

Letian (Letty) Yang

letianya@usc.edu | [LinkedIn](#) | [Personal Website](#)

EDUCATION

University of Southern California

Ph.D. in Statistics (GPA: 3.95/4.00)

Aug 2023 – May 2028 (expected)

Los Angeles, CA

Southern University of Science and Technology

B.S. in Statistics and Data Science (GPA: 3.78/4.00)

Aug 2019 – Jun 2023

Shenzhen, China

Exchange Student, University of California, Berkeley (GPA: 4.00/4.00)

Jan 2022 – Jul 2022

PROFESSIONAL EXPERIENCE

Roblox

Data Scientist PhD Intern

May 2026 – Aug 2026

Ecosystems Research and Learning Platforms

- Diagnosed a 50–60% over-forecast in the production model that predicts Peak Concurrent Players (~20M at platform peak), which cloud infrastructure teams use to plan server capacity. The model combines Prophet for baseline traffic with a random forest for game-event spikes; traced it to compounding error in the baseline imputation and fixed it, restoring the normal 5–6% error (WMAPE).
- Cut under-forecasting, which risks server shortages at peak hours, by a further 25% (WMAPE 4.35% to 3.24% in backtests) by modeling how game events ramp up and linger instead of treating each event as a single start time.
- Added audit tables logging every intermediate step of the production pipeline to check it against the prototype, uncovering three silent bugs (wrong join in lag features, truncated inputs, static game list dropping new games).

Peking University Shenzhen Hospital

Data Science Intern

Jun 2021 – Aug 2021

Shenzhen, China

- Identified 60 candidate prognostic genes for head-and-neck cancer from TCGA and GEO data via differential expression and survival modeling, giving clinicians a shortlist for biomarker validation.
- Designed and ran an oral-health survey (1,000+ responses); age-stratified analysis informed patient education programs.

RESEARCH EXPERIENCE

Covariate Adjustment in High-Dimensional Randomized Experiments

Graduate Researcher

Apr 2024 – Present

University of Southern California

- Proposed a partially regularized OLS interpolator for covariate adjustment in experiments with more covariates than samples ($p > n$); proved the treatment-effect estimator unbiased, consistent, and asymptotically normal.
- Characterized when the estimator beats difference-in-means in efficiency (narrower confidence intervals), giving practical guidance for high-dimensional A/B tests.

Text-Aware Matrix Completion with LLM Embeddings

Graduate Researcher

May 2026 – Present

University of Southern California

- Developing a rating-matrix completion method that uses LLM embeddings of free-text reviews as side information; formulated a joint loss that models the interaction between text and ratings.

GeneNetting: Context-Specific Gene Prioritization

Researcher

Jun 2022 – Present

University of California, Berkeley

- Built a weakly supervised deep learning framework that prioritizes genes for specific biological processes by combining Gene Ontology labels with GTEx RNA-seq data (20,000+ genes), despite label noise and missing annotations.

MANUSCRIPTS

- L. Yang**, D. Shen. Algebraic and Statistical Properties of the Partially Regularized Ordinary Least Squares Interpolator. *arXiv preprint*, [arXiv:2411.06593](#).
- L. Yang**, P. Ding, D. Shen, D. Song. Covariate Adjustment in Randomized Experiments via Partially Regularized, Overparametrized Regression. *In preparation*.
- L. Yang**, S. Li, Y. Ye, E. Theusch, M. Medina, R. Shamir, H. Huang. GeneNetting: A Deep Weakly Supervised Framework for Context-Specific Gene Prioritization. *In preparation*.

TEACHING AND AWARDS

- Instructor, BUAD 425 Introduction to Business Analytics, USC Marshall, Fall 2026 (two sections, ~80 students).
- Marshall Mini Grant, USC (2026); National Scholarship, Ministry of Education of China (2022).

SKILLS

Programming and Tools: Python (NumPy, scikit-learn, PyTorch, TensorFlow), R, SQL, Spark, Git, Linux, AWS, GCP

Methods: Causal inference and A/B testing, time-series forecasting, high-dimensional regression, deep learning