KALYAN RAM GORIPARTHI

Forward Deployed Engineer | AI Solutions Architect | Full Stack & Applied AI
Horsham, PA, USA | 716-247-3202 | itsgkram@gmail.com | linkedin.com/in/itsgkram | github.com/itsGKRam

SUMMARY

Forward Deployed Engineer and AI Solutions Architect with 4+ years of software engineering, including 2+ years delivering for enterprise customers at an AI-powered HR tech SaaS company. Turn ambiguous customer requirements into working integrations, data pipelines and AI tools across 20+ enterprise deployments, from discovery and solution design through go-live. Ship LLM applications with Python, TypeScript, React and Node.js, including RAG pipelines, agents and local and on-device models, and evaluate new models and agent tools hands-on before building with them. M.S. in Computer Science.

SKILLS

AI and LLMs: LLM Applications, RAG, AI Agents, Prompt Engineering, Claude API, Gemini, Vercel AI SDK, LangGraph, MCP, ChromaDB, Ollama, llama.cpp

Languages: Python, TypeScript, JavaScript, SQL, Swift, Java

Frameworks: React, Next.js, Node.js, Express, FastAPI, Flask, Tailwind CSS, Playwright

APIs and Data: REST, SOAP, XML, JSON, GraphQL, OData, OAuth 2.0, SAML, ETL, Data Pipelines, SFTP, Postman, OpenAPI

Enterprise Systems: Workday, SAP SuccessFactors, Oracle Taleo, iCIMS, UKG, Salesforce Administration

Cloud and Architecture: AWS, Docker, PostgreSQL, MongoDB, Supabase, Apache Kafka, System Design, Multi-tenant SaaS, Microservices, Kubernetes Architecture, CI/CD, Git

EXPERIENCE

Associate Solution Architect **Aug 2024 - Present**

Phenom People, Inc. | Horsham, PA

- Lead technical delivery for 20+ enterprise customers in healthcare, financial services, manufacturing, logistics and hospitality: run discovery, design the solution, build the integration and support go-live.
- Work in both solution architect and Forward Deployed Engineer (FDE) roles; completed Phenom's FDE bootcamp with on-site project delivery and build custom tools when configuration alone isn't enough.
- Build bi-directional integrations and data pipelines between Phenom and enterprise systems such as Workday, SAP SuccessFactors and Oracle using REST, SOAP and OData APIs, OAuth 2.0 and SAML, file feeds and ETL for historical data migration.
- Built 10+ internal tools adopted across implementation and solution architecture teams, cutting typical tasks from about 5 hours to under an hour (80-90% faster); one cut business process change analysis from hours, sometimes days, to minutes.
- Built LLM applications for the team: a Claude-based knowledge pipeline with vector search and a knowledge graph, enterprise AI agents on Atlassian Rovo, an agentic template platform with human approval and a Chrome extension that writes and lints SQL with a local model.
- Reconciled hiring records between two systems for a health system, surfacing 186 mismatches and 48 unmatched records before go-live, and wrote an implementation guide estimated to cut effort on later projects by 80%.
- Scope solutions with sales, review statements of work, answer technical RFP questions and mentor new hires.

Senior React JS Developer **Mar 2022 - Oct 2022**

DivIHN Integrations LLP | Remote

- Built responsive React and Redux interfaces and integrated them with Java Spring Boot REST APIs.
- Improved code quality and release practices, contributing to a 30% reduction in deployment time.

Full Stack Web Development Head (Intern) **Apr 2022 - Dec 2022**

Engineers Cradle | Remote

- Led a team of frontend and backend intern developers delivering client web applications in React and Node.js.

Backend Node.js Developer (Intern) **Jul 2021 - Mar 2022**

Engineers Cradle | Remote

- Optimized server-side processing, reducing API response times by 25%.

Frontend React.js Developer (Intern) **Jan 2021 - Jun 2021**

Engineers Cradle | Remote

- Applied frontend performance practices that reduced page load times by 30%.

PROJECTS

AI-Assisted Data Mapping Tool | Enterprise data onboarding | *React*, *TypeScript*, *Supabase*, *Gemini*

- Built a tool where teams upload customer data files, map columns by hand or with Gemini suggestions ranked by confidence, validate the data and export clean files; adopted by implementation teams.

Knowledge Ingestion Pipeline | RAG and knowledge graph for team knowledge | *Python*, *Claude API*, *ChromaDB*, *NetworkX*, *Pydantic*

- Built an event-driven pipeline that watches for new documents, extracts structured knowledge with Claude and indexes it three ways: linked notes, a knowledge graph and vector search for retrieval.

Applied AI Research and Prototyping | Hands-on model and tooling evaluation | *Open-weight LLMs*, *Speech-to-Speech Models*, *LLM Gateways*, *Ollama*, *llama.cpp*, *Docker*

- Evaluate new models and AI tooling hands-on: ran local evaluations of a full-duplex speech-to-speech model, self-hosted a multi-provider LLM gateway in Docker and documented deployment options for a frontier open-weight model.
- Choose between local (Ollama, llama.cpp) and hosted models (Claude, Gemini) per build based on privacy and cost, for example local-by-default SQL and typing assistants.

itsTypo | Context-aware typo correction for macOS | *Swift*, *macOS*, *llama.cpp*, *On-device LLM*

- Built a system-wide macOS utility in Swift that resolves each typed word against sentence context at word boundaries with an on-device LLM, replacing text only when confidence is high.

itsRAM | Voice assistant for macOS | *Python*, *FastAPI*, *PyObjC*, *Ollama*, *LLM*

- Built a push-to-talk voice assistant that classifies intent, risk and target app with typed confidence scores, routes commands in code and answers through local (Ollama) or cloud LLMs; packaged as a native macOS app.

ITS Curio | Content curation agent in Instagram DMs | *Python*, *LangGraph*, *Claude*, *FastAPI*, *PostgreSQL*

- Built an agent that turns shared reels into searchable summaries and tags (caption, Whisper transcript, frame vision) and retrieves them from plain-language requests.

Tadka | Creator-first recipe platform for chefs | *React*, *TypeScript*, *Supabase*, *PostgreSQL*

- Built a mobile-first recipe platform with authentication, recipe authoring, collections, favorites, reviews and creator analytics on Supabase, with 16 PostgreSQL migrations and an edge function for onboarding email.

EDUCATION

Master of Science in Computer Science **May 2024**

University at Buffalo, SUNY | Buffalo, NY

Bachelor of Engineering in Electronics and Communication **May 2023**

Vellore Institute of Technology | Vellore, India