

SMIT SOJITRA

AI Engineer · Bangalore, India

+91 95746 26040 | smit.sojitra.dev@gmail.com | smit.in
github.com/smit153 | linkedin.com/in/smit-sojitra | dev.to/smit153

Summary

Software developer with 2+ years building full-stack products and, more and more, the AI systems inside them: LLM features, agent pipelines and the APIs, data and infrastructure around them. I own work from architecture to deployment, and I write about how LLMs behave in production at smit.in/writing.

Experience

Updot

Jun 2024 – Present

Software Developer

Bangalore, India

- Lead **5+ client products** end to end, from scoping and system architecture through launch and post-launch fixes, working directly with designers and stakeholders.
- Architect and scale a client platform to **10k+ active users**, owning the APIs, data model and infrastructure behind it.
- Build **RAG chatbots and AI search** over client content, using embeddings and retrieval to answer questions and recommend content from each client's own data.
- Ship **LLM document-extraction pipelines** that turn PDFs and forms into validated structured data, replacing manual data entry.
- Build **agent-based automations** for repetitive client workflows, saving teams **5+ hours a week** of manual work.
- Set up headless CMSs (**Payload, Strapi, Hygraph, Sanity**) so client teams publish without a developer, and own CI/CD, releases and page-speed tuning after launch.

Projects

BeyondDiff | *Python, FastAPI, Neo4j, LangGraph, tree-sitter, SCIP, OpenTelemetry* github.com/smit153/beyonddiff

- GitHub App that flags what a PR left unfinished *outside* the diff (breaking callers, stale docs, companion files, tests worth running) in one comment per PR.
- Keeps a **Neo4j code graph** current on every push via SCIP with a tree-sitter fallback; the static graph holds **70%** of the calls Click's test suite really made.
- Runs a **six-stage LangGraph agent pipeline** with a skeptic agent and mechanical checks on every citation, at roughly **50–150k tokens per PR**; checkpoints resume crashed workers.
- Backed by **225 pytest tests** against a real Neo4j, including fixture repos with every expected finding and planted traps written down in advance.

ModelMux | *Python, FastAPI, asyncio, Pydantic, Docker, JSON Schema, Prometheus* github.com/smit153/modelmux

- Self-hosted **OpenAI-compatible Chat Completions API** over **2 coding CLIs** (Claude Code, Codex) with SSE streaming, so OpenAI SDK, LangChain and LiteLLM apps plug in unchanged.
- Validates **tool calls and JSON Schema output** (strict mode), with **1 bounded corrective retry**.
- Sandboxes each request in a private workspace and supervised process group, and kills it on any execution event, with queue limits, output caps and idle/total timeouts.
- CI enforces **90%+ test coverage** on every server layer and tests the CLI on **3 OSes** (Linux, macOS, Windows).

Technical Skills

AI / LLM: LangGraph, LangChain, Vercel AI SDK, MCP, RAG, embeddings, structured outputs, evals, multi-agent systems, guardrails, LLM observability (OpenTelemetry, Phoenix)

Languages: TypeScript, Python, JavaScript, SQL, Go

Backend: Node.js, FastAPI, Express, NestJS, Hono, GraphQL, tRPC, Pydantic, Socket.IO, REST

Frontend: React, Next.js, React Native, Tailwind CSS, Astro

Data & Cloud: PostgreSQL, MongoDB, Redis, Neo4j, Pinecone, Docker, Kubernetes, AWS, GCP, GitHub Actions

CMS: Payload, Strapi, Hygraph, Sanity

Education

Amity University Jaipur

B.Tech in Computer Science

2020 – 2024

Jaipur, India