

MANAS CHOUDHARY

Seeking a Software Engineering Internship, Summer 2027 (May – July 2027)

Backend and AI-infrastructure engineer. At Atlan I build the platform that runs AI agents for 16+ engineering teams: code review, reliability and LLM cost. Five internships and one startup, all alongside full-time study.

EXPERIENCE

Atlan (Series C, \$80M ARR) Jan 2026 – Present
Associate Software Engineer, Intern Remote

Internal AI-agent platform used by 16+ engineering teams. Python, FastAPI, PostgreSQL, Kubernetes.

- **Security code review:** built Atlan's AI code-review service end to end. It is on by default across about 300 repositories, has run **14,500+ security reviews**, and blocks merges with high or critical issues.
- **Reliability:** raised the platform's success rate **from about 50% to 90%+** by fixing the root cause of 10 bugs, then led a reliability sprint that cut failed runs from about 75 a week to about 5 a day.
- **Support automation:** built the integration that hands each Zendesk support ticket to an AI agent, which investigates and posts its findings back on the ticket. Demoed it to the whole company.
- Showed the platform's success dashboard was undercounting by 2 to 3x, and added automatic retries that now rescue **about 100 runs a week**.
- Removed agent start-up delays of up to 9 minutes with a pool of pre-warmed containers, and halved worst-case crash recovery from 14 to 7 minutes.
- **Cost:** traced **72% of LLM spend** across 45 agents to one pipeline, and found why prompt caching failed for every model behind the gateway (0.1% hit rate against 92.5%).
- A streaming component I built was adopted by another team as their reference implementation.

Samsung R&D Institute India Jul 2025 – Jan 2026
R&D Intern, AI and Design Automation Remote

- Built an AI agent that turns Figma designs into React and TypeScript code (Claude Agent SDK, FastAPI, Playwright).
- Wrote 4 custom **MCP tools** over a ChromaDB component library, so the agent decides when to reuse existing code instead of writing new code. GitHub and Vercel MCP give one-click deploys.
- Each build is rendered in a headless browser and scored against the design, then fixed and retried until it passes.

IIT Roorkee, VITAL Lab Feb 2026 – Jul 2026
Research Intern, LLM Bias Remote

- **Finding bias:** built a MAP-Elites search (Python, LangGraph) that keeps the most biased prompt for every bias-category and context pair. It found **192 bias categories**, against 9 in the standard BBQ benchmark.
- **Mitigating bias:** trained a small adapter per bias category and fused them (AdapterFusion). Testing showed most of the measured drop came from the model refusing to answer, not from less stereotyping, which became the basis of a paper draft.

IIT Jodhpur May 2025 – Jul 2025
Summer Research Intern, Robotics Jodhpur, India

- Built a 4-node ROS2 robot stack in simulation: object detection, tracking, depth-camera 3D localisation and navigation.

Tribal Research and Training Institute Feb 2025 – May 2025
Web Developer Intern Govt. of Chhattisgarh

- Built the Tribal Freedom Fighters Museum website (tffm.cg.gov.in) with React and Node.js. **Inaugurated by the Prime Minister of India.**

Sigmoyd Mar 2025 – Feb 2026
Co-founder Raipur, India

- Co-founded an AI automation startup (sigmoyd.in) with **400+ customers**. Built its website, docs site and mobile app (Next.js, React Native) and a voice-agent prototype on Twilio.

TECHNICAL SKILLS

Languages: Python, C++, TypeScript, JavaScript, Rust, SQL. **Backend and Cloud:** FastAPI, Node.js, PostgreSQL, pgvector, TimescaleDB, Redis, Docker, Kubernetes, AWS, Cloudflare, Linux.
CS: data structures, algorithms, operating systems, object-oriented design, distributed systems, concurrency, agile. **AI:** LLM agents, LangGraph, RAG, vLLM.

PROJECTS

AgentOS Rust, Python, TypeScript
manasdev.codes/projects/agent-os

Problem: every AI agent needs model serving, memory, safety and code execution, and teams rebuild them for each agent. AgentOS gives any agent all four through one URL change, on hardware you own.

- **Gateway:** Govrix Scout (below) governs every call the agent makes.
- **Inference** (vLLM): priority classes with preemption and fractional-GPU placement. Two vLLM engines share one RTX 5090.
- **Memory:** four tiers, from the context window to pgvector, Postgres facts and a TimescaleDB hypertable archive. Keyword and vector search fused by rank (RRF); benchmarks picked Postgres over Qdrant and Neo4j.
- **Sandbox:** any chat agent can run code; the gateway scans it, then runs it in a locked-down Docker container.

ChunkScope Python, TypeScript
github.com/manaspros/chunkscope

Problem: choosing how to chunk, retrieve and rerank documents for RAG is guesswork.

- Profiles each document on about 20 signals and recommends a pipeline, weighting techniques by published research and explaining the rejected ones.
- LLM-as-judge evaluation written from scratch: faithfulness by claim decomposition and entailment, relevancy, context precision and recall, MRR.
- Visual builder (FastAPI, React Flow, pgvector with HNSW) that exports a runnable Python and Docker project.

Govrix Scout 70+ stars, 3 forks | Rust
github.com/manaspros/govrix-scout | govrix.dev

Problem: nobody knows which agent called which model, what it cost, or whether it leaked data.

- One door every agent request crosses: the meter, the guard and the flight recorder. Change one base URL, no SDK (Rust, Tokio, hyper).
- Per-agent kill switch, PII checks and a SHA-256 hash-chained audit log, written through a bounded queue so the database never sits on the request path.
- Carried **383,000 requests with zero errors**; the GPU, not the proxy, was the bottleneck.

Google DeepMind Concordia 1.7k stars | Python
github.com/google-deepmind/concordia

Problem: memory lookups in DeepMind's agent-simulation library slowed down sharply as memories grew.

- 7 changes merged, including a memory-store rewrite from quadratic to linear time (10s to 0.1s for 1,000 records).

EDUCATION

IIIT Naya Raipur Aug 2024 – Apr 2028
B.Tech, Data Science and AI Class XII: 93%

ACHIEVEMENTS

- **Rank 9**, Amazon ML Challenge 2026 (public leaderboard), entity resolution.
- **1st Place**, Hack-a-Sol 2025 AI/ML Track. 2nd Place, E-Summit Hackathon.