

LEAD ENGINEER – GENERATIVE AI & DEVOPS

GENERATIVE AI SPECIALIST, LLM DEPLOYMENT, FINE-TUNING, RAG SYSTEMS, AI AGENTS, KUBERNETES AND AWS CLOUD ADMINISTRATION WITH DEVOPS WORKFLOW WITH AI INTEGRATION

Ashwani Kumar V
Mobile: (+91) 7989 121 975
avardhineni7@gmail.com

avardhineni.netlify.app
[linkedin.com/in/avk-sd](https://www.linkedin.com/in/avk-sd)
github.com/SimplifiedDevOps

Professional Summary

A results-driven professional with **11.5 years of IT experience**, specializing in **Generative AI** (2+ years) and **DevOps Engineering** (6+ years). Expertise in **LLMs (Large Language Models)**, **RAG (Retrieval-Augmented Generation)**, **AI agents**, and **LLMOps**, with a proven track record in leading AI-driven projects and integrating them with DevOps workflows. Skilled in **fine-tuning models** and developing **AI applications** to deliver customized, innovative solutions.

Proficient in **AWS Cloud Administration**, **Terraform**, **Ansible**, and **Kubernetes**, focusing on building scalable, secure cloud-based solutions. Demonstrated success in fostering collaboration between **AI and DevOps teams**, optimizing cloud deployments, and driving operational efficiency. Seeking a **Lead Engineer** role to leverage expertise in AI project management, cloud technologies, and DevOps practices

Key Skills

Generative AI, Deep Learning & NLP

- **Large Language Models (LLMs):** GPT, BERT, custom AI models
- **Retrieval-Augmented Generation (RAG)**
- **Fine-Tuning and Transfer Learning**
- **Natural Language Processing (NLP):** Text Summarization, Sentiment Analysis, Question Answering Systems
- **Deep Learning Frameworks:** TensorFlow 2, Keras, PyTorch, Hugging Face Transformers
- **Conversational AI:** RASA-based chatbot development
- **Data Preprocessing and Pipeline Automation**
- **Model Evaluation and Optimization**
- **AI Observability:** LangSmith, LangChain, Prometheus, Grafana
- **Graph Databases:** Neo4j, GraphDB queries
- **Vector Databases:** Pinecone for similarity-based searches
- **AI Agents and Agentic Patterns:** Multi-agent frameworks, LangGraph
- **AI Model Deployment and Automation**
- **AI Ethics and Bias Mitigation**

DevOps & Cloud Engineering

- **AWS Cloud Administration:** EC2, S3, RDS, IAM, VPC, Lambda, API Gateway
- **Infrastructure as Code (IaC):** Terraform, CloudFormation
- **Configuration Management:** Ansible, Packer
- **CI/CD Pipeline Setup:** Jenkins, GitLab CI/CD, GitHub Actions
- **Containerization & Orchestration:** Docker, Kubernetes, Helm, AWS EKS

- **Monitoring & Logging:** Prometheus, Grafana, ELK/EFK Stack, distributed tracing with Jaeger and Kiali
- **Version Control:** Git, GitHub, Bitbucket
- **Security and Compliance:** DevSecOps, AWS Security Services, Istio Service Mesh
- **GitOps:** ArgoCD for automated infrastructure and application deployments
- **Automation Scripting:** Python, Bash
- **Disaster Recovery & High Availability Architecture**
- **Project Management:** Jira, Agile/Scrum methodologies

Education

- Master of Science (MSc.) in IT, 2011 from University of East London, London.
- PG Diploma (Computer Systems & Networking), 2009 from University of Greenwich.
- B. Tech (ECE), 2006 from T.R.R College of Engineering, Hyderabad, Affiliated to Jawaharlal Nehru Technological University.
- Board of Intermediate (M.P.C), 2002 from J.V.S Jr. College, Vishakhapatnam.
- Xth CBSE, 2000 from Kendriya Vidyalaya No.1, Nausenabaugh, Vishakhapatnam.

Work Experience

Enterprise Minds India Pvt Ltd, Hyderabad,

Jul'23 – Till Date

Position: Lead Engineer – Generative AI & DevOps

Job Profile: AI Model Development, LLMOps & RAG Pipelines, Cloud Infrastructure Automation, AI-DevOps Integration, Kubernetes Admin and Project Management.

- **LLM Development and Fine-Tuning:** Developed and fine-tuned **Large Language Models (LLMs)** for enterprise applications, improving AI-driven solutions by **40%** in accuracy and relevance.
- **Retrieval-Augmented Generation (RAG) Systems:** Implemented **RAG systems**, enhancing response times by **25%** and improving the contextual accuracy of AI outputs.
- **GPT-based Model Optimization:** Fine-tuned GPT models to enhance **customer interaction systems**, achieving a **30% increase in user satisfaction scores**.
- **NLP for Sentiment Analysis and Content Generation:** Applied advanced **NLP techniques** to deliver personalized, context-aware AI interactions through **sentiment analysis** and **content generation**.
- **Graph Database Integration:**
 - Designed and deployed **Neo4j-based graph databases** to model relationships between data points, enabling **knowledge management** and **AI-driven insights**.
 - Leveraged **GraphDB queries within LangChain workflows** to improve the contextual understanding of **AI-powered incident response systems**.
- **LangSmith & LangGraph Integration:**
 - Applied **LangSmith** to **monitor, optimize, and debug multi-agent workflows**, ensuring robust **AI observability**.
 - Utilized **LangGraph** to **visualize complex agent workflows**, track task dependencies, and ensure seamless collaboration between AI and DevOps teams.

- **RASA for Conversational AI:** Built and deployed **context-aware chatbots** using RASA to automate customer interactions, reducing response times and enhancing **customer satisfaction** by 30%.
- **AI Model Development with TensorFlow 2 and Keras:** Developed and fine-tuned **deep learning models** using TensorFlow 2 and Keras for NLP tasks such as **text classification** and **sentiment analysis**.
- **AI Ethics and Responsible AI Practices:**
 - Incorporated **AI ethics principles** into LLM workflows, focusing on **bias mitigation**, **privacy preservation**, and ensuring **transparency** of AI outputs.
 - Conducted **ethical audits** to ensure AI applications meet fairness and **compliance standards**.
- **Agentic Patterns in AI Applications:** Designed and deployed **multi-agent frameworks** using LangChain, automating **multi-step workflows** for **incident management**, **IaC generation**, and **log analysis systems**.
- **AI Observability with LangChain & Prometheus:** Developed real-time monitoring systems for AI workflows, ensuring early detection of issues and **reducing downtime** by 30%.
- Proficient in LLMOps for seamless **AI model deployment and lifecycle management**.
- **AI-DevOps Integration:** Led AI-DevOps integration by automating **CI/CD pipelines** with **Jenkins, Docker, and GitHub Actions**, ensuring seamless **model deployment**.
- **Kubernetes and Istio Management:** Managed **Kubernetes clusters** and **Istio service mesh** to enable scalable, secure deployment of AI services in production.
- **Cloud Infrastructure Automation:** Automated infrastructure management with **Terraform and Ansible**, optimizing resource utilization on AWS Cloud.
- **Monitoring and Observability:** Configured **Prometheus, Grafana, and EFK Stack** to monitor AI models and infrastructure, ensuring high performance and **fast troubleshooting**.
- **Cloud-Native Deployment:** Transitioned AI models to **cloud-native environments** using **Kubernetes and Helm** for scalability and reliability.
- **DevSecOps Practices:** Integrated **DevSecOps practices** to secure deployments and ensure compliance throughout the **AI project lifecycle**.
- **Jira Project Management:** Managed **Jira workflows and boards**, applying **Agile/Scrum methodologies** to ensure efficient project tracking and delivery.
- **Domain Knowledge:**
 - Working with **Viasat**, a **telecommunications provider**, specializing in **high-performance internet services**.
 - Gained working knowledge across multiple domains, including **finance, healthcare, retail, and e-commerce**, ensuring a versatile approach to AI and DevOps solutions.

Project 1: Proactive Incident Response with AI Agents and RAG

Overview:

Developed an automated incident response system integrating AI Agents, LangChain, and RAG. This system uses a vector database for real-time retrieval of past incidents, improving response workflows and providing automated resolutions.

Impact:

Reduced incident response times by **30%**, enhanced resolution accuracy by **25%**, and improved operational efficiency by **40%** through automation and advanced AI workflows.

[GitHub Link](#)

Project 2: Log Analytics and AI Agents**Overview:**

Built an AI-driven log analysis system using NLP and LangChain to streamline error detection and troubleshooting. This project incorporated RAG, Elasticsearch, and Pinecone for efficient log querying and real-time insights, accessible through Slack for easy operator interaction.

Impact:

Reduced troubleshooting time by **35%**, improved error resolution accuracy by **30%**, and enhanced collaboration by **25%** with automated insights and context-aware suggestions.

[GitHub Link](#)

Project 3: AI-Driven Infrastructure-as-Code (IaC) Management**Overview:**

Developed an AI-powered system for managing Infrastructure-as-Code (IaC) using LangChain, RAG, and OpenAI to automate Terraform template generation, validation, and deployment. The system ensures configuration consistency across sessions and automates multi-step deployments.

Impact:

Reduced IaC deployment time by **40%**, ensured **100%** configuration consistency across environments, and lowered manual errors by **30%** through automated workflows and continuous learning from operator feedback.

[GitHub Link](#)

Infostretch Corporation (India) Pvt. Ltd. (Apexon), Hyderabad, Sep'22 – Apr'23

Position: Cloud/DevOps Lead Engineer

Job Profile: Infrastructure & Deployment Automation, and Kubernetes Admin.

- **CI/CD Migration:** Led the migration from **Jenkins** to **GitHub Actions**, reducing maintenance efforts, costs, and improving workflow efficiency.
- **Streamlined DevOps Workflow:** Achieved **50% reduction** in build and release time by optimizing processes with **GitHub Actions**.
- **Aurora MySQL Upgrade:** Executed an **in-place upgrade** of Aurora MySQL using **Terraform scripts**, ensuring minimal downtime and meeting End-of-Life requirements.
- **Database Performance Tuning:** Optimized **Aurora architecture**, including query execution, cache management, and read replica configuration.
- **Proactive Monitoring:** Used **CloudWatch** and **RDS Enhanced Monitoring** to detect and resolve issues **80% faster**, improving database reliability.
- **Database Security and Compliance:** Managed **encryption, authentication, and auditing** to ensure **Aurora database integrity** and regulatory compliance.

- **Infrastructure Automation:** Integrated **Terraform** with CI/CD pipelines to automate infrastructure creation and application testing, reducing manual intervention by **80%**.
- **Serverless Development:** Developed **scalable serverless applications** with **AWS Lambda** and **API Gateway** to meet dynamic business needs efficiently.

Visionet Systems Pvt. Ltd. Bengaluru,

July'21 – July'22

Position: Senior Cloud/DevOps Engineer

Job Profile: Build and Release Management, Deployment Automation, and Kubernetes Administration.

- **GitOps with ArgoCD:** Implemented **GitOps pipelines** using **ArgoCD**, automating infrastructure and application deployments with version control, resulting in **faster rollbacks** and improved **deployment consistency**.
- **Kubernetes Orchestration:** Managed complex containerized applications with **Kubernetes**, ensuring **scalability and fault tolerance**.
- **Networking and Security:** Managed **container networking** with load balancers, ingress controllers, and network policies, implementing **Docker and Kubernetes security best practices** to prevent unauthorized access.
- **Troubleshooting Expertise:** Debugged **Docker/Kubernetes issues**, analysed logs, and resolved production issues by **90%**, improving system reliability.
- **CI/CD Pipelines and Automation:** Developed and managed **Jenkins pipelines**, and automated deployments with **Terraform and Ansible**, reducing manual tasks by **70%**.
- **AWS EKS Management:** Managed **EKS clusters**, worker nodes, and integrated them with AWS services for efficient cloud operations.
- **Monitoring and Observability:** Implemented centralized logging with **EFK**, **Prometheus**, **Grafana**, and distributed tracing with **Jaeger** and **Kiali** for real-time visibility and performance monitoring.
- **DevSecOps Strategy:** Assisted in developing and implementing a **DevSecOps strategy**, integrating security controls throughout the SDLC.

TCS, Hyderabad,

Jan'18 – July'21

Position: Cloud/DevOps Engineer

Job Profile: AWS Administration, Infrastructure Automation & DevOps Engineer

Payroll: Ivytel Technologies Pvt. Ltd. Bengaluru,

- **AWS Cloud Administration:** Managed **EC2, RDS, S3, VPCs, Lambda, API Gateway**, and other AWS services for efficient cloud operations.
- **AWS Security and Compliance:** Expertise in **IAM, KMS, AWS Config, GuardDuty, and Inspector** to maintain secure and compliant cloud environments.
- **Networking and Monitoring:** Managed **VPCs, subnets, security groups**, and used **CloudWatch/CloudTrail** for monitoring and troubleshooting AWS resources.
- **DevOps and CI/CD Tools:** Proficient with **Git, Jenkins, Ansible, Terraform**, reducing deployment time by **50%** and ensuring fast, reliable releases with **Jfrog** and **SonarCloud** integration.
- **Containerization and Orchestration:** Used **Docker and Kubernetes** to build, deploy, and scale applications, improving platform performance and reducing manual processes.

- **Training and Collaboration:** Conducted **DevOps best practices training** for development and operations teams, fostering collaboration and operational efficiency.

Mindtree LTD. Hyderabad,

Feb'15 – Dec'17

Position: Module Lead

Job Profile: AWS Cloud Administration & Infrastructure Automation

- **AWS Infrastructure Management:** Managed **EC2, RDS, and S3** for a high-traffic e-commerce site, optimizing costs and performance.
- **Networking and Security:** Configured **VPCs, subnets, and security groups** to ensure optimal performance and robust security.
- **Fault Tolerance and Scalability:** Implemented **ELBs, Auto Scaling Groups, and custom AMIs** for high availability during traffic spikes.
- **IaC Automation:** Developed reusable **Terraform modules** and deployed infrastructure across environments, reducing manual efforts by **70%**.
- **Monitoring and Performance:** Configured **CloudWatch alarms** to track key metrics, enabling early detection and quick issue resolution.

Aster Minds Enterprise Solutions Pvt. Ltd. Hyderabad

May'14 – Feb'15

Position: Messaging Server Consultant (Client: Microsoft BLR, C2H)

Job Profile: Exchange & O365 Migration

Zenith InfoTech Pvt. Ltd. Mumbai

July'12 – May'14

Position: Server Engineer (L3)

Job Profile: Exchange Server Administration