

Omkar Kabde

☎ 8500-383-520 | ✉ omkarkabde@gmail.com | 🌐 omkar-kabde | 🎧 omkar-334 | 📁 Portfolio

EXPERIENCE

Indian Institute of Science

Bengaluru, KA

Research Intern

Jun 2025 – Present

- Developed conditional diffusion models for time-series forecasting, anomaly detection, imputation and classification. (submitted to the AAAI 2026 Conference)
- Achieved state-of-the-art performance among 15+ models across all tasks.
- Implemented and benchmarked multiple model architectures - DL, transformers, KAN, diffusion models

Indian Institute of Technology, Indore

Indore, MP

Research Intern

Oct 2024 – Mar 2025

- Developed a two-stage multimodal contrastive learning framework and ImpliHateVid, a large-scale benchmark dataset of 2,009 videos for hate speech detection. (Accepted at ACL 2025 Main Conference)
- Worked on Semi-supervised crop and weed disease detection. (Submitted to Information Fusion Journal)
- Worked on knowledge distillation and diffusion models for crop disease segmentation and classification.

Bhabha AI

Remote

AI Research Intern

Apr 2024 – Oct 2024

- Designed scalable data pipelines for Indic languages, incorporating translation, backtranslation, prompt rephrasing, and scoring for dataset creation, with multiprocessing to parallelize tasks.
- Automated data scraping workflows (300,000+ records) using Selenium & Playwright for AI forecast models
- Implemented research papers for data quality & synthetic data generation.

Google Summer of Code

Remote

Open Source Contributor

May 2024 – Aug 2024

- Enhanced the **xarray-subset-grid** library for high-resolution **STOFS** ocean data subsetting.
- Improved subsetting operations by **70%** and created comprehensive documentation using **readthedocs**.
- Performed benchmark evaluations for subsetting & chunking methods with **Dask**, **Zarr**, and **Xarray**.

Go India Advisors

Mumbai, MH

Python Developer Intern

Jan 2024 – Jun 2024

- Automated the extraction of 10 years of stock exchange data (2M records) using Python, AWS EC2, and Docker. Managed 20+ SQL tables and pipelines for data ingestion and transformation.
- Improved similarity indexing, pdf ingestion and fuzzy matching algorithms for shareholder and investor data
- Enhanced pipeline performance by dockerizing multiple AWS schedulers and implementing multiprocessing.

EDUCATION

Chaitanya Bharathi Institute of Technology | Hyderabad, TG

2022 – 2026

Bachelor of Engineering in Computer Science

GPA: 8.24/10

- **Relevant Courses:** Object-Oriented Programming, Data Structures & Algorithms, Design & Analysis of Algorithms, Database Management Systems, Computer Networks, Operating Systems, AI & Machine Learning, Natural Language Processing, Deep Learning

SKILLS

Languages: Python , R, SQL, C, C++, Javascript, HTML, Rust

Libraries: Numpy, Pandas, FastAPI, BeautifulSoup, Playwright, Selenium, Matplotlib, Scikit-Learn, Pytorch, TensorFlow, Langchain, Xarray, Dask, Llamaindex, Huggingface, Transformers, NLTK, Keras

Technologies: MS Excel, Power Automate, GCP, Git, Github, Docker, AWS, MongoDB

PROJECTS

[MedRaga](#) - personalised medical assistant with latest research access (FastAPI, Langchain, Qdrant)

[ScholarSearch](#) - Publication aggregation platform and Summary generator (Python, App Engine, React)

[Social.AI](#) - Social Network simulation with AI agents (Python, Huggingface, OpenAI, Gradio)