

# Mayur Kailas Shirsat

Pune, Maharashtra 411023 | 9067611982 | mayurshirsat1999@gmail.com | LinkedIn: [mayur-shirsat-422757151](https://www.linkedin.com/in/mayur-shirsat-422757151) | Github: <https://github.com/Mayur-shirsat>

## EDUCATION

### SYMBIOSIS INTERNATIONAL UNIVERSITY

Master of Science Data Science and Spatial Analytics

Pune, Maharashtra

July 2023 – July 2025

### BHARATI VIDYAPEETH DEEMED UNIVERSITY

Bachelors of Science Computer Science

Pune, Maharashtra

July 2017 – November 2020

## SKILLS AND STRENGTHS

**Generative AI & LLMs:** GraphRAG (FAISS, NetworkX), LoRA/QLoRA, Unsloth, Fine-tuning, vLLM, Ollama, llama.cpp, LangChain, Prompt Engineering, Hugging face transformers.

**AI & Computer Vision:** Machine Learning, Deep Learning, NLP, YOLO (Ultralytics), CNN-LSTM, Roboflow, Image Segmentation, Action Recognition.

**Languages & Frameworks:** Python (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow, OpenCV), FastAPI, R, SQL, HTML/CSS.

**Data Architecture & Analytics:** PostgreSQL, MongoDB, Data Quality & Governance, EDA, Statistical Modeling, Predictive Analytics, Web Scraping (Selenium).

**Visualization & BI:** Power BI, Tableau, MS Excel, Matplotlib, Seaborn.

**Cloud & Deployment:** Microsoft Azure, Render, Docker, Linux, n8n (Agentic Workflows), Agile, Git.

## WORK EXPERIENCE

### AI ML Engineer - RapportSoft

December 2025 – Present

- **Vision & Action Recognition:** Architected end-to-end vision systems utilizing **YOLO11 and YOLO26** for object detection and **CNN-LSTM** frameworks to model complex action recognition and sequence-based temporal behaviors.
- **Data Engineering & Optimization:** Orchestrated robust data pipelines, including automated collection, **EDA** for class imbalance resolution, and advanced **augmentation** techniques to maximize model generalization and robustness.
- **VLM Deployment & Research:** Developed high-throughput **OCR** and visual analysis pipelines using **LLMs/VLMs (Qwen3vl:2B)**, exposing model inference via **FastAPI** while conducting research-driven experiments to optimize real-world performance.
- **Model Alignment:** Spearheaded the fine-tuning of SLMs via **LoRA** on custom datasets, significantly reducing **hallucination** and enhancing response reliability in production-grade RAG pipelines.
- **Production Efficiency:** Spearheaded model optimization initiatives that **slashed GPU overhead from 95% to 15%** through **TensorRT** acceleration and **ONNX** runtime optimizations, maintaining model accuracy while drastically reducing operational cloud costs

### AI ML Engineer - Rajyug Solutions

July 2025 – November 2025

- Developed **multi-agent AI systems** using **n8n** and Vision Transformers for complex behavioral analytics and OCR tasks.
- Built and integrated **RAG-based conversational agents** into Streamlit interfaces, utilizing FastAPI for backend model serving.
- Elevated deep learning recognition accuracy by 20% by architecting custom **Roboflow** pipelines and implementing ensemble training methods to reduce model variance.

### Data Science and AI intern - Canspirit.ai

May 2024 – November 2024

- **Containerized** and launched AI/ML projects using **Docker** and Render, streamlining deployment workflows and improving system efficiency by 30%.
- Automated lead generation and SEO workflows using **LLMs**, resulting in a 25% increase in acquisition and 40% higher user engagement.

## PROJECTS

### 1. Multimodal LLM and Computer Vision-Based Agricultural Assistance System

- **Multimodal Architecture:** Spearheaded a cross-modal advisory platform synthesizing **text, image, and speech** inputs via **LLM-driven reasoning**; achieved **95% accuracy** in plant disease identification and treatment recommendations.
- **Deployment & Integration:** Developed a recommendation engine integrated with **Hugging Face** text-to-speech and **FastAPI** to provide real-time access to market data and **5 government welfare schemes** via official portals.

### 2. Smart Examination System using AI

- **LLM and VLM-based microservices** where Gemma-3-4B generates question papers from examiner-provided syllabi.
- Used **Qwen-2.5-VL** to extract and structure student responses from scanned answer sheets using visual-language understanding.
- Architected **LLM-based evaluation pipeline** in which a GPT-OSS 20B model analyzes student answers against the question paper and marking scheme to automatically assign scores through coordinated microservice architecture.

### 3. Snag Detection System (Real-Estate Damage Detection)

- computer vision-based snag detection system using YOLO models over 90% Mean Average Precision to identify structural and electrical defects in real-estate buildings, including wall cracks, fire damage, metal rusting, and wire failures.
- Performed **end-to-end dataset creation**, including **web scraping** for image collection, manual annotation, and label assignment to train and **fine-tune the detection models**.
- Designed a damage-response pipeline where detected defects trigger real-time alerts, and YOLO outputs are passed to a **VLM model** for damage verification and automating draft email generation to notify builders or maintenance teams.

### 4. Executive Dashboard for Project Performance Analysis:

- Built a **Power BI** dashboard using SQL to assess project performance across departments, roles, and audiences. Provided insights on revenue by project category, roles, and audience, identifying corporate projects as top contributors (\$510.75M).
- Extracted actionable insights through **departmental analysis**, identifying international projects as the primary revenue drivers (\$159.06M). Open-sourced the project on **GitHub** to facilitate peer review and data-driven strategy development.

### 5. RAG-Based Document Information Extraction System:

- Retrieval-Augmented Generation (RAG) system using Meta LLaMA to extract and summarize data from 5+ document types, including links, DOC, PDF, CSV, and text files and integrated it with **NetworkX for Knowledge Graph based RAG**.
- With 90% accuracy rate in multi-format data retrieval, enhancing document analysis by 40%. Optimized query handling to support complex information requests, improving data accessibility for users by 35%.

---

## CERTIFICATION AND ACHIEVEMENTS

ISRO AI/ML for Geodata analysis in August 2024 | Python certification from **Infosys springboard** in October 2023 | **Prompt Engineering** by Great learning