SONALI VEDARAJU

sonalivedaraj@gmail.com | (682)-374-6951| <u>LinkedIn</u> | <u>Github</u> | Pennsylvania

PROFESSIONAL SUMMARY

AI-focused Full-Stack Engineer with nearly 6 years of experience building scalable microservices and high-performance SPAs. Proven expertise in designing and deploying LLM-powered agentic AI systems for reliable, enterprise-grade outputs. Strong React/Next.js foundation with a track record of performance optimization and bundle efficiency. Skilled in cloud-native development, AWS integrations, and streamlining workflows through automated CI/CD pipelines.

PROFICIENCY

Generative AI: LLMs, Agentic AI, LangGraph, RAG, Prompt Engineering, HuggingFace, VectorDatabases (FAISS, OpenSearch, Pinecone)
Languages: Python, TypeScript, Java
Frontend & APIs: React, Next.js, RESTful APIs, Microservices
Data & ML: Databricks (Spark SQL, Model Serving), MLflow
Databases: PostgreSQL, MySQL, MongoDB, Redis
Cloud & DevOps: AWS (EC2, S3, Lambda, IAM, Bedrock, Rekognition), Docker, Kubernetes, GitLab CI/CD, Kafka

WORK EXPERIENCE

Software Developer | Paycom, USA
Present - Jun 2024

- Integrated an **LLM-powered conversational agent** into HR and Payroll systems, enabling natural language employee self-service within production workflows.
- Designed **structured prompt templates** and implemented deterministic JSON schema validation to ensure reliable, business-rule–compliant LLM outputs for downstream payroll API integration.
- Built an **AI-to-UI orchestration layer** dynamically mapping structured LLM responses to **reusable React components**, bridging generative AI with enterprise-grade frontend systems.
- Resolved **rendering bottlenecks** in the Employee Directory data grid via **row virtualization and memoization**, reducing memory usage by 60% and improving render time by 20%.
- Reduced UI Library **bundle size by ~25%** through dependency optimization, t**ree-shakeable ES module exports** and used bundle analysis to validate.
- Implemented **cross-repository GitLab CI/CD workflows** to automate build, versioning, and containerized deployments without manual intervention.
- **Collaborated with clients, product, and design teams** to deliver customizable advanced filtering capabilities across HR modules.

Research Assistant | University of Texas at Arlington, USA
May 2024 - Aug 2023

- Conducted model experimentation using **pre-trained transformer models** from **Hugging Face**, running **inference** on university datasets and comparing performance metrics such as accuracy and latency to determine the most suitable models for different use cases.
- Assisted professor in **evaluating model outputs** and documenting **performance trade-offs** to support research analysis and informed model selection decisions.

Software Development Engineer | TravTech Inc, Bangalore, India
Jul 2022 - Aug 2018

- Implemented microservices-based **ETL workflows** to process heterogeneous cruise pricing and itinerary feeds from multiple vendors, ensuring validated and normalized data prior to database persistence, significantly improving booking data integrity and reliability.
- Implemented event-driven booking workflows using **Apache Kafka** to publish booking lifecycle events for email notifications, loyalty point updates, and analytics processing, thereby decoupling microservices and improving system reliability under peak traffic.
- Migrated booking invoice storage from **on-prem file systems to Amazon S3**, enabling secure customer access via pre-signed URLs, thereby eliminating storage constraints and enhancing document security.

EDUCATION

Master of Science in Computer Science - *University of Texas at Arlington* *May 2024 - Aug 2022*
Bachelor of Engineering in Information Science - *Visvesvaraya Technological University* *Jun 2018 - Aug 2014*

CERTIFICATIONS

- Generative AI Fundamentals - Databricks *Oct 2025*
- AWS Certified Solutions Architect, AWS *Mar 2025*
- Certified Kubernetes Application Developer (CKAD) , Cloud Native Linux Foundation *Dec 2024*

PROJECTS

- **AWS CDK Image Analysis Pipeline** -Built a serverless image analysis pipeline using **AWS CDK**, with S3-triggered Lambdas analyzing images via **Amazon Rekognition.** Stored extracted metadata in DynamoDB, enabling scalable, event-driven processing with IaC best practices.
- **AI Travel Advisory Agent** - Built LLM-based agent on **Databricks** with LangGraph tool routing, combining **Spark SQL** queries for flights and RAG-based retrieval for policies, tracked via **MLflow.**
- **Chatbot using OpenAI API** - Built an AI-powered chatbot using Node.js and the **OpenAI API** to enable real-time conversational responses. Designed for easy application integration, providing a scalable and reusable intelligent assistant.