

SUBRAMANYA CHOWDARY KALLURI

+1 (937) 956-4105 | Skalluri04@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

GenAI-focused DevOps & Platform Engineer with 5+ years of experience architecting scalable cloud infrastructure and pioneering AI-driven automation solutions. Expert in building LLM-powered code review agents, legal AI assistants, and workflow orchestration systems using Temporal. Proven track record in Kubernetes orchestration, CI/CD pipeline optimization, and implementing cutting-edge GenAI solutions including Claude Code and agentic AI systems. Passionate about leveraging AI to transform traditional DevOps practices into intelligent, self-healing infrastructure operations.

CORE COMPETENCIES

GenAI & AI Engineering

- Large Language Models (LLMs): Claude (Sonnet, Opus), GPT-4, Code Llama, DeepSeek for enterprise applications
- AI Agent Development: Claude Code, Claude Agent SDK, LangChain, CrewAI, multi-agent orchestration systems
- RAG Architecture: Vector databases, semantic search, document retrieval systems for legal & code analysis
- MLOps & LLMops: Model deployment, monitoring, A/B testing, prompt engineering, fine-tuning workflows

DevOps & Platform Engineering

- Container Orchestration: Kubernetes, Helm, Istio service mesh, custom operators
- Workflow Orchestration: Temporal.io, Apache Airflow, Argo Workflows for durable execution
- Infrastructure as Code: Terraform, Ansible, CloudFormation, Pulumi for multi-cloud deployments
- CI/CD & GitOps: Jenkins, GitHub Actions, GitLab CI, ArgoCD, Tekton for ML/AI pipelines
- Cloud Platforms: AWS (EKS, Lambda, SageMaker, EC2, S3), GCP; F5 Load Balancer, HAProxy
- Observability: Prometheus, Grafana, ELK Stack, Datadog, OpenTelemetry for AI system monitoring

Programming & Development

- Languages: Python, Go, Java, TypeScript/Javascript, Bash scripting
- Frameworks: Spring Boot, FastAPI, Flask, Node.js for microservices development
- Databases: PostgreSQL, Redis, MongoDB, Pinecone, ChromaDB (vector stores)

PROFESSIONAL EXPERIENCE

Senior DevOps & GenAI Platform Engineer

Wells Fargo, Fremont, CA | October 2022 – Present

- **Architected and deployed LLM-powered Code Review Agent** using Claude API and GitHub Actions, automating PR analysis for 50+ repositories, reducing manual review time by 60% and improving code quality metrics by 35%
- **Built Legal AI Assistant platform** leveraging RAG architecture with Claude for contract analysis and compliance review, processing 1,000+ legal documents monthly with 95% accuracy in risk identification
- **Implemented Temporal.io workflows** for durable DevOps automation, orchestrating complex multi-step deployments across 96 VMs with automatic failure recovery and state persistence
- **Pioneered Claude Code adoption** for infrastructure-as-code development, enabling AI-assisted Terraform module creation and Ansible playbook generation, accelerating IaC development by 40%
- **Designed Kubernetes-native MLOps pipeline** using KubeFlow and ArgoCD for model deployment, achieving 99.9% uptime for production AI inference services
- Managed mission-critical infrastructure supporting 8+ high-availability web applications on Red Hat Linux with F5 load balancer configurations and automated failover
- Developed custom AI-powered monitoring agents using Prometheus and Grafana with LLM-based anomaly detection, reducing incident response time by 45%
- Created Ansible automation playbooks integrated with AI suggestion systems for configuration management, patching, and software deployment across enterprise infrastructure

Tech Stack: Claude API, LangChain, Temporal.io, Kubernetes, Terraform, Ansible, Python, Go, Jenkins, GitHub Actions, Prometheus, Grafana, AWS, Docker, Red Hat Linux

DevOps Engineer / Java Developer

Teki Smarg, Hyderabad, India | May 2020 – July 2021

- Led technical teams in designing and implementing CI/CD pipelines for Java/J2EE applications, reducing deployment time from hours to minutes
- Architected containerization strategy using Docker, migrating legacy monolithic applications to microservices architecture
- Implemented infrastructure automation using Ansible and Chef, managing configuration across 96+ virtual machines
- Established SLA-driven support processes, achieving 99.5% uptime across all supported applications
- Designed release management workflows ensuring zero-downtime deployments for customer-facing applications

Tech Stack: Java/J2EE, Spring, Hibernate, Docker, Ansible, Chef, IBM WebSphere, DB2, Jenkins, Linux

KEY GENAI PROJECTS

Autonomous Code Review Agent

- Developed multi-model code review system using Claude Sonnet and GPT-4 with automatic PR summarization, security vulnerability detection, and actionable code suggestions
- Integrated with GitHub Actions for automated triggering on PR events, with results posted as inline review comments
- Implemented RAG-based context retrieval from codebase documentation and historical PR patterns

Legal Document AI Assistant

- Built enterprise legal AI platform using Claude for contract analysis, risk assessment, and compliance verification
- Designed semantic search pipeline with vector embeddings for cross-referencing legal clauses across document corpus
- Achieved 95% accuracy in identifying contractual risks and suggesting standardized legal language

Temporal-Powered DevOps Orchestration

- Architected durable execution framework using Temporal.io for complex infrastructure provisioning workflows
- Implemented parent-child workflow patterns for EKS cluster deployment with automatic rollback capabilities
- Built human-in-the-loop approval workflows combining automation with manual oversight for sensitive deployments

EDUCATION

Master of Science in Computer Science – New England College, NH (Jan 2022 – Dec 2022)

Bachelor of Science in Computer Science Engineering – Koneru Lakshmaiah Educational Foundation, India (2016 – 2020)