and GDPR, and for the AI side of enterprise customers' security reviews.
- **Reliability and cost.** Set the cost, latency and quality SLOs the platform runs to, backed by an evaluation harness for regression testing and end-to-end tracing. The platform has served **200,000+ agent runs** at **99.9% reliability**, and **AI spend fell by over 50%** through context and caching strategy and a move from monolithic agents to supervisor and sub-agent orchestration.
- **Data platform and analytics.** Architected the Databricks lakehouse (Unity Catalog, Delta Lake, medallion) that consolidated analytics across operational databases, source systems and CRMs, plus a low-latency read layer serving every tenant. Designed a Neo4j knowledge graph of the FSM domain that grounds natural-language self-service BI: reports that carried a **two-week SLA** from the reporting team, customers now build **in minutes**.
- **Applied ML.** Directed the predictive models behind scheduling (engineer-to-job assignment, duration and ETA prediction, SLA-breach and rework risk), all running live in dispatch. Now leading on-device inference with quantised Gemma on LiteRT and fine-tuned action models.
- **Leadership and governance.** Reporting to the CTO, grew the AI/Data organisation past **20 engineers** across ML, LLM, data engineering and analytics, and own its hiring bar. Run architecture review for every significant design and set the platform standards other teams build on. Remain hands-on in the .NET product codebase.
- **Vendors and partners.** Own relationships with AWS, Databricks and AI model and voice providers: subprocessor agreements, security and infra reviews, and commercial terms, including better Bedrock pricing on volume.

Automation Developer, Integrations and Marketplace — *Sep 2023 – Feb 2025*

- Designed the Marketplace's **composable integration model**: each accounting and business-system integration is an independently deployable, API-first service on Azure Functions and App Service, contract-defined in OpenAPI and built to reconcile correctly after late or partial runs. The model became the foundation of the AI platform's extension SDK and remains in production.
- Standardised 50+ automations on shared async, retry and observability patterns, so the whole integration estate fails, recovers and is monitored the same way.

– Built the migration pipelines that bring customers over from competing platforms, reconciling mismatched source schemas with validation reports, and cut onboarding for **hundreds of customers** from days to hours. Productised the public API as a Zapier integration.

Independent Consulting Remote
ML Engineering and Systems Interoperability *Mar 2021 – Mar 2023*

– Delivered a production signal processing and ML suite for a client: FFT/STFT denoising, feature engineering and classical ML models implemented from first principles in C/C++, behind a C# desktop application and CLI.
– Reverse-engineered proprietary protocols and binary formats to recover data from closed legacy and industrial systems, delivering documented, tested, production-ready tooling.

Selected Projects

Tessellate *composable, multi-tenant AI agent platform* Architect

– Businesses publish agents that their own customers enable with their own accounts, so each agent acts for one person under that person's consent. Agents are declarative, sealed into a version-pinned plan at publish, and run durably on Temporal through crashes and long waits, with streaming, guardrails, memory, and event, poll and schedule triggers. New models, tools and connectors plug in with no change to the core, proven when an AI agent built a Stripe connector from the docs alone. Credentials stay behind an egress proxy, outside the agent path.

Ariadne *runtime-aware code analysis for coding agents* Research

– A code graph combining static analysis with runtime tracing flags code an agent wrote but never wired into the product, scoring **F1 0.80 against 0.04** for the closest open-source tool in a pre-registered evaluation. Agent tools built on the graph made a coding agent **2–6× more efficient** than a text-only baseline.

Infraio *AI-native SSH and DevOps management platform* Founder

– Browser-based SSH for server fleets with shared, recorded terminals, runbook automation, and a lightweight agent deployed on customer machines. The AI operator requires risk analysis and human approval before any destructive command and builds a memory of each host, on a multi-tenant backend with SSO, RBAC and usage-based billing.

Technical Skills

Architecture: MACH, composable architecture and packaged business capabilities, API-first, headless design, multi-tenant SaaS, event-driven and durable workflows, build-vs-buy evaluation, architecture governance
AI and Agentic Systems: Multi-agent orchestration, capability registries, RAG and knowledge graphs, context engineering, evaluation and observability, LangGraph, Claude Agent SDK, AWS Bedrock, Pipecat
Machine Learning: Gradient boosting, time series forecasting, PyTorch, scikit-learn, LoRA/QLoRA fine-tuning, on-device inference and quantisation
Security and Compliance: Prompt injection defence, guardrails, tenant isolation, PII redaction, data residency, ISO 27001, GDPR and UK DPA
Data: Databricks, Delta Lake, Unity Catalog, Spark, Neo4j, SQL Server, ETL/ELT
Cloud and Delivery: Azure (Container Apps, Functions), AWS, Bicep, Azure DevOps, Temporal, Grafana
Languages: Python, C# and .NET, TypeScript and React, C/C++, SQL

Education

BSc Computer Science, National University of Computer and Emerging Sciences Islamabad, Pakistan