Jaydeep Solanki

AI & Backend Engineer | Python, LLM Systems

Bangalore, India · contactjaydeepsolanki@gmail.com · jaydeepsolanki.vercel.app · solanki-jaydeep · whoisjayd

Summary

Python engineer focused on LLM-powered tools and the backend systems around them. I build, publish and maintain my own software: two packages on PyPI, an IoT backend, and flight and vision software for national and international robotics competitions.

Projects

NoteWise: LLM study-notes CLI GitHub | Docs | PyPI

- Built and maintain an open-source async Python CLI that turns YouTube lectures and playlists into structured Markdown, HTML, PDF and DOCX notes with quizzes. 16 releases since March 2026 (v1.0 to v1.7), 258 commits, 10 stars and 13 forks.
- Solved long-input limits by splitting transcripts into 4,000-token chunks with 200-token overlap, then stitching them at heading and code-block boundaries so the output reads as one document.
- Routes requests across LLM providers (Gemini, Claude, OpenAI, Groq, local Ollama) through LiteLLM, with exponential backoff and jitter, token and cost tracking, and a SQLite cache that skips repeat transcript fetches.
- Ships like a product: typed codebase, tests with CI and coverage tracking, docs site, standalone installers and a Docker image.

GitResume: repository-grounded resume assistant Project page

- Built an AI tool that takes a GitHub repository and a job description and produces resume bullets, project summaries and interview questions grounded in the code, without inventing achievements.
- Used Tree-sitter to extract structure and stack from real code instead of keyword matching; streamed long analyses over WebSockets, added GitHub OAuth for private repos, Redis for session state and swappable model providers.

BlazeServe: high-throughput HTTP file server GitHub | PyPI

- Built and published a Python HTTP/1.1 file server that keeps memory flat on multi-GB files: zero-copy sendfile for full downloads, 256 MB sliding-window mmap for range requests and a bounded buffered fallback.
- Implemented the protocol details instead of skipping them: resumable single and multi-range downloads (RFC 7233), conditional caching with ETag and 304 responses, and per-IP token-bucket rate limiting with 429 responses.
- Added Prometheus metrics plus built-in doctor and benchmark commands; checked with Ruff, strict mypy and pytest.

Experience

Deloitte · *Engineering Intern, Data & Backend* Jan 2026 – Present

- Built Python ETL pipelines covering ingestion, transformation and validation during the data engineering phase (Jan-Jun 2026), alongside SQL and analytics training.
- On-site in Bangalore since July 2026 in a structured backend program focused on Python API development and data services.

Ellipsis Infotech · *Project Intern* May 2025 – Jun 2025

- Co-built an IoT platform in a 2-person team for centralized control of multi-brand AC units, using ESP8266 devices, FastAPI, PostgreSQL, MQTT, WebSockets and a Next.js dashboard.
- Wrote the async Python backend: device provisioning, JWT/device auth, MQTT command routing, live state updates, telemetry history.
- Deployed on Google Cloud Run; hardened device connectivity with MQTT over TLS, reconnect with backoff and OTA updates.

Team Dyaus, Nirma University · *Co-Founder & Software Lead* Nov 2022 – Jan 2026

- Co-founded a student aerospace and robotics team and led its software. Results: 2nd at IN-SPACe CanSat India 2024, 9th internationally at TEKNOFEST 2024, Flipkart GRID 5.0 national finals. Mentored 20+ members whose teams placed 7th globally (1st in Asia) at TEKNOFEST 2025.
- Built Python ground-station and flight software: live telemetry and command/control over serial and WebSocket, concurrent sensor workers on a Raspberry Pi, fault detection and black-box logging.
- Built the vision stack for an autonomous pick-and-place robot: custom YOLO object detection, Intel RealSense depth-based 3D localization and inverse kinematics.

Skills

Languages: Python, SQL, Go (basic), Bash

LLM & AI: LiteLLM, multi-provider routing (Gemini, Claude, OpenAI, Groq, Ollama), long-document chunking, token and cost tracking, Tree-sitter code parsing, YOLO object detection

Backend & Data: FastAPI, Flask, Django REST Framework, AsyncIO, WebSockets, MQTT, JWT and OAuth, PostgreSQL, Redis, SQLite, MongoDB, ETL pipelines

Tools: Git, Docker, Linux, pytest, Ruff, mypy, Prometheus, Grafana, GCP, AWS

Education

Nirma University · B.Tech in Electronics and Communication Engineering, Minor in Data Science Aug 2022 – May 2026

CGPA: 8.0/10.0