

SAI NIVEDH V

✉ sainivedh26@gmail.com — 📞 8300874166 — 🌐 Github — 🌐 LinkedIn — 🌐 sainivedh26.live

Education

Amrita Vishwa Vidyapeetham

B.Tech. in CSE with specialization in Artificial Intelligence

Aug 2023 – June 2027

CGPA: 8.01/10

Technical Skills

- **Languages and frameworks:** Python, Java, C, SQL, JavaScript, Django, Flask, FastAPI, React, Tailwind, Bootstrap, Material3, Go, Git, Postman, Docker, Jenkins, CI/CD, REST API, GraphQL, Azure, GCP, AWS, MySQL, SQLite, MongoDB, PostgreSQL, Redis, Firebase, Neo4j, ChromaDB, Supabase, TensorFlow, PyTorch, scikit-learn, Langchain, Langgraph, LlamaIndex, Haystack, Snowflake, Autogen, CrewAI, ADK, Unsloth, Ollama, n8n workflows

Work Experience

Tech Profuse (Hyderabad, India)

June 2025 – Present

AI Developer Intern (Remote)

- Developed and Deployed Production Grade Voice Agent for Various clients varying from Restaurant to Saloon. Built with Agent Development Kit(google), Inferred Agent Streaming Via Livekit websockets using Go and python for high concurrency support. Deployed call inference using Exotel for Indian and international number support

IdeaboxAI (North Carolina, US)

March 2025 – June 2025

Backend Developer (Remote)

- Developed and productionized scalable backend APIs using FastAPI and integrated Langgraph agents for task automation. Collaborated with a distributed team to build an internal RAG tool supporting 1000+ documents specifically for invoices. Deployed agent observability using Langsmith

InceptAI (Coimbatore)

Feb 2025 – March 2025

Research Intern – AI Security Toolkit Development (Hybrid)

- Developed RAG-based enhancements to improve prompt handling in Security Copilot workflows. Designed proof-of-concept to detect prompt injection attacks in LLM-based development environments. Evaluated the performance of tool using metrics like BLEU, ROUGE, METEOR.

Binari Intelligence Systems Pvt. Ltd. (Salem)

Nov 2024 – Jan 2025

Software Development Intern (Onsite)

- Designed a system to automate interview scheduling and skill assessments for candidates. Gained hands-on experience in collaborative version control using Git and Agile workflows.

Achievements

Winner in AWS Global SuperOps Hack

Nov 2025

Finalist in Google Agentic AI Hackathon

June 2025

NextGen Hack 2025 Winner

May 2025

GenderXAI Hackathon 2025 Winner

Jan 2025

Web3Conf Hackathon 2024 Winner

Dec 2024

Projects

Thala - Intelligent Incident Management | FastAPI, MSK, Redis, Opensearch, Bedrock

Nov 2025

- Developed an ITSM agent plugin in team discussion platforms providing integration support with Slack and Jira for automated incident detection, classification, and resolution tracking using Kafka (MSK), Redis, and WebSockets for real-time streaming at scale. Built async FastAPI services handling 10000+ concurrent streams and used AWS Stacks like Lambda, Bedrock, Textract, OpenSearch, and Dockerized microservices deployed on AWS Elastic Container Registry. Latest version supports integration with Zoho Cliq Platform

Accessible AI | Flask, Google Cloud, React, Pinata, Stanford Parser

Dec 2024

- Engineered an innovative application for generating Indian Sign Language (ISL) videos from diverse media formats using Google Cloud Speech-to-Text for real-time transcription. Implemented NLP for ISL syntax translation and MoviePy to compile animated videos with accurate sign representation for Hearing Impaired

VerifyIt - Chrome Extension | Flask, Google Custom Search API, Crawl4AI, and Gemini

Nov 2024

- Developed an adaptive search and Multi-layer content verification system, providing real-time credibility analysis. Designed an intuitive, lightweight UI for seamless user experience without disrupting browsing flow

Deepfake Detection - Intel OneAPI Challenge | Huggingface, Intel OneDNN, Streamlit, Openvino

Mar 2024

- Designed and implemented a deepfake detection application using the ResNeXt-50 model, trained on the Celeb-DF (v2) dataset with 6,000+ samples, achieving 85% detection accuracy. Optimized inference speed by 50% and reduced latency by 40% using Intel's oneDNN and OpenVINO toolkit for efficient AI processing on GPUs and CPUs

Courses and Certificates

- AWS Certified Cloud Practitioner | [Certificate](#)
- Microsoft — Azure AI-900 Fundamentals | [Certificate](#)
- Oracle Cloud Infrastructure Foundation Associate [[Certificate](#)] and Data Science Professional [[Certificate](#)]
- Microsoft certified Fabric Data Engineer Associate | [Certificate](#)
- Neo4j Graph Data Science Certification | [Certificate](#)