

# Shubham Chaudhari

Los Angeles, CA • [shubhamc@usc.edu](mailto:shubhamc@usc.edu) • [linkedin.com/in/shubham-ch](https://www.linkedin.com/in/shubham-ch) • [github.com/Esoteriikos](https://github.com/Esoteriikos) • [My Portfolio](#) • +1 (213) 275-7114

## SKILLS

- Languages: Python | Core: AI/ML Systems Design, MLOps/LLMOps, Computer Vision, GenAI, Data Science, AI Ethics, Transformers, XGBoost, PyTorch, TensorFlow, Scikit-learn, OpenCV, NumPy, Pandas
- LLM & Generative AI: RAG, AI Agents, (LangChain, LlamaIndex, LlamaParse), RAG Evals (Ragas), LangGraph, Fine-Tuning, LoRA, QLoRA, Prompt Engineering, Multi-modality, GANs, Azure Vision Studio, VectorDB, Guardrails
- MLOps & DevOps: Docker, Git, GitHub Actions, DVC, FastAPI, Flask, Streamlit, MLflow, Postman, Linux
- Databases: PostgreSQL, MongoDB, Redis, MySQL | Orchestration & ETL: Kafka, Airbyte, n8n, Airflow
- Monitoring & BI: Metabase, Grafana, Plotly, Seaborn | Optimization: ONNX, TensorRT, OpenVINO

## EDUCATION

### UNIVERSITY OF SOUTHERN CALIFORNIA

*Master of Science in Computer Science - Data Science Specialization (GPA 3.65/4.00)*

Los Angeles, CA

January 2025 - December 2026

### UNIVERSITY OF MUMBAI

*Bachelor of Engineering, Computer Engineering (CGPA 9.14/10)*

Mumbai, India

July 2017 - June 2021

## EXPERIENCE

### HERE TECHNOLOGIES, TECHNOLOGY INNOVATION LABS

Mumbai, India

#### AI ML Engineer II

August 2024 - December 2024

- Architected and developed an end-to-end GenAI assistant with LangChain and CodeQwen, leveraging a custom-configured vector database to query internal Java SDKs
- Built a high-fidelity data ingestion pipeline using LlamaParse and custom PyPDF parsers, optimizing document chunking strategies which enhanced LLM processing efficiency by 40%
- Engineered a hybrid retrieval system with a cross-encoder re-ranker to boost contextual accuracy by 25% over baseline vector search, delivering superior context to the generator model
- Established a robust LLMOps framework: utilized Apache Airflow for data orchestration, Docker for containerization, and MLflow for experiment tracking, leading to a 30% rise in developer productivity

### TATA CONSULTANCY SERVICES, RETAIL STRATEGIC INITIATIVES LAB

Mumbai, India

#### AI ML Engineer

July 2021 - July 2024

- Architected a multi-camera vision pipeline integrating OpenPose, OpenCV with data from IoT load sensors using Pub/Sub messaging system to track end-to-end customer journey in an autonomous store, reducing checkout times by over 80%
- Developed two targeted solutions on Azure Vision Studio: 1) A custom shelf-monitoring model that cut stockout and spoilage delays by 60%, and 2) A CCTV video summarization service that reduced incident review time by 70%
- Pioneered a generative AI pipeline for instant e-commerce cataloging; integrated Super-Resolution CNNs and GANs to perform virtual clothing try-ons from a single 2D image, cutting catalog production time from weeks to hour
- Optimized models for edge deployment on NVIDIA Jetson devices, using ONNX, TensorRT and model quantization to boost real-time item recognition to achieve 7x faster inference
- Spearheaded the 'TechFM' knowledge-sharing program, training 400+ colleagues and 1000+engineering students in practical AI/ML concepts, fostering a culture of technical excellence and continuous learning

### TATA CONSULTANCY SERVICES

Mumbai, India

#### Data Scientist Intern

February 2021 - May 2021

- Engineered predictive models for Materiovigilance, utilizing Lasso regression for feature selection and XGBoost for high-accuracy prediction, which boosted the early detection of adverse events by 25%
- Led the design and deployment of a Python-based Streamlit analytics dashboard, providing healthcare professionals with real-time, actionable insights from MongoDB data and accelerating clinical response times

## PROJECTS.

- Data Streaming Pipeline: Architected Airbyte -> Kafka -> PostgreSQL -> Metabase stream for unified analytics
- Monocular Depth Estimation: Engineered U-net pipeline trained on KITTI/Cityscapes with Object detection, and 3D scene reconstruction

## PUBLICATIONS AND RECOGNITION.

- Published Whitepaper: 'A Hybrid approach towards Signal Management in PV' ([link](#))
- 'A Synopsis of Monocular Depth Estimation' (Presented at ICDSMLA, Published by Springer) ([link](#))
- AIP Anchor Award | Innovista Top Seeds & Sparkles | CLP Faculty Award | BPS Champions League Runner-up