VIGNESH V

Team Lead | Applied AI & Full Stack Engineering

Bengaluru, India • +91 79752 75944 • okayvignesh12@gmail.com • linkedin.com/in/okayvignesh • github.com/okayvignesh • Portfolio

PROFESSIONAL SUMMARY

Team lead with over four years of experience building enterprise SaaS, now focused on AI systems that run in production. I built SimplifyQA's AI assistant, an LLM agent with 106 tools, hybrid search and self-hosted models, and led a team to ship Scroll, a realtime documentation platform with AI built in. I work in Python (FastAPI) and TypeScript (Node.js, React), and I own the search, safety checks, model serving and Kubernetes deployment behind these features.

TECHNICAL SKILLS

AI & LLM	LLM agents and tool calling, RAG (hybrid dense and BM25 search, reranking), embeddings, MCP, prompt injection and hallucination guardrails, multi-provider routing
Models & Serving	OpenAI, Azure OpenAI, Anthropic, Gemini, vLLM, Hugging Face Transformers, MLX, Qwen3, BGE, Whisper
Vector & Data	Qdrant, pgvector, MongoDB Atlas Vector Search, MongoDB (change streams), PostgreSQL, Redis
Languages	Python, TypeScript, JavaScript, SQL
Backend	FastAPI, Node.js, Express, Celery, BullMQ, RabbitMQ, SSE, WebSockets, REST
Frontend	React, Angular, Next.js, Tiptap, Yjs, Tailwind CSS
Cloud & DevOps	Kubernetes (EKS, AKS), Docker, GPU containers, Jenkins, GitHub Actions, Azure, AWS, Nginx

PROFESSIONAL EXPERIENCE

Simplify3x Software, Bengaluru | Team Lead — Nov 2023 – Present

- Lead the engineering team on SimplifyQA, an enterprise test automation platform used by banks, insurers and healthcare companies. I am responsible for architecture, sprint planning, code reviews and releases across the AI and platform services.
- Built the platform's AI layer: the AI assistant, the self-hosted model server, and the AI test design service, where I am the top contributor. That service reads requirement documents, Figma screens and Jira or Git data, and generates epics, stories and test cases using multiple agents and RAG.
- Wrote the AI infrastructure cost study that chose Azure OpenAI GPT-4o-mini over ₹8.5 lakh of GPU hardware, projected at about ₹15,000 a month for 1,000 daily users while meeting India's data residency (DPDP) rules.
- Built an internal RAG assistant over company documents that cut the time teams spent searching for content by about 60%.
- Set up RabbitMQ processing for long-running workflows and standardised REST contracts and front-end state across teams, which cut rework by about 40%.

Kandra Professionals, Bengaluru | Software Engineer — Dec 2022 – Nov 2023

- Built and maintained three production web applications used by around 1,500 active users, working with React, Redux and Node.js.
- Reorganised the front end into modular components with client-side routing and a central Redux store, which brought average page load time down by about 30%.
- Integrated more than eight third-party REST APIs through a shared service layer with proper error handling and retries, and kept Jest coverage on shared utilities at around 75%.

PROJECTS

SimplifyQA AI Assistant

Simplify3x | Python, FastAPI, Qdrant, vLLM, RabbitMQ, Kubernetes

- **What:** A chat assistant built into the SimplifyQA test platform. QA leads can ask about sprints, test runs, defects and coverage, and get the answer back with live charts.
- **How:** Designed an LLM agent that sends each question to one of **106 typed tools** across three product areas, so the model never writes database queries on its own. Search combines dense vectors and BM25 on Qdrant, with a BGE reranker on top. MongoDB change streams feed RabbitMQ and an embedding worker, so new data shows up in answers within moments.
- Built the self-hosted model server (Qwen3-8B, BGE embeddings and reranker, Whisper for voice) with swappable vLLM, Hugging Face, MLX and Azure OpenAI engines, and GPU images for both NVIDIA and AMD.

- Added the safety layer needed for production: search results stay inside each customer's own data, every ID the model cites is checked against real tool output, prompt injection is filtered, and repeat questions are served from a Redis semantic cache.
- **Value:** QA leads get release readiness, flaky test and coverage answers in a chat instead of building reports by hand. The whole stack can run on-premise for banks and insurers that cannot send data to a public LLM.

Scroll: Realtime AI Documentation Platform

Simplify3x | Team lead | React, Tiptap, Yjs, Node.js, PostgreSQL, pgvector, Redis

- **What:** A Confluence-style workspace inside SimplifyQA for test plans, release notes and runbooks. Pages can embed live test cases and reports, and users sign in with their SimplifyQA account, so customers do not need a separate wiki.
- **How:** Realtime co-editing with Tiptap and Yjs (Hocuspocus), kept in sync across server pods through Redis. The Express and PostgreSQL back end has around 310 endpoints and 52 tables, with role-based access, version history with page diffs, and import from Word, PDF, Notion and Confluence.
- AI features include hybrid search (full text plus pgvector), question answering over a team's spaces, an AI writing assistant, and AI-generated slide decks.
- **Leadership:** Planned the architecture and sprints, reviewed and merged the team's work, and wrote about 40% of the code myself. Fixed a cross-tenant SSO flaw and a private space access bug over WebSockets, solved lost edits during simultaneous editing (only 1 of 4 edits survived before the fix, all 4 after), and set up more than 360 automated tests with a 60% coverage gate in CI.

SimplifyQA: Enterprise Test Automation Platform

Simplify3x | Angular, Node.js, MongoDB, Redis, RabbitMQ, Socket.IO, Kubernetes

- **What:** Simplify3x's core product, a no-code test automation platform for web, mobile, API, desktop, SAP and mainframe applications. It is used by banks, insurers and healthcare companies, either as SaaS or installed in their own cloud.
- **How:** Worked across the Angular front end and the Node.js microservices (gateway, project, test, execution, reporting, scheduling and integration services) running on MongoDB, Redis and RabbitMQ in Kubernetes. Built the drag-and-drop workflow canvas with Gantt and timeline views, much of the defects and reporting modules including PDF export, and faster live streaming of test steps during execution.
- Rebuilt the Azure DevOps integration with durable webhooks, pause and resume for syncs, and a preflight self-test, extended two-way Jira sync, and built the team's shared Angular rich text editor library.
- **Value:** 15+ enterprise clients design and run test workflows without writing code, and background processing handles more than 5,000 jobs a week. I have made more than 1,500 commits across 19 service repositories.

Legacy to Microservices Data Migration

Simplify3x | Node.js, MongoDB, Docker

- **What:** A tool that moves customers from the legacy SimplifyQA, which kept everything in one database, to the new microservice platform with separate project and test management databases.
- **How:** Built a **31-step pipeline in 8 phases** that streams data in batches with memory-aware garbage collection, maps old IDs to new ones, and can safely be re-run. Mapped **438 legacy automation actions** onto the new templates, and wrote reconciliation reports in Excel that compare old and new data, along with scripts to repair or roll back skipped records.
- **Value:** Designed to handle millions of records in a single migration, including up to **200,000 test cases and 1 million test steps**, with no data loss on core entities and historical execution reports kept for audit. Used across 12+ client programmes, and I am the top contributor on the codebase.

ClubAdmin: Sports Academy Management SaaS (clubadmin.domainrental.in)

Personal product | React, Node.js, MongoDB, Docker

- Designed, built and sold a multi-tenant SaaS for sports academies, covering role-based access, member management, attendance and payments, deployed with Docker and Jenkins CI.
- **Value:** Sold three commercial licences and earned recurring revenue from the first release.

EDUCATION

Master of Computer Applications (MCA), IGNOU — 2023 – 2026

Bachelor of Computer Applications (BCA), St. Francis College, Bangalore — 2018 – 2021