

Chirag Singhal

Software Engineer — SAP CPQ / Lead-to-Cash · Backend, Full-Stack & Applied AI

Ghaziabad, India · +91 74284-49707 · chirag@oriz.in · github.com/chirag127 · linkedin.com/in/chirag127 · oriz.in

SUMMARY

Software engineer with 3+ years across backend, full-stack, and applied AI, currently on SAP's Lead-to-Cash (CPQ) platform at Tata Consultancy Services, where I keep enterprise quoting systems running and build AI tooling that speeds real work. Alongside my day job I design and ship oriz.in — a self-run family of free, privacy-first web tools, live-data APIs, and open-source Model Context Protocol (MCP) servers used by AI agents — which keeps me fluent end-to-end, from backend architecture and data pipelines to frontend, DevOps, and the AI layer that ties it together. A JEE Advanced 2020 All India Rank 11870 (top 1%) and college topper, I move fast on hard problems and care about shipping. My vision is to build AI-native products that make powerful tools free and effortless for everyone — the way oriz.in already does at small scale.

EXPERIENCE

3+ years shipping backend, full-stack, and applied-AI systems — from enterprise SAP CPQ at TCS to a self-run family of production web tools and open-source libraries.

Software Engineer — Tata Consultancy Services

Bhubaneswar · Jun 2025 – Present

- Support the SAP CPQ (Configure-Price-Quote) Lead-to-Cash platform across 20+ functional areas — Bill of Materials & Pricing, Opportunity Management, Quote Info & Terms, Deal Reviews & Approvals, Provisioning, Contracting, OBV (Order-Based Value), and Public Cloud Order Management.
- Built a hybrid-RAG search and documentation system over the CPQ onboarding knowledge base (100+ videos, 40+ decks) — dense embeddings + BM25 + reranking on SAP-approved Hyperspace AI — that cuts ticket-resolution and onboarding time; currently completing the Cloud Quote KT knowledge base.
- Automate pricing and configuration validation, testing, and CI/CD workflows in Python; build React dashboards giving real-time visibility into backend configuration state.
- Diagnose and resolve configuration, pricing, and quote-lifecycle issues across a large enterprise codebase in an Agile team.

Software Developer — Full Stack — QRsay.com

Remote · Jul 2023 – May 2025

- Built and maintained a high-traffic food-commerce platform end to end with Python, Node.js, and MongoDB.
- Designed an event-driven order-processing pipeline on Apache Kafka to synchronise orders across distributed outlets without data loss under peak load.
- Integrated a Razorpay payment gateway with HMAC-verified webhooks and secure authentication flows.
- Cut API response times through query optimisation, indexing, and Redis caching; built modular React frontend components.

OPEN-SOURCE PROJECTS & PRODUCTS

oriz.in · flagship · Astro · React · TypeScript · Cloudflare

oriz.in

- A growing family of free, privacy-first, client-side web tools and live-data APIs; all compute runs in-browser or on Cloudflare Workers — no server needed for end users. Publishes reusable @chirag127/* npm packages (design-system atoms, keyless multi-provider LLM client) that power every site.

Clear-Thought MCP Server · 36★ · TypeScript

chirag127/Clear-Thought-MCP-server

- Model Context Protocol server exposing 11 structured-thinking tools (sequential thinking, mental models, design patterns, programming paradigms, debugging approaches, collaborative reasoning, decision frameworks, metacognitive monitoring, scientific method, structured argumentation, visual reasoning) to MCP-compatible AI clients; a lean, typed toolset built on the MCP SDK.

bookmark-mind · 10★ · JavaScript

chirag127/bookmark-mind-bs-ext

- Chrome/Firefox extension that auto-categorises bookmarks with an LLM and offers AI chat + multi-model comparison over your saved links. Bring-your-own-key over any OpenAI-compatible endpoint (Groq, OpenRouter, Gemini, Mistral, local LM Studio / Ollama) through one uniform provider adapter — keys stored client-side, plus user-added custom providers.

OmniDistribute · TypeScript

chirag127/OmniDistribute

- Resilient, idempotent content-distribution engine that publishes a single Markdown source to 30+ platforms; adapter pattern with circuit-breaker + retry for at-least-once delivery.

agent-forge · Python · multi-agent · evals

[chirag127/agent-forge](#)

- Multi-agent LLM orchestrator — planner/executor/critic architecture with a pluggable tool system (JSON-schema-validated function calling), provider failover, and an LLM-as-judge eval harness; OpenTelemetry-instrumented, 67 tests.

rag-lens · Python · RAG · vector search

[chirag127/rag-lens](#)

- Production hybrid-RAG service: dense embeddings + BM25 → RRF fusion → cross-encoder reranking over a vector DB (LanceDB), exposed via FastAPI + Docker; retrieval eval suite reports Recall@k / MRR / nDCG, OpenTelemetry spans throughout.

deploy-kit / go-vault · Kubernetes · Terraform · Go

[deploy-kitgo-vault](#)

- **deploy-kit:** reference GitOps deploy — Kustomize + Helm manifests (HPA, probes, NetworkPolicy), Terraform IaC, Argo CD, Prometheus/Grafana. **go-vault:** Go gRPC microservice (Protocol Buffers) with Postgres (pgx), Redis read-through cache + rate limiting, Prometheus metrics, graceful shutdown.
- **Also:** Stochastic-Thinking MCP Server (probabilistic reasoning) · YouTube AI Navigator (Gemini chapter summaries) · olivia (Python voice assistant with fully offline speech recognition via local faster-whisper) · sponsorblock-ai (ML sponsor detection) — github.com/chirag127

SKILLS

Languages: Python (advanced), TypeScript, JavaScript, Go, SQL, Bash

Backend: Node.js, FastAPI, Flask, Express.js, REST + gRPC APIs, Protocol Buffers, microservices, event-driven architecture (Apache Kafka), authentication (OAuth2)

Frontend: React, Astro, Tailwind CSS, HTML5/CSS3, PWAs, browser extensions (Chrome/Firefox)

AI / LLM: MCP (Model Context Protocol) server development, multi-agent orchestration (planner/executor/critic, tool-calling), RAG pipelines (dense + BM25 + RRF fusion + cross-encoder reranking), vector search (LanceDB), LLM evals (Recall@k / MRR / nDCG, LLM-as-judge), multi-provider LLM integration (Gemini, Groq, Mistral, NVIDIA NIM, Cohere), prompt engineering, web scraping (Playwright / httpx / BeautifulSoup)

Databases & Storage: PostgreSQL (Neon / Supabase / pgx), Turso (SQLite edge), Redis / Upstash, MongoDB

Cloud & DevOps: Docker, Kubernetes (Kustomize, Helm, HPA), Terraform / IaC, Argo CD / GitOps, Prometheus / Grafana, OpenTelemetry, Cloudflare Workers & Pages, Firebase, GitHub Actions CI/CD, Linux, cron-scheduled pipelines

Payments & Integrations: Razorpay (orders, webhooks, HMAC verification), Clerk auth, sops + age secrets

Practices: System design (HLD/LLD), distributed systems, OOP & SOLID, TDD, unit + integration testing, observability, Agile, Git

QUALIFICATION

B.Tech, Computer Science & Engineering — AKTU (A.P.J. Abdul Kalam Tech. Uni- Uttar Pradesh · 2020 – 2024
versity)

- College Topper (Rank 1).
- Senior Secondary (CBSE) — 97%.

Honors & certifications:

- **JEE Advanced 2020** — All India Rank 11870 (top 1%, 1M+ aspirants).
- **Claude Certified Associate — Foundations**, Anthropic (2026).
- 20+ Udemy Business courses (via TCS) across AI/LLM (Claude Code, MCP, Generative AI), Machine Learning, Python, EU AI Act compliance, and Agile.