

Pranav Sharma

New Delhi, India | Email: pranavsharmaa1908@gmail.com | GitHub: github.com/pranavsharmaa1908

Summary

I'm Pranav, a developer specializing in **automation, AI integration, and system development**. I build **fast, privacy-first solutions** that simplify workflows, enhance productivity, and scale efficiently. My work spans **Python scripting, AI-driven automation, intelligent agents, and containerized deployments**, combining tools like **LangChain, Hugging Face, OpenAI API, Gemini, Docker, and Linux** to deliver lightweight, maintainable, and user-friendly systems.

Skills

Programming & Automation: Python, Bash, Selenium, Streamlit

AI & ML Tools: LangChain, Hugging Face, OpenAI API, Gemini, BLIP2, FAISS, Llama.cpp

DevOps & Systems: Docker, Docker Compose, Arch Linux, Linux CLI, GCP

Other: Git, Prompt Engineering, OCR, Data Processing (NumPy, Pandas)

Projects

PlatoAI – Philosophy-Based Life Coach (RAG Pipeline)

July 2025 – August 2025 | Built an **end-to-end Retrieval-Augmented Generation (RAG) pipeline** using LangChain and FAISS.

- Implemented **sentence embedding retrieval** with Sentence Transformers.
- Developed a **React + Tailwind frontend** and **Express backend** with JWT authentication.
- Added **offline LLM inference** with Llama.cpp for privacy-first processing.

Minisite Automation Tool – Large Scale Workflow Automation

August 2025 – September 2025 | **Reduced minisite creation time from 12–14 hours to just 30 minutes** per site.

- Automated design, deployment, and content integration using **Python and Selenium**.
- Streamlined deployment through **Dockerized environments** for scalable workflows.

Project Alfred – Offline LLM Assistant (Android)

June 2025 – Present | **Kotlin-based AI assistant** for Android enabling **offline LLM inference** with Llama.cpp.

- Modular design for tasks like **summarization, reminders, translation** without cloud dependency.
- Optimized performance using Llama.cpp's Java bindings.

Education

Vivekananda Institute of Professional Studies – BCA

May 2022 – June 2025 | *Delhi, India*

Core focus: Systems Design, AI, Web Development

Capstone: Lightweight LLM Assistant for Android

Certifications

Machine Learning Specialization – Stanford Online (Coursera)

Instructor: Andrew Ng | Completed: March 2023

Topics: Supervised Learning, Neural Nets, Recommenders, Reinforcement Learning