

Govind Sharma

Founding Engineer · AI Products from Zero to Scale

govindgovind852@gmail.com · Mumbai, India · govindio.com · linkedin.com/in/govind-io · github.com/govind-io

Open to founding engineer roles at early-stage AI companies · Remote (US/EU overlap) or India

SUMMARY

Founding engineer with **5+ years** taking products from first commit to revenue. One of Stamina.io's first engineers: built the core platform, AI SDR, calling and workflow systems while ARR grew from **\$250K to \$2M+** in **9 months**, then led engineering. Solo founder of FastBowlingLab, an AI coaching app with **1,200+ real users** and paying customers. Owns the whole path from product scope to infrastructure, and keeps cost and speed in view.

EXPERIENCE

Staff Engineer — Head of Engineering · Full-time Contract | [Stamina.io](https://stamina.io) | Jun 2025 – Oct 2026

Promoted: Founding Backend Engineer — Lead (Jun 2025) → Staff Engineer — Head of Engineering (Apr 2026)

AI sales, CRM and marketing SaaS · **1,000+ customers** · **50+ countries** · **\$2M+ ARR** · New York, United States · Remote

- › Joined as one of the first engineers and built the core platform (CRM data model, outreach, AI and workflows) while ARR grew from **\$250K to \$2M+** in 9 months and the team grew from 4 to 10; promoted to Staff Engineer and Head of Engineering.
- › Designed and built **Zara**, Stamina's AI SDR agent (**LangGraph**, tool calling, structured outputs): enriches **20K leads/day**, doubled reply rates and cut lead response time from 24 hours to instant for 1,000+ customers in 50+ countries.
- › Architected the workflow automation engine (**5,000+ workflows/day**, 50+ triggers and actions, 10+ integrations) to start exactly one run per event on at-least-once delivery: transactional outbox to **Google Pub/Sub**, layered dedup in **PostgreSQL**, a durable step machine with leases, version pinning and at-most-once CRM writes, and stateless, autoscaling workers with connection pooling to protect the database.
- › Designed the credit ledger behind usage billing: serializable, idempotent spends across concurrent workers, expiring grants spent soonest-first, an append-only ledger enforced in the database, and **Stripe** sync that survives late and duplicate webhooks.
- › Built the **Twilio** power dialer used by **200+ customers making 5,000+ calls/day**: row-locked lead claims so no lead is dialled twice, signed per-call tokens re-checked in the voice webhook, and per-brand Twilio subaccounts.
- › Kept the platform cheap and safe to change: per-task model routing with LLM evals (**70%** lower LLM cost per lead), 40% lower cloud spend by scaling idle queue workers to zero, per-PR preview environments, and a PR check that boots the full service fleet before merge.

Software Engineer II — Engineering Lead | Better Software · Embedded at [NeshAI](https://nesh.ai) | Jul 2024 – Jun 2025

B2B AI SaaS for Fortune 500 sales teams · Remote

- › Engineering lead for a **5–6 engineer** team embedded at [NeshAI](https://nesh.ai); built its AI recommendation engine with **LLM-powered retrieval** over 1M+ documents for **Blue Origin, Chevrolet, DKSH, WRGrace**.
- › Held p95 retrieval under **800ms** at 85% answer accuracy on the eval set, ingesting 1,000+ new docs/day through **Kafka** pipelines.
- › Shipped **MCP servers**, OpenAI Assistants API integrations and AI workbooks rated **5 stars**; built Salesforce, Google Drive, S3 and Azure Blob connectors and the **SOC 2** controls clients required.

Sr. Software Engineer — Tech Lead | Loktantra Mediatech | Jan 2022 – Jun 2024

Promoted twice: Software Engineer (Jan 2022) → Team Lead (Jul 2022) → Sr. Software Engineer — Tech Lead (Jan 2023)

Social and live-streaming platform (Khulke) · Bengaluru, India (Hybrid)

- › Built a self-hosted WebRTC SFU on **LiveKit**, a signaling server and a React SDK from scratch, replacing a **\$2,000/month** vendor; scaled to 100,000 concurrent users across the platform at 1,000 req/sec, with rooms of 50+ speakers and 500+ listeners.
- › Cut page load from **4.2s to under 1s** and API p95 from 900ms to 200ms through Redis caching, query tuning, code splitting and optimistic updates.

› Hired and led a **5–7 engineer** team; introduced testing practices and code reviews that cut bugs per sprint from 20+ to 5; wrote the TDDs/RFCs.

Full Stack Developer · ICLEAR CCTV & Security System, Delhi (Remote) · May 2021 – Jan 2022 · Co-built the React frontend of a CCTV livestreaming app; repeat customers rose 30%.

PROJECTS & OPEN SOURCE

- › [FastBowlingLab](#) · Founder, Aug 2026 – Present — AI fast-bowling coach built solo: **MediaPipe** pose estimation on phone video (33 body landmarks, 19 measurements), annotated video and adaptive training plans. **1,200+ real users**, 300+ monthly actives, ~20 paying customers, 2,500+ deliveries analysed. Next.js, React Native (Expo), tRPC, PostgreSQL, Inngest, AWS ECS.
- › [livekit/components-js](#) — Merged contribution to LiveKit's official React components, the library behind the WebRTC platform I built at Loktantra.

SKILLS

Product & 0 to 1:	MVP scoping, shipping to first users, payments (Stripe , Razorpay), cost-aware architecture, hiring the first engineers
AI / LLM:	Generative AI , LLMs , agentic AI , AI agents , LangGraph , embeddings, vector search, tool calling, MCP servers , LLM evaluation, guardrails, prompt engineering, OpenAI, Anthropic Claude API, computer vision , pose estimation (MediaPipe)
Full Stack:	TypeScript , Python, Node.js , NestJS , tRPC, Next.js , React, React Native (Expo) , PostgreSQL, Redis, Inngest, video processing
Infra:	AWS (ECS, EKS, Lambda, RDS, SQS/SNS), Kubernetes , Terraform , Docker, Kafka, Google Pub/Sub, OpenTelemetry, Prometheus, Grafana, CI/CD
Security:	OAuth2, SSO, RBAC, SOC 2