Sharvil Oza

sharvil010@gmail.com | github.com/so-19 | so-19.github.io | LinkedIn

Education

Dhirubhai Ambani Institute of Information and Communication Technology Gandhinagar, India
Bachelor of Technology in Computer Science and Engineering *Oct. 2022 – May 2026*

- **Relevant Coursework:** GPU Architecture, Deep Learning, Computer Vision, Quantum Machine Learning, Reinforcement Learning, Large Language Models, Natural Language Processing, Time Series Analysis, Machine Learning, Data Structures and Algorithms

Work Experience

Research Intern June 2025 – Present
Technion - Israel Institute of Technology

- Conducting research on predictive learning algorithms within reinforcement learning frameworks, investigating emergent latent representations and their applications to artificial intelligence systems.
- Analyzing high-dimensional simulated neural spike train data to model cognitive processes, applying techniques like Principal Component Analysis (PCA) and Variational Autoencoders (VAEs) to uncover low-dimensional neural manifolds.
- Developing and implementing novel deep learning architectures (e.g., contrastive learning, self-attention) for extracting meaningful latent representations from complex neural data, focusing on unsupervised learning techniques.

Research Intern December 2024 – April 2025
University of New South Wales, Business School

- Developed deep learning models for supply chain optimization that reduced forecasting errors , leveraging advanced algorithms to enhance operational efficiency .
- Applied 5+ statistical techniques, including correlation analysis, VCF, and regression models, to derive actionable insights that improved decision-making accuracy by 32% in supply chain management.
- Conducted crop yield prediction using ensemble machine learning algorithms, achieving 91% accuracy and optimizing resource allocation .

Research Intern May 2024 – Aug 2024
Georgia Tech Financial Services Innovation Lab (FSIL)

- Developed and backtested 7 quantitative trading strategies, including Pairs Trading and statistical arbitrage, achieving a 14.3% average improvement in Sharpe ratio compared to benchmark strategies.
- Strengthened algorithm robustness by designing a Risk Metric class encompassing 25+ risk metrics, optimizing portfolio allocations that outperformed benchmark returns .
- Built a custom data wrapper for Polygon API integration that reduced data processing time , enabling real-time data fetching for 200+ financial instruments simultaneously.

Projects

LMFusion Paper Implementation | *Pytorch* March 2025 – March 2025

- Implemented the LMFusion framework to extend pretrained language-only LLMs with multimodal generative capabilities, integrating modality-specific attention and feedforward modules for text and image processing.
- Developed custom dataloaders and loss functions to support joint training on image-caption datasets using both cross-entropy for text and diffusion-based objectives for image generation.
- Engineered a modular multimodal training pipeline, preserving language capabilities by freezing text modules and training image modules, achieving efficient cross-modal attention and high-fidelity image generation.

Trading Engine | *Python, MySQL* June 2024 – July 2024

- Developed user profiles and portfolios with personalized features, enhancing user experience and enabling seamless tracking of trades and holdings across 50+ securities.
- Engineered core trading functionalities by designing and implementing an L3 Order Book .
- Created efficient order-matching algorithms using Queue Data Structures that improved execution speed by .

Investment Portfolio Optimization | *Python* Jan. 2025 – Feb. 2025

- Developed a Modern Portfolio Theory (MPT)-based asset allocation model that improved risk-adjusted returns by 16.4% across 8 diverse market scenarios.
- Designed and implemented Monte Carlo simulations with 10,000+ iterations to assess portfolio risk, reducing Value-at-Risk (VaR) estimates by 22% compared to traditional methods.
- Automated financial reporting processes with dynamic dashboards that reduced analysis time and enabled real-time tracking of performance metrics.

Quantitative Research Project | *Python* Jan. 2025 – Feb. 2025

- Conducted in-depth analysis of emerging market inefficiencies across 15+ markets, identifying pricing discrepancies that yielded 7.2% annualized alpha with a Sharpe ratio of 1.8.
- Applied advanced statistical arbitrage techniques on 50+ asset pairs, generating 11.3% average returns with 63% lower volatility compared to traditional long-only strategies.
- Created a factor-based investment approach integrating 8 macroeconomic indicators and 12 technical signals that outperformed market benchmarks .

RAG Pipeline with Local LLM | *Python, PyTorch* December 2024 – January 2025

- Developed a Retrieval-Augmented Generation (RAG) pipeline that processed 1000+ pages of PDF documents into 12,000+ text chunks with 92% retrieval accuracy.
- Integrated Google/GEMMA-2B-IT local language model that reduced inference time while maintaining response relevance .
- Optimized the RAG pipeline's performance by implementing vector search techniques that improved query processing .

Reinforcement Learning-Based Drone Stabilization Simulation | *Python, PyBullet*October 2024 – December 2024

- Developed a physics-based drone simulation in PyBullet that achieved 94.7% success rate in autonomous recovery from inverted positions .
- Engineered a RL-based reward algorithm that reduced stabilization time through optimized control inputs based on real-time environmental feedback.
- Created a detailed URDF file for the drone model and enhanced the RL algorithm with 13 state variables, achieving faster convergence during training .

Weather Prediction Model | *Python, TensorFlow* August 2024 – September 2024

- Developed an advanced weather prediction model for tropical Indian climate using LSTM and Attention mechanisms, achieving 93.8% accuracy across 7 climate variables.
- Collected and preprocessed 10+ years of daily weather data .
- Successfully deployed the model as a Flask API that handles 500+ concurrent requests with average response time of 120ms.

Technical Skills

Languages: Rust, Python, C/C++, SQL (Postgres, MySQL), CUDA
Libraries: PyTorch, TensorFlow, NumPy, Pandas, scikit-learn, Matplotlib, Seaborn
Developer Tools & Platforms: Git, Linux/Bash, Jupyter Notebooks, Flask, FastAPI

Extracurricular Activities

Solvay Business Game Brussels, Belgium

Selected Participant *March. 2025*

- Collaborated on McKinsey's pitch challenge focusing on workforce strategy, developing innovative solutions for skill gaps and preparing the workforce for Industry 4.0.
- Conducted comprehensive financial modeling for the CACEIS challenge, analyzing break-even analysis for an incubation hub and creating performance projections that identified key growth opportunities.
- Participated in BDO's negotiation challenge, demonstrating strategic communication skills and successfully reaching win-win agreements in complex business scenarios.