SOHAIL HARESH GIDWANI

Los Angeles, CA
+19736525842 | sohailgidwani15@gmail.com | https://linkedin.com/in/sohail-gidwani/
https://github.com/SohailGidwani | https://sohailgidwani.app/

SUMMARY

Software engineer with expertise in building scalable infrastructure and ensuring reliability of mission-critical systems. Skilled in Python, Java, C++, Node.js and cloud environments (AWS, Azure) with hands-on experience in Docker and serverless architectures. Delivered AI agent builder platform and internal data chatbots, improving operational efficiency and system robustness. Seeking a software engineering internship to contribute to foundational infrastructure projects.

EDUCATION

University of Southern California **Aug 2025 - May 2027**
Masters of Science, Computer Science *Los Angeles, CA*
- **GPA:** 3.5/4.0
- **Coursework:** Analysis of Algorithm, Information Retrieval And Web Search Engines

University of Mumbai **Aug 2019 - May 2023**
B.E., Computer Engineering *Mumbai, India*
- **GPA:** 9.05/10
- **Coursework:** Artificial Intelligence, Machine Learning, Advanced DBMS, Data Structures & Algorithms, Operating Systems, Software Engineering, Cloud Computing, Object-Oriented Programming, Big Data Analytics, Computer Networks, Cryptography & System Security, Blockchain

WORK EXPERIENCE

Keck School of Medicine, USC **Oct 2025 - Present**
Student Worker / Research Assistant *Los Angeles, CA*
- Build VLM experimentation workflows for multimodal healthcare tasks: dataset preparation, training runs, and evaluation.
- Support research connecting visual + text representations; implement reproducible pipelines and result tracking for model iterations.

Insaito, Inc. **May 2025 - Jul 2025**
Senior Software Engineer - I *Remote*
- AI Agent Builder Platform: Led the architecture and development of an AI agent builder platform, enabling users to create advanced AI agents with custom workflows and deep third-party integrations. Built core infrastructure for open-source LLM deployment, OAuth for 100+ apps, and Model Context Protocol (MCP) servers.

IIFL Finance Ltd. **Jun 2023 - May 2025**
Full Stack - Software Developer *Mumbai, India*
- Built an internal employee support chatbot using NLP, Python, and Flask. Integrated with Qdrant vector database, Azure OpenAI service, and Zoho ticketing system, streamlining internal processes with strong analytical skills. (Certificate of Achievement)
- Engineered an AI-powered Gold Loan Image Audit App using models like GroundingDINO and Swin-Transformer, enhancing fraud detection and reducing potential loan fraud by 15% while upholding a passion for quality software.
- Designed and implemented CapitalGenie, an automated user support system leveraging internal APIs and GPT-4o to fetch user data, diagnose issues, and generate personalized responses, accelerating resolution by 70% and demonstrating expertise in large-scale systems.

SKILLS

Programming: Python, TypeScript, JavaScript, Java (intermediate), C/C++ (Basic), SQL
Full-Stack: React, Next, Node.js, Express, Hono, FastAPI, Flask, RESTful APIs, HTML5, CSS3, Tailwind CSS, WebRTC
AI/ML: TensorFlow, Keras, Scikit-Learn, NLP, Computer Vision, OpenCV Vector DBs, LLMs, CNN, LSTM, Transformer, DBs, LLMOps, N8N, MCP
Databases & Cloud: MySQL, PostgreSQL, MongoDB, Oracle, SQLAlchemy, Azure, AWS, Docker, Git, Linux, Serverless, Large-scale Systems
Methodologies & Soft Skills: Agile, Test-Driven Development, Code Reviews, Backend, End-To-End, Analytical Skills, Passion For Quality Software

PROJECTS

Image Feature Detection & Captioning: Developed an AI app using CNN/VGG-16 for image feature extraction and an LSTM/Transformer-based captioning model achieving a BLEU score of 0.80, with a Streamlit interface for user-friendly interaction.
Knowledge Hub (Personal Semantic Search & Study Assistant): Built a local-first portal that ingests PDFs/images/handwritten notes; runs OCR + chunking; and delivers hybrid search (FTS + semantic) with citations and an LLM-backed answer API via Ollama.
Project-TechUpdates: Built a headline aggregation and classification tool using Python for scraping, TypeScript for data handling, and a minimal frontend to deliver ad-free, categorized tech news updates.

EXTRACURRICULAR ACTIVITIES

Event Leadership
Organized "Blind Code ," a sub-event of ASCENT 2022 (ISTE).
Competition Success
Won Tech-A-Thon at IIFL. Runner-up at Crack The Code (Jai Hind, 2018). Participated in Feynwick (KJSIEIT, 2021) and Trident (TSEC, 2019).
Courses
0-100 Cohort 2.0 (Harikart Singh), TensorFlow for deeplearning.