

SAURAV CHOURASIA

Bhopal, India

+91-9993465530 | saurav04042004@gmail.com | linkedin.com/in/saurav-chourasia | github.com/Saura-4

Education

VIT Bhopal University May 2027
Bachelor of Technology in Computer Science and Engineering (Cyber Security and Digital Forensics) CGPA: 8.48 / 10

Experience

FOSSEE Summer Fellowship, IIT Bombay May 2026 – July 2026
Software Engineering Intern – eSim AI Chatbot Remote | [GitHub](#)

- Redesigned the chatbot inference pipeline by extracting deterministic circuit facts before LLM inference, replacing raw netlist-to-prompt processing with structured circuit analysis.
- Built a rule-based NgSpice error analysis engine that prioritizes errors by causal dependency, enabling identification of root causes before downstream failures.
- Developed a fully offline circuit analysis pipeline that generates deterministic circuit summaries without requiring an LLM.
- Shifted the LLM's role from guessing circuit behavior to explaining deterministic circuit analysis, reducing hallucinated reasoning and improving explanation reliability.

Projects

Modular RAG Evaluation Framework
Python | Ollama | ChromaDB | Cross-Encoder Reranking [GitHub](#)

- Built a modular Retrieval-Augmented Generation (RAG) framework with interchangeable indexing, chunking, retrieval, reranking, generation, and evaluation modules for end-to-end experimentation.
- Implemented and benchmarked 4 chunking strategies and 4 retrieval pipelines on the BEIR SciFact dataset.
- Implemented 4 retrieval evaluation metrics (Recall@k, Precision@k, Mean Reciprocal Rank (MRR), and NDCG@k) from scratch to systematically compare retrieval pipelines.
- Evaluated Baseline RAG, Chain-of-Verification (CoVe), Cited-Span Generation, and Abstain Generation while developing a RAGTruth-inspired hallucination evaluation pipeline.

LLM ChatBot Backend
FastAPI | PostgreSQL | SQLAlchemy | Redis | Docker | JWT [GitHub](#) | [Demo](#)

- Designed and built a production-ready backend supporting authenticated multi-user conversations, persistent chat history, and REST APIs using FastAPI and PostgreSQL.
- Implemented JWT authentication, SQLAlchemy ORM, Alembic database migrations, conversation ownership validation, and a layered CRUD architecture for secure and maintainable backend development.
- Developed real-time streaming responses using Server-Sent Events (SSE), integrated Redis caching, containerized the backend with Docker, and deployed the backend for production use.

Snip & Ask
JavaScript | Chrome Extension | OCR | Multi-LLM APIs [GitHub](#) | [Chrome Web Store](#)

- Built and published an AI-powered Chrome extension with 540 organic users, enabling users to capture any region of the screen and receive contextual AI assistance without leaving the current webpage.
- Implemented Guest Mode through a self-hosted Cloudflare Worker using Groq's free-tier API, eliminating API-key setup and providing immediate access for new users.
- Designed Auto Mode that performs OCR before inference, routes extracted text through lightweight language models, and automatically falls back to vision models only when required to reduce cost and mitigate API rate-limit issues.

Technical Skills

Languages: Python, C++ **Backend:** FastAPI, REST APIs, SQLAlchemy, PostgreSQL, Redis, JWT Authentication
AI / LLM: RAG, Ollama, ChromaDB, Sentence Transformers, BM25, Cross-Encoder Reranking
Developer Tools: Git, Linux

Achievements

- Selected for the FOSSEE Summer Fellowship 2026, a nationally competitive open-source software fellowship conducted by IIT Bombay.
- Published Snip & Ask on the Chrome Web Store, organically reaching 540 users.