

Rishav Raj

✉ rairishav221@gmail.com | 📞 +91-9798600997 | 📍 Hyderabad, India
🌐 rishavraj.info | 🐙 github.com/rishavraj221 | 🔗 [linkedin.com/in/rishav-raj](https://www.linkedin.com/in/rishav-raj)

SUMMARY

I take problems end to end: system design, proof of concept, UI, backend, cloud setup. Recent work is mostly making a language model answer real questions from a company's own documents and show where the answer came from.

EDUCATION

B.Tech, Indian Institute of Technology Madras

2019 to 2023

EXPERIENCE

Evernorth Health Services (Cigna Group), Hyderabad

Mar 2025 to Present

Software Engineer, AI and Cloud [read more](#)

- **Retrieval.** Three searches run in parallel: LLM article routing, Azure AI Search chunks, and a curated OKF bundle. A reasoning model reads them through tools and answers only with a citation. [read more](#)
- **Ingestion.** Crawls seed articles and sub links to depth k, parses xlsx, docx, pdf and ppt attachments, and emits a routing index, search chunks and an OKF bundle. [read more](#)
- **Regression testing.** Months of real mailbox threads became a golden set, scored by a judge model. Caching verdicts on a content hash made the numbers repeatable run to run. [read more](#)
- **Caching and limits.** Two caches keyed on question, content set and corpus version, so re-ingestion retires stale answers. Token bucket at the gateway holds bursts inside model quota. [read more](#)
- **Admin and scale.** Feedback review, ingested content as tree and graph, a retrieval test bench, analytics, each behind its own permission. Sized for 8k daily users. [read more](#)

Flexday AI Hyderabad

Feb 2024 to Feb 2025

Software Developer, AI and Cloud [read more](#)

- **RBAC service.** Multi tenant, Node and Express on MongoDB, owning users, roles and machine identities. Clients sat on AWS, Azure and their own hardware, so cloud IAM could not cover them. [read more](#)
- **Hybrid checks.** Coarse permissions in the JWT validate locally, resource level checks hit a central endpoint. A per tenant policy version lands a role change on the next request rather than at expiry. [read more](#)
- **Tropicana chatbot.** [Customer facing](#), GPT-4 and Elasticsearch, on AWS and the client's own servers. Cleared their security review, firewall rules and VPN setup first.

Fastail Hyderabad

2023 to 2024

Software Engineer [read more](#)

- **Social commerce feed.** Reels from local brands with ordering in the feed. Delivery area picks the candidates before engagement ranks them, using GeoJSON polygons on a 2dsphere index cached per location cell. [read more](#)
- **Payments and POS.** Razorpay webhooks as the source of truth rather than the browser callback, with ranked states so a late event cannot move an order backwards. Orders push into the restaurant's Petpooja POS keyed on our order id, so a retry cannot print twice. [read more](#)
- Node on Lambda, MongoDB, Cloudinary. Two engineers, I led the system design and built most of the frontend.

PROJECTS

Kureita, node based AI video studio

kureita.com [read more](#)

- **Runner.** Wired nodes resolve in topological order. FastAPI on Lambda self invokes for a fresh time budget, since generation outlasts the API Gateway timeout. [read more](#)
- **Webhooks and billing.** Provider callbacks are ED25519 verified and idempotent, so a retry cannot double charge. 15k lines of Python. [read more](#)

SKILLS

Search and AI: RAG pipelines end to end, crawling and parsing a corpus, chunking, hybrid and graph search, re ranking, evals and LLM as judge, guardrails, Azure AI Search, Weaviate, OpenAI, Claude, Gemini

Backend: Python, FastAPI, Node.js, Express, TypeScript, MongoDB, PostgreSQL, Redis, Elasticsearch

Infrastructure: AWS (Lambda, ECS Fargate, EC2, S3, API Gateway, SQS, ElastiCache, CloudFront, Secrets Manager), OpenShift, Terraform, Docker, SAM, GitHub Actions, Auth0, Okta

System design: rate limiting (token bucket, sliding window), caching and invalidation, idempotency, queues for load levelling, sharding and replication, consistency trade offs, multi tenant isolation, RED metrics

BIT OF PERSONAL LIFE

Training most days, gym and runs and cycling. Badminton and swimming. Out with run clubs, cycling groups and weekend treks, marathons when the calendar allows, Hyrox still pending. Road trips and long rides in between. Happy to have company for any of it.