

— PRINCIPAL ENGINEER & TECHNICAL LEAD

SHAIK MUZAMMIL AHMED

Building & leading production AI and full-stack systems · 10+ years

IMMEDIATE JOINER

GENERATIVE AI · FULL-STACK ENGINEERING · AZURE & MULTI-CLOUD · ENGINEERING LEADERSHIP



📍 Abu Dhabi, UAE · open across UAE & GCC 📞 +971 52 595 6123 · WHATSAPP ✉ encodingmuzzu8.0@gmail.com

🌐 muzammilahmed.com 🐙 github.com/encoding-muzzu 🔗 linkedin.com/in/encoding-muzzu

PROFILE

Principal Software Engineer and Technical Lead with **10+ years** architecting, building, and leading the delivery of production-grade AI and full-stack systems. Has shipped enterprise platforms at serious scale — document models **trained for 120+ countries** and **processing 3M+ documents/month** in production, and identity systems trusted by **leading public- and private-sector banks** — including **SBI Cards, VISA, and Standard Chartered** — and founded AI products end-to-end as a solo engineer. Blends deep, hands-on engineering across **Generative AI**, full-stack development, and **cloud & GPU infrastructure** with proven leadership — mentoring engineers, setting technical direction, and turning ambiguous, high-stakes requirements into reliable, scalable solutions delivered in secure, **regulated (RBI, SOC)** environments.

ENGINEERING RANGE Fluent across generations and scales of compute — from Windows 2000-era 128 MB systems to Apple M4 silicon and 256 GB+, many-core enterprise servers; equally at home in traditional infrastructure, cloud, high-performance computing, and GPU-accelerated AI environments.

CAREER HIGHLIGHTS

- ◆ **Architected and delivered enterprise-grade AI and full-stack platforms** processing **3M+ documents/month** — with document models **trained for 120+ countries** and **live across 10+** — owning the full lifecycle from front-end to cloud.
- ◆ **Pioneered production Generative AI capabilities** — RAG, agentic and multi-agent orchestration, fine-tuning, and LLM evaluation — taking them from prototype to dependable, monitored production systems.
- ◆ **Drove substantial gains in cloud-cost efficiency and scalability** through architecture optimization and modern containerized, microservices-based deployment.
- ◆ **Led and mentored high-performing cross-functional teams** — **35+ engineers at peak**, spanning front-end, back-end, database, QA, and DevOps — establishing engineering, security, and quality best practices that raised delivery speed and product reliability.
- ◆ **Trusted technical advisor to clients and senior stakeholders**, consistently turning ambiguous, high-stakes requirements into well-architected, defensible outcomes.

CORE COMPETENCIES

Generative AI & ML

LLMs (Claude, OpenAI, **Azure AI**, open-source) · RAG (vector search, reranking, evaluation) · Agentic & multi-agent orchestration (MCP, LangGraph) · Fine-tuning & domain adaptation (PyTorch, TensorFlow) · Prompt engineering · LLM evaluation & monitoring · LangChain · LlamaIndex · Vector DBs (PGVector, FAISS) · OCR / document intelligence · Voice AI agents (real-time STT/TTS — Whisper, ElevenLabs) · GPU inference optimization (vLLM)

Full-Stack Engineering	Python (FastAPI, Flask, Django) · Node.js · Express.js · NestJS · React.js · Next.js · TypeScript · JavaScript (ES6+) · Responsive UI · Reusable components · Redux · Context API · Tailwind CSS · Styled Components · Material UI
Backend & Architecture	Scalable RESTful APIs · Microservices · Distributed systems · Third-party integrations · Authentication (OAuth, JWT, SSO, Keycloak) · Event-driven patterns · Backend-as-a-Service (Supabase, Directus) · System design
Cloud & DevOps	Microsoft Azure · AWS · GCP · Docker · Kubernetes · CI/CD · Linux · On-prem / hybrid / cloud deployment · Observability & monitoring · DevSecOps
Data	PostgreSQL · SQL Server · MongoDB · Elasticsearch · Snowflake · Data modeling · Stored procedures
Quality & Tooling	Playwright (end-to-end automation) · Jest · Unit testing · Performance tuning · Debugging · Code reviews
AI Tools & Agents	Cursor · Windsurf · Claude Code · GitHub Copilot · Cline · ChatGPT · LangGraph · CrewAI · OpenAI Agents SDK · Lovable · Bolt.new · n8n
Leadership	Team leadership & mentoring · Technical direction & architecture · Engineering best practices · Stakeholder & client engagement · Agile / Scrum · Documentation

PROFESSIONAL EXPERIENCE

Principal Software Development Engineer — M2P Fintech

APR 2024 – PRESENT

Senior technical leadership — architecture, AI delivery, and engineering standards across full-stack teams.

- Led the **architecture and end-to-end delivery** of **production AI and full-stack products (Python, React, Node.js)** for identity-verification and document-automation platforms serving a **global user base**.
- Pioneered **Generative AI capabilities** — **RAG pipelines, agentic and multi-agent workflows** (MCP, LangGraph), **model fine-tuning**, and rigorous **LLM evaluation and monitoring** (Claude, OpenAI, Azure AI).
- Set **technical direction** and **mentored engineers** across the full stack, enforcing engineering, security, and testing **best practices** within **regulated (RBI, SOC)** environments.
- Served as a **trusted technical point of contact** for stakeholders and clients, translating ambiguous requirements into **well-architected, reliable solutions**.

Senior Engineering Manager — M2P Fintech

APR 2023 – MAR 2024

People and delivery leadership across backend, web, mobile, and QA on real-time verification platforms.

- Led and grew **cross-functional engineering teams** across backend, web, mobile, and QA, owning **delivery, quality, and release readiness** for real-time verification and onboarding platforms.
- Drove significant **cloud-cost reduction and scalability improvements** through architecture optimization and **containerized, microservices-based deployment** (AWS / GCP, Kubernetes).
- Built **production AI/ML systems** — **OCR, document intelligence, face matching, and anti-spoof detection** — instrumented with **monitoring and operational diagnostics**.

Lead Software Development Engineer — M2P Fintech

FEB 2019 – MAR 2023

Hands-on technical lead — full-stack platform delivery and engineering modernization across teams.

- Architected and built **full-stack platforms with Python and React** — large-scale **PII-scanning and document-automation** tooling spanning **PostgreSQL, MongoDB, Elasticsearch, and S3**.
- Spearheaded **engineering modernization** across teams — migration to **Linux / AWS** and adoption of **Docker, CI/CD**, and code-quality standards.

Full-Stack Developer — LaunchMyICO

JAN 2018 – FEB 2019

End-to-end product development for a trading / transaction platform.

- Built **end-to-end web applications** with **React, Node.js, Express, and MongoDB**, including **secure transaction workflows** and **smart-contract integration**.

Web Developer — Trescon

2016 – 2017

- Delivered **responsive, high-performance websites** with **front-end optimization, SEO, and third-party API integrations**.

KEY PROJECTS

◆ Document Intelligence Platform · M2P Fintech

TECHNICAL LEAD

- Led architecture and delivery of a **multi-model document intelligence pipeline** with models **trained to support documents from 120+ countries** — now **in production across 10+ countries** at **3M+ documents/month** and **99% + extraction accuracy** on clean inputs.
- Enabled **training of any new document type in under 2 hours** via a **CNN classifier + LayoutLMv3 fine-tuning** (PyTorch / TensorFlow), allowing rapid onboarding of new regions and formats.
- Built a **hybrid search and retrieval layer (OpenSearch + FAISS)** with storage on **AWS S3 + PostgreSQL** for scalable ingestion across high-volume production workloads.
- Augmented extraction with **LangChain, LlamaIndex, OpenAI, AWS Bedrock & Gemini** for LLM-powered understanding; served models via **FastAPI** microservices on **Docker / AWS**.

Stack Python · LayoutLMv3 · PyTorch / TensorFlow · OpenSearch + FAISS · LangChain / LlamaIndex · Bedrock / OpenAI / Gemini · FastAPI · AWS S3 · PostgreSQL · Docker

◆ Face Match & Anti-Spoofing Platform · Banking & Fintech

TECHNICAL LEAD

- Led delivery of a high-performance facial recognition platform supporting **1:1 verification, 1:N identification**, and **passive anti-spoofing** for enterprise KYC and identity workflows.
- Developed an **iBeta-certified anti-spoofing** solution detecting **digital-replay, print, and 3D-mask attacks**, significantly reducing identity fraud and account-takeover risk.
- Engineered a face search engine scanning **2M+ facial records in under 1 second** at **100+ TPS** through optimized embeddings, vector indexing, and GPU-accelerated inference.
- Built embedding storage and similarity search on **PostgreSQL + PGVector**; optimized inference with **PyTorch + ONNX Runtime**, deployed as **FastAPI / Flask** microservices on **AWS / Docker**.

Stack PyTorch · ONNX Runtime · PostgreSQL + PGVector · FastAPI / Flask · AWS · Amazon S3 · Docker · GPU servers

◆ Video KYC & Digital KYC Suite · Banking & Fintech

PRINCIPAL SDE

- Led architecture and delivery of a **Video KYC / Digital KYC platform** enabling **fully remote onboarding** for **SBI Cards, VISA, Standard Chartered, UCO Bank & Unity SFB**.
- Scaled the platform to **80,000+ KYC verifications/month**, eliminating branch visits and physical paperwork while improving verification success rates.
- Integrated **WebRTC-based live video** for secure customer-agent sessions, with automated document capture, **OCR**, audit trails, and compliance checks.
- Architected cloud-native microservices in **React, Node.js, .NET, Python, MySQL & Docker**, aligned with **RBI and financial regulatory** guidelines; reduced onboarding turnaround time.

Stack React · Tailwind CSS · Node.js · .NET · Python · MySQL · WebRTC · Docker

◆ PII Discovery & Data Privacy Scanner · Enterprise Data Security

PRINCIPAL SDE

- Led delivery of a large-scale **PII discovery platform** scanning **900M+ records** across enterprise databases and storage for centralized visibility into sensitive data assets.
- Built detection and classification engines for **Aadhaar, PAN, credit/debit cards, bank accounts, email & mobile**, with configurable rules, pattern matching, **checksum validation**, and risk-based scoring.
- Engineered a **distributed scanning framework** across **PostgreSQL, MySQL, MongoDB, MSSQL, Elasticsearch & S3** with minimal impact on production workloads.
- Delivered centralized **reporting and analytics** for security and compliance teams; architected scalable **Python + Docker** microservices for horizontal scale.

Stack Python · Docker · Amazon S3 · PostgreSQL · MySQL · MongoDB · MSSQL · Elasticsearch · Distributed processing · Microservices

INDEPENDENT VENTURES

◆ PerfectClip · AI Video Generation

SOLO DEVELOPER

- **Solely built and operated** an AI video platform end-to-end — frontend, backend, infrastructure, DevOps, QA, and production ops — generating **4M+ personalized videos in 30 days**.
- Built AI voice pipelines with **text-to-speech and voice cloning** for human-like personalized audio at scale, plus **Whisper**-based speech-to-text for automated transcription.
- Engineered high-throughput rendering with **FFmpeg, GPU acceleration, and distributed processing**; optimized inference via **PyTorch, CUDA & ONNX Runtime** to cut latency and cost.

- Designed a highly available architecture on **Docker, NGINX & GPU servers** with monitoring, fault tolerance, and auto-recovery for high-volume workloads.

Stack Python · Flask · PyTorch · CUDA · ONNX Runtime · FFmpeg · Whisper · WebRTC · PostgreSQL / Supabase · Amazon S3 · Docker · NGINX

◆ Pixalmania · AI Event Photo Delivery

FOUNDER · PIXALMANIA.IN

- **Founded and solely built** an AI event-photo platform — product, architecture, AI/ML, full-stack, and ops — that auto-delivers personalized galleries to attendees via facial recognition.
- Engineered a face search engine scanning **1M+ facial records in under 1 second**, enabling near real-time photo discovery and delivery.
- Developed **face clustering, grouping, and duplicate detection** to automatically organize event photos by attendee, eliminating manual sorting for photographers.
- Built recognition with **InsightFace, CNN & TensorFlow** and a responsive **Next.js / Vite** experience on **Node.js, Python, PostgreSQL / Supabase, Docker & Linux** with privacy-first access controls.

Stack Node.js · Python · Next.js · Vite · InsightFace · TensorFlow · CNN · PostgreSQL / Supabase · Docker · Linux

◆ Token Distribution System (TDS) · Web3

SMART CONTRACTS

- Designed an end-to-end **token distribution platform** supporting **allocation, vesting, cliffs, airdrops, staking rewards, and treasury** distributions via smart contracts.
- Built configurable **vesting engines** (linear, milestone-based, cliff, and custom) and automated **airdrop / reward** workflows for large-scale on-chain transfers.
- Integrated **MetaMask / Web3** wallet auth and **OpenZeppelin** security patterns, with **blockchain event listeners** and audit trails for full transparency and traceability.
- Built real-time allocation dashboards backed by **Redis** caching/queues on **AWS EC2 / S3 / Docker** for performance and reliability.

Stack Node.js · React · Web3.js · Solidity · OpenZeppelin · MetaMask · Redis · MySQL · AWS EC2 / S3 · Docker

KEY ACHIEVEMENTS

- ★ Designed and delivered a complete **banking portal — 100+ Figma screens in two nights** — to meet an aggressive project deadline.
- ★ Drove the organization's **migration from Windows to Linux infrastructure**, improving stability and operational efficiency.
- ★ Championed and rolled out a **modern tech stack across the Hyderabad office** — Node.js, React.js, Python, MongoDB, PostgreSQL, Docker, and Kubernetes.
- ★ Planned and executed the **migration of 30+ production servers to AWS in 5 days** with minimal downtime.
- ★ **Regained control of a compromised production website**, restored operations, and hardened security to prevent recurrence.

EDUCATION & RECOGNITION

B.Com

Suresh Gyan Vihar University, Jaipur · 2025

Intermediate (MPC)

Osmania College, Kurnool · 2012

★ **High Availability Award** — Syntizen, 2023

★ **Engineering Champion** — Trescon, 2017

LANGUAGES English, Urdu, Telugu & Hindi — read, write & speak · Arabic — read & write.

INTERESTS Competitive bodybuilding — **4th place, Mr. Rayalaseema Championship (55 kg category).**