

VANSHIKA PARKAR

+1 (585) 981-1828 | vp1837@g.rit.edu | [LinkedIn](#) | [GitHub](#)

WORK EXPERIENCE

Research Intern / Assistant

Rochester Institute of Technology

Jan 2025 – Aug 2025

New York, NY

- Led concrete foundation and promising results for AI-ASL system developed using Large Vision Language models, collaborating with PhD candidates as co-author of the research project with clear documentation and agile practices.
- Engineered structured ETL data pipelines in Python to process 4K+ video-annotation pairs of multimodal ASL data collected using web scraping methods and stored as a Vector Database.
- Identified the need to improve input QA data quality for LLM training and used RAG framework with Hugging Face language models to produce quality text embeddings which increased the input data efficiency by 35%.
- Performed evaluation benchmark analysis, testing data on state-of-the-art LVLMs like OpenAI GPT 4.1, LLaVA, and Qwen using API integration through Python-written pipeline scripts.
- Executed pre-training and fine-tuning of LLaVA-OV 7B model in PyTorch with LoRA efficiency using hyperparameter tuning and up-graded the model architecture adding a faster vision encoder saving 25% of processing time.
- Optimized GPU performance by model parallelism choosing effective variants that uplifted performance metrics by 14%. Developed automated reporting workflows using Weights & Biases and Git CI/CD.

Associate Data Scientist

Acars Realty Ltd.

July 2022 – Aug 2023

Mumbai, India

- Used predictive analysis with machine learning algorithms on curated AWS database of 5+ years of market trends as consistent contribution in property valuation engine that prevented 20+ low-yield acquisitions.
- Migrated over 80K+ scattered records into Snowflake as a part of weekly tasks by rebuilding ETL/ELT SQL pipelines to be at least 38% more concrete for company's data foundation.
- Designed customer segmentation pipelines for analyzing buying and interaction behavior in Python and integrated those insights into CRM helping sales team drive an average 12% increase in conversion rates.
- Participated actively in multiple cross-functional collaborations presenting the statistical and business KPIs built in Tableau and put efforts to bridge the gap in conveying them to non-technical staff.

Data Engineer

ACL Digital

Jan 2022 – June 2022

Bengaluru, India

- Built production ML pipelines for 12+ client projects in retail and commerce businesses using Docker and implementing CI/CD workflows, with A/B testing under senior guidance that improved deployment efficiency for cycles by 32%.
- Created feature engineering and processing pipelines in PySpark for 50GB+ daily data from company analytics platforms, integrating MLflow for experiment tracking with better inference APIs using FastAPI.

EDUCATION

Rochester Institute of Technology

Master of Science in Information Technology and Analytics (STEM) | GPA: 3.9/4.0

Rochester, NY, USA

Aug 2023 – Dec 2025

- Relevant coursework: Applied Statistics, Database Management, Visual Analytics, Text Mining, Neural Networks

University of Mumbai

Bachelor of Engineering in Electronics and Telecommunication | GPA: 3.2/4.0

Mumbai, India

Aug 2017 – Dec 2021

- Relevant coursework: OOP, Applied Mathematics, Big Data Analytics, Operating Systems, MATLAB

ACADEMIC PROJECTS

Generative BI Analyst | Python, LangChain, SQLite, Streamlit

- Simplified data analytics by training a Text-to-SQL agent using LangChain and Llama 3 making natural language processing tasks easier across complex relational schema with 92% syntax accuracy on multiple-table joins.
- Utilized RAG pipelines for data retrieval feature in the system, reducing time taken to generate insights by 40% as compared to manual SQL queries by handling tasks like aggregations, window functions and nested queries.
- Introduced a self-correcting execution loop within Streamlit dashboard that did the work of parsing database errors to automatically debug bad queries, resolving 85% syntax errors without the need of user intervention.

TECHNICAL SKILLS

Programming Languages: Python, Java, SQL, R, Shell Scripting, C++

Machine Learning & AI Frameworks: PyTorch, Scikit-learn, Hugging Face, OpenCV, LangChain, RAG Pipelines, LLMs/LVLMs

Data Science & ML: PyTorch, Scikit-learn, Hugging Face, LangChain, RAG Pipelines, LLMs/LVLMs, GenAI, Text-to-SQL, LLM Fine-tuning (LoRA), Pandas, NumPy, Spark, Matplotlib

Cloud & Databases: AWS S3, GCP (BigQuery), Snowflake, Databricks, MySQL, Oracle, SQLite, FastAPI, RESTful API Design

Tools: Docker, Git, CI/CD Pipelines, MLflow, Weights & Biases, Apache Airflow, Tableau, Power BI, Looker, Streamlit