

# Zikang (Eric) Huang

[zkhuang521@gmail.com](mailto:zkhuang521@gmail.com) | (510) 289-2042 | [linkedin.com/in/zikang-huang/](https://www.linkedin.com/in/zikang-huang/)

## EDUCATION

**University of California, Los Angeles**

*B.S. in Data Theory, Minor in Data Science Engineering*

Cumulative GPA: 3.7/4.0, Dean's Honor List

*Los Angeles, CA*

**Expected Graduation: December, 2025**

## PROFESSIONAL EXPERIENCE

**UCLA Library Data Science Center**

*Data Science Consultant*

*Los Angeles, CA*

**June 2024 - Present**

- Developed content-validation tool using Llama's API to efficiently check researcher-developed Carpentry lessons for quality control, reducing manual review time by 30% and contributing to a 20% increase in project completion rates
- Optimized database management using MongoDB and Apache Spark, supporting researchers in building and analyzing large-scale unstructured data, implemented user access logs and monitored data retrieval speeds, resulting in a 25% increase in data usability and tripled the number of data publications of [DSC Dataverse](#)
- Automated the consultation reservation process by developing serverless Python app using AWS Lambda, integrating Calendly events with LibInsight statistics. Tracked consultation booking in real-time, enabling detailed analysis of engagement trends, contributing to a 15% increase in departmental engagement and 12% rise in campus consultations

**China Construction Bank Fintech**

*Big Data Development Intern*

*Shenzhen, China*

**July 2024 - September 2024**

- Spearheaded the integration of over 100 terabytes of multi-source financial data into the data center, improving data processing efficiency by 17% through optimized ETL pipelines and data warehousing strategies
- Collaborated with cross-functional team to develop and deploy machine learning models that increased loan approval accuracy by 15%, utilizing big data analytics and advance algorithms for risk assessment
- Addressed critical issues related to mis-sampling and database inconsistencies in encrypted PKCS 12 files during data center migration, ensuring data integrity and seamless web server integration to minimize service disruptions

**UCLA Library Data Science Center**

*Data Science Consultant*

*Los Angeles, CA*

**June 2024 - Present**

- Improved 35% algorithm performance through refining risk-score triggering algorithm of the flagship accident and driving habit prediction system. Leveraged time-series and physics-based data to develop KPIs that gauged data quality, enabling advanced analytics for improved driving risk assessments
- Increased 15% user retention and reduced 13% false positive driving events alerts by enhancing SVM classification models to annotate 100+ driving events, resolving inconsistencies in risk score calculations during driving scenarios
- Achieved 10% rise in expected DAU by providing safer driving experiences through designing a CNN based-image and video analysis pipeline from AWS for accurate recognition of relevant traffic signs and signals, increased the precision of object detection by 20% and improved algorithm's real-time decision-making

## ACTIVITIES

**The Data Science Union at UCLA**

*Senior Project Lead*

*Los Angeles, CA*

**September 2022 - Present**

- Cleaned and formatted raw social media text and applied K-Means clustering and LDA for preliminary pattern recognition. Enhanced feature extraction with pre-trained BERT model, subsequently classifying MBTI types with a logistic regression model, culminating in a 74% accuracy rate
- Led client projects facing non-profit organizations, offered financial insights on future fundraising direction, developed interactive Tableau dashboard, helped secured the largest corporate grants in organization's history

## PROJECTS & RESEARCH

**Predicting Furniture Style using Zero-Shot Classification** | *Python*

**September 2023 - December 2023**

- Evaluated and benchmarked multiple pre-trained Natural Language Processing models and Large Language Models, developed and tuned deep learning model from scratch using the Ludwig framework, enhancing model prediction accuracy from 21% to 91%

**Dr. Willy Hugo Lab** | *Cellpose, StarDist, Python*

**March 2024 - Present**

- Designed and analyzed efficient data pipelines and workflows using the StarDist model on H&E tissue images, significantly accelerating wet lab analysis process. Generated accurate masks and labeling, organized comprehensive datasets for future process, enhancing the understanding of drug resistance mechanisms in melanoma

## SKILLS

**Languages and frameworks:** Python, C++, R, SQL, Redis, Tableau, MongoDB, PostgreSQL, Streamlit, Spark, Hadoop

**Libraries:** Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, Scipy, Beautiful Soup, Keras, Plotly, Tensorflow, Pytorch

