Omkar Deepak Shinde

GitHub | LinkedIn | Portfolio | Email | Phone

Experience

CGI | **Software Engineer**

September 2025 – Present
Pune, Maharashtra

Codebase Intelligence System — Graph RAG Pipeline for Legacy C++ Repository

- Architected an **LLM-driven Graph RAG pipeline** combining knowledge graph construction and retrieval-augmented generation to semantically navigate a 10,000+ line legacy C++ codebase, cutting developer onboarding time by 40% on undocumented systems.
- Built a **knowledge graph in Neo4j** mapping 100+ code entities (classes, functions, dependencies) across 50+ modules, enabling graph-traversal-based context retrieval and natural language queries over undocumented legacy code.
- Built a **dependency traversal engine** using AST-based call detection to retrieve full function chains across 3–5 dependency levels and inject complete context into LLM prompts, reducing hallucination rate in generated explanations by over 50%.

Report Migration Pipeline — Crystal Reports to Jasper via LLM

- Designed an **LLM-driven formula migration pipeline** converting Crystal Reports to Jasper Reports via prompt engineering, eliminating manual formula rewriting and reducing migration effort by 75%.
- Orchestrated metadata extraction into **structured JSON representations** fed into engineered LLM prompts, achieving consistent and scalable formula conversion across 200+ report templates.

Pull Request Review Bot — Code Analysis & Automated Test Generation

- Developed an **LLM-based GitHub PR Review Bot** that analyzes changed files, detects missing test coverage, and delivers structured review feedback directly on pull requests, saving 2–3 hours of manual review per sprint.
- Developed **intelligent test gap detection** that identifies untested functions across PR diffs and auto-generates accurate unit tests via LLM prompts, reducing manual reviewer overhead by 60%.

Stakeholder Report Generator — Jira Hierarchy Summarization Pipeline

- Built a **Jira REST API automation pipeline** that recursively traverses full release hierarchies — epics, subtasks, and bugs — aggregating data across 50+ tickets per release cycle, eliminating 3+ hours of manual reporting per sprint.
- Engineered **structured LLM prompt templates** consolidating hierarchical Jira data into coherent summarization inputs, cutting report generation time from 2 hours to under 2 minutes per release.

ELIXIR.AI | **Data Science Intern**

March 2024 – July 2024
Pune, Maharashtra

- Introduced an **OCR-based automation pipeline using PyTesseract and OpenCV**, improving text recognition by 20% for medical document processing.
- Engineered section-wise text extraction covering 8+ document fields, increasing data extraction precision by 30% and accelerating downstream processing.
- Trained a **custom neural network with TensorFlow and OpenCV**, achieving 95% precision in cooler compliance monitoring using a dataset of 4,000+ images.
- Delivered a **real-time compliance monitoring solution** with TensorFlow and Roboflow, processing 500+ cooler audits and improving operational compliance rates by 35%.

Projects

Resume Classification & Parsing System Aug 2025 – Present
Python, spaCy, TensorFlow, APILayer, NLP

- Architected an **NLP-based resume parser** classifying resumes into 9 job categories with 85%+ accuracy using TensorFlow neural networks and spaCy.
- Scripted **synthetic dataset generation** producing 100+ labeled JSON resumes across 9 categories, cutting data preparation time by 90% and enabling model training without manual labeling.
- Streamlined PDF and DOCX text extraction via APILayer API, achieving 95%+ extraction reliability across document formats.

Video Recommendation System Apr 2024 – May 2024
Python, FastAPI, TensorFlow, Pandas, NumPy, Scikit-learn, SQLite

- Designed a **deep learning-based recommendation engine** using ReLU and Sigmoid activations with collaborative filtering, processing 5,000+ videos across 10,000+ data points.
- Optimized API response time to <500ms via Scikit-learn preprocessing and similarity ranking, boosting engagement by 20%.

Technical Skills

Languages	Python, Java, HTML/CSS, JavaScript, SQL
AI / ML	LLM Integration, Retrieval-Augmented Generation (RAG), Prompt Engineering, LangChain, FAISS, Vector Embeddings, Semantic Search, Knowledge Graphs, Neo4j, spaCy, TensorFlow, Keras, Scikit-learn, NLP Pipelines, OCR (PyTesseract)
Frameworks / Libs	FastAPI, Flask, React.js, Node.js, NumPy, Pandas, OpenCV, Roboflow
Databases	MySQL, MongoDB, SQLite
APIs & Integrations	GitHub API, Jira REST API, APILayer
Developer Tools	VS Code, GitHub, Cloud Platforms, API Development

Education

Manipal Institute of Technology, Bangalore Aug 2021 – July 2025
B.Tech in Computer Science with Specialization in Artificial Intelligence

Relevant Coursework

Reinforcement Learning · Introduction to Big Data · AI for Everyone · Front-end Development · Programming with JavaScript · UI/UX Design

Achievements

- Solved **200+ problems on LeetCode** focusing on arrays, hashing, sliding window, binary search, trees, and dynamic programming — building strong algorithmic foundations for technical interviews.

Extracurricular

Treasurer | Manipal Open Source Community (MBOSC) Apr 2023 – Apr 2024
Manipal Institute of Technology

- Managed the community's budget, handling all financial aspects and ensuring successful execution of multiple events.
- Coordinated with the executive board to execute key initiatives, improving community operations by 20%.
- Drove chapter service and unity-focused programs, enhancing community engagement and overall member experience.