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 **LLM-driven** FastAPI service utilizing a multi-agent AI platform for intelligent task automation.
- Engineered an RAG pipeline with LangChain and PGVector, with advanced prompt engineering strategies for context retrieval.
- Orchestrated automated CI/CD workflows for containerized deployment to AWS EC2 with comprehensive integration testing.

Analytical Scientist Intern **June 2023 – December 2023**
FICO, San Diego, California
- Implemented an **adaptive time-series algorithm** to monitor the output 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
- Deployed an **AI station** for automated drop testing of flour packets. Developed **cloud-based** modules for Spillage Detection, Orientation Checks, Depth and Pressure Sensors, and other real-time insights from **live video streams**.
- 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 products, enabling automated quality control on assembly lines. Achieved **95% accuracy** and a mean inference time of **2.5 seconds per board**.

Machine Learning Intern **May 2021 – December 2021**
Hypertechpreneurs, Haryana, India
- Productionized a Vehicle Damage Detection Model utilizing Mask R-CNN for Instance Segmentation to automate inspections.
- Implemented robust OCR and Object Detection systems achieving **90%+** accuracy in dynamic real-world environments.

PROJECTS

GPT Document Proofreader App **August 2024 – September 2024**
- Developed a Python GUI application for multi-format document proofreading and grammar correction using *GPT-4o mini*.
- Integrated customizable context-aware prompts, asynchronous processing, and efficient caching and rate-limiting mechanisms.

Rubik's Cube 3D Visualizer & Deep Reinforcement Learning Solver **August 2024 – September 2024**
- Developed NxN Rubik's Cube 3D visualizer and implemented Monte Carlo Tree Search algorithm optimized with a Deep Reinforcement Learning ResNet. Achieved 71% solution rate for 9-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 model performance on MMLU-Medical with 0.05% of the data size.

RESEARCH EXPERIENCE AND PUBLICATIONS

Research Fellow under Prof. H.C. Taneja, Delhi Technological University **September 2021 – May 2022**
- Sood, S., Jain, T., Batra, N., Taneja, H.C. (2023). Black–Scholes Option Pricing Using Machine Learning. [ICDSA 2023]

Research Assistant under Prof. Anurag Goel, Delhi Technological University **February 2021 – August 2021**
- S. Sood and Y. Ahuja, "Selective Lossy Image Compression for Autonomous Systems." [STSIVA 2021]

TECHNICAL SKILLS

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