

SHAURYA PATHAK

Cerritos, CA · shauryapathak24@gmail.com · shauryapathak.xyz · 562-330-8614

EDUCATION

University of California, Riverside

Riverside, CA

BS Computer Science with Business Application *GPA: 3.82 (Major GPA:4.0)*

Sep 2021 - Dec 2024

Student Assistant for Hamed Mohsenian-Rad

Relevant Coursework: Data Analysis, Machine Learning, Artificial Intelligence, Databases

PROFESSIONAL EXPERIENCE

NVIDIA | Innovation Engineering Team

Remote

Data Engineer for NeMo LLM

August 2024 - Present

- Designed dashboards to track annotation quality and throughput for Nvidia LLM Data Factory, enabling leadership to monitor the performance of 860 annotators processing over 500K multimodal annotations monthly.
- Implemented GPT detection system to ensure annotator compliance and assess GPT-related biases, flagging 3,000+ potential GPT-assisted annotations.
- Developed 15 linguistic metrics to assess annotation similarity and consistency, supporting annotators' progression toward more linguistically accurate submissions.

Skills Developed: *Data Engineering, LLM Bias Detection, Dashboarding, Analytical Reporting*

Tools Used: *Python, Data Visualization Libraries, SQL, Large Language Model Frameworks, NLTK*

NASA | City of Los Angeles

Los Angeles, CA

Student Researcher

September 2021 - Present

- Utilized satellite data to estimate PM2.5 levels in Los Angeles, without the usage of ground level sensors.
- Integrated a pipeline to display machine learning models on ai-aq.com.
- Created a live traffic count model by combining static traffic count data with real-time traffic flow.

Skills Developed: *Data Analysis/Modeling, Machine Learning Implementation, Website Integration*

Tools Used: *Python, Keras, Convolutional Neural Networks, ArcGIS, Tensorflow, Scikit, Matplotlib, AWS (Sagemaker, EC2, Lambda), CesiumJS*

AllCare Therapies

Cerritos, CA

Machine Learning Intern

June 2024 - August 2024

- Automated 90% of fax handling utilizing Serverless frameworks.
- Created algorithm to automate multi-classification document splitting.
- Built a classification logistic regression algorithm to categorize faxes into distinct buckets.

Skills Developed: *Cloud Architecture, Data Engineering, Machine Learning, Automation*

Tools Used: *AWS Lambda, Serverless Framework, DynamoDB, AWS Step Functions, Python*

PROJECTS

Chess State Recognition *Computer Vision, Data Engineering, Machine Learning*

- Utilized Unity to generate artificial data and annotations for various chess positions, creating a rich dataset tailored for training machine learning models in chess state recognition.
- Implemented YOLO object recognition technology to pinpoint the positions of each chess piece on the board.
- Integrated the object recognition output with a system designed for live chess game transcription.

Chess Bot: Adaptive Gameplay through Hyperparameter Optimization *Python, Machine Learning Algorithms*

- Tuned hyperparameters (regularization, nodes) to mimic human gameplay levels.
- Built a machine learning model to analyze human chess strategies.