

Yiran (Jenny) Shen

(984) 837-3114 | jes038@ucsd.edu | [Personal Homepage](#) | [LinkedIn](#) | [GitHub](#)

Education

University of California, San Diego

PhD in Computer Science

Advisor: Prithviraj (Raj) Ammanabrolu

La Jolla, CA

Sep 2024 - Present

Duke University

Master of Science in Interdisciplinary Data Science

Honors: Dean's Research Award

Durham, NC

Aug 2022 - May 2024

The University of Melbourne

Bachelor of Commerce in Economics and Finance, Minor in Statistics (First-Class Honors)

Honors: Melbourne International Undergraduate Scholarship (50% fee remission), International Graduate Merit

Scholarship, Melbourne Graduate Scholarship, Leaders in Communities Award, DJI RoboMaster 2021 University AI

Challenge third prize

Melbourne, VIC

Mar 2019 - Dec 2021

Research Interests

My research interests lie within the area of trustworthy AI, with a particular focus on the interpretability, fairness, reasoning, and robustness capabilities of different kinds of ML models. Particularly, I am passionate about large language models and developing advanced multi-agent systems, exploring their synergies to enhance collaborative problem-solving, and investigating their social impacts and ethical implications in decision-making.

Research Experiences

Dynamic Reasoning of Language Modeling through Multi-Role Discussion

Advised by [Dr. Liuyi Yao](#) from Alibaba Group

- Propose a differentiated multi-role LLM discussion system to enhance problem decomposition, reasoning, and answering capabilities, implementing structured role-based interactions (e.g., teacher-student) with automated scene generation, ensuring focused discussions and improved error correction across iterations.
- Contributed to [Agentscope](#), an open-source project with 5,000+ GitHub stars

Interpretable Alignment of LLMs with Multi-Dimensional Reward Signals

Advised by [Prof. Brandon T. Fain](#) from Duke Computer Science Department

- Developed a novel approach to align LLMs using interpretable methods. Leveraged representative sub-dimensions and a performant out-of-the-box LLM to generate rewards, reducing reliance on expensive human-curated preference data. Fine-tuned and evaluated two models using the new approach and traditional black-box rewards. Demonstrated maintained performance with superior explainability and flexibility through advanced LLM-based evaluation methods, enhancing transparency in RLHF.
- **Related Publication:** Yiran Shen*, Aditya John*, Brandon Fain. Explainable Rewards in RLHF Using LLM-as-a-Judge. *Under Submission*.

Causal Shapley Value for Federated Learning Incentive Mechanism

Advised by [Dr. Yaliang Li](#) from Alibaba Group

- Proposed a novel causal valuation function to fairly measure participant contribution to federated learning while preserving local data privacy. Leveraged direct linear non-Gaussian acyclic model to identify the causal graph, and distinguish between direct and indirect effects among participants and the aggregation server.
- Introduced a contribution score function, directing Gaussian noise-based reward allocations, with empirical validation on two real-world datasets confirming its enhanced capability in federated learning reward distribution.
- **Related Publication:** Yiran Shen, Jinduo Liu, Shunqiang Liu, Liuyi Yao, Keqin Zhong, Chenglin Miao, Yaliang Li. Incentive Mechanism in Federated Learning with Causal Shapley Value. *Under Submission*.

Simulating and Optimizing Hospital Scheduling Through Machine Learning

Advised by [Prof. Ethan Fang](#), [Prof. Yehua Wei](#), and [Prof. Bruce W. Rogers](#) from Duke Rhodes Information Initiative

- Leveraged machine learning methods like XGBoost to predict surgical case length, and effectively extracted and analyzed historical time schedules from previous surgical data. Employed the simulated annealing optimization technique to finetune time schedules, targeting a reduction of ~20% in expected bed usage variance.

- Designed and implemented an interactive user interface, allowing stakeholders to simulate, visualize, and evaluate time schedule modifications. Empowered domain experts to assess the impact of various scheduling policies, laid the foundation for an automated and intelligent scheduling system at Duke University Hospital.

Technical Projects

LLM Network Acceleration and Compression Research (1yr Capstone Project with Proofpoint)

Advised by [Dan Salo](#), [Ryan Skinner](#), and [Altamash Rafiq](#) from Proofpoint ML Lab

- Finetune and conduct experiments on neural network model compression techniques such as quantization, distillation, and pruning on LLMs with ~3B parameters to increase inference speed by 50% and prevent an accuracy loss greater than 2%. The goal of the compressed model is to generalize the detection of cyber threats in emails, documents, and other communication channels.

Domain Adaptive Ophthalmic Image Segmentation using Transfer Learning [\[report\]](#)

- Improved Diabetic Macular Edema detection using transfer learning on limited OCT images. Applied finetuning on pre-trained models from MM-Segmentation library, achieved enhanced model performance with a mDice score of 79.38% and mIoU of 70.64%, with the final UNet model incorporating DeepLabV3 heads.

Real-Time Facial Emotion Detection from Partially Obstructed Faces using CNNs [\[code\]](#)

- Utilized the VGG-16 model, outperforming ResNet50 by 7%, to detect emotions from partially obstructed faces. Applied saliency maps to enhance interpretability and deepen insights into human emotion perception. Integrated the trained VGG-16 model with OpenCV's haar-cascade to craft a real-time emotion detection system.

Customer Profitability Analysis and Predictions using LSTM for Targeted Marketing [\[code\]](#)

- Classified customer profitability into three categories using K-means clustering on standardized monthly sales data. Implemented a customized LSTM model to predict consumers' profitability, achieved an accuracy rate of 96.65% and improved targeted marketing strategies for retailers to increase revenue growth.

Dynamic On-Campus Cup Circulation Web App with Real-Time Analytics [\[code\]](#)

- Developed a robust Streamlit web application for on-campus coffee cup circulation, with the backend database hosted on AWS RDS for enhanced scalability. Introduced a real-time dashboard for providing dynamic visual insights into cup usage, which led to a 20% reduction in cup shortages during peak hours.

Professional Experiences

Alibaba Group

Institute for Intelligent Computing, Research Intern

Hangzhou, CN

Jan 2024 – Present

- Enhance problem-solving capabilities within multi-agent systems through role-playing simulations; work on human-agent cooperation aimed at generating high-quality queries to maximize the value of human feedback.

Duke University

Teaching Assistant

Durham, NC

Jan 2023 – Sep 2023

- Tutored and graded assignments for over 400 students in two statistics courses (STA101 and DECISION518Q) covering topics such as inference, regression modeling, A/B testing, and time series in R.

Innoway

Quantitative Research Intern

Beijing, CN

Apr 2022 - Jul 2022

- Advised top global tech firms on long-term localization strategies through customer industry analysis. Automated quarterly competitive evaluations to track investment and innovation trends, reducing turnaround time by 40%.

Yarra Riverkeeper Association

Business Analytics Intern

Melbourne, VIC

Aug 2021 - Nov 2021

- Developed a scalable web crawler, harvested data from 50,000 web pages (10M rows) and derived key insights for informed habitat preservation decisions.
- Led attribute analysis through roundtable events, engaging 15+ corporate partners to shape ESG strategies.

Skills and Interests

Computer Proficiency: Python, Java, R, SQL, MATLAB, Tableau, JavaScript, CSS, HTML, PHP

Frameworks: Sklearn, PyTorch, Tensorflow, Numpy, Pandas, Dask, Flask, MySQL, Git, OpenCV, AWS, Streamlit

Interests: Hiking, traveling, watching soccer, sudoku, taekwondo, documentaries