

Anubhav Choubey

+91 0000000000 | choubey.anubhav253@gmail.com | anubhavchoubey.com | [LinkedIn](#) | [GitHub](#)

TECHNICAL SKILLS

Languages & Core: Python (Expert), TypeScript/JavaScript, SQL, HTML/CSS, Dart, OOP, AsyncIO

Backend & Data Systems: FastAPI, Django REST, PostgreSQL, ClickHouse (OLAP), Redis, Typesense, Celery

Autonomous AI & Agents: Model Context Protocol (MCP), Agent Runtimes, LangGraph, LangChain, PyTorch, Diffusion/VLM

Cloud & DevOps Infra: Kubernetes (GKE), Docker, Terraform (Multi-Region), Helm, Gateway API, SSHPiper, GCP, AWS

Concepts & Security: Distributed Systems, System Design, Microservices, OAuth 2.1, CASL RBAC, Multi-Tenancy

EXPERIENCE

Founding Engineer

Mar 2026 – Present

Urumi.ai

Goa, India

- **In-Pod Agent Runtime:** Built a stateful Codex agent runtime inside customer tenant pods with disconnect-resilient turns; designed an autonomous Planner → Implementer → Reviewer → Fixer state machine manipulating container filesystems via Model Context Protocol (MCP) tools.
- **Multi-Tenant Isolation:** Resolved concurrent multi-user editing in single pods by orchestrating per-user git worktrees, container UID isolation, dynamically routed preview docroots, cross-site cookie isolation, and Gateway API health sentinels.
- **Deterministic Figma-to-Code:** Replaced vision hallucinations with a DOM-based convergence loop, extracting vector tokens via custom Figma MCP and iteratively measuring computed CSS/DOM metrics against live renders to achieve pixel convergence.
- **MCP Fabric & OAuth 2.1:** Authored a production OAuth 2.1 authorization server, DPoP-validating resource server, and centralized tool fabric orchestrating external tools with granular permission tiers across regional tenant pods.
- **Multi-Region Infrastructure (IaC):** Provisioned duplicate GCP infrastructure via Terraform across London and Virginia regions, orchestrating GKE clusters, Cloud SQL, Gateway API, Workload Identity, and SSHPiper for multiplexed container SSH routing.
- **Zero-Trust Auth & RBAC:** Architected partial-JWT step-up authentication for TOTP 2FA, sliding 24-hour rate limiters, reverse proxy route sentinels, and a 22-rule CASL permission matrix replacing static roles.

SWE Intern

Jul 2025 – Oct 2025

Rozana.in

Bengaluru, India

- **Automated Replenishment Engine:** Designed and implemented a full-stack automated replenishment engine, integrating ClickHouse for high-throughput analytical aggregations, PostgreSQL for ACID compliance, and Redis caching for sub-millisecond retrieval.
- **Hybrid Search Engine:** Built search infrastructure using Typesense, combining keyword matching with dense semantic embeddings to deliver high-relevance hybrid textual and semantic search across the catalog.
- **Recommendation Engine (POC):** Architected a multi-stage recommendation pipeline pairing semantic vector similarity for contextual relevance with category-based filtering for low-latency retrieval.
- **Generative AI Catalog Pipeline:** Streamlined asset ingestion by chaining generative vision models (Flux, Qwen Image Edit, Vision LLMs) with automated evaluation gates to produce metadata, descriptions, and enhanced product imagery.

AI Intern

Sep 2024 – Jul 2025

Devnex Technologies

Remote

- **Distributed Pub/Sub Inference Engine:** Architected an asynchronous event-driven pub/sub system to broker, queue, and dispatch concurrent inference requests across heterogeneous AI models deployed across disparate remote worker clusters.
- **Multi-Tool Autonomous AI Agent:** Engineered a system-aware autonomous agent orchestrating a modular tool harness synthesizing real-time analytical telemetry and multimodal workflows via stateful multi-step reasoning and schema validation.
- **Cost-Optimized Vision Pipelines:** Built end-to-end pipelines for virtual try-on, semantic garment masking, and localized diffusion inpainting; fine-tuned and quantized open-source models to run co-located on single-GPU VMs to minimize compute costs.

PROJECTS

- PyroKube** | *Kubernetes, Python, FastAPI, Helm, Docker, Traefik, PostgreSQL* | [Details](#)
- **Managed Cloud Orchestration:** Built an on-demand infrastructure orchestrator using the Python Kubernetes SDK to provision dynamic, isolated workload namespaces without manual YAML overhead.
 - **Polymorphic Service Factory:** Engineered a factory driver pattern to instantiate catalog datastores (PostgreSQL, Redis, MongoDB, RabbitMQ, Garage S3) with dynamic Helm templating, resource quotas, and credential rotation.
 - **Dynamic Ingress & Routing:** Configured Traefik/Nginx ingress controllers with dynamic wildcard subdomain routing (*.pyrokube.local), automated cluster tear-down hooks, and gateway health sentinels.

- Upfler** | *FastAPI, Python, PostgreSQL, Redis, Celery, Docker, Locust* | [Details](#)
- **High-Throughput Backend:** Architected a modular REST API using FastAPI with sliding-window Redis rate limiters, strict Pydantic v2 validation, and sub-50ms p95 latencies under Locust-simulated concurrency spikes.
 - **Dual-Layer Caching & Async Pipelines:** Utilized Redis Sorted Sets (ZSET) to serve precomputed global leaderboards with sub-millisecond reads; offloaded reward distributions and feed aggregations to asynchronous Celery worker queues.
 - **Zero-Downtime Deployment:** Hardened PostgreSQL with connection pooling and transactional isolation, deploying containerized runtimes onto KVM instances with automated CI/CD rolling reload.

- AnubhavFlix** | *FastAPI, React, FFmpeg, HLS/DASH, Redis, Docker, Tailwind CSS* | [Details](#)
- **Adaptive Streaming Engine:** Built automated FFmpeg transcode pipelines converting media into multi-bitrate streams (1080p/720p/480p) with adaptive HLS/DASH chunking modulated by real-time client bandwidth telemetry.
 - **Access Control & Frontend:** Engineered JWT-authenticated API endpoints managing watchlists, per-user bandwidth quotas, and playback state persistence, paired with a custom low-latency React video player.

EDUCATION

Jabalpur Engineering College (JEC) <i>B.Tech in Artificial Intelligence and Data Science; CGPA: 8.55</i>	Jabalpur, MP <i>Jul 2022 – Apr 2026</i>
--	--

ACCOMPLISHMENTS

YouTube Content Creator: 26k+ subscribers & 8.6M+ views on technical coding channel.
Hackathons: 2x SIH National Finalist (2024, 2025) | Winner at CodeKumbh Hackathon.
DSA Proficiency: Solved 500+ LeetCode problems.
Certifications: Python, AI/ML, Tech Support Fundamentals. [View](#)