

NIKHIL SHINGADIYA

MACHINE LEARNING ENGINEER

Ahmedabad,Gujarat,India| P: +91 6354071470 | shingadiyanikhil1634@gmail.com | [Linkedin](#) | [Github](#)

About Me ([Portfolio](#))

I am a Machine Learning Engineer with 4+ years of experience building and deploying AI-driven solutions with a focus on large language models. My work blends deep learning, NLP, and data science to deliver scalable systems for intelligent content generation and storytelling. I specialize in fine-tuning LLMs, optimizing model performance, and creating real-world applications that turn complex data into impactful outcomes.

TECHNOLOGY

AI & Machine Learning: Generative AI, Transformers, LLMs, HuggingFace, Fine-tuning LLMs, Tool-augmented LLMs, Agentic Workflows & Complex Multi-Agent Systems, Prompt Engineering, Statistics & Probability, Machine Learning (ANN, CNN, RNN), Exploratory Data Analysis & Data Visualization.

Data & Databases: Python, PySpark, Web Scraping, Vector Databases & Knowledge Retrieval Pipelines, MySQL, MongoDB.

Software & Development: Object-Oriented Programming (OOP), FastAPI, Celery, Redis, Docker, Docker-compose, AWS (EC2, S3, ECR, Lambda), Familiar with ML-flow.

Conversational AI: Conversational AI Design.

Libraries: LangChain,LlamaIndex,TensorFlow,Keras,OpenCV, SciPy, Scikit-learn, Statsmodels, Dash Plotly,BeautifulSoup, Selenium, Numpy, Pandas, Matplotlib, Seaborn.

IDE: Visual Studio Code, Jupyter Notebook, Spyder, Google-Colab, And Pycharm.

Version Control : Git & Github **Operating System:** Linux (Ubuntu 20.04 LTS), Windows, Mac Os.

WORK EXPERIENCE

Rishabh Software - (200-250 employees)

Vadodara, Gujarat

MACHINE LEARNING ENGINEER

Present – Aug 2025

★ Project 1 : BI-Agent

● Objective:

Developed an Multi User System **AI-driven Business Intelligence** (BI) Agent to automate client database schema extraction, metadata generation, and natural language-to-SQL translation—enabling intuitive data querying, visualization, and analytics for enterprise environments with RBAC Implementations.

● Tech Stack:

FastAPI, PostgreSQL, **SQLAlchemy**, Celery, Redis, **LangGraph**, Milvus, MinIO, Python 3, OpenAI,Docker & Docker-Compose. **Intel Opea Architecture** .

- **Key Features & Achievements:**

- Achieved **85% accuracy** in NLP-to-SQL generation, enabling precise, natural-language database querying.
- Applied **Elbow Method on retrieval scores** to optimize similarity thresholds, cutting token costs by **30%** and reducing noise in document searches.
- Utilized **Intel's Opea Architecture** for intelligent schema extraction and metadata automation with reusable, scalable data pipelines.
- Built a **relationship-driven FAQ Generator** using LangGraph, enabling **complexity-controlled agents** that generate contextual FAQs and training samples for improved SQL understanding.
- Engineered a **modular FastAPI backend** with Celery and Redis for distributed AI processing and metadata workflows.
- Leveraged **Milvus** for vector-based metadata storage and **MinIO** for secure, on-premise object storage.
- Developed a **custom LLM-powered visualization module** to extract meaningful graph columns and relationships for deeper data insights.
- **Deployed on self-hosted Rishbh servers**, ensuring data privacy, scalability, and complete operational control.

Unlimited WP - (100-150 employees)
AI/ML LEAD DEVELOPER

Ahmedabad, Gujarat
Present – Dec 2023

- ★ **Project 1:** [WeamAI](#)

- **Objective:**

Develop an **AI-powered platform** that enables companies to register, onboard employees, and engage in collaborative communication with integrated company documents—streamlining business operations through advanced AI.

- **Tech Stack:**

FastAPI, Celery, Redis, Docker-compose, LangChain, Llmindex, Ray framework, Pinecone, Qdrant, Generative AI, LLM models, OpenAI, HuggingFace, Anthropic, Streamlit, EC2 Instance.

- **Key Features & Achievements:**

- Engineered robust containerization with **Docker** to ensure scalable and reliable deployments.
- A **Layered Architecture** was designed that enhanced maintainability and scalability.
- Developed a comprehensive **CI/CD** pipeline to automate testing, deployment, and monitoring, reducing release cycles.
- Built a custom **Retrieval-Augmented Generation (RAG)** pipeline, significantly improving the accuracy and relevance of Q&A responses.
- Integrated **LangChain** agents to facilitate intelligent document interactions and guide users through sequential actions.
- Employed **Pinecone** and **Qdrant** for efficient storage and retrieval of embedding vectors, optimizing AI model performance.
- Utilized the Ray framework to maximize **GPU utilization**, ensuring efficient training and inference of custom LLMs.
- Integrated models from **OpenAI, HuggingFace, Gemini and Anthropic** to enhance the platform's AI capabilities.

- Successfully deployed the entire AI backend on **Amazon EC2** for robust performance.
- Developed custom metrics for scaling **Celery workers** via **CloudWatch**, improving resource efficiency.
- Created a custom module to resolve **LangChain & FastAPI** bugs, bolstering overall system stability.

★ Project 2: GARUDA AI

● Objective:

Develop an AI-driven blog writing engine designed to boost website blog rankings by generating **SEO-Optimized Content**.

● Tech Stack:

Google Keyword Planner, Ahrefs, Generative AI, LLM models.

● Key Features & Achievements:

- Utilize **Google Keyword Planner** to conduct in-depth keyword search volume analysis and identify high-impact keywords.
- Scrape data from targeted websites using **Ahrefs** to gather comprehensive and relevant information.
- Automatically generate five optimized blog titles, complete with keyword tags, to enhance **SEO performance**.
- **Produce high-quality**, SEO-friendly content based on the generated titles using advanced generative AI and **LLM models**.

F(x) Data Labs PVT LTD. (40-50 employees)

MACHINE LEARNING ENGINEER

Ahmedabad, Gujarat

Dec 2023 – Jan 2022

- Built a strong foundation in **statistics and probability**, specializing in hypothesis testing, A/B testing, regression analysis, and Bayesian testing.
- Expert in translating business requirements into precise **mathematical metrics** for robust analysis.
- Proficient in data **cleaning** and **transformation** to ensure high **data quality** and **reliability**.
- Proficient in a wide range of machine learning techniques, such as **classification**, **clustering**, and **Genetic algorithm**.
- Knowledge in state-of-the-art **generative image models**, such as **diffusion models** and controlnet.
- Skilled in constructing Large Language Models (**LLMs**) and fine-tuning them for precise applications, specifically in generating stories and **summarizing** complex narratives.
- Proficiency in deep learning models, including **CNNs**, **RNNs**, and **Attention Networks(Transformers)**.
- Familiarity with **MLflow** and **DVC** (Data Version Control) for efficient **data versioning** and model tracking.
- Experience with **Docker** and **Docker Compose** for containerization and deployment.
- Familiar with cloud platforms such as **Azure** and **AWS** for model deployment and scalability.

★ Project 1 : The GMR Group - MCP Prediction(Time Series Forecasting)

- **Objective:**

Developed a predictive model for forecasting the **Market Clearing Price (MCP)** in the **Day-Ahead Market (DAM)** using a Random Forest algorithm, achieving an accuracy of **70% ($\pm 6\%$)**.

- **Key Features & Achievements:**

- Integrated data from diverse sources—including **IEX Market**, weather data, and Google Trends—using Selenium for web scraping, followed by comprehensive data cleaning and transformation.
- Designed a **confidence metrics** model to help traders assess the reliability of MCP forecasts for each time block.
- Employed **Monte Carlo simulation** to demonstrate that our model-based trading strategy achieved an **11% improvement** in profitability compared to random trading strategies.

★ Project 2: Video Redering Wowsly (Computer Vision)

- **Objective :**

Develop an API using **FastAPI** that allows multiple users to upload a **Video** and a corresponding CSV file. Based on the CSV, the system automatically edits the video by overlaying text on specified frames for set durations by using **OpenCV**, stores the edited videos in an **S3 bucket**, and generates a tracking token for users to monitor task progress.

- **Key Features & Achievements:**

- Edited videos by overlaying text on video clips at predetermined positions and durations based on CSV inputs.
- Utilized **Docker & Docker Compose** to manage microservices including FastAPI, Celery, Redis, and Flower.
- Deploy this whole system on **AWS Lambda (AWS,EC2,Lambda)** for scaling up system

EDUCATION

L.D.College Of Engineering (Gujarat Technological University)

Ahemadabad, Gujarat

Bachelor of Engineering

May 2022 - July 2018

Major in Computer Engineering ,Cumulative CGPA: 8.62/10

Certifications & Training: Online Course in [Statistics with Python by Michigan University \(Coursera\)](#), [Mathematics for Machine Learning by Imperial College London](#), Computational Thinking and Data Science(MIT).

Languages: Fluent in Gujarati, and Hindi; Conversational Proficiency in English.