SHREYAN SOOD
92092, La Jolla, California, USA
shreyansood@gmail.com | +1 8583196576 | Portfolio | GitHub | LinkedIn

EDUCATION

University of California San Diego **September 2022 – December 2024**
Master of Science, Data Science *GPA (3.98/4.0)*
Teaching Assistant for DSC-261: Data Ethics, DSC-291: Statistical Models

Delhi Technological University, New Delhi **August 2018 – June 2022**
Bachelor of Technology, Mathematics and Computing Engineering *CGPA (8.73/10)*

WORK EXPERIENCE

Machine Learning Engineer **September 2024 – Present**
Prompt Inversion AI, Dover, Delaware
- Scaled and optimized an AI driven API service for automating candidate-job matching and sending recommendation emails to employers, leveraging crewAI platform for LLM agents and RAG tools, with Pinecone vector database.
- Refactored integration tests for the LLM agent pipeline, FastAPI endpoints, database interactions and resolved critical bugs.

Analytical Scientist Intern **June 2023 – December 2023**
FICO, San Diego, California
- Devised and implemented **adaptive time-series algorithm** to monitor the latent features of a State-of-the-Art **fraud detection neural network** and **trigger real-time alerts** for significant shifts in distributions. Validated the algorithm for **15 major clients**.
- Developed an ETL pipeline to compute and visualize the distributions of terabyte-scale transaction datasets using PySpark.
- Conducted calibration experiments to simulate drastic shifts in customer behavior and cluster sophisticated fraud schemes.

Research Engineer **January 2022 – September 2022**
Collablens, Haryana, India | Funded by MIT Media Lab
- Developed and deployed an **AI station** for automated drop testing of flour packets. Integrated dynamic **cloud-based** modules for Spillage Detection, Pose Estimation, Orientation Checks, Depth Sensors, and other real-time insights from **live video footage**.
- Helped secure a contract to deploy the system in **50 factories**. Helped raise over **$200,000** in investment offers.
- Prototyped a versatile **Computer Vision System** for real-time defect detection in laser-engraved wooden boards on a moving assembly line with cloud-based result logging. Achieved 95% accuracy and a mean inference time of 2.5 seconds per board.

Machine Learning Intern **May 2021 – December 2021**
Hypertechpreneurs, Haryana, India
- Developed and productionized Vehicle Damage Detection Model utilizing Mask R-CNN for Instance Segmentation to automate vehicle inspections. Extrapolated it to a Severity and Cost Estimation pipeline. Helped raise over **$50,000** in funding.
- Developed systems for OCR and Object Detection in dynamic environments while maintaining a minimum accuracy of **90%**.

RESEARCH EXPERIENCE AND PUBLICATIONS

Research Fellow under Prof. H.C. Taneja, Delhi Technological University **September 2021 – May 2022**
- Outperformed the Black-Scholes Model for option pricing using LSTM, MLP, XGBoost and SVM leveraging real market data.
- **Sood, S., Jain, T., Batra, N., Taneja, H.C. (2023). Black–Scholes Option Pricing Using Machine Learning.**

Research Assistant under Prof. Anurag Goel, Delhi Technological University **February 2021 – August 2021**
- Integrated State-of-the-Art CNN based Object Detection Networks (CenterNet, Faster R-CNN) with self-devised algorithms for selective lossy image compression techniques to enhance the storage and processing efficiency in autonomous systems.
- **S. Sood and Y. Ahuja, "Selective Lossy Image Compression for Autonomous Systems."**

PROJECTS

Rubik's Cube 3D Visualizer & Deep Reinforcement Learning (DRL) Solver **August 2024 – September 2024**
- Developed a NxN Rubik's Cube visualizer with quaternion-based rotations and implemented a Monte Carlo Tree Search algorithm augmented with a DRL network, achieving 97% solution rate for 6-move scrambles and sub-second solving times.

MediLoRA: LLM for medical Q&A with QLoRA **October 2023 – January 2024**
- Fine-tuned OpenHermes-2.5-Mistral-7B with Q-LoRA on 300M medical text tokens. Improved PubMedQA and MedQA accuracy by over 20% and matched State-of-The-Art 70B open models on MMLU-Medical with 0.05% of the data size.

TECHNICAL SKILLS

- **Programming** **:** Python, SQL, R, C++, MATLAB, JavaScript, HTML, CSS
- **Technologies** **:** Pandas, PyTorch, Keras, TensorFlow, OpenCV, AWS, PostgreSQL, Bash, PySpark, Docker, GCP
- **Skills** **:** Data Science, Deep Learning, Computer Vision, Natural Language Processing, Hypothesis Testing, MLOps