

Parth Srivastava

+91 8887664156

parthsrivastava6112004@gmail.com

LinkedIn

GitHub

Skills

Programming: Python, SQL, KQL (Kusto Query Language)

ML & AI: Machine Learning, Deep Learning, NLP, RAG, LLM Fine-tuning, AI Agents, Synthetic Data

Frameworks & Libraries: PyTorch, Hugging Face Transformers, PEFT, LangChain, LangGraph, FastAPI, FAISS, Scikit-learn, XGBoost, LightGBM

Deployment & Tools: Linux, Docker, vLLM, Apache Airflow, RunPod, Novita AI, Git, GitHub, Prometheus, Locust, Axiom (AQL), MongoDB, Jupyter Notebook, Streamlit, Kaggle

Experience

AI/ML Engineer, 4Way Technologies

Nov 2025 – Present

- Deployed 20+ AI models on RunPod and Novita using Docker, vLLM, and FastAPI.
- Built Apache Airflow ETL pipelines with Bronze, Silver, and Gold data lake architecture, processing 10K+ records daily.
- Built a RAG-based chatbot with real-time knowledge synchronization, supporting dynamic document ingestion, updates, and deletion.
- Developed a multilingual AI assistant integrated with RAG for context-aware conversations across multiple languages.
- Performed API load testing using Locust and monitored latency, throughput (RPS), error rates, and concurrent user performance with Prometheus.
- Developed image generation workflows using Stable Diffusion, ComfyUI, and Automatic1111, serving 200+ users.

Projects

LLM Fine-Tuning Pipeline

Source Code

Python, Hugging Face Transformers, PEFT, PyTorch, QLoRA, RunPod, vLLM

- Fine-tuned Qwen2.5-0.5B using Full Fine-Tuning and QLoRA on RunPod GPUs for structured JSON generation.
- Built a synthetic data generation pipeline to create 10K+ instruction samples for fine-tuning.
- Implemented custom tokenization and label masking, training only on assistant responses by masking prompt tokens (-100) for supervised fine-tuning.
- Organized the dataset into multiple progressive stages (P0, P1, ...), iteratively removing conflicting samples to improve data consistency and instruction quality.

Synthetic Data Generator

Source Code

Python, LangGraph, Groq LLaMA 70B, SDV TVAE, Pydantic

- Built a synthetic data generator that creates realistic datasets from one or multiple CSV files using LangGraph workflows and LLMs.
- Integrated SDV TVAE with validation and automatic retry mechanisms to preserve outliers and improve the realism of generated data.
- Performed automated EDA to extract statistical metrics, feature distributions, and correlations, using them to guide synthetic data generation.
- Applied feature categorization and schema validation with Pydantic to preserve data types, relationships, and diversity across generated datasets.

Awards & Achievements

Student Leader, SIH 2025

2025

- Led the team in Smart India Hackathon 2025, contributing to solution design and technical decision-making.

Education

Raj Kumar Goel Institute of Technology

Aug 2022 – Sep 2026

Bachelor of Technology in Computer Science & Engineering (Data Science)

CGPA: 7.4

Ghaziabad, Uttar Pradesh