

Vivek Jadav

jadavvivek0501@gmail.com • vivekjadav.com • [linkedin](#) • [GitHub](#) • Ahmedabad, India

EDUCATION

Government Engineering College, Bhavnagar

May 2022 – April 2026

B.Tech, Information Technology | CGPA: 8.51/10 | DSA, DBMS, Linear Algebra, Probability

EXPERIENCE

AI Engineer — AST Consulting

May 2026 – Present

- Developed AI-powered content creation workflows within a CMS platform used by media organizations and digital creators.
- Built and integrated LLM-driven features for article generation, story creation, SEO metadata generation, and content automation.
- Collaborated with product and engineering teams to deploy AI features into production workflows.
- Led the platform's internationalization effort, implementing multilingual support across the entire application
- Built an asynchronous image optimization pipeline for the news platform, improving page load performance

Full-Stack Developer Intern — AUM Imagineering

Jan 2026 – April 2026

- Engineered a production Document Management System end-to-end — architected document upload, versioned storage, retrieval, and role-based access control across frontend, backend, and database layers.
- Shipped core features using Next.js, Node.js, and MongoDB; owned the full development lifecycle from schema design to deployment.

PROJECTS

Scientific Co-Pilot — Autonomous Research Agent | Python, FastAPI, LangGraph, pgvector, Docker 2025 – Present ([Live](#))

- Built an AI research platform for literature discovery, hypothesis generation, and multi-agent debate using Next.js, FastAPI, PostgreSQL, and Ollama.
- Deployed a public demo on Cloudflare Pages; validated with 218 automated tests.

Memorization vs. Generalization in Small Transformers | Python, PyTorch

- Trained a 50M-parameter decoder-only transformer from scratch on algorithmic tasks (addition, sorting, reversing); designed controlled experiments with data duplication and output-only loss masking to isolate memorization behavior.
- Built custom evaluation suite — n-gram overlap, likelihood divergence, exact-match eval — revealing near-zero OOD accuracy despite strong in-distribution performance, quantified generalization failure modes.

SQL Performance Tuner | TypeScript, Next.js, SQLite, Lamatic.ai ([Live](#))

- Built an agent-driven SQL tuning application that tests index suggestions and query rewrites in isolated SQLite copies, compares results, and benchmarks performance.
- Integrated Lamatic strategist and reviewer workflows; contributed the kit through a merged PR to Lamatic's AgentKit repository.

AI-Powered Multimodal Study Assistant | Python, Whisper, Llama 3.2, LoRA

- Built a voice-interactive AI tutor combining Whisper ASR and Llama 3.2; deployed RAG pipeline over 4,800+ textbook pages with LoRA-fine-tuned TinyLlama-1.1B achieving 80% MCQ relevance.
- Awarded Best Technical Solution at college hackathon (1st of 65 teams).

TECHNICAL SKILLS

AI & ML: PyTorch, HuggingFace Transformers, LangChain, LangGraph, RAG, LoRA, Whisper, FAISS, Ollama

Backend & Web: Python, FastAPI, Flask, Node.js, Next.js 14, TypeScript, React, Tailwind CSS

Databases & DevOps: MongoDB, PostgreSQL, pgvector, Prisma ORM, Docker, Git

OPEN-SOURCE CONTRIBUTION

Lamatic AgentKit — SQL Performance Tuner | [Merged PR #391](#)

- Contributed a SQL Performance Tuner built with Next.js, TypeScript, and SQLite, using Lamatic-powered AI agents to evaluate query optimization strategies.

