

# HIMANSHU RAJ

Software Developer / AI Enthusiast

+91 8709947443   himanshuraj.cse@gmail.com   github.com/himaenshuu  
Bengaluru, India   linkedin.com/in/himaenshuu   hub.docker.com/u/himaenshuu



## SUMMARY

Full-Stack AI Engineer building AI-powered web applications with React/Next.js, FastAPI, vector databases, multimodal RAG workflows, Docker, AWS, and CI/CD orchestration.

## EXPERIENCE

### Software Development Intern

Growyt Jul 2025 – Dec 2025

- Improved contextual retrieval accuracy by 35% using multimodal RAG workflows with LangChain, Gemini API, and Qdrant.
- Reduced semantic search latency by 40% through optimized embedding ingestion, vector indexing, and Dockerized deployment workflows.

### Software Engineer Intern

Lifesync Feb 2025 – May 2025

- Increased chatbot workflow reliability by 30% through secure Python-based backend API integrations.
- Reduced backend communication failures by 45% by debugging distributed API workflows and improving engineering collaboration with Git workflows.

## PROJECTS

### ShopSense – AI Product Intelligence Platform

- Increased product discovery accuracy by 32% using semantic recommendation workflows with vector retrieval.
- Reduced product search latency by 40% using FastAPI services, Pinecone indexing, and optimized frontend integration.

### Multimodal RAG System

- Improved retrieval relevance by 35% using LangChain orchestration and Gemini-powered contextual reasoning.
- Reduced query latency by 40% through optimized Qdrant indexing and FastAPI retrieval services.

### SmartVault – AI Knowledge Vault

- Improved enterprise document retrieval efficiency by 38% using embedding-based semantic search.
- Reduced document processing effort by 35% through FastAPI orchestration and vector indexing pipelines.

### Pinecone Semantic Search Platform

- Increased semantic retrieval relevance by 40% using vector similarity search workflows.
- Improved query performance by 30% using optimized Pinecone indexing and scalable AI search APIs.

## EDUCATION

### B.Tech in Computer Science

Indian Institute of Information Technology, Nagpur July 2022 – June 2026

## CODING PROFILES

LeetCode   Codeforces

## ACHIEVEMENTS

- Codeforces Specialist with a maximum rating of 1556.
- Solved 320+ DSA and competitive programming problems on LeetCode.
- Contributed 800+ open-source commits across 40+ public GitHub projects.

## CORE SKILLS

### AI & LLM Stack

|               |          |            |
|---------------|----------|------------|
| Python        | PyTorch  | TensorFlow |
| LangChain     | RAG      | Gemini API |
| Vector Search | Pinecone | Qdrant     |

### Full Stack Development

|               |           |            |
|---------------|-----------|------------|
| React         | Next.js   | JavaScript |
| TypeScript    | FastAPI   | Node.js    |
| Microservices | REST APIs |            |

### Data & Infrastructure

|         |                |        |     |
|---------|----------------|--------|-----|
| MongoDB | SQL            | Docker | AWS |
| CI/CD   | GitHub Actions | Linux  | Git |

## FULL STACK AI CAPABILITIES

- AI-powered frontend and backend application development
- RAG-based intelligent product workflows
- FastAPI microservices and REST API orchestration
- React/Next.js frontend integration with LLM pipelines
- Vector search and embedding retrieval systems
- Dockerized deployment and scalable AI workflows
- Cloud-native deployment and CI/CD automation

## CERTIFICATIONS

### OCI Generative AI Professional

Oracle • 2025

### OCI Data Science Professional

Oracle • 2025

### IBM Docker Essentials

IBM SkillsBuild • 2025

### Career Essentials in GitHub

GitHub • Nov 2025

## SOFT SKILLS

|                         |       |
|-------------------------|-------|
| Technical Communication | ★★★★★ |
| Team Leadership         | ★★★★☆ |
| Problem Solving         | ★★★★★ |
| Adaptability            | ★★★★☆ |

## EXTRACURRICULARS

- Core member, Coding Club, IIIT Nagpur.