

PARITOSH

kparitosh760@gmail.com | 9327027422 | LinkedIn | GitHub

Education

Kalinga Institute of Industrial Technology, B.Tech in Computer Science

July 2023 – July 2027

Experience

SDE Intern — Agnost.ai

March 2026 – June 2026

- Built **OpenTelemetry instrumentation** for LLM pipelines across multiple agent frameworks (Mastra, Volt, OpenAI SDK, LangChain, CrewAI, DSPy), enabling standardized observability for production AI agents.
- Developed an **MCP server** from an OpenAPI specification and built developer-facing **Chrome extensions**, including demo integrations (Arcjet, WorkOS, Sentry) for developer relations work.

AI / Document Intelligence Intern — ONGC

May 2025 – July 2025

- Built an internal **document intelligence pipeline** to index and query 50+ technical drilling reports using a RAG architecture with **LLM APIs** and **ChromaDB** vector storage.
- Implemented **document preprocessing** and chunking strategies for LLM context ingestion; reduced document search time by 40%.
- Integrated **OpenAI embeddings** for semantic retrieval, enabling structured information extraction from unstructured PDF reports.

Research Intern — KIIT University

Jan 2025 – Apr 2025

- Performed large-scale data analysis on 54 African countries (2000–2022) using Python; applied **PCA** for dimensionality reduction and correlation analysis on derived components.
- Collaborated on research under faculty mentorship, contributing to architectural decisions on data pipeline design.

Projects

vid_translate — Japanese/English/Spanish Video Translation App

github

- Built a cross-platform **Tauri** desktop app (Linux, Windows) that translates video content between **Japanese, English, and Spanish**.
- Automated the cross-platform release pipeline with **GitHub Actions CI/CD** using a matrix strategy, producing **AppImage, .deb, .msi, and NSIS** artifacts.

Multimodal Document Intelligence RAG — Open-Source Stack

github

- Built an end-to-end **document intelligence pipeline**: parsed PDFs with **marker** (OCR/layout-aware), then used **qwen3-vl** (VLM) to auto-generate descriptive captions for embedded images enriching context for downstream LLM retrieval.
- Vectorized preprocessed documents with **bge-large-en-v1.5** into **ChromaDB**; served open-source **gpt-oss:120b** for response generation fully local, production-grade RAG stack.

Technologies

Languages: Python | JavaScript | PowerShell

LLM & AI: LLM APIs (OpenAI, Claude) | LangChain | RAG Pipelines | OpenTelemetry for LLM Pipelines | MCP Server Development | OCR (marker) | vLLM | Embeddings (bge, OpenAI) | VLMs (qwen3-vl)

Infrastructure & Tools: ChromaDB | Docker | AWS EC2 | MySQL | Postman | Claude Code / Codex