JEYA VARSHINI BHARATH

📞 +1 (857) 829-0729 | ✉ jvarshini99@gmail.com | 🔗 linkedin.com/in/jvarshini99 | ⭘ github.com/jvarshini99

EDUCATION

Master of Science in Artificial Intelligence *January 2024*
Boston University, Boston, MA
Bachelor of Technology in Computer Science and Engineering *July 2021*
Vellore Institute of Technology, Vellore, India

SKILLS

Languages & Softwares: Python, Java, JS, SQL, C++, GCP, AWS, Azure, Kubernetes, Docker, Vertex AI, GIT
Tools & Frameworks: PyTorch, TensorFlow, NLTK, Langchain, Numpy, Pandas, Keras, OpenCV, FastAPI

EXPERIENCE

Terragrit Corp. - Remote, USA *January 2025 - Present*
Software Engineer - AI
- Working on building the AI insights feature of company's flagship product - digital twin.

Quantiphi Inc. - Boston, USA *March 2024 - December 2024*
Machine Learning Engineer Intern
- Designed a **multi-agent Retrieval-Augmented Generation (RAG)** implementation using **Langchain**, **LangGraph** and **Gemini** for a Sales Assistant Agent. Used **AlloyDB** for the **vector embeddings**. Leveraged **Grounding API** for web search and **NL2SQL** for information retreival. Containerized and deployed using **FastAPI** on **GCP Vertex AI endpoint**, enabling seamless, real-time user interactions, achieving precision in response generation. Built the **RAG evaluation** pipeline using **Ragas**, incorporating metrics such as **BLEU**, **ROUGE**, and **cosine similarity**.
- Developed an end-to-end CV pipeline using **OpenCV** to integrate **GroundingDINO** and **Segment Anything** for object detection and bounding box generation. Deployed it using **Docker & Vertex AI** for an advertisement usecase.
- Built robust **MLOPs** pipelines with **Terraform**, integrating frameworks like **PyTorch, Tensorflow, XGBoost**, and **Scikit-learn** with **GCP Services, MLFlow, Kubeflow, Vertex AI, Google cloud storage**, and **BigQuery** enabling automated model training and both batch and online prediction for diverse business use-cases. Developed **CI/CD** pipelines with **GitHub Actions** to streamline the deployment of AI infrastructure and models.
- Developed an automated system for a life insurance company utilizing Vertex AI's **cloud functions** and **Doc AI API** to extract and organize vital information from medical records along with prompt engineering using **Gemini Pro 1.5** and **Gemini Flash 1.5** models, enhancing data accuracy and processing speed.

Honeywell - Bangalore, India *January 2021 - August 2022*
Application Dev Specialist - Cloud Infrastructure
- Enhanced reliability and performance of 10+ legacy applications by migrating and optimizing them for deployment on the **Azure cloud** platform, resulting in improved reliability and performance. Set up **Virtual Private Clouds (VPCs)**, **Load Balancers**, and **Firewalls** in Azure to ensure secure, scalable, and robust cloud infrastructure for the applications. Collaborated with development teams to identify and resolve issues in the **.NET (C#)** codebases, ensuring seamless production deployment, and successful implementation of scalable solutions for the enterprise applications.

ACADEMIC PROJECTS

AI-Powered Health Assistant and Emotional Support Chatbot
- Developed 'AskFabian', an AI-powered chatbot, aimed at bridging the gap in accessible health information and emotional support using **NLTK**, utilizing **LLMs** and deployed on **HuggingFace**.(GitHub)
- Utilized **GPT-2**, and **BERT** models for sentiment analysis & **Llama 2**, and **Mistral 7B** models for their refined response generation. Integrated these models via **Chainlit** and **Langchain** for interface development.
- Successfully optimized response time via model optimizations like **LoRA** and **quantizations**, achieving real-time engagement capabilities. Internal evaluation indicated enhanced accuracy in medical query handling and emotional support.

CERTIFICATIONS

Google Professional Machine Learning Engineer (link), **Google Associate Cloud Engineer** (link), **Snowflake Snowpro Core Certification** (link)