

# LOU ZHOU

Houston, TX

Email: [lz80@rice.edu](mailto:lz80@rice.edu)   LinkedIn: [www.linkedin.com/in/lou-zhou/](https://www.linkedin.com/in/lou-zhou/)   Website: [lou-zhou.github.io/](https://lou-zhou.github.io/)   U.S. Citizen

## Education

### Rice University

Expected Graduation May 2027

Houston, TX

Bachelors of Arts in Statistics, Sport Analytics, and Mathematics (3.8 GPA)

**Relevant Coursework:** Advanced Sport Analytics, Introduction to Program Design, Linear Regression, Real Analysis, Advanced Statistical Methods, Stochastic Methods, Tools and Models for Data Science, Practical Machine Learning

**Languages & Tools:** Python(scikit-learn, pandas, NumPy, PyTorch, Keras, BeautifulSoup4), R(tidyverse, ggplot, face, glmnet, shiny, caret, rstan), XGBoost, Java, Regex, SQLite, Git, EC2, EMR, Hadoop, PySpark

## Experience

### Rice Athletics

August 2025 - Present

Student Technical Projects Co-Lead | Git, Python, Pandas, Shiny

Houston, TX

- Co-leading a 13-person engineering team in building a Shiny app that generates scouting reports for Rice Basketball
- Liaison to non-technical stakeholders, scoping features with scouting leads and translating requirements into clear specs and timelines in plain language

### Rice University

August 2024 - Present

Research and Teaching Assistant | Python, pandas, NumPy, PyTorch, Git, Matplotlib, SQLite

Houston, TX

- Conducting research under Dr. Scott Powers in partnership with TSG Hoffenheim's Research Lab, evaluating how virtual reality training impacts in-game decision-making in soccer
- Reduced data processing time by 83% using pandas by redesigning a data processing pipeline that generated features from 300+ matches of soccer tracking and event data
- Expanded data availability for ML training by broadening data pipeline support to 3 new soccer data provider formats(SkillCorner, Sportec, Hawkeye) with pandas/NumPy
- Refined soccer pass quality estimations, adding features(e.g. player velocity) to a PyTorch CNN to allow the model to more accurately capture game context
- Achieved a 94% student satisfaction rate from semester survey by guiding 17 students as the sole TA in an applied data science course through in-class help and office hours

### Carnegie Mellon University

June - July 2025

Undergraduate Researcher(CMSACamp) | R, tidyverse, ggplot, Git, XGBoost

Pittsburgh, PA

- Fully-funded research experience in sports analytics under the mentorship of Dr. Karim Kassam and Quang Nguyen
- Matched public research deep learning accuracy (59.9% vs. 59.8%) on NFL throw target prediction using simpler XGBoost model with two novel features describing potential separation and QB vision

### St. Jude Children's Research Hospital

June - July 2024

Biostatistics Research Assistant | R, tidyverse, ggplot, face, plink

Memphis, TN

- Created ALS patient-specific disease evolution projections by building models to forecast ALS progression using functional PCA on longitudinal patient data in R
- Contributed to understanding of ALS genetic risk factors by identifying 3 progression markers with Genome-Wide Association Studies on genotype and PCA data using HPC resources

## Projects

### GPS: A Metric for Evaluating Goalkeeping Positioning

October 2025 - January 2026

- Extended soccer goalkeeper evaluation beyond shot-stopping by building an interpretable framework which quantifies positioning decisions that shape danger before shots occur
- Modeled positional danger using XGBoost and contextualized goalkeeper positioning with a PyTorch spatial CNN, jointly estimating both impact and expectation from tracking data
- Presented at the 2026 American Soccer Insights Summit

### gg-pyscraper: A Python Library for Parsing Esports Data

July 2025

- Expanding public ease of access to esports data by building a generalized BeautifulSoup4/regex parsing library that standardizes community-wiki pages into structured datasets for 55+ video games
- Presented poster at the 2025 Carnegie Mellon Sports Analytics Conference

### Team Coin Flip: Travel Fatigue and Performance

January 2024

- Created XGBoost models with scikit-learn and a modified ELO ranking to assess travel impact on performance, placing 2nd of 59 teams at the 36-hour 2024 Rice Datathon

## Additional

Activities: · Tracks Organizer, 2025 Rice Datathon · Mentee, 2025 MIT Sloan Sports Analytics Mentorship Program