Jasmine Lo

jasmine.lo@columbia.edu | jasminelo2020.github.io/jasmine-lo

EDUCATION

Columbia University New York, NY
Master of Science in Data Science, GPA: 3.79/4.0 Dec 2025
- Courses: Machine Learning, Prob & Stats, Forecasting, Finance & Structuring, Deep Learning, Computer Vision

University of California, San Diego La Jolla, CA
Bachelor of Science in Data Science, GPA: 3.86/4.0 Jun 2024
- Courses: Data Structures & Algorithms, Statistics, Data Management, Data Visualization, Recommender Systems

EXPERIENCE

Aiello Research Group, Columbia University Irving Medical Center New York, NY
Graduate Research Assistant Jan 2025 - Present
- Designed and implemented RAG AI agent that retrieves and synthesizes public health data from research databases, improving information retrieval efficiency and reducing search time
- Optimized Whisper machine learning model to automate scoring of thousands of audio files with 91% accuracy
- Executed data auditing across 2 projects and assisted with quality control to ensure data integrity of transcriptions

NextLabs San Mateo, CA
Marketing Analytics Intern Jun 2025 - Aug 2025
- Built Google Analytics dashboard with top 8 KPIs for website traffic analysis to provide insight on business needs
- Blended multiple sources of data from LeadInfo, Google Search Console, and LinkedIn into interactive dashboard
- Extracted actionable insights to support digital marketing team in optimizing targeted B2B campaigns and SEO

Catalis Alpharetta, GA
Data Analyst Intern Jun 2024 - Aug 2024
- Implemented data management techniques with Python, MS Excel, and Power BI to reconcile data between CRM (Salesforce) & ERP (NetSuite) business applications, building the Customer Master data structure
- Designed a Python/Pandas pipeline to automate cleaning, transforming, and mapping for 72K+ accounts
- Wrote 5+ scripts to quickly scrape and generate datasets, reducing time spent on generating unique IDs by 80%

Halicioglu Data Science Institute, UC San Diego La Jolla, CA
Undergraduate Teaching Assistant Jan 2023 - Jun 2024
- Collaborated on development of babypandas library, and created 2 new projects for UCSD data science courses
- Updated infrastructure of course problem set site, practice.dsc10.com, to allow topical searching for 10+ topics
- Held weekly office hours to provide a deeper understanding of lectures and labs, catering to 500+ students

Ayers Lab, Veterans Medical Research Foundation La Jolla, CA
Research Assistant Jun 2023 - Apr 2024
- Calculated and inputted t-scores into REDCap to allow statistical analysis on hoarding treatment for 100+ patients

PROJECTS

Large Language Model Based Active Learning Sep 2025 - Present
- Designed and set up end-to-end LLM-based pipeline for counterfactual data augmentation to enhance decision making in active learning

Robust and Interpretable NN Models for CV and NLP Sep 2023 - Mar 2024
- Dissected image classification deep neural network with CLIP-Dissect, analyzing 4,000+ neurons to understand neuron activations and interpretability, and mapping key neurons to facial emotion concepts
- Applied research to refine VGGNet-based deep learning model for emotion recognition, achieving 0.11% increase in test accuracy by applying strategic neuron weight modifications

Shazam 2.0 Jan 2023 - Jun 2023
- Led a team of three and coordinated between team and project directors to run project smoothly
- Implemented AWS's EC2 instance to transfer 300GB of data onto remote clusters and constructed a CNN model classifying songs by genre, fine-tuning parameters to increase accuracy by 12%
- Created a Streamlit web app taking in real-time raw audio and returning top 3 song genres and information

SKILLS

Languages: Python, Java, SQL (Postgres), JavaScript, HTML/CSS, R
Developer Tools: Git, Jupyter Notebook, VS Code, RStudio, Power BI, Google Analytics 4, Looker Studio, Excel
Libraries: Pandas, Scikit-learn, PyTorch, Beautiful Soup, Seaborn, d3, Streamlit