

VIJAYSHREE VAIBHAV

Portfolio: vijayshree.xyz

GitHub: github.com/vijayshreepathak

Email: vijayshree9646@gmail.com

Mobile: +91 7667627647

LinkedIn: linkedin.com/in/vijayshreevaibhav

EDUCATION

Thapar Institute of Engineering and Technology

B.Tech in Computer Science

Patiala, India

Aug 2023 – May 2027

EXPERIENCE

Rimba (YC X25)

Software Engineer Intern

Remote

July 2026 – Present

- Built **Ask Rimba**, a multi-turn, context-aware AI agent using **Model Context Protocol (MCP)** and read-only data tooling, enabling natural-language queries over customer data with tenant-aware SQL validation, schema-constrained execution, persistent context, citations, and follow-up workflows.
- Improved agent reliability and latency by diagnosing schema-hallucination and tool-call failures and implementing **rejection-aware retry handling** that reuses the permitted schema instead of blindly retrying, reducing failed turns without weakening data-access boundaries; hardened the runtime with MCP readiness checks, **fail-closed routing**, fallback handling, and structured tool-lifecycle logging.
- Improved **Composer V3** performance and UX: refined chart rendering/loading behavior and readability, added per-chart date-range selectors with explicit Apply controls, and fixed reporting/export issues including Excel decimal formatting.
- Built, deployed, and stabilized an internal **analytics dashboard** (fixing environment/API routing issues), and delivered an end-to-end **45Z GREET ethanol calculator** demo, translating domain requirements into a customer-facing prototype for stakeholder demonstrations.
- Strengthened release quality via automated validation — **100+ MCP tests** and **120+ Ask Rimba backend tests**, plus frontend type-checking, production-build verification, and zero ESLint errors.

Airmeet

SDE & AI Intern

Bengaluru, India

Sept 2025 – March 2026

- Built **PromptLab**, a multi-LLM orchestration and evaluation platform executing prompts across provider APIs and comparing outputs via embeddings and cosine similarity; engineered async **FastAPI** + **asyncio** pipelines for concurrent multi-model requests, plus reusable prompt templating and context orchestration for consistent structured outputs.
- Developed **MCP** server pipelines for AI content automation and auto-posting across WordPress and external publishing systems; worked with the **Claude SDK** and Crawllee-based agents to extract structured website data, generate insights, and automate workflows.
- Built **Playwright** automation for workflow validation and regression testing; contributed full-stack features in **Next.js**, **React**, **Node.js/Express**, and **PostgreSQL**, with structured logging, retries, and rate limiting for release stability.

Snap Inc.

Software Engineering Extern

USA (Remote)

Sept 2024 – Nov 2024

- Developed scalable backend features impacting 500k+ users and improving system reliability.

TECHNICAL SKILLS

- **Languages:** Python, JavaScript, TypeScript, C++, SQL
- **AI Agents & LLM Systems:** Model Context Protocol (MCP), AI agents, tool/function calling, multi-turn context management, RAG, embeddings & vector search, prompt & context engineering, structured outputs, guardrails, LLM evals, agent observability
- **LLM Platforms & AI Tooling:** Claude SDK / Claude Code, Anthropic API, OpenAI API, Codex, Cursor, LangChain, Hugging Face, Streamlit
- **ML/Data:** PyTorch, Scikit-Learn, XGBoost, Pandas, NumPy, Explainable AI (SHAP), Anomaly Detection, Ensemble Learning
- **Backend:** FastAPI, Node.js/Express, REST APIs, WebSockets, asyncio, PostgreSQL, MongoDB, Redis
- **Frontend:** React, Next.js, TypeScript, Tailwind CSS, Tiptap
- **DevOps & Testing:** Docker, CI/CD, Git, AWS, Playwright, Pytest/Jest, structured logging, retries, rate limiting

PROJECTS

- **AI Meeting Intelligence Copilot** — AI system converting meeting transcripts into summaries, action items, risks, and follow-ups; Slack-integrated for real-time automation.
- **Conversational Browser Agent** — AI agent executing browser tasks from natural language via LLM tool-calling + Playwright.
- **QuantAlert** — **Real-time Alerts** — Built a real-time alerting system using FastAPI, WebSockets, and PostgreSQL.

CERTIFICATIONS

- MCP (Model Context Protocol) · OpenAI API: Agents · Intro to LLMs · NVIDIA: Deep Learning · CI/CD with Docker

ACHIEVEMENTS

- Hack Finance IIIT Delhi Winner · 7× National Hackathon Winner · National Winner, Startupthon (AIR 1) · Judge — Hackathon, Canadian University Dubai