

SUVRAT JAIN

Rochester, NY | 585-710-8813 | suvijain@gmail.com | simplysuvi.com | github.com/simplysuvi

WORK EXPERIENCE

DIESELCORE

Houston, TX

AI/ML Engineer

May 2024 – Present

- Developed & deployed an AI chatbot using RAG pipelines (LlamaIndex, FAISS, Pinecone, ChromaDB), implementing BM25, Cohere Reranker, ColBERT for retrieval, and integrating LLMs on Microsoft Teams, reducing document lookup time by 60%.
- Built document parsing and embedding pipelines, processing documents at scale, with Python, LlamaIndex, OpenAI embeddings, and optimizing retrieval precision via reciprocal rank fusion (RRF) & hybrid search.
- Designed a vision-based part identification system for edge devices (tablets/phones), enabling real-time classification from live camera streams using YOLOv8, CLIP, similarity matching, achieving ~92% accuracy.
- Integrated vision AI with ERP systems, automating warehouse part tracking & logging, reducing manual errors by 40%, via FastAPI, OpenCV, PyTorch, REST APIs.

GOLISANO INSTITUTE FOR SUSTAINABILITY, RIT

Rochester, NY

Machine Learning Engineer

May 2022 – May 2024

- Developed & deployed automated vision system for sorting remanufactured parts (YOLOv8, Siamese Networks, PyTorch, TensorFlow, Docker) achieving 95% accuracy, 5s cycle time. reducing manual labour time by ~50%.
- Built a garment classification & segmentation system (YOLOv8, UNet, SAM, PyTorch, NIR imaging) integrating GCode generation for automated laser cutting, optimizing material recovery by 30%.
- Designed a hybrid YOLO + similarity matching system, preventing unknown misclassification, increasing identification accuracy by 30%. (PyTorch, FAISS, OpenCV)
- Designed scalable ML pipelines, collecting 18K+ images, expanding datasets by 30%, boosting accuracy by 12%, and deploying cross-system solutions with Docker, FastAPI, MLflow.

EDUCATION

ROCHESTER INSTITUTE OF TECHNOLOGY

Rochester, NY

Master of Science (M.Sc.) in Data Science

2021 - 2023

Coursework: Data Analysis, Software Engineering; Time Series Analysis & Forecasting; Database Design

PROJECTS

- [HUE Vision](#): Developed a real-time eye-tracking web app using TensorFlow.js, clmtrackr.js, and JavaScript, integrating ML-based gaze prediction & heatmap visualization for in-browser inference. ([Demo](#))
- [Eventify-U \(BrickHack 9 2023\)](#): Developed a web app for personalized event recommendations using AngularJS, Node.js, Spring Boot, MySQL, and NLP-based interest matching, integrating custom APIs for event data and location services.

PUBLICATIONS

Islam, A., Jain, S., Nenadic, N.G., Thurston, M.G., Greenberg, J., & Moss, B. (2024). Image-based machine learning in automotive used parts identification for remanufacturing. In N. Nasr (Ed.), Technology innovation for the circular economy (Chapter 39). <https://doi.org/10.1002/9781394214297.ch39>

SKILLS

Programming: Python, Java, C++, SQL, HTML, CSS, JavaScript, jQuery

Machine Learning & AI: Scikit-learn, NumPy, Pandas, OpenCV, PyTorch, TensorFlow, Keras

Tools & Platforms: Tableau, Power BI, PySpark, FastAPI, Flask, Selenium, Jupyter, Streamlit, Salesforce Cloud, AWS, GCP, Azure, Docker, Git, CI/CD, MLflow, LangChain, LlamaIndex, Vector database