

Dhruv

Research Assistant – Robotics & Reinforcement Learning

San Jose, CA, United States | [LinkedIn](#) | [GitHub](#) | [Personal Website](#)

ML engineer with 3+ years building production AI systems, distributed data pipelines, and on-device models across Samsung, Bain, and robotics research.

EXPERIENCE

Research Assistant – Robotics & Reinforcement Learning San José State University Built scalable ML training and evaluation infrastructure for RL policies on Unitree robots with automated sim-to-real transfer pipelines and W&B experiment tracking. Designed modular pipeline architecture that separated training, evaluation, and hardware deployment, eliminating manual reconfigurations and reducing experiment-to-hardware handoff cycles by 3x.	Jan 2026 — Present <i>San Jose, California, United States</i>
President Applied Intelligence Systems (AIS) Club Applied Intelligent Systems at SJSU As President of the AIS Club, I drive initiatives that enhance member engagement and professional growth. • Spearheaded the planning and execution of events and workshops tailored to member interests and industry trends. • Fostered collaboration with faculty and alumni to provide networking opportunities and professional development resources. • Developed strong communication strategies to gather feedback and streamline event planning processes.	Dec 2025 — Present
Senior Associate Software Engineer (AI/ML) Bain & Company Fine-tuned Mistral-8B with RAG on Bain's presentation corpus to build an AI-driven slide assistant, delivering 40% productivity gains. Designed and operated distributed PySpark ETL pipelines processing 1TB+ daily into Amazon Redshift with idempotent jobs and schema-enforced quality gates, cutting pipeline failure rework from ~3 hours to <30 minutes. Engineered the full-stack AI-based Map GL Classifier application resulting in \$150 million in yearly revenue and reducing processing time by 3 weeks per case.	Apr 2024 — Jul 2025 <i>New Delhi, Delhi, India</i>
Research And Development Engineer Samsung R&D Institute India Led development of an instruction-based image editing system using LLMs and Stable Diffusion, reducing inference time by 250ms. Developed an NLP proof of concept for real-time English-to-Hindi call translation with 450ms inference latency. Orchestrated ProtoTile for Galaxy Watch 6 Contacts, designing cloud backup and restore pipelines. Optimized a production LLM image editing microservice through request batching and async processing, reducing p99 latency from 1.2s to 480ms.	Jun 2022 — Apr 2024 <i>Noida, Uttar Pradesh, India</i>
Software Engineer Intern Wipro Limited	Jun 2021 — Jul 2021 <i>India</i>
Data Analyst Intern Edulyt India	Jun 2020 — Aug 2020
Research Intern Sardar Vallabhbhai National Institute of Technology, Surat	Dec 2019 — Feb 2020 <i>Greater Surat Area</i>

EDUCATION

San José State University <i>Master of Science - MS in Applied Data Intelligence</i>	Aug 2025 — May 2027
San Jose State University <i>M.S. in Applied Data Science</i> GPA: 4.0 / 4.0	Aug 2025 — May 2027
Sardar Vallabhbhai National Institute of Technology Surat <i>B.Tech in Electronics & Communication</i> GPA: 3.7 / 4.0	Jun 2018 — May 2022
Sardar Vallabhbhai National Institute of Technology, Surat <i>Bachelor of Technology - BTech in Electronics and Communication Engineering</i>	Jan 2018 — Jan 2022

PROJECTS

Data pipeline ingesting job listings with analytical views on skill demand and salary trends, The Last Degree
– Graduate ROI Intelligence

Built a self-directed Airflow pipeline extracting 1M+ daily job listings from Adzuna API with rate-limit handling and upsert-based Snowflake loading. Engineered dbt transformations producing 5 analytical views including Skill Demand Index, salary trends, and ROI forecasts. Exposed results via REST API to a React frontend with failure alerting and data-quality checks.

Full-stack microservice system for AI-powered vehicle repair discovery and booking, AutoQuote: AI Vehicle Repair Finder

Independently architected a full-stack microservice with circuit-breaker patterns, two-tier caching, and timeout handling to defend against high-latency dependencies. Integrated AI image analysis for vehicle damage, shop discovery, voice-call automation via VAPI, and appointment booking into a single orchestrated backend with structured logging and request tracing.

Two-DAG Airflow architecture for OHLCV data ingestion and per-symbol ML forecasting, Stock Price Forecasting & Analytics Pipeline

Built a two-DAG Airflow system handling daily OHLCV ETL from yfinance with transactional idempotent loads into Snowflake. Implemented dynamic task mapping with Airflow Pools for controlled Snowflake concurrency and per-symbol model training. dbt layering separated RAW and ANALYTICS with training views and final tables merging historical and forecasted values.

End-to-end vector search pipeline with SentenceTransformers embeddings, Semantic Search Engine – Pinecone + Airflow

Built a self-directed Airflow-orchestrated semantic search pipeline covering ingestion, preprocessing, and SentenceTransformer embedding with all-MiniLM-L6-v2 model. Implemented idempotent Pinecone vector upserts with index-reset capability to prevent duplication. Containerized via Docker Compose for reproducibility.

CERTIFICATIONS

Mastering Model Context Protocol (MCP) , LinkedIn	Sep 2025
Lean Six Sigma for a Sustainable Future , LinkedIn	Aug 2023
Software Architecture Foundations , LinkedIn	Aug 2022
Applied Data Science Capstone , Coursera	May 2020
IBM Data Science Professional Certificate , Coursera	May 2020
Data Visualization with Python , Coursera	May 2020
Structuring Machine Learning Projects , Coursera	Apr 2020
Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization , Coursera	Apr 2020
Machine Learning(Stanford Online) , Coursera	Apr 2020
Data Analysis with Python , Coursera	Apr 2020
Programming Foundations with JavaScript, HTML and CSS (with Honors) , Coursera	Apr 2020
Databases and SQL for Data Science , Coursera	Jan 2020
Machine Learning with Python , Coursera	Jan 2020
Python for Data Science and AI , Coursera	Dec 2019
Python for Data Science and AI , Coursera	Dec 2019
Google Cloud Platform Fundamentals: Core Infrastructure , Coursera	Dec 2019
Technical Support Fundamentals , Coursera	Dec 2019
Neural Networks and Deep Learning , Coursera	Dec 2019
Data Science Methodology , Coursera	Dec 2019
Data Science Methodology , IBM	Dec 2019

SKILLS

Problem Solving · Visual SLAM · Docker · React.js · Kotlin · Django · Flask · Spring Framework · JavaScript · Grafana · Amazon Web Services (AWS) · MongoDB · Python (Programming Language) · C++ · C (Programming Language) · Java · Microsoft Office · SQL · HTML · Microsoft PowerPoint · Matlab · SLAM · Motion Control ·

Robotics · Snowflake · Systems Design · Natural Language Processing (NLP) · Computer Vision · Research and Development (R&D) · Large Language Models (LLM) · Distributed Systems · Generative Adversarial Networks (GANs) · Object-Oriented Programming (OOP) · Algorithms · Software Development · Model-view-viewmodel (MVVM) · Android Development · Microservices · Project Management · Site Reliability Engineering · Physical AI · ROS2 · Data Build Tool (DBT) · Retrieval-Augmented Generation (RAG) · Generative AI · Stable Diffusion · AI Software Development · AI Agents · C# · TypeScript · Data Structures · Graph Algorithms · Dynamic Programming · Design Patterns · S3 · Lambda · DynamoDB · SQS · SNS · CloudWatch · Glue · Amazon Redshift · ECS · IAM · PySpark · Apache Airflow · ETL/ELT Pipelines · Fault Tolerance · Scalability · PyTorch · TensorFlow · SentenceTransformers · Pinecone · Weights & Biases · Quantization · Reinforcement Learning · Spring Boot · FastAPI · Node.js · Express · Kubernetes · PostgreSQL · CI/CD · GitHub Actions · Distributed Tracing · OAuth 2.0 · REST APIs · Git · Linux