

Vivekanand Choudhary

(+91) 9399225149 | chyvivekanand@gmail.com | [linkedin](#) | [github](#) | [leetcode](#)

EXPERIENCE

Software Engineer

April 2025 – Present

Proficonlabs Pvt Ltd. (Simplo.ai)

Hyderabad, Telangana

- Designed and implemented per-tenant rate limiting and Kafka queue isolation across Spring Boot microservices, isolating high-throughput tenants (e.g., high-volume jewelry retail customers) from impacting 20+ shared tenants, reducing Kafka consumer lag by 80% during peak traffic and maintaining consistent performance platform-wide.
- Built a per-page SEO (Search Engine Optimization) configuration API using Java, Spring Boot and cache-aside architecture, supporting meta title, description, canonical URL served SEO metadata in under 20ms for client-generated websites.
- Built a production-grade CRM module for a SaaS platform, enabling customers to manage client details, schedule meetings, and receive real-time notifications. Implemented a file storage service using AWS S3 to securely store client documents and images
- Implemented JWT-based authentication and role-based access control (Admin, Manager, Agent) across 8+ microservices using Spring Security, with field-level permissions so users only see data relevant to their role.
- Build real-time capabilities using Kafka and WebSockets, enabling instant UI updates and stream-based processing for internal analytics pipeline
- Mentored and onboarded 5+ interns on Java, Spring Boot, microservices, and development workflows, helping them understand the codebase, follow engineering practices, and independently contribute production-ready changes.

Software Intern

Aug 2024 — Mar 2025

Proficonlabs Pvt Ltd.

Hyderabad, Telangana

- Cut open bug backlog by 25% in 4 weeks by systematically resolving defects across 6 Spring Boot microservices covering MongoDB query inefficiencies, Redis cache stampedes, and REST contract mismatches.
- Maintained 80%+ unit-test coverage using JUnit and Mockito across all services, blocking 3 regression bugs from reaching production.

PROJECTS

ARX-24+ Model-AI-Chat-Platform | *Python, Next.js, TypeScript, ChromaDB, LiteLLM,*

[GitHub](#)

- Built a multi-LLM chat platform delivering hallucination-resistant answers grounded in private documents, 24+ models across 11 providers, sub-300ms streaming, and a 3-layer hybrid RAG pipeline (semantic caching + vector search + re-ranking).
- Designed a zero-restart model registry that allows new AI models to go live without code changes, supporting 29 models across a 6-stage pipeline (security, quota, RAG, routing, inference, and streaming); exposed RAG search, ingestion, and routing as structured MCP tool calls instead of prompt-based integration.

Event-Triggered Automation Agent | *Python, FastAPI, Redis, PostgreSQL, gRPC, GraphQL, Docker*

[GitHub](#)

- Built a backend system that listens for events (like a payment, a GitHub update, or a Slack message), figures out what to do about it, and does it automatically using FastAPI, Redis/Celery, gRPC, and PostgreSQL. It also retries failed actions and safely stores ones that keep failing, so nothing gets lost.
- Eliminated missed and duplicate processing of failed payments, support tickets, and form submissions by replacing manual handling with an automated workflow: validate, queue, decide (rules-based or LLM-based), and execute exactly once; enabled teams to track live progress via GraphQL, removing the need for manual status checks.

TECHNICAL SKILLS

Languages: Java, Python, JavaScript, TypeScript, SQL

Backend Technologies: Spring Boot, Microservices, REST APIs, JPA/Hibernate, gRPC, OAuth2, JWT, Spring Security

Frontend Technologies: React.js, Next.js, HTML, CSS

Databases: MySQL, PostgreSQL, MongoDB, Cassandra, ChromaDB

Messaging & Caching: Apache Kafka, Redis, WebSockets

Architecture Patterns: Idempotency, REST API Design, Cache-Aside, Rate Limiting, Pub-Sub Messaging

Cloud and DevOps: Docker, Kubernetes, Jenkins, Git, CI/CD Pipelines, AWS (EC2, S3, IAM, VPC, Lambda)

AI & LLM: LangChain, RAG, MCP, Vector Databases, Azure OpenAI, Tool Calling, Agentic AI

Testing & Monitoring: JUnit, Mockito, SonarQube, Grafana, Prometheus

EDUCATION

PG-Diploma in Advanced Computing

KNOW-IT, Authorized Training Center (ATC) of C-DAC ACTS

Pune, Maharashtra

Sep 2023 - Mar 2024

Bachelor of Technology - Electrical and Electronics Engineering

Bhilai Institute of Technology

Durg, Chhattisgarh

Jul 2018 - Jun 2022