

Sohan Venkatesh

Email: soh.venkatesh@gmail.com LinkedIn: [sohan-venkatesh](https://www.linkedin.com/in/sohan-venkatesh) GitHub: [sohv](https://github.com/sohv) Website: sohv.github.io Phone: [9611768593](tel:9611768593)

Education

Manipal Institute of Technology Bengaluru

Bachelor of Technology in Computer Science (Specialization in AI)

Relevant Coursework: Neural Networks, Computer Vision, NLP, Pattern Recognition, Machine Learning

Expected August 2026

CGPA: 8.40 / 10

Experience

Precog-at-IIITH

Research Intern

May 2025 – Present

Hyderabad, India

- Studying the effect of narrow fine-tuning on LLM personas under the guidance of Prof. Ponnuram Kumaraguru.
- Fine-tuned Qwen3 1.7B and 0.6B on a misaligned reasoning dataset using LoRA, identifying misaligned behavior in CoT reasoning traces.

Application Development Cell, MITB

Software Developer

September 2024 – March 2025

Bengaluru, India

- Worked on an automated tool to submit college data to various ranking and accreditation bodies
- Performed data modeling that simplified database design by **30%** and created comprehensive documentation with a user manual to help end-users navigate the tool seamlessly.

Fellowship.ai

Data Science Fellow

July 2024 – September 2024

San Francisco, USA

- Designed a Jupyter Copilot extension to assist data scientists with coding, debugging, and testing, improving efficiency by **30%**.
- Engineered CrewAI agents for data preprocessing and model training using GPT-4, reducing latency by **30%** and enhancing code readability by **50%**.

Infosys

Project Intern

May 2024 – July 2024

Bengaluru, India

- Built an ML model using PyCaret and Scikit-learn to predict obesity risk across multiple categories, achieving **83%** accuracy and **82%** precision.
- Applied Pandas and Matplotlib for data analysis, streamlining model training by **30%** and developed a Streamlit web app for real-time prediction.

Projects

Multi Code Review and Generation (MCRAG) | OpenAI, Gemini, FastAPI, React

- Built automated multi-LLM pipeline for code review and generation using GPT-4o, Gemini 2.5-pro, and Deepseek R1, reducing manual review time by **30%**.
- Improved code quality by **50%** using LLM critic feedback and consensus-based refinement.

Scorj – Resume Evaluator | GPT-4o, PyPDF, BeautifulSoup

- Developed an AI tool to evaluate resumes against LinkedIn job descriptions, scoring them out of 100 and generating detailed feedback.
- Adopted a hybrid LLM + embedding-based approach for scoring, reducing API cost by **40%** and improving processing speed by **50%**.

Financial Compliance Monitoring | Neo4j, Gemini, Knowledge Graphs

- Developed a semantic + rule-based system to monitor financial transactions for compliance using Neo4j graphs.
- Used Gemini 2.0 Flash to extract rules with **80%** extraction accuracy and integrated them into the knowledge graph.

SQL Chatbot | Llama-3.3, Groq, Streamlit

- Built a real-time SQL chatbot using fine-tuned Llama-3.3 on Groq, improving query generation accuracy to **91%**.
- Deployed via Streamlit, enabling fast and interactive query responses with optimized latency.

Technical Skills

Programming Languages: Python, JavaScript, C, SQL

Machine Learning & AI: PyTorch, Scikit-learn, HuggingFace Transformers, PEFT, OpenAI API, Ollama, LangChain, LlamaIndex, Pandas, Numpy, Matplotlib, OpenCV, Neo4j

Full-Stack Development: React, Next.js, Flask, FastAPI, TypeScript, PyTest, MongoDB, MySQL, SQLite

Developer Tools: AWS, Azure, Docker, Git, Jupyter Notebook, VS Code

Achievements

- Selected to attend the **Amazon ML Summer School 2025** program with a selection rate of **5%**.
- Awarded **Cerebras Fellowship** for the project titled 'Efficient Document Retrieval Using Knowledge Graphs'.