

# CHETAN SANJAY SHIVADE

+91 7378518714 ♦ Pune, Maharashtra

[chetanshivade34@gmail.com](mailto:chetanshivade34@gmail.com) ♦ [linkedin.com/ChetanShivade](https://www.linkedin.com/ChetanShivade) ♦ [github.com/ChetanShivade](https://github.com/ChetanShivade)

## SUMMARY

Innovative AI Engineer at TCS with 11 months of experience in building Retrieval-Augmented Generation (RAG) systems, LangChain pipelines, and Knowledge Graph-based search architectures. Skilled in integrating LLMs with enterprise data, designing secure APIs, and optimizing semantic retrieval performance using Python, Neo4j, and FastAPI. Strong background in Java backend systems and microservices architecture, enabling full-stack understanding of intelligent systems.

## EDUCATION

**B. Tech in Computer Engineering**, R. C. Patel Institute of Technology, Shirpur Jul, 2020 - Jul, 2024

Relevant Coursework: Java, Python, DSA, DBMS, OOPS, SDLC

CGPA: 8.05/10

## SKILLS

<b>AI Technologies</b>	LangChain, Hugging Face, ChromaDB, RAG Pipelines, Prompt Engineering
<b>Technical Skills</b>	Java, Python, SQL, JavaScript, Spring Boot, FastAPI, Django, Microservices
<b>Tools &amp; Frameworks</b>	PostgreSQL, MySQL, Neo4j (Graph DB), Git, GitHub, Gitlab CI/CD, REST APIs
<b>Security</b>	OAuth2, JWT, Spring Security

## EXPERIENCE

**Software Engineer** Jan, 2025 - Present

Tata Consultancy Services (TCS) *Pune, India*

- Engineered a Retrieval-Augmented Generation (RAG) system for workflow automation, integrating LangChain, ChromaDB, and Azure OpenAI, reducing processing cycles from 2 years to near-instant automation.
- Designed and implemented hybrid retrievers combining semantic vector search and knowledge graph traversal, achieving 45% higher retrieval relevance.
- Built FastAPI microservices exposing secured REST endpoints with JWT/OAuth2 authentication for internal AI assistants.
- Integrated Neo4j-based Knowledge Models enabling semantic context linkage and domain-specific reasoning across multiple enterprise systems.
- Deployed RAG pipelines through GitLab CI/CD, ensuring zero-downtime releases and 30% faster iteration cycles.
- Collaborated with cross-functional teams to integrate LLM-powered context enrichment into enterprise search and automation systems.

**Java Developer Internship** Aug, 2022 - Sep, 2022

R3 Systems India Private Limited *Nashik, India*

- Developed a high-performance web application using JSP, Servlets, and MySQL, supporting 5,000+ concurrent users.
- Enhanced backend performance by 30% through optimized Java logic and multi-threaded handling.
- Led a team of 4 to deliver a production-ready web application achieving 99.9% uptime.

## PROJECTS

**Knowledge Graph RAG Chatbot (FastAPI, Python, LangChain, Neo4j, ChromaDB, OpenAI, JWT)**

- Built a Knowledge Graph-enhanced RAG pipeline integrating LangChain retrievers and Neo4j for structured semantic retrieval.
- Leveraged ChromaDB embeddings and Azure OpenAI GPT models for contextualized Q&A and insight generation.
- Secured endpoints using OAuth2/JWT and deployed via GitLab CI/CD for internal testing.
- Improved contextual accuracy by 60% through hybrid graph-vector retrieval optimization.

**Candidate Assessment Tool (Java, Spring Boot, JSP, MySQL)**

- Architected a layered Spring Boot backend with MVC pattern, reducing development time by 30%
- Automated candidate evaluation workflows, improving HR efficiency by 25%.
- Implemented Spring Data JPA for efficient MySQL access, reducing development time by 30%.

**Black Spot Detection System (Java, Spring Boot, Google Maps API, MySQL)**

- Built a web dashboard enabling 20% faster emergency response using real-time location data.
- Integrated Google Maps APIs for visualization and SQL optimization for faster data access.

## CERTIFICATES AND ACHIEVEMENT

- LangChain for LLM Applications – DeepLearning.AI
- Java, Spring Boot- Wipro TalentNext
- SQL- Hackerrank
- Solved 400+ DSA problems across platforms