

SAURABH GAUR

AI ENGINEER & FULL-STACK DEVELOPER | LLM EVALUATION · RAG SYSTEMS · MULTI-AGENT WORKFLOWS | AZ-400 CERTIFIED

Rudrapur, Uttarakhand, India | saurabhgaur122000@gmail.com | saurabhgaur.world | [GitHub](https://github.com/saurabhgaur) | [LinkedIn](https://www.linkedin.com/in/saurabhgaur) | Open to Remote Roles

PROFESSIONAL SUMMARY

AI Engineer with a programming foundation dating back to 2016 (self-taught, starting with C++) and 1+ year of professional experience evaluating frontier LLM outputs for RLHF training pipelines at Scale AI. Independently designed and shipped 5 production AI systems — including a multi-agent hiring SaaS, a hybrid RAG document-intelligence engine, and a full-stack SSR property management platform — using LangGraph, LangChain, FastAPI, and Gemini/OpenAI APIs. Microsoft AZ-400 certified (900/1000). Seeking senior AI engineering, full-stack, or LLM-focused roles requiring deep technical ownership, end-to-end production deployment, and strong fundamentals.

EXPERIENCE

AI Code & Reasoning Evaluator | LLM Quality Analyst

May 2025 – May 2026 (1 year)

Scale AI · Freelance | Remote

- Evaluated frontier LLM outputs across code generation, reasoning, and instruction-following tasks for RLHF and preference-learning training pipelines.
- Applied structured evaluation rubrics to assess response quality, factual accuracy, code correctness, logical consistency, and safety alignment.
- Contributed to LLM quality-improvement workflows supporting large-scale model training at a leading AI data platform.

Independent AI Product Developer

May 2026 – Present (2+ months)

saurabhgaur.world · Self-Employed | Remote

- Designed and shipped 5 production AI systems end-to-end: hybrid RAG pipeline, multi-agent SaaS, SSR full-stack applications, and a hardware-monitoring daemon.
- Built deployable, revenue-generating AI products using LangGraph, LangChain, FastAPI, and Gemini/Claude APIs.
- Achieved Microsoft AZ-400 DevOps certification (900/1000) and completed GitHub Copilot Agent Mode, Azure DevOps, and Git Enterprise training.

Quality Assurance & Quality Control Engineer

Oct 2020 – Mar 2021 (6 months)

Interarch Building Products Ltd · Full-time | On-site, India

- Managed QA/QC processes for pre-engineered metal building production, ensuring compliance with structural quality standards and production timelines.

Independent Technical Development (Career Break)

Apr 2021 – Jun 2024 (3 years 3 months)

Self-Directed Study | Rudrapur, Uttarakhand, India

- Pursued a self-directed transition from mechanical engineering into software and AI engineering, building hands-on proficiency in Python, C++, and blockchain fundamentals through continuous independent study and project-based learning.
- Completed Artificial Intelligence with Python and Blockchain Basics certifications (Great Learning, Feb 2023), strengthening foundations in machine-learning workflows and distributed-ledger concepts.
- Maintained continuous programming practice since 2016 (starting with C++), building a cumulative technical foundation that enabled the transition into professional AI evaluation and product development roles.

PROJECTS

HireOS — AI Resume Intelligence Platform — Python, FastAPI, LangGraph, Gemini 2.0 Flash, Razorpay, Vercel, Railway

Links: [Live Demo](#) · [GitHub Repo](#)

- Architected a multi-agent LangGraph supervisor workflow orchestrating resume analysis, ATS keyword injection, and real-time AI coaching agents; deployed as a freemium SaaS supporting 500 concurrent users.
- Built an async FastAPI backend with rate limiting and security hardening; integrated Razorpay payments and deployed via Vercel (frontend) plus Railway (4-worker backend infrastructure).
- Developed during Microsoft Agents League Hackathon 2026, implementing circuit breakers, state rollback, and an observability dashboard.

RAGForge — Intelligent Multi-Document Q&A System — Python, LangChain, ChromaDB, BM25, Gemini API,

OpenAI, Streamlit, Docker, Pytest

Links: [GitHub Repo](#)

- Built a production-grade RAG pipeline with hybrid BM25 + ChromaDB vector search via LangChain's EnsembleRetriever, LLM-driven query expansion, post-generation hallucination validation, and a keyword-overlap reranker.
- Implemented multi-turn conversational memory, JSON session persistence, live token-cost tracking (tiktoken), and exact source citations; shipped with a full Pytest suite and Docker deployment with health checks.

MakaanBook — Smart Rent & Electricity Bill Manager — TypeScript, TanStack Start, Supabase, PostgreSQL,

Gemini 1.5 Flash, Tailwind CSS v4, Vercel

Links: [Live Demo](#) · [GitHub Repo](#)

- Engineered a production SSR SaaS with Supabase PostgreSQL (Row-Level Security), parallelized data fetching via Promise.all across 8+ pages, and Vite vendor code splitting for optimized load performance.
- Built a Hinglish AI assistant with live database context injection, fault-tolerant split-meter electricity billing, jsPDF monthly statement export, and full WCAG accessibility compliance.

KundaliAI — Vedic Astrology × AI Readings — TypeScript, TanStack Start, Supabase, Gemini AI, Tailwind CSS, Zod

Links: [GitHub Repo](#)

- Built a Vedic astrology planetary-geometry computation engine integrated with Gemini AI to generate personalized chart readings from real birth date, time, and location data.
- Designed a custom CSS design system and implemented Supabase Auth with server-rendered route protection in a full-stack SSR TypeScript application.

LaptopPulse — AI Predictive Hardware Monitor (Beta) — Python, Gemini API, Claude API, psutil, WMI, Windows

ACPI, Flask, SQLite

Links: [GitHub Repo](#)

- Developed a silent Windows daemon with <0.3% CPU footprint using trend-based anomaly detection to predict hardware failures 30–60 days in advance.
- Integrated Gemini/Claude AI for on-device, plain-language health reports with a 0–100 composite health score; local web dashboard at localhost:5747 with zero cloud dependency.

SKILLS

AI / ML: LLM Evaluation, RLHF, Prompt Engineering, RAG Systems, Multi-Agent Systems, LangChain, LangGraph, ChromaDB, BM25, EnsembleRetriever, HuggingFace Transformers, Gemini API, OpenAI API, Claude API, vLLM, GPU Optimization for LLM Inference

Languages: Python, TypeScript, JavaScript, C++, SQL

Backend: FastAPI, REST APIs, Supabase, PostgreSQL, Row-Level Security, Async Programming, Zod, Flask

Frontend: React, TanStack Start (SSR), TanStack Query v5, Vite, Tailwind CSS v4, Streamlit, Framer Motion, Radix UI (shadcn/ui)

DevOps / Cloud: Docker, Vercel, Railway, GitHub Actions, Microsoft Azure, Azure DevOps, Azure Repos, GitHub Copilot Agent Mode

Tools & Libraries: tiktoken, Pytest, jsPDF, Razorpay, psutil, WMI

EDUCATION

Bachelor of Technology — Computer Science & Engineering

Aug 2020 – May 2025

Dr. A.P.J. Abdul Kalam Technical University (AKTU), Lucknow | CGPA: 6.78

Diploma — Mechanical Engineering

Aug 2016 – Jun 2020

Board of Technical Education, Uttar Pradesh (BTEUP) | CGPA: 7.8

CERTIFICATIONS

AZ-400: Designing & Implementing Microsoft DevOps Solutions — Microsoft | **Score: 900/1000**

Jun 2026

AZ-104: Prerequisites for Azure Administrators — Microsoft

Jun 2026

GitHub Copilot Fundamentals — Parts 1 & 2 — Microsoft Learn

Jun 2026

AI Skills Fest 2026 — Microsoft

Jun 2026

Microsoft AI Product Manager Professional Certificate (5 courses) — Microsoft / Coursera

Jun 2026

GPU Optimization for LLM Inference — AMD AI Academy

Jun 2026

Hugging Face on AMD — AMD AI Academy

Jun 2026

Google AI Essentials (7-course bundle) — Google / Coursera

May 2026

AI Fundamentals — Google / Coursera

May 2026

AI for Brainstorming and Planning — Google / Coursera

May 2026

AI for Research and Insights — Google / Coursera

May 2026

Artificial Intelligence with Python — Great Learning

Feb 2023

Blockchain Basics — Great Learning

Feb 2023

ACHIEVEMENTS

- Microsoft Agents League Hackathon 2026 — Built HireOS, a multi-agent hiring automation system (LangGraph + Claude API) with circuit breakers, state rollback, and an observability dashboard.
- AZ-400 DevOps Certification — Scored 900/1000 (passing threshold: 700) on first attempt, June 2026.
- Microsoft AI Skills Fest 2026 — Awarded digital badge for AI applications and AI ethics.
- 5 Production AI Systems Shipped — End-to-end delivery spanning RAG, multi-agent SaaS, SSR full-stack, and system-level AI products.