

# Karan Kalyan

San Jose, CA

karankalyan006@gmail.com

linkedin.com/in/karan-kalyan — github.com/KalyanKaran — karankalyan.com

## Summary

AI Engineer with expertise in **Machine Learning, Deep Learning, and MLOps**, focusing on **scalable AI solutions**. Passionate about **AI research, cloud infrastructure, and automation**. Proven ability to develop **multimodal models, optimize cloud-based ML workloads, and implement MLOps best practices**. PhD candidate in **Information Technology**.

## Skills

- **AI/ML:** Deep Learning, Transformers, Neural Networks, PyTorch, TensorFlow, OpenAI API
- **MLOps:** Kubernetes, Docker, Terraform, CI/CD, Model Deployment, AWS SageMaker
- **Cloud DevOps:** AWS (EKS, EC2, S3, Lambda), Google Cloud, Azure, Helm, Prometheus, Grafana
- **Programming:** Python, Go, C++, Java, Bash, TypeScript, Node.js
- **Certifications:** AWS Certified Solutions Architect - Professional, Kubernetes Certified Administrator (CKA), Terraform Associate

## Experience

### Kaiser Permanente, Pleasanton, CA

*AI Engineer - Machine Learning*

Apr 2023 - Present

- Developed **multimodal language models** using **Transformers and Diffusion architectures**.
- Designed and optimized **AI model pipelines** for **scalability and efficiency**.
- Implemented **real-time inference systems** using **MLOps principles**.

### LIXIL, New York, NY

*Platform Engineer - AI/ML*

Oct 2023 - Apr 2024

- Designed **high-performance ML infrastructure**, optimizing GPU/CPU workloads.
- Automated cloud resource provisioning with **Terraform and Kubernetes**, reducing setup time by 30%.
- Integrated **Dockerized ML workflows** into high-performance computing (HPC) clusters.

### Savvas Learning Company, Arizona, USA

*AWS Cloud Engineer*

Oct 2022 - Sep 2023

- Designed and deployed **scalable EKS clusters** with Kubernetes **auto-scaling and rolling updates**.
- Implemented **RBAC security policies** and optimized Kubernetes workload isolation.
- Developed **AWS CDK and Helm charts** for streamlined cloud deployments.

### Nomi Health, Austin, TX

*AWS Infrastructure Engineer*

May 2021 - Sep 2022

- Built and maintained **CI/CD pipelines** using AWS CodePipeline, CodeCommit, and CodeBuild.
- Integrated **serverless AWS Lambda functions** with API Gateway.

- Developed **Terraform modules** for **automated provisioning** of AWS resources.

**KLA, Milpitas, CA**

*Site Reliability Engineer*

Aug 2020 - Apr 2021

- Defined and implemented **SLOs and SLIs** for system reliability.
- Developed monitoring solutions using **Prometheus, Grafana, and Datadog**.
- Designed and maintained **AWS disaster recovery solutions** with multi-region backups.

## Education

**University of the Cumberlands** – PhD in Information Technology (2024 - 2027)

**California State University, Fresno** – M.S. in Data Science and Machine Learning (2018 - 2020)

**UBDT College of Engineering, Davangere** – B.E. in Computer Science (2013 - 2017)

## Projects

**AI-Driven Cybersecurity System** – California State University, Fresno

- Developed a **machine learning-based threat detection system** using **TensorFlow and AWS SageMaker**.
- Achieved **real-time network intrusion prevention** by training deep learning models.

**Serverless IoT Data Pipeline** – California State University, Fresno

- Designed an **IoT data pipeline** using **AWS Lambda, Kinesis, DynamoDB, and S3**.
- Enabled real-time **sensor data ingestion** with **fully serverless deployment**.

## Awards and Certifications

- AWS Certified Solutions Architect - Professional
- Kubernetes Certified Administrator (CKA)
- Certified Terraform Associate
- Recognized as **Infrastructure as Code (IaC) Wizard** at Nomi Health for excellence in cloud automation.
- Finalist, **TechX Hackathon 2021** – Built an AI-powered chatbot for healthcare automation.