

Arpit Chauhan

☎ +91 9082784190 ✉ arpitchauhanofficial@gmail.com 🌐 [arpit-chauhan](https://arpit-chauhan.github.io) 🔄 [botARPIT](https://botARPIT.github.io) 🌐 portfolio.arpitdev.site

EDUCATION

A.P. Shah Institute of Technology, Mumbai University
Bachelor of Engineering in Computer Engineering — CGPA: 8.3/10

Maharashtra, India
August 2020 – May 2024

TECHNICAL SKILLS

Backend: Python (FastAPI), Node.js (Express, Hono), TypeScript
Applied AI: LLM agents (Google ADK, LangGraph, LangChain), RAG pipelines, agent orchestration, vector DBs (FAISS, Chroma, Pinecone), Gemini / Qwen
Infra & DevOps: Docker, Docker Compose, GitHub Actions CI/CD, Nginx, AWS EC2 + S3 + Secrets Manager, Cloudflare Workers
Databases: PostgreSQL (Alembic migrations), MongoDB, Redis (cache-aside, pub/sub)
Observability: Prometheus, Grafana (provisioned dashboards), structured logging, distributed tracing (OpenTelemetry, Tempo)
Security: JWT, OAuth2 PKCE, CORS/CSRF, Argon2, Web Crypto API, secrets on tmpfs RAM

PROJECTS

Blogify AI — Production Multi-Agent Blog Generation System blogifyai.arpitdev.site | 2026

Python, FastAPI, Google ADK, Gemini, PostgreSQL, Redis, Docker, AWS EC2

- Designed and deployed a **5-agent pipeline** (intent → research → outline → writer → editor) using Google ADK with sequential and conditional routing, retry logic, and structured output validation
- Engineered full production infrastructure: GitHub Actions CI/CD → Docker Hub → AWS EC2, Alembic migrations, 3-worker horizontal scaling, and secrets stored exclusively in **tmpfs RAM** — never written to container filesystem
- Generates ready to blogs in markdown format in approx. 40 secs from intent to final draft phase.
- Instrumented **distributed tracing** (OpenTelemetry → Tempo sidecar) across the 5-agent pipeline by manually creating span context

Async Job Processing System — Fault-Tolerant Distributed Worker [github repo](#) | 2026

Python, FastAPI, PostgreSQL, Redis, Prometheus, Grafana, Docker

- Architected a **lease-based async job system** treating PostgreSQL as authoritative state and Redis as transport — preventing dual execution, recovering from mid-flight worker failure, and distinguishing execution vs. transport failures
- Instrumented a reaper process for orphaned leases, latency decomposition metrics (queue time, execution time, total time), and Prometheus dashboards for **burst-load observability**
- Authored a formal engineering design doc: failure matrix, job lifecycle spec, and worker execution contract enforcing approx. 20 sec p99 latency under load.

Pinpoint PDF — Layout-Aware RAG Document Chatbot [github repo](#) | 2025

Python, FastAPI, Celery, Pinecone, PostgreSQL, Redis, S3/GCS, Gemini, React

- Architected a **3-stage async ingestion pipeline** (Celery workers: preprocess → chunk+summarize → embed+index) with layout-aware PDF parsing, section-level chunking, and table extraction; outputs stored in S3/GCS
- Built **multi-query RAG** with cross-encoder reranking (Sentence Transformers), Reciprocal Rank Fusion (RRF), and **Pinecone** with per-document namespacing for isolated retrieval
- Used **3 isolated PostgreSQL databases** (user profiles, document/chunk metadata, chat history) and summary-first indexing to improve retrieval precision on dense technical documents

E-Library REST API — Scalable Backend with Observability elib.arpitdev.site | 2025

Node.js, Express, TypeScript, MongoDB, Redis, Prometheus, Docker, AWS EC2

- Developed a modular monolithic backend with **JWT + Argon2** auth, role-based access control, Redis cache-aside strategy, and large file upload handling (PDFs to S3)
- Instrumented Prometheus metrics, structured logging, and containerised CI/CD to AWS EC2 with rollback support

Blogify Edge — Globally Distributed Blogging Backend blogify.arpitdev.site | 2025

Hono, Cloudflare Workers, PostgreSQL, Prisma Accelerate, TypeScript

- Built a **stateless edge backend** on Cloudflare Workers with OAuth2 PKCE, Web Crypto API (~2–3× faster sign-in vs. bcrypt.js), PostgreSQL via Prisma Accelerate, and sub-minute global CI/CD deployments

ADDITIONAL

Recognition: Collaborator on WalkSense (AI assistive navigation); architecture contributions selected in **Amazon AI for Bharat Hackathon**, top 30 teams nationally

Currently: Building systems programming intuition through C — concurrency primitives, memory model, process isolation (github.com/botARPIT/systems-learning)