

VISHAL BHOSLE

Head of Software Engineering | AI & Cloud Architect

Mumbai, India (Open to Relocation)

+91 8850013526

vishalbhosle83@gmail.com

<https://www.linkedin.com/in/vishal-bhosle>

PROFESSIONAL SUMMARY

Head of Software Engineering with 7+ years of experience architecting **0-to-1 AI products** and scaling global platforms across **the US, EU, SE Asia, India, and the GCC**. Expert in **Vision Intelligence (MediaPipe)**, **RAG-driven diagnostics**, and high-concurrency serverless systems. Proven impact: achieved an **80% global latency reduction** via Lambda@Edge, secured **IMDA/Deloitte accreditation** for Government Cloud (GCC) deployment, and transformed a 10-person engineering org to a **3x shipping velocity** (shifting from fortnightly to 2–3 ships per week). Expertly balances rapid AI innovation with the high-compliance security and scalability standards required for diverse international markets.

TECHNICAL SKILLS

- **AI & Vision Intelligence:** Multimodal LLMs, LLM Orchestration (LangChain), RAG Architecture, Vector Databases (Pinecone, Weaviate), MediaPipe (BlazePose), Stable Diffusion, CUDA/GPU Orchestration.
- **Cloud & Distributed Systems:** AWS (ECS/Fargate, Lambda@Edge, CloudFront, ElastiCache), Vercel, Neon (Serverless Postgres), Docker, Kubernetes, Terraform, **SSE & WebSockets**, **BLE Protocols**.
- **Systems & Architecture:** Node.js (TypeScript), Python, Go, Flutter (Native Modules), PostgreSQL, Redis, MongoDB, Microservices, RESTful & GraphQL API Design, **Performance Tuning (p99 Optimization)**.
- **Leadership & Governance:** Technical Strategy & Roadmapping, **IMDA/Deloitte Audit (Green Lane)**, SOC2 Compliance, Multi-region Team Leadership (India & Singapore), Technical Hiring & Mentorship.
- **Productivity & Frontend:** AI-Assisted Development (**Cursor/Copilot**), CI/CD Automation, TDD (Jest, Cypress), Next.js, React.js, Mobile Cross-Platform (Flutter, React Native).

WORK EXPERIENCE

Head of Software Engineering (Technical Lead)

EliteFit.AI — January 2024 - Present

- **AI Health Engine & Vision Intelligence:** Architected a **Flexibility, Strength & Balance (FSB)** assessment engine and Physio AI utilizing **BlazePose (MediaPipe)** and **LLMs** to automate movement diagnostics for 50,000+ users. Developed a **"Mobility Age"** diagnostic result that analyzes biomechanical errors to provide prescriptive feedback and **RAG-powered chat support**, allowing users across SE Asia, India, and the US markets to query report details and specific improvement strategies.
- **Real-time IoT Integration:** Spearheaded the end-to-end integration of **BLE (Bluetooth Low Energy)** protocols for real-time heart rate monitoring and synchronized biometric tracking within the workout ecosystem.
- **High-Stakes Compliance & Governance:** Led the technical roadmap for **IMDA Green Lane accreditation**, navigating rigorous security, reliability, and maintainability audits by **IMDA and Deloitte** to enable procurement and deployment within Singapore's **Government Commercial Cloud (GCC)**.
- **B2B AI Ops & Engineering Velocity:** Engineered a **B2B AI Reporting Assistant** using **RAG** to query live databases for corporate insights, and transformed a **10-person multi-region team** by integrating **AI-assisted development (Cursor/Copilot)** and CI/CD automation. Accelerated release frequency by **3x**, shifting from fortnightly cycles to **2–3 ships per week**.

Senior Software Engineer

EliteFit.AI — May 2021 - December 2023

- **Infrastructure & Global Performance:** Led a high-stakes, zero-downtime migration from legacy EC2 to containerized **ECS Fargate**, slashing monthly spend by **30%** while maintaining 99.99% availability. Integrated **AWS CloudFront** and **Lambda@Edge** to optimize content delivery, resulting in an **80% reduction** in global page load times across the US, EU, and GCC.
- **Real-time Systems & Public Infrastructure:** Engineered an **SSE-driven (Server-Sent Events)** live leaderboard for **SportSG** kiosks at **Decathlon**, delivering instant health analytics and competitive rankings to public users.
- **Data Throughput & Scalability:** Optimized PostgreSQL indexing and Redis caching strategies to handle a sustained throughput of **5M+ daily event records**. Improved API responsiveness by **120ms** during peak traffic bursts, ensuring system stability during high-concurrency public events.

Full-Stack Developer

Inovant Solutions — January 2020 - April 2021

- **Enterprise 0-to-1 Builds:** Engineered and launched 10+ enterprise SaaS platforms across EU and GCC markets, including 4+ ground-up builds for multi-tenant e-commerce and high-concurrency booking apps
- **System Reliability & Compliance:** Designed robust, multi-tenant backend architectures using **Node.js (TypeScript)** and **PHP (Yii2)** that maintained 99.9% uptime and ensured **SOC2-compliant** data handling for high-compliance international clients.

PROJECTS

PrescriptoAI (prescriptoai.com)

Independent Build | AI-SaaS & HealthTech Architecture

- **Vision Intelligence & Data Integrity:** Engineered a medical-grade OCR platform utilizing **Next.js** and **GPT-4o Vision** to transform complex, handwritten prescriptions into structured JSON; achieved **99% extraction accuracy** in internal benchmarking across a curated dataset.
- **Serverless Architecture:** Architected a high-concurrency **Vercel** backend leveraging **Neon (Serverless Postgres)** and **Auth0**; engineered external-facing APIs via **Serverless Functions** to support third-party user integrations.

QRCodeFox (youtube.com/watch?v=NjcxdrueDQ)

Independent Build | Generative AI & GPU Infrastructure

- **Generative AI & GPU Orchestration:** Built a **CUDA-accelerated** image pipeline utilizing **Stable Diffusion** and **ControlNet** on **AWS EC2 (G4dn/G5)** to generate functional, high-aesthetic AI QR codes via prompt engineering.
- **Cost & State Optimization:** Developed an **intelligent instance manager** to trigger GPU workers only during active requests, eliminating idle spend; integrated with **Firebase** for low-latency state management and a decoupled, event-driven worker tier.

EARLY CAREER & INTERNSHIPS

- **Full-Stack Developer Intern:** *Devtaa Group — January 2019 - December 2019*
- **Software Engineer Intern:** *Medinfi Healthcare — June 2018 - October 2018*

EDUCATION

- **Bachelor of Technology in Computer Science**
Shivaji University, India — 2017 - 2020
- **Diploma in Computer Engineering**
Maharashtra State Board of Technical Education (MSBTE), Mumbai, India — 2014 - 2017