

# SRIJAN

+91 8434279769 | [srijandev01@gmail.com](mailto:srijandev01@gmail.com) | [portfolio.bysrijan.com](http://portfolio.bysrijan.com) | [github.com/Srijan-t0d0](https://github.com/Srijan-t0d0)

## SUMMARY

---

- Wrote the first commit at Vubo and built the codebase from scratch; promoted to lead after 4 months. Now own product architecture, code review, and technical direction for an AI video platform integrating 45+ models (Kling, Seedance, Gemini, Flux Pro) and serving thousands of creators.
- Strong in DevOps and infrastructure as well: built preview deployments with Docker, Caddy, and GitHub Actions that sleep and wake on request, and wrote Terraform for on-demand staging infra that cut AWS cost by 30%.
- Full-stack across TypeScript, Next.js, Python, and SQL: from product features and API design to AWS Lambda rendering pipelines, BullMQ job queues, PostgreSQL, and Redis.
- Strong in system design and developer tooling: type-safe RouteBuilder across 70+ API routes, Turborepo monorepo migration, and a RAG-powered knowledge graph on Cloudflare's free tier.
- Shipped 8 production SaaS products end-to-end; most recent (Social Claw) includes multi-platform post scheduling, keyword-triggered DM automation, 4-tier subscription billing via Polar, and 80+ backend tests in CI.

## EDUCATION

---

### Indian Institute of Technology Madras

*Bachelor of Science in Data Science and Applications*

Chennai, India

Aug. 2022 – Sep. 2026

Online degree program for working professionals, currently pursuing while working full-time as founding engineer at Vubo

**Certification:** CS50x, Harvard University's Introduction to Computer Science

## EXPERIENCE

---

### Founding Engineer → Lead Engineer

Oct. 2024 – Present

*Vubo, AI Video Generation Platform ([vubo.ai](http://vubo.ai))*

Remote

- Wrote the first commit and built the codebase from scratch; promoted to lead after 4 months, now owning architecture, code review, and technical direction for a 7-person engineering team serving thousands of creators
- Built preview deployment system with Docker, Caddy, and GitHub Actions so ephemeral environments sleep and wake on request, keeping review apps available without paying for idle compute
- Wrote Terraform for staging infrastructure that can be created or destroyed on demand, reducing AWS spend by 30% versus keeping staging always running
- Orchestrated async AI video pipeline using fal.ai, Replicate, and Remotion on AWS Lambda; integrated 45+ models (Kling, Seedance, Gemini, Flux Pro) with prompt engineering to ensure consistent generation quality across 1,000+ processed videos
- Developed distributed background job system using Inngest and BullMQ for media processing, credit management, and workflow orchestration at scale
- Shipped backend features in TypeScript: subscription logic, atomic credit deduction, rate limiting, referral system, and internal admin tooling; backed by PostgreSQL (RDS) and Redis (ElastiCache)
- Built social scheduling with OAuth across 4 platforms (Meta, TikTok, YouTube), enabling creators to publish directly from the app and eliminating dependency on external scheduling tools
- Engineered a type-safe RouteBuilder abstraction reused across 70+ API routes, centralising auth, workspace scoping, rate limiting, and tracing while eliminating per-route boilerplate across the entire API surface
- Integrated OpenTelemetry tracing and PostHog analytics; optimized CI/CD with GitHub Actions caching and automated CloudFront invalidations; migrated to Turborepo monorepo with React 19 + Partial Pre-Rendering

### Full-Stack Engineer

Oct. 2024 – Dec. 2024

*MrXP, Home Services Marketplace ([mrxp.com](http://mrxp.com))*

Remote

- Optimized slow PostgreSQL queries on expert search and dashboard endpoints (indexing, query rewrites, N+1 elimination), cutting P95 response times and database load on the highest-traffic paths
- Added multi-layer caching (build-time static generation for categories/services, search result caching, server-side) that reduced runtime API calls by ~100x and measurably lowered infrastructure cost

- Converted expert profiles, search results, and dashboard from CSR to SSR, eliminating redundant re-renders and improving perceived load times on high-traffic pages
- Shipped SEO fundamentals: dynamic meta tags on expert and search pages, XML sitemap, category landing pages for organic indexing, server-side pagination, and DNS prefetching
- Migrated codebase to TypeScript and restructured Navbar, search, and category components into proper Next.js patterns

## Freelance Full-Stack Engineer

Feb. 2024 – Jan. 2025

*Self-Employed*

*Remote*

- Designed and built client websites end-to-end, from UI implementation to production deployment and custom domain configuration
- Set up CI/CD pipelines (GitHub Actions) for client projects, automating build, test, and deployment workflows to reduce manual release overhead
- Managed web hosting infrastructure for multiple clients: server provisioning, SSL/TLS setup, Nginx reverse proxying, and ongoing uptime monitoring

## PROJECTS

---

### Social Claw | *Next.js 16, Convex, Better Auth, Turborepo, Polar, Cloudflare R2* | [Live](#) 2025

- Built a full production social media management SaaS solo: multi-platform post composer with Cloudflare R2 media uploads, keyword-triggered DM automation engine with follow-gating and tracked links, and interactive scheduling calendar
- Integrated 4-tier subscription billing via Polar; shipped with 80+ backend tests (Vitest) and a full CI/CD pipeline for production-grade quality from day one

### Internet Mindmap | *Next.js, Hono, Neon Postgres, pgvector, Workers AI, Vercel AI SDK, WXT* | [Live](#) 2025

- Built a full RAG pipeline: Chrome extension captures pages, a Cloudflare Workflow extracts content, generates 768-dim embeddings via EmbeddingGemma-300M, chunks long articles (6k chars / 500 overlap), and stores vectors in Neon Postgres with HNSW indexing for cosine similarity search
- Implemented HyDE (Hypothetical Document Embeddings) for chat queries, CRAG retry with query rewriting when relevance drops below threshold, and tag-graph hops to surface related items, all streamed via Vercel AI SDK with bracket-reference source citations
- Designed a migration-proof embedding registry with active/shadow provider pattern: new models embed in parallel without downtime, then a config flip switches query-time provider with zero code change and zero user-visible downtime

### Reddit MCP Server | *TypeScript, MCP SDK, Hono, Cloudflare Workers, Zod, Bun* | [Live](#) 2025

- Built a Model Context Protocol server exposing Reddit data through 7 tools (search, posts, subreddits, users, multi-feed aggregation) with Zod-validated schemas across 4 deployment targets: stdio, Streamable HTTP, REST API, and Apify Actor
- Engineered stale-while-revalidate caching with in-flight request deduplication and metrics tracking; deployed to Cloudflare Workers edge with cross-runtime support (Bun, Node.js, CF Workers)

### Visual App Builder | *React, React Flow, Hono, Mastra, OpenRouter, Zustand, Drizzle, SQLite* 2025

- Visual workflow builder that converts drag-and-drop canvases into production Next.js 16 projects via a 6-agent code generation pipeline with fan-out parallelization (concurrency 4) and rule-based skill dispatch, runs on free-tier LLMs (\$0 cost)
- Two-pass validation (static import/dep checker + in-memory bash-tool sandbox) with SSE progress streaming and per-agent SQLite telemetry tracking model, tokens, latency, and finish reason

### Emotion Classifier | *PyTorch, Transformers, BERT, DeBERTa, BiLSTM, Gradio* | [Live](#) | [GitHub](#) 2025

- Compared 4 NLP models (BERT, DeBERTa, BiLSTM-Attention, custom TextCNN) on multi-class emotion classification. Deployed the best as a Gradio app on HuggingFace Spaces

### Self-Hosting Infrastructure | *Go, Docker, Nginx, Prometheus, Grafana* | [GitHub](#) 2024

- Runs multiple side projects on a home server: Docker Compose for containers, Nginx as reverse proxy, a custom DDNS updater written in Go, and Prometheus + Grafana for monitoring

## TECHNICAL SKILLS

---

**Languages:** TypeScript, JavaScript, Python, Go, SQL, HTML/CSS, Bash, C

**Frameworks:** Next.js 16, React 19, Astro, Hono, Express.js, Flask, Tailwind CSS v4, tRPC, GraphQL, Remotion

**Backend & Data:** PostgreSQL, Drizzle ORM, Prisma, Convex, Redis, BullMQ, Inngest, MongoDB, Socket.IO, Cloudflare D1, KV Storage

**AI/ML:** PyTorch, Vercel AI SDK, LangChain, MCP SDK, Mastra, Cloudflare Workers AI, RAG, HyDE, pgvector, Embeddings, scikit-learn, OpenAI, Gemini, Mistral, OpenRouter, fal.ai, Replicate, Prompt Engineering, Browser Extension APIs

**Infrastructure:** AWS (EC2, ECS, EKS, Lambda, S3, ECR, RDS, CloudFront, IAM, VPC, ElastiCache, ALB, WAF, CloudWatch), Cloudflare Workers, Terraform, Docker, Kubernetes, Nginx, Caddy, Turborepo, GitHub Actions, Linux

**Observability & Auth:** OpenTelemetry, Prometheus, Grafana, CloudWatch, PostHog, Weights & Biases, Better Auth, Auth.js, OAuth