

# Vansh Pratap Singh

Pune, India | Open to Relocation & Remote

+91-8299225810 | vanshispratap@gmail.com | linkedin.com/in/vanshpratapsingh | github.com/llxSkyWALKeRxll | vanshpratapsingh.com

## SUMMARY

Backend engineer with 4 years building high-concurrency distributed systems and AI developer tooling in Go, Java, Python and Rust. Shipped real-time platforms serving 50K+ concurrent users and production LLM infrastructure used by enterprise customers.

## EXPERIENCE

**Bito** Oct 2024 – Present  
Software Development Engineer II Pune, India (Remote)

- Built core backend services for **AI Architect**, which indexes customer codebases into a cross-repo knowledge graph and serves it to coding agents (Cursor, Claude Code, Codex) over **MCP**; used by enterprise customers including Swiggy, PubMatic and Kredivo.
- Cut LLM operational costs by up to **60%** through hot-path re-architecture, request batching, and migrating workloads to cheaper models with no measurable regression in output quality.
- Designed a **CI-integrated regression suite** that replays PRs through the AI Code Review Agent and scores findings against expected results, replacing manual QA and cutting release cycle time **20%**. Won 1st place, company hackathon.
- Diagnosed and resolved production bottlenecks across the code-review pipeline, and ran competitive benchmarking against rival agents to inform roadmap priorities, supporting enterprise retention and new client wins.
- Tech:** Go, Python, MCP, LLMs & prompt engineering, AWS, Docker, CI/CD, SQL, Git, Linux.

**Adda247** Jul 2022 – Oct 2024  
Software Development Engineer II Gurugram, India

Promoted twice in two years: SDE Intern (Jul 2022) to SDE-I (Jul 2023) to SDE-II (Apr 2024)

- Built the real-time YouTube live-class platform (polls, chat, leaderboards, hand-raise, doubt resolution) sustaining **50K concurrent users at 2,000 messages/second**; drove **18% conversion** from YouTube and a **17% engagement lift**.
- Led the in-house migration of a contractor-run microservice, re-architecting its APIs and adding **request coalescing** on the read path, which **reduced server and database load by 40%** during peak class hours.
- Fixed a JVM bottleneck in a core microservice, eliminating CPU utilization spikes and delivering a **4x throughput increase**.
- Built search, popular-courses and faceted-filter APIs for the paid content platform on Elasticsearch and Redis, making purchased courses discoverable and lifting content engagement.
- Tech:** Java, Spring Boot, Go, Lua, Kafka, Redis, Elasticsearch, MongoDB, SQL, Docker, Kibana, AWS, Linux.

**Samsung R&D Institute India** Nov 2021 – May 2022  
Research Intern Remote

- Built **ns-3** simulations in C++ modeling 5G RAN energy consumption across deployment scenarios, validating theoretical efficiency models against simulated network behavior.
- Authored an internal report defining energy-efficiency KPIs for RAN network elements and surveying optimization strategies, circulated to inform R&D direction.

## PROJECTS

**Stimulus: Gym & Training** [Play Store](#)  
React Native (Expo), TypeScript, Supabase, PostgreSQL, Cloudflare Workers [Live on Google Play](#)

- Designed, built and shipped a strength-training tracker end to end: mobile app, database, auth, marketing site and store release. Live with 18 monthly active users and growing, shipping regular updates.
- Modeled a normalized Postgres schema with row-level security and 7 versioned migrations; built passwordless email/phone OTP auth, progress analytics (estimated 1RM, session volume) and a Play-compliant data deletion flow.

**Clerva AI: Interview Preparation Platform** Personal Project  
Rust, Axum, Tokio, SvelteKit, PostgreSQL, LLMs

- Built an async **Rust** backend (Axum + Tokio) generating targeted interview questions from a job description and resume, with session handling, secure user management and payment gateway integration.
- Added multi-provider LLM routing (Anthropic, OpenAI, Gemini) with per-action model selection, plus Argon2 password hashing, rate limiting, OAuth login and scheduled cleanup jobs.

## TECHNICAL SKILLS

**Languages:** Go, Java, Python, Rust, SQL, C++, TypeScript, Lua

**Backend & Data:** Spring Boot, Kafka, Redis, Elasticsearch, PostgreSQL, MongoDB, Distributed Systems, System Design

**Infrastructure:** AWS, Docker, Linux, Git, CI/CD, Kibana

**AI:** LLM integration, Prompt Engineering, Model Context Protocol (MCP), Code Indexing & Retrieval, AI Evaluation Harnesses

## EDUCATION & AWARDS

**Chandigarh University**, Bachelor of Engineering, Computer Science Jul 2019 – Jul 2023 | Punjab, India

- 1st Place, Bito Company-Wide Hackathon** (2025): for the AI Code Review Agent evaluation suite, since adopted internally.
- Biomkesh Bakshi Award**, Adda247 (2024): for technical contributions and delivery speed on reliability-critical systems.