Vidit Naik

3429 Canyon Crest Dr, Riverside, CA | (951) 425-7229 | viditnaik@gmail.com | linkedin.com/in/viditnaik | viditnaik.netlify.app

Education

University of California, Riverside   March 2025
Master of Science in Computer Science (GPA: 3.63)
- **Relevant Coursework:** Operating Systems, Agile Development, SDLC, Data Structures and Algorithms, Object Oriented Programming, Machine Learning, Artificial Intelligence, Networking, Design Patterns, Distributed Systems

Vellore Institute of Technology, Chennai   August 2023
Bachelor's in Computer Science and Engineering

Skills

Languages Java, Python, JavaScript, C++, SQL

Database MySQL, MongoDB, NoSQL

Frameworks and Libraries React.js, React Native, Hadoop, Node.js, Django, Express, RESTful APIs, Flask, SpringBoot

Testing PyTest, JUnit, Selenium

AI and ML Large Language Models (Qwen, BERT, Llama3), RAG, PyTorch, Fine-Tuning (LoRA), Ollama

Cloud and DevOps AWS (S3, Glue, Lambda, EC2), Docker, Kubernetes, CI/CD Pipelines, Jenkins, Git/GitHub, Jira

Tools and Methodologies Microservices, Tableau, Distributed Systems, RESTful API, Agile/Scrum, Jira, OAuth 2.0, Linux, Unix Shell Scripts

Experience

Shifa Precision Inc.   July 2025 - Present
Software Engineer - AI/ML   *Boston, MA*
- Leading the backend development of Project Oasis - a platfrom leveraging AWS and Neo4j to generate real time patient "digital twins".
- Engineering a scalable data pipeline to unify biomedical data (ClinVar, PubMed, openFDA) into a knowledge graph with over 1M+ nodes and 10M+ relationships
- Implemented backend services integrating LLM APIs for biomedical data extraction and supported deployment of a GNN-based prediction engine

Center for Robotics and Intelligent Systems   October 2024 - March 2025
Student Researcher   *Riverside, CA*
- Developed an AI-powered system that integrates drones with Large Language Models (LLMs) for user-driven control through natural language.
- Utilized Retrieval-Augmented Generation (RAG) to provide the LLM with contextual knowledge from drone manuals, increasing command accuracy by 68%.
- Collaborated in a Scrum-based Agile development process to design features with a core commitment to quality and scalability.

StuDetails   July 2020 - February 2021
Software Engineer   *Noida, India*
- Architected and developed a full-stack web application using React and Flask, enabling 50+ enterprise clients to manage scheduling and inspections, resulting in a 20% increase in operational efficiency, supporting over 5,000 monthly active users.
- Led cross-functional collaboration to optimize AWS S3 data workflows using AWS Glue, improving data accuracy by 25% and enabling real-time integration with client-facing systems.
- Implemented automated testing and CI/CD pipelines, cutting deployment failures by 30% for apps processing 1TB+ data monthly.

Projects

InsureSearch: RAG based AI Chatbot
- Developed and deployed a full-stack AI chatbot using a React frontend, a Node.js/Express.js backend, and a MongoDB database.
- Optimized the backend to reduce token usage by 65% through a top-k BERT search implementation, significantly improving response times.
- Enhanced LLaMA-3 8B with 4-bit quantization and LoRA adapters, fine-tuning on insurance policy data to boost accuracy for user prompts. Handled the source control of the project through Git.

CitySafe: Chicago Crime Insights Dashboard
- Built an analytics dashboard processing 1.5M+ records using SQL, Spark, and PostGIS with ETL pipelines for automated data extraction and crime trend insights.
- Designed geospatial visualizations with Python libraries and React.js, enabling self-service monitoring of crime hotspots.
- Implemented a Python-based RESTful API with optimized data models and indexing, which reduced query response time by 25% and efficiently served data to a React.js frontend.

CreateFlow - CalHacks 11.0
- Designed a multi-agent system with LangGraph linking AI agents for content, scheduling, & analytics, cutting creation time by 50%
- Enhanced LLaMA-3 8B with 4-bit quantization and LoRA adapters, fine-tuning on LinkedIn data to boost engagement predictions