Shiva Thavani

thavani.shiva3@gmail.com +91-7901913699
linkedin.com/in/shiva-thavani-916177167 https://shiva-rag-bot.vercel.app/

AI Engineer · 5+ YOE · Agentic AI | MLOps | GenAI | RAG | GPU Inference

WORK EXPERIENCE

Atlassian | Jun 2025 - Present
Machine Learning Engineer

- Built and deployed an Agentic AI Customer Support (CSM-Rovo) Agent leveraging **LangGraph** and **Model Context Protocol (MCP)** for dynamic tool orchestration and context aware workflow automation, reducing manual intervention by 45%+.
- Implemented hybrid orchestration and semantic query pipelines using LlamaIndex, A/B tested retrieval strategies, improving response relevance by 35%.
- Optimized and deployed AI models using ONNX and TFLite runtimes, enabling efficient low-latency cross-platform inference.
- Developed GraphRAG based guidance workflows for knowledge graph-driven context retrieval, leveraging GPT-5.2 and Gemini-3-Flash models
- Led rollout of the CSM AI Agent across GTM and support workflows, enabling **10k+** AI assisted resolved conversations.

Autodesk | Jul 2024 - June 2025
Senior AI/ML Engineer

- Deployed GenAI Models in Autodesk product suite by creating **AI Inference Service** with Go in ML Platform built on top of **AzureML**.
- Optimized GPU inference workloads using CUDA-based containers, improving throughput by 2.5x and reducing GPU utilization costs by 30%.
- Led AI inference service architecture and managed scalable ML workflows using Terraform, MLflow, and Kubeflow for ML model lifecycle management across enterprise-scale applications.
- Developed Agentic AI incident-resolution workflows leveraging LangChain, RAG, Pinecone, FAISS, OpenAI models, and ReAct/CoT reasoning

Thomson Reuters | Sept 2022 - Jul 2024
Senior AI Engineer | Aug 2023 - Jul 2024
AI ENGINEER | Sept 2022 - July 2023

- Built AI Platform for various stages of the Data Science Lifecycle, supporting 200+ data scientists and ML engineers across **AWS** & **Azure** Services like **SageMaker** studio, S3, Lambda, **AzureML** studio, **Azure AI** studio, Azure **OpenAI**, Blob Storage, **GCP Vertex AI** etc.
- Engineered the MLOps Platform using IaC - Terraform, CloudFormation templates.
- Managed end-to-end MLOps lifecycle, data processing, model development, deployment to prod, and monitoring - across **AWS** (SageMaker Pipelines, Lambda, ECR, DynamoDB) and **Azure** (Azure ML, Kubernetes, Kafka).
- Deployed and served LLMs like LLaMA2&3, Claude3, and GPT4, on AWS Bedrock, Azure AI Studio.
- Led the AI Platform service end-to-end, from model evaluation and optimization to driving platform adoption.

Ernst & Young - EY GDS | April 2021 - Aug 2022
Data Scientist - Internship & FTE

- Trained an AI Model for solving the use case of **Document AI** using **TensorFlow Object Detection** achieving 90%+ object detection accuracy.
- Containerized and deployed Document AI models on **Azure Kubernetes Service** (AKS) using Azure ML deployment pipelines.

EDUCATION

Thapar Institute of Engineering and Technology, Patiala
Dual Degree B.E. - M.B.A Computer Science Engineering
BE - Computer Science Engineering CGPA - 9.26/10
MBA - Finance & Big Data Analytics CGPA - 8.91/10

TECHNICAL SKILLS

- **Generative AI** - **LangGraph**, LangSmith, GraphRAG, MCP, RAG, Vector Database, OpenAI APIs, Claude, Prompt Engineering (CoT/ReAct)
- **AI Infrastructure** - **Pinecone**, Weaviate, ChromaDB, **FAISS**, Azure AI Search, Redis, Kafka
- **Programming & Scripting Languages** - C, C++, **Python**, JavaScript, R, SQL, Shell Script, **Golang,** Kotlin
- **ML Frameworks** - Flask, OpenCV, **TensorFlow**, **PyTorch**, YOLO, CUDA GPU, Nvidia GPU Toolkit, **FastAPI**, **ONNX, Hugging Face**
- **MLOps** - **Git**, Docker, **Kubernetes**, MS Azure Cloud, **Azure** ML, Google Cloud, **MLflow**, AWS Sagemaker, **Kubeflow**, **PySpark,** Datadog
- **Infrastructure as Code** - **IaC -Cloudformation** templates, **Bicep** templates, AWS WAF, Route 53, Azure Service principal , **Terraform**

ACHIEVEMENTS

Hackathons

- Winner **Smart India Hackathon 2020**. Theme - Open **Innovation**.
- 1st Prize Winner in **Hackpions-EY GDS** Hackathon 2021. Theme - Intelligent **Automation** with AI
- 1st Prize Winner in **Bosch AI Hackathon** 2021. Theme - Smart **AI** Robot
- 1st Prize Winner in **Xebia Xe-Thon** 2021. Theme - Intelligent hiring using **AI**

Competitions

- **Gold** Medalist in Academic BE-MBA Degree Qualification
- **First** Rank in Academic year at TIET 2019-2020 **and** 2018-2019

Scholarships & Programs

- **Merit Scholarship** holder at TIET for the year 2019-2020.
- **Merit Scholarship** holder at TIET for the year 2018-2019.