Vidit Naik

1628 Alemany Blvd, San Francisco, CA | (951) 425-7229 | viditnaik@gmail.com | linkedin.com/in/viditnaik | viditnaik.netlify.app

Education

University of California, Riverside <div align="right">March 2025</div>
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 <div align="right">August 2023</div>
Bachelor's in Computer Science and Engineering

Skills

Languages and Databases Java, Python, JavaScript, C++, SQL, MySQL, MongoDB, NoSQL
AI and Machine Learning: Large Language Models (Claude, Gemini, Qwen), RAG, Fine-Tuning (LoRA), PyTorch, Ollama
Frameworks & Platforms: React.js, Playwright, Node.js, Express, SpringBoot, Flask, Django, RESTful APIs, Hadoop, PyTest, JUnit, Selenium
Cloud and DevOps AWS (S3, Glue, Lambda, EC2), GCP, Docker, Kubernetes, CI/CD Pipelines, Jenkins, Git/GitHub, Jira

Experience

Checksum AI <div align="right">November 2025 - Present</div>
Forward Deployed Engineer <div align="right">*San Francisco, CA*</div>
- Leading end-to-end delivery of AI products into client software, managing design, launch, and enablement.
- Architecting self-healing QA pipelines by integrating LLM models like Claude to automatically analyze DOM changes and patch broken selectors.
- Developing internal AI tools using LLM agents to ensure uniform code quality across multiple codebases, ensuring minimal failures on cloud.
- Acted as the primary technical liaison for enterprise clients, leading technical design sessions, live demos, and workshops to translate complex requirements into scalable solutions.

Shifa Precision Inc. <div align="right">July 2025 - October 2025</div>
Software Engineer - AI/ML <div align="right">*Boston, MA*</div>
- Led the backend development of Project Oasis - a platform leveraging GCP and Neo4j to generate real time patient "digital twins".
- Owned critical systems end-to-end, architecting a scalable API-first data pipeline to unify biomedical data into a Neo4j knowledge graph with 10M+ relationships.
- Built high-throughput backend microservices to process unstructured biomedical text, utilizing LLM-based entity extraction to populate a Neo4j Knowledge Graph with 95% accuracy.

Center for Robotics and Intelligent Systems <div align="right">October 2024 - March 2025</div>
Student Researcher <div align="right">*Riverside, CA*</div>
- Developed an AI-powered system that integrates drones with Large Language Models (LLMs) for user-driven control through natural language.
- Engineered a RAG pipeline using LangChain and vector embeddings to semantic search drone manuals, reducing hallucination rates and increasing command execution accuracy by 68%.
- Collaborated in a Scrum-based Agile development process to design features with a core commitment to quality and scalability.

StuDetails <div align="right">July 2020 - February 2021</div>
Software Engineer <div align="right">*Noida, India*</div>
- Directed the full requirements lifecycle, from gathering insights from 15+ stakeholders to authoring technical specifications, culminating in the successful on-schedule launch of a feature that improved user engagement by 20%.
- 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.