

GAURAV GULATI

+91 9899559207 | gauravrunsthis@gmail.com | [Portfolio](#) | [LinkedIn](#) | [Github](#)

TECHNICAL SKILLS

Programming Languages	: Python, C++, SQL
Web Development	: HTML, CSS, Flask, FastAPI
Data Science/Machine Learning	: Langchain, Pandas, NumPy, Matplotlib, Seaborn, Pytorch
Database and ORM	: MySQL, MongoDB, PostgreSQL, SQLAlchemy
Misc	: Linux, Git, Docker, AWS, Streamlit, BeautifulSoup, Postman, Gradio

EXPERIENCE

AI Engineer Intern

Aug 2025 – Oct 2025

Sopra Steria

Noida, India

- Developed advanced **Retrieval-Augmented Generation (RAG)** architectures for domain-specific chatbots, enhancing contextual accuracy and response relevance.
- Optimized RAG pipelines using **LlamaIndex, LangChain, ChromaDB, and FAISS** for high-performance retrieval and search.
- Containerized **FastAPI**-based RAG services using **Docker** for efficient deployment and scalable inference in production environments.

Generative AI Intern

Jun 2024 – Aug 2024

Sansoftech Services Private Limited

New Delhi, India

- Completed a 2-month **Generative AI** internship at **IGDTUW** in partnership with **Sansoftech Services Pvt. Ltd.**
- Engineered a full-stack **Video Transcript Summariser and Audio Summariser** pipeline leveraging the **Gemini API**, including data extraction, chunking, summarization, and UI integration.
- Built modular experimentation pipelines with **OpenAI, Anthropic Claude, Gemini, and Hugging Face Transformers** to compare LLM behavior and output quality.

PROJECTS

MediVox: Medical AI Assistant

[[Live](#)] [[GitHub](#)]

- Developed an AI assistant integrating patient speech, medical images, and **RAG** over PDFs for context-aware responses.
- Used **FastAPI** for backend integration, **LangChain** and **Groq** for RAG pipeline and gTTS for speech processing.
- Applied **Transformers, PyTorch, and FAISS** for similarity search, and **PyPDF2** for PDF extraction and fast performance.

Social Media API

[[Live](#)] [[GitHub](#)]

- Built a scalable **REST API** with **FastAPI, PostgreSQL, SQLAlchemy, and Alembic** supporting full **CRUD, JWT auth**, and a **voting/like system**.
- Developed modular endpoints for **posts, users, auth, and votes** with secure token-based access.
- Implemented **CI/CD** via **GitHub Actions**, cutting deployment time by **80%**.

Vehicle Insurance MLOps Pipeline

[[GitHub](#)]

- Designed an end-to-end **production-grade MLOps pipeline** for vehicle insurance data using **FastAPI**.
- Integrated **MongoDB Atlas** for scalable data storage and automated model storage/retrieval using **AWS S3**.
- Implemented **CI/CD automation** using **Docker, GitHub Actions**, and deployed on **AWS EC2** through **ECR**.

EDUCATION

Degree/Certificate	Institute/Board	Grade	Graduation
B.Tech - CSE	Dronacharya College Of Engineering	76%	Ongoing
Senior Secondary	Lawrence Public School	82.8%	May 2022
Secondary	Lawrence Public School	90.8%	May 2020

CERTIFICATIONS & ACHIEVEMENTS

- Solved 200+ coding problems across LeetCode & GeeksforGeeks [[Profile](#)]
- Earned NPTEL Elite + Silver in Introduction to Large Language Models [[Badge](#)]
- IR4.0 Technologies - Programme of Microsoft, SAP, and Edunet Foundation [[Certificate](#)]
- Open-Source Models with Hugging Face - DeepLearning.ai [[Certificate](#)]