

# SURAJ CHOUHAN

surajchauhan8349@gmail.com · +91 8349123662 · LinkedIn · Indore, India

## PROFESSIONAL SUMMARY

---

Python & AI Developer with 3+ years of experience in backend development, AI/ML solutions, and scalable API-driven systems. Strong expertise in Python, LLM evaluations (SFT/RLHF), RAG pipelines, and dataset creation. Skilled at writing high-quality, tested Python code with concurrency, debugging, and async programming. Experienced in conducting model evaluations, benchmarking outputs, and providing structured feedback to improve LLM alignment.

## EDUCATION

---

- **Medi-Caps University**, Indore, MP  
M.Tech in Computer Science (Artificial Intelligence) Aug 2023– May 2025  
GPA: 9.2
- **Indore Institute of Science & Technology**, Indore, MP  
B.Tech in Computer Science Aug 2019– Jun 2023  
GPA: 8.59

## EXPERIENCE

---

**Python & AI Developer**, Gems Essence Infotech Pvt. Ltd. Remote  
Aug 2025 – Present

- Designed and implemented **LLM evaluation pipelines** (OpenAI, Hugging Face, Groq) to benchmark reasoning, factuality, and clarity.
- Created and maintained **SFT datasets** by curating high-quality code prompts, questions, and solutions.
- Conducted **RLHF evaluations**, ranking and comparing model outputs with detailed rationales for alignment.
- Developed robust **Python scripts with unit/integration tests** to automate embeddings, FAISS/Pinecone indexing, and semantic retrieval.
- Applied **multi-threading and async programming** in Python to optimize API-based inference pipelines.
- Debugged and optimized Elasticsearch queries, ensuring low-latency document retrieval.

**Associate Software Engineer**, Techinfini Solutions Pvt. Ltd. Indore, MP  
Aug 2023 – Aug 2025

- Integrated LLMs (OpenAI, Groq, Hugging Face) into business apps for chatbots, summarization, and sentiment analysis.
- Designed image-based diagnostic models with OpenCV and CNNs for disease prediction from medical imagery.
- Delivered NLP-based process automation and developed HIPAA-compliant healthcare tools.
- Automated e-commerce product feeds (Google Merchant Center, Facebook) to optimize digital marketing workflows.
- Built custom Shopify apps and fine-tuned GraphQL/REST APIs for seamless product and order synchronization.

**Software Engineering Virtual Intern**, Forage Foundation Remote  
Mar 2021 – Apr 2022

- Gained practical experience in software design and database modeling through simulated e-commerce systems.

## PROJECTS

---

**Medical Literature Summarizer** — AI Research Assistant (Jan 2025 – Apr 2025)

- Implemented RAG-based summarization using embeddings + vector stores (FAISS) to condense PubMed research articles.
- Reduced literature review time by 45% for medical researchers.

**GenAI-Powered Ad Generator** — LLM-Based Marketing Automation Tool (Oct 2024 – Dec 2024)

- Created a tool that scrapes domain URLs and generates tailored ad titles/descriptions using Gemini and diffusion models for visuals.
- Deployed a cross-platform solution using Flask (backend), HTML/CSS + JavaScript (frontend), and Docker.

**AI - Medi Boat** — Multimodal Medical Assistant

(Jul 2024 – Sep 2024)

- Engineered an assistant that handles both image and voice queries using Groq, Gradio, PyDub, PyAudio, and GTTS.
- Delivered real-time interaction through speech recognition and multimodal response generation.

**Production Data ETL Automation** — Industrial Data Pipeline

(Feb 2024 – Jun 2024)

- Designed Python + SQL ETL workflows to process and clean manufacturing sensor data.
- Reduced manual reporting time by 60% by automating feature extraction and dashboard updates.

**REST API Integration with Vector Database**

(Mar 2024 – Jun 2024)

- Developed REST APIs to serve embedding-based search queries from FAISS/Elasticsearch.
- Enabled secure enterprise knowledge retrieval system for distributed teams.

**Quality Inspection Automation using OpenCV**

(Nov 2023 – Feb 2024)

- Built a Python + OpenCV system to detect defects in assembly line images.
- Integrated results with Ignition dashboard for real-time quality alerts.

**RESEARCH**

---

**Multimodal Fusion of CT and MRI for Liver Tumor Segmentation and Classification Using Attention-Based CNNs**

Published in: *Journal of Neonatal Surgery, Vol. 14 No. 31S (2025)*

Focused on enhancing liver tumor detection through multimodal medical image fusion and deep learning using attention-based CNNs.

**TECHNICAL SKILLS**

---

- **Languages:** Python, SQL, JavaScript
- **Frameworks/Tools:** Flask, FastAPI, PyTorch, Hugging Face Transformers
- **AI/ML:** LLM Evaluations (SFT, RLHF), RAG pipelines, embeddings, vector stores (FAISS, Pinecone), NLP, Computer Vision
- **Databases:** MySQL, PostgreSQL, MongoDB, Elasticsearch
- **Python Ecosystem:** Async programming, multi-threading, unit/integration testing (pytest), debugging, modular design
- **Cloud:** AWS (EC2, S3, Lambda), Azure, GCP
- **Other:** Docker, Git, Power BI, Databricks

**ACHIEVEMENTS**

---

- Member, Google Developer Student Club (GDSC), IIST.
- Completed Walmart Global Tech Virtual Experience in Advanced Software Engineering.
- Earned certifications in AWS Machine Learning, AWS Cloud Foundations, and IBM Web Development.