

Teja Vardhan Reddy Avuthu

+1 (571) 386-9672 | ✉ tejavardhanr01@gmail.com | in LinkedIn

Experience

George Mason University, *Software Engineer 2*

May 2025 – Present | Remote, USA

- Built a browser-based carpal tunnel assessment app for diagnosed patients to track their recovery, pairing MediaPipe JS hand tracking with a four-stage signal-processing engine validated across 200+ patient recordings.
- Developed a HIPAA-safe patient AI assistant with PageIndex vectorless retrieval and PII masking, explaining assessment results in plain language and **cutting clinician follow-up questions by 30%**.
- Shipped a WebSocket and Celery/Redis pipeline that pushes each patient's assessment results the moment processing finishes, **replacing 1.5-second polling with near-instant delivery**.
- Architected the FastAPI, async SQLAlchemy, and PostgreSQL backend with JWT auth, audit logging, and per-patient RBAC across 40+ REST endpoints, meeting HIPAA access-control and audit requirements for patient health data.
- Deployed the app to Azure Container Apps with Application Insights and Prometheus monitoring and TOTP 2FA, serving 50+ clinicians and patients in production.
- Introduced test-driven development across the application with SonarQube quality gates in CI, **raising automated test coverage to 90%**.

Blue Horse Digital, *Software Engineer*

Jun 2024 – Apr 2025 | Remote, USA

- Built the AI Case Generator on Next.js and async FastAPI with Azure Service Bus background jobs, delivering 200+ instructional cases to clients weekly.
- Architected a multi-agent CrewAI pipeline with LiteLLM routing each agent to a cost-appropriate model and an evaluator grading every case on an 8-category rubric, **catching 95%+ of hallucinated content**.
- Designed a RAG layer on PostgreSQL with pgvector HNSW indexing and re-ranking, grounding generated cases in retrieved source material and cutting retrieval **latency from 800ms to under 200ms**.
- Implemented token-usage and cost tracking for every case generation run, giving leadership per-client AI spend visibility and **cutting Azure OpenAI costs by 25%**.
- Standardized GitHub Actions CI/CD to Azure App Service with Cypress E2E gates, cutting deploys from 1 hour to under 10 minutes.

IQVIA, *Software Engineer*

May 2021 – Dec 2022 | Bengaluru, India

- Developed an event-driven push-notification microservice in Java and Spring Boot, pushing 10K+ clinical record updates daily between EHR platforms and Salesforce.
- Containerized the microservice with Docker and deployed it to Azure Kubernetes Service (AKS) through Azure DevOps CI/CD pipelines, shipping **zero-downtime rolling releases**.
- Automated release validation for cross-system data integrations, **cutting regression testing from 3 days to 4 hours** per release.
- Built JUnit, Mockito, and Postman API test suites for regulated clinical data systems, reaching 85%+ code coverage and tracking every defect in JIRA to production sign-off.

Skills

Languages: Python, Java, TypeScript, JavaScript, SQL

Backend & Data: FastAPI, Spring Boot, Spring JPA, Next.js, Node.js, async SQLAlchemy, Alembic, REST APIs, WebSockets, Event-Driven Microservices, Celery, PostgreSQL, Redis

Cloud & DevOps: Azure (App Service, Container Apps, AKS, Service Bus, PostgreSQL Flexible Server, Blob Storage), Docker, Kubernetes, Azure DevOps, GitHub Actions, Application Insights, Prometheus

AI & LLM: Azure OpenAI, LiteLLM, CrewAI, Multi-Agent Workflows, RAG, Vectorless Retrieval (PageIndex), Vector Search (pgvector, HNSW), Re-ranking, Prompt Engineering, LLM Cost Tracking, MediaPipe JS

Testing & Quality: Test-Driven Development, SonarQube, Cypress (E2E), JUnit, Mockito, Postman, JIRA

Security & Compliance: HIPAA, PII Masking, JWT Auth, TOTP 2FA, RBAC, Audit Logging, Rate Limiting

Education

M.S. Computer Science, George Mason University

Jan 2023 – Dec 2024 | VA, USA

Related Coursework: Advanced Data Structures & Algorithms, Software Systems Design & Integration, Cloud Computing