

Devaneyan Palanisamy

Director of Engineering | AI Engineer

Bengaluru, India | neyacse@gmail.com | hi@neya.one | Portfolio: neya.one
LinkedIn: linkedin.com/in/devaneyan-palanisamy | GitHub: github.com/devacse

SUMMARY

Director of Engineering with 10+ years in SaaS, payments and cloud, and the last two years focused on AI engineering. At Omnify I lead a team of 12 engineers on a platform used by 500+ businesses. I build LLM products with RAG, claim checking and LLM-as-judge evaluation, use multi-agent systems to help my team ship, test and document faster, and run the AWS infrastructure underneath, where I cut costs by 50%.

TECHNICAL SKILLS

AI / LLM: LLM applications, RAG, vector search (pgvector), AI agents, multi-agent orchestration, LLM evaluation (LLM-as-judge), prompt engineering, Claude, AWS Bedrock, Strands Agents, OpenAI, LiteLLM, MCP

Machine Learning: Vertex AI, AutoML, Kubeflow Pipelines, TensorFlow, computer vision (ResNet50, VGG16), clustering, explainable AI

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

Frameworks: FastAPI, React, Next.js, NestJS, Spring Boot

Databases: PostgreSQL, MongoDB, BigQuery, pgvector, dbt

Cloud & DevOps: AWS (Lambda, ECS Fargate, CloudFront, SES, RDS, IAM, EventBridge, WAF), GCP (Cloud Run, Pub/Sub), Terraform, AWS CDK, Docker, Kubernetes, GitHub Actions, CI/CD

Payments & Security: Stripe Connect, PayPal, Razorpay, SOC 2, multi-tenant data isolation

PROFESSIONAL EXPERIENCE

Director of Engineering

Omnify, multi-product SaaS for 500+ businesses (~500K monthly visits) | 2023 – Present

- Lead a team of 12 engineers across a 10–15-product platform, owning architecture, delivery and production reliability.
- Built an LLM documentation and knowledge platform (Python, FastAPI, pgvector, React) in which a writer model cites every sentence, Claude Opus verifies each claim, and a RAG chat answers questions over the docs.
- Selected the production model through a blind LLM-as-judge evaluation with two judge models: AI-written pages scored +1.0–1.42 points above existing pages.
- Designed a multi-agent engineering system (Claude Code, MCP, custom skills and hooks): 14 agents documented 500+ APIs and 444 database tables in parallel, 10–15 agents run browser QA, and a nightly CI agent keeps documentation in sync with code.
- Architected multi-account AWS in Terraform (SSO, OIDC, ECS Fargate, WAF), cutting AWS costs by 50%; shipped customer custom domains on CloudFront and multi-region email (SES) with per-tenant sender reputation.
- Redesigned Stripe Connect payments with webhook-driven dispute handling, a Stripe-compatible API gateway (Go, FastAPI) for enterprise customers, and a pluggable payment-gateway layer.
- Led security and reliability: AWS organization audit and recovery, cross-tenant data-access reviews, outage root-cause analysis, and the V2 to V3 platform launch.

Senior Tech Lead

Omnify | Oct 2022 – 2023

- Rebuilt the Omnify storefront (v2) with embeddable widgets, white-labelled custom domains and themes.
- Built Omnify Custom Payments on Stripe Custom Connect.

Tech Lead

Omnify | 2020 – 2022

- Moved the full stack to Docker and Kubernetes and cut developer onboarding from 2 days to 2 hours.
- Led SOC 2 compliance and the AngularJS to Next.js storefront migration.

Senior Software Engineer

Omnify | 2018 – 2020

- Built the payment system processing over \$1M per month (Stripe, PayPal, Razorpay) and email delivery at 500K+ messages per month.

Co-founder & CTO

Edudharma | 2015 – 2018

- Built and led engineering for an education crowdfunding platform (Java, Spring Boot) from zero.

AI & Cloud Platform

Enterprise identity, compliance and reporting SaaS (Freelance)

- Led the design and delivery of an AI agent used by about 10 enterprise customers across three products (Python, Strands Agents, AWS Bedrock), with natural-language queries over live data, drill-down dashboards, licence-savings reports and what-if role simulation.
- Architected the agent's AWS runtime: an isolated Lambda MicroVM per user session streaming responses over WebSocket, deployed per product with Terraform; enforced tenant isolation in PostgreSQL with per-user read-only roles and security-barrier views.
- Migrated a multi-tenant reporting product to a second cloud provider on one codebase with zero-drift Terraform, and hardened it for a regulated enterprise client through their security review.
- Led AWS platform work in CDK: container-image Lambdas, private RDS, a Fargate scheduled-report pipeline with SES, a bastion VPC, and an event-driven approval workflow with EventBridge.

AI / ML Engineer

Storyteller, AI creative intelligence for Meta ads (Freelance)

- Built ad-data ingestion from the Meta Graph API with rate-limit-aware backoff, Fivetran and 117 dbt models on BigQuery for ROAS, CTR and video-completion metrics.
- Built AI tagging for ad images, video and copy (Cloud Vision, Video Intelligence, Natural Language) and clustered creatives from ResNet50/VGG16 embeddings on Vertex AI.
- Shipped explainable KPI prediction: Kubeflow pipelines train AutoML models on 21 creative features, and an LLM (PaLM) turns feature attributions into next-step recommendations.

PROJECTS

CarKadai | FastAPI, PostgreSQL, Next.js

- Multi-tenant SaaS for auto salvage yards and used-parts sellers with inventory, per-tenant storefronts and checkout, using FastAPI microservices and PostgreSQL row-level security.

EDUCATION

PG Program in Artificial Intelligence & Machine Learning

University of Texas at Austin, McCombs School of Business | 2023 – 2024

Bachelor of Engineering, Computer Science

Anna University | 2010 – 2014