

VISHNU N V

AI Engineer | DevOps & Platform Engineering

+91-8281294818 | vishnunvnair@gmail.com | LinkedIn | Bangalore, India

PROFESSIONAL SUMMARY

AI Engineer with 6+ years in financial-services technology, including 4 years at JPMorgan Chase, specializing in **agentic AI systems, LLM tool-calling, and cloud-infrastructure automation**. Designed and shipped a production agentic-AI developer platform (VS Code + Claude Opus) that autonomously plans, invokes 12+ tools, and remediates Terraform/AWS infrastructure at enterprise scale. Strong DevOps and Platform Engineering foundation across AWS (30+ services), EKS/ECS, Terraform Enterprise, and CI/CD, with a consistent record of cost optimization, automation, and turning senior SRE expertise into reusable, AI-driven tooling.

TECHNICAL SKILLS

AI / ML Engineering: Agentic AI, LLM Tool-/Function-Calling, ReAct & Multi-Turn Orchestration, Retrieval-Augmented Generation (RAG), Vector Embeddings, Semantic Search, Prompt Engineering, Context-Window Engineering, Intent Classification, LLM Safety & Guardrails, Prompt-Injection Defense, Human-in-the-Loop Approval

Models & AI Frameworks: Claude Opus (Anthropic), Amazon Bedrock, ChromaDB, sentence-transformers (MiniLM), onnxruntime, VS Code Language Model API

Languages: Python, TypeScript, JavaScript, SQL, HCL (Terraform), PowerShell, Bash

Cloud & DevOps: AWS (EKS, ECS, RDS, Lambda, S3, IAM, VPC, CloudWatch, CKMS), Terraform / Terraform Enterprise, Docker, Kubernetes, OpenShift, Cloud Foundry, Spinnaker, Bitbucket, CI/CD Pipelines, Ansible

Observability & Data: Dynatrace, Grafana, CloudWatch, Splunk, Geneos, Oracle SQL, SQLite

Web / Full-Stack: React, Node.js, Express, Vite, FastAPI / Flask, REST API Design, nginx

PROFESSIONAL EXPERIENCE

JPMorgan Chase & Co.

Sales & Research Technology · Bangalore, India

Feb 2022 – Present

AI & Platform Engineer (Associate Vice President)

2024 – Present

- Built and shipped **TerraDune v2**, an agentic-AI developer platform (detailed under Key Projects) that converts natural-language intent into safe, audited Terraform/AWS actions, compressing day-long upgrade workflows to under 10 minutes.
- Spearheaded **S3 lifecycle optimization** for the IODS platform, cutting daily S3 cost from \$19 to \$0.38 and delivering ~\$4,600 in net annual savings.
- Designed, developed, and deployed an **Unused VPC Endpoint deletion tool** for recurring cloud-resource cleanup, improving security posture and cost efficiency — recognized by leadership.
- Automated **CKMS certificate renewal and fresh-certificate creation** across RDP, MCP, ICE, and Research Metrics, and established **SCX secrets-manager rotation** for Contact Master, removing manual effort and improving compliance.
- Led **Terraform module-version upgrades and workspace decommissioning** across CIB Activity Hub, ICE, Contact Master, and CHIMEUI; resolved the majority of divest/prohibited-module FARMs and AO communications to maintain platform hygiene and stability.
- Built a **React-based SRPM Data-Quality Dashboard** consolidating incident/problem/change governance into a single view with automated email reporting, eliminating manual weekly reporting for team leads.
- Delivered a firm-wide training and demo on CKMS automation in the **Ignite Automation Community**; core member of the Bengaluru Ignite Automation team and contributor to Ignite Spark 2024.

Production Support Engineer (Associate)

Feb 2022 – 2024

- Provided L1/L2 production support across Sales & Research Technology applications during APAC hours, sustaining SLA-driven service quality while leading L1/L2 knowledge-consolidation and transfer sessions.
- Automated application installation and operational tasks via Python/ENAML tooling, built Geneos alert monitoring, and deployed upgrades through Jules and Cloud Foundry (cf CLI) pipelines.

Tata Consultancy Services Ltd.

Operations / Support Engineer · Client: Deutsche Bank (BFSI) · Bangalore, India

Jan 2020 – Feb 2022

- Supported RPA (Blue Prism), risk-analytics fraud-detection, and OTP authentication platforms across production and UAT for a global investment bank, maintaining 100% SLA on production-issue resolution.
- Automated software installation, certificate renewal, and IIS configuration via **Ansible** (DBRUN) and PowerShell; performed Geneos monitoring, Splunk log analysis, and Oracle SQL data extraction.

KEY PROJECTS

TerraDune v2 — Agentic AI Copilot for Cloud Infrastructure

2025 – Present

TypeScript · Python · Claude Opus · ChromaDB · VS Code Extension API · boto3 · Terraform / HCL

- Architected an end-to-end **agentic AI platform** delivered as a VS Code chat extension that turns natural-language intent into safe, audited actions across AWS, EKS/ECS, RDS, Terraform Enterprise, Bitbucket, Spinnaker, Dynatrace, and Grafana — owning the full stack from prompt layer to tool layer to retrieval layer.
- Implemented a multi-turn **Gather→Synthesize tool-calling loop** (up to 10 reasoning rounds) orchestrating 12+ schema-validated tools, plus a **hybrid intent classifier** (weighted-regex + LLM fallback) that eliminates ~80% of unnecessary LLM calls and reduces latency and cost.
- Built a **ChromaDB-backed RAG layer** (6 collections, MiniLM embeddings, per-team sharding) for self-healing fix retrieval and natural-language infra lookup, and a modular **skill registry** that dynamically composes domain prompts and scoped tools per query.
- Engineered an **8-phase self-healing /upgrade agent** that scans Terraform modules, risk-scores version selection, rewrites HCL, runs `terraform plan`, auto-diagnoses failures, re-prompts the LLM for fixes, re-plans, and opens PRs — with bounded-retry anti-loop guards.
- Designed defense-in-depth **AI safety**: read-only allow-listed tools, every mutating action gated behind explicit user consent, OS-keychain secret storage, 200K-char context budgeting with truncation and dedup caches, and prompt-injection-aware tool I/O.
- Impact:** ~30× MTRR reduction for EKS/ECS/RDS triage and 30–100× faster Terraform module upgrades, scaling senior SRE expertise to every engineer through a single installable extension.

SCOPE — Internal Developer Tools Portal & Plugin Marketplace

2024 – 2025

React 18 · Vite · Node.js · Express · SQLite (sql.js) · Docker · nginx · Cloud Foundry · Bitbucket REST API

- Designed and shipped a **full-stack internal developer portal** that catalogs, versions, and distributes VS Code extensions (.vsix) and team documentation, replacing ad-hoc Bitbucket/email sharing and cutting tool-publish time from days to minutes.
- Built a Node/Express REST API over embedded SQLite with BLOB-backed binary delivery, and a **3-tier self-healing persistence pipeline** (in-process DB → debounced JSON snapshot → automated Bitbucket commit + Cloud Foundry restage) guaranteeing zero data loss on a stateless PaaS.
- Containerized via a multi-stage Docker build behind nginx with SPA routing and an automated publish-to-prod CI loop; adopted as the team's canonical distribution channel for developer-productivity tools.

EDUCATION & TRAINING

Bachelor of Technology, Mechanical Engineering

LBS College of Engineering, Kasaragod, Kerala, India

2014 – 2018

Software Engineering Trainee (Python) — ART Technology and Software India Pvt. Ltd.

2019

CERTIFICATIONS

- AWS Certified Cloud Practitioner** — Amazon Web Services (2023)
- IBM Cloud Essentials & IBM Developer Skills Network** — Kubernetes, Containers, Infrastructure Services (2021)
- In progress:** AWS Certified DevOps Engineer – Professional · AWS Certified AI Practitioner

Languages: English, Hindi, Malayalam