Trishala Thakur

Chicago, IL (Open to relocation) | +1 (720) 490-0543 | trrshla@gmail.com | GitHub | LinkedIn

PROFESSIONAL EXPERIENCE

Machine Learning Data Scientist, University of Chicago **June '25 - Present**

- Built an agent-orchestrated pipeline **(Gemini 2.5 Flash)** to automate face-name matching at 92% F1 score, eliminating manual annotation
- Developed page **layout classifier (DNN, OCR)** to segment 1TB+ yearbook corpus by layout complexity, achieving 71% accuracy
- Prompted **GPT-5-nano** to extract insights from 1.2M papers, enabling computational social science analysis and saving 12k review hours

Data Scientist, Schneider Electric **August '21 - July '23**

- Built **LightGBM** classifier predicting equipment failures 5 days in advance at 87% precision, reducing downtime across 6 assets
- Integrated **predictive maintenance** model into mobile app used by 10+ field engineers, reducing troubleshooting time by 30 min/case
- Engineered features from high-frequency sensor data (lag, rolling features, failure patterns) and built **SQL** pipeline for scalable training

Data Science Intern, Schneider Electric **February '21 - July '21**

- Created 10+ **Tableau** dashboards tracking failure metrics and sensor anomalies, enabling root-cause analysis for 3 critical failures
- Implemented RaspberryPi sensor system to **stream real-time data** into Digital Twin model via SQL pipeline for predictive maintenance
- Led talks and Q&A sessions with tech leaders and lab engineers, identifying three opportunities for data-driven product improvement

RESEARCH

Graduate Research Assistant, CIRES **May '24 - August '24**

- Identified drought signals in satellite-derived data by linking anomalies in monthly regional averages to observed drought events
- Designed a **Siamese CNN** architecture on rasterized monthly-averaged satellite data to forecast regional droughts with 71% recall

Graduate Student Assistant, CIRES **January '24 - April '24**

- Trained four **XGBoost** models on **DistilBERT embeddings**, achieving up to 78% F1 across climate change text classification tasks
- Deployed models as **Flask RESTful API** integrated in Label Studio, automating annotation and saving 2 hours/week in manual effort

PROJECTS

Automated Lost-Item Retrieval System **March '25 - May '25**

- Developed a **multimodal search application** using **CLIP-ViT-B/32**, performing embedding-based similarity matching for image and text
- Deployed on **GCP** to support scalability, stored embeddings with **FAISS** and created a **Streamlit UI** for secure public access

Mock Interview Agent for Personal Growth **December '24 - May '25**

- Built a **retrieval-augmented** based interview agent that generates role-specific mock questions grounded in curated job descriptions
- Leveraged **Chain-of-Thought prompting** to evaluate responses and deliver personalized feedback, improving user experience

Remote Fine-tuning Framework **September '24 - December '24**

- Built a **cloud-based fine-tuning** platform with Streamlit UI to perform real-time model tuning on GPU VMs without local compute
- Enabled **rapid deployment** of fine-tuned models by exposing them into an **API**, minimizing integration effort for downstream pipeline

EDUCATION

Master of Science in Data Science | University of Colorado Boulder, USA

Data Mining, Machine Learning, Deep Learning, Cloud Computing, Natural Language Processing, Computer Vision, Data Structures

SKILLS

Programming: Python, R, NumPy, pandas, PyTorch, TensorFlow, HuggingFace, scikit-learn, SciPy, OpenCV, Matplotlib, Seaborn, Plotly

ML Techniques: LLMs, RAG (Retrieval-Augmented Generation), Chain-of-Thought prompting, CNNs, time-series forecasting, XGBoost

Cloud & Tools: Google Cloud Platform, AWS, HPC (High-Performance Computing), RESTful API, Git, SQL, Linux, Docker (learning)