

# JANYS (JIAYANG) LI

(424) 440-3485 | jli25@g.ucla.edu | [www.linkedin.com/in/janys-li](https://www.linkedin.com/in/janys-li) | Los Angeles, CA

## EDUCATION

University of California, Los Angeles (UCLA) | GPA: 3.98 / 4.0

Los Angeles, CA

B.S. in Statistics and Data Sciences & Cognitive Science, with Specialization in Computing

Expected August 2025

- **Honors:** Dean's List; **Dean's Prize** for Excellence in Research and Creative Inquiry (top 20 out of 1,658)
- **Relevant Courses:** Machine Learning, Data Mining, Computer Science, Linear Algebra, Multivariable Calculus, Probability, Linear Models, Computing & Optimization, Experiment Design & Analysis, Computer Vision

## SKILLS

- **Programming:** Python (Pandas, Numpy, PyTorch, Tensorflow, SKLearn), R, SQL, Excel, C++, MATLAB
- **Data Tools:** Databricks, Tableau, Power BI, Azure, Spark, LaTeX, AWS, Git, OpenAI, AsyncIO
- **Data Science Methods:** A/B Testing, Data Mining, Machine Learning, Database Management, Data Visualization, Time Series Analysis, User Experience (Figma, HTML, Javascript)

## PROFESSIONAL EXPERIENCE

Ekimetrics

Hong Kong SAR

Data Science Consultant Intern

Jun 2024 – Dec 2024

- Provided marketing budget allocation suggestions that optimize return on investment for a Fortune 500 client through building 10+ **Bayesian regression models** with Python to predict sales from budget allocation scenarios
- Built Retrieval-Augmented Generation (RAG) pipeline with **Azure** and **OpenAI** to automatically extract numerical information from PDF documents, increasing sustainability report-screening efficiency by 70% for an ESG client
- Boosted processing speed by 210% with parallel optimization and distributed processing on **Spark** and **Databricks**
- Processed 10,000+ rows of data and visualized time-series market sales volume trends using automated, interactive **Power BI** dashboards to demonstrate actionable budget optimization insights for external stakeholders

The Adaptive Cognition, Memory, and Emotion Lab @ UCLA Psychology

Los Angeles, CA

Research Assistant, supervised by Prof. David Clewett and Mason McClay

Mar 2022 – Sep 2024

- Segmented participants into 3 different cohorts based on symptom levels with **K-Means clustering** and discovered a significantly lower likelihood of emotional transition among groups with high psychological symptom levels
- Modeled latent emotional states and the probability of changing between states with **Hidden Markov Models**
- Designed 300 participant-experiment to study relationships between depression, PTSD, and emotional fluctuations
- Conducted in-person **usability testing** & interviews with 10+ participants to test the usability of the interactive experimental tool, identifying 20+ improvement insights that increased user confidence by 24%

Ozcan Research Lab @ UCLA Electrical and Computer Engineering

Los Angeles, CA

Research Assistant, supervised by Prof. Aydogan Ozcan and Hanlong Chen

Sep 2023 – Jun 2024

- Improved convolutional networks using **Pytorch** to enhance training efficiency of medical image reconstruction, providing fast reconstruction of medical images that support instant endoscope imaging for medical diagnosis
- Implemented physics-informed **loss function** that potentially improves training efficiency by 50%
- Improved processing efficiency of training dataset by 150% with Harris Corner detection-based program in **Python**

Ipsos

Shenzhen, China

User Experience Research Intern

Jul 2022 – Sep 2022

- Summarized and visualized 200+ gain/pain points from 1000+ pages of user interview results to highlight key user characteristics and areas of product improvement for smart cooking appliances with **Tableau** and PowerPoint
- Led and presented competitive **user research** on 10 location-based services to 30 clients at innovation workshop
- Investigated daily consumption and online shopping trends of high-end consumers based on 20+ industry reports

## PROJECTS

Alcohol Intoxication Status Prediction Kaggle Competition

September 2023 - December 2023

- Predicted patient intoxication status based on biological and demographic data to inform timely health intervention
- Deployed **Random Forest** Classifier variants in R to select the most robust variants using 10-fold cross-validation
- Constructed ensemble model with **XGBoost** and **LightGBM** classifiers to enhance prediction accuracy by 3%
- Improved model performance by 3% through imputing missing values based on other known measures
- Ranked 1st of 38 teams with intoxication status classification accuracy = 0.74 (baseline = 0.5)