NIRVISHA SONI

352-682-9917 | nsoni@ufl.edu | linkedin.com/in/nirvishasoni | github.com/Nirvisha82 | nirvishasoni.dev | Gainesville, FL

EDUCATION

Master of Science - Computer Science, University of Florida, Gainesville, FL   GPA 3.9/4 | **August 2024 – December 2025**
Coursework: Advanced Data Structures, Distributed Operating Systems Principles, Analysis of Algorithms, NLP.
Bachelor of Technology - Information Technology, VIT, Pune, India   CGPA 9.62/10 | **August 2019 – May 2023**
Coursework: DSA, OOP, DBMS, Computer Networks, Data Science, Machine Learning, Image Processing and Computer Vision.

TECHNICAL SKILLS

Languages: Python, Go, JavaScript, TypeScript, HTML5, CSS, SQL
Frameworks & Libraries: Node.js, NestJS, React.js, Flask, LangChain, LlamaIndex, PyTorch, NumPy, Pandas, Tensorflow
Databases & Tools: MySQL, PostgreSQL, MongoDB, Git, Docker, RESTful APIs, Figma
Development Practices: Software Development, Agile (Scrum, Kanban), CI/CD, Software Architecture, Version Control

EXPERIENCE

AI Software Developer | UF IC3, Gainesville, FL   **August 2025 – Present**
  – Driving design and implementation of an end-to-end RAG pipeline leveraging UF Shands data to power a virtual clinical avatar for Hospital Digital Twin project.
  – Building a modular ingestion and indexing framework (CSV, DOCX, PDF, TXT, Images) with ChromaDB.
  – Integrating retrieval framework with web search to deliver accurate, context-aware responses for clinical support.
  – Implementing PHI redaction, hallucination detection, and HIPAA-aligned guardrails, ensuring regulatory compliance.

Software Engineer Apprentice | American Express, FL   **August 2025 – Present**
  – Building a GitHub automation bot in Go that streamlines development workflows by auto-linking issues, generating PR templates, creating branches, and assigning appropriate reviewers.
  – Implementing repository-wide governance checks (naming conventions, approvals, testing requirements) to enforce contribution standards across teams.
  – Investigating advanced automation features such as flaky test detection and duplicate issue resolution, with planned extension to GitLab.

Graduate Analyst | Barclays, Pune, India   **July 2023 – July 2024**
  – Automated post-trade validation by designing a serverless AWS Lambda pipeline that integrated with internal systems, reducing manual intervention by 60% and accelerating settlement workflows.
  – Wrote optimized SQL scripts to audit and cleanse high-volume trade data, ensuring regulatory compliance and enabling reliable downstream reporting.

Machine Learning Intern | AlgoAnalytics Private Limited, Pune, India   **January 2023 – April 2023**
  – Conducted topic modeling on 5,000+ financial news articles, enabling website feature for topic-based document search.
  – Built transformer-based summarization models to condense financial news, improving clarity by 10%.
  – Developed a sentiment-stock movement analysis model with rolling windows, achieving a Pearson correlation of 0.7.

PROJECTS

BioAssist - Biomedical Web-Search Enabled RAG Chatbot (GitHub)   **March 2025 – Present**
  – Built a biomedical RAG chatbot with multimodal document ingestion, FAISS-based vector retrieval and MiniLM sentence embeddings for domain-specific Q&A.
  – Designed fallback retrieval logic that integrates local knowledge with real-time web search.
  – Implemented domain-specific guardrails, query rewriting, and hallucination detection.
  – Developed an interactive Streamlit UI supporting source traceability, chat persistence, and exportable conversations.

Roamio: The Ultimate Travel Companion (GitHub)   **January 2025 – April 2025**
  – Designed a full-stack platform enabling users to browse, manage, and share personalized travel itineraries using React.js and Go.
  – Implemented dynamic web forms for itinerary posting and editing, with backend integration and secure storage.
  – Integrated RESTful APIs to fetch and display destination-specific content and interactive location-based recommendations.
  – Built a responsive UI with Tailwind CSS & modular React components to deliver a seamless cross-device user experience.

Re:Gen: AI-Powered Gmail Reply Assistant (GitHub)   **February 2025 – May 2025**
  – Built a real-time system that listens for incoming emails and auto-generates personalized draft replies using Gmail API and LLM-powered prompt engineering.
  – Implemented a secure, multi-user backend using NestJS and SQLite, with OAuth2-based authentication and token management.
  – Integrated Google Cloud Pub/Sub to receive instant push notifications and ensure low-latency, event-driven reply generation.