# Rachel Longjohn

Website: rlongjohn.github.io
Email: rlongjoh@gmail.com

## EDUCATION

**Ph.D. in Statistics** Sep 2019 - Present
University of California, Irvine
Advisor: Padhraic Smyth (joint faculty in Statistics and Computer Science)

**M.S. in Statistics** Sep 2019 - Jun 2021
University of California, Irvine

**B.S. in Applied and Computational Mathematics** Aug 2015 - May 2019
University of Southern California
Specialization in Computer Programming

## RESEARCH INTERESTS

My research interests include forensic statistics, uncertainty quantification, and algorithmic evaluation, explainability, and fairness in machine learning. I am also interested in supporting efforts in data curation and reproducible research.

## RESEARCH OUTPUT

### Publications

1. Longjohn, R., & Smyth, P. (2024). Likelihood ratios for changepoints in categorical event data with applications in digital forensics. *Journal of Forensic Sciences*. https://doi.org/10.1111/1556-4029.15512
2. Longjohn, R., Smyth, P., & Stern, H. S. (2022). Likelihood ratios for categorical count data with applications in digital forensics. *Law, Probability and Risk*. https://doi.org/10.1093/lpr/mgac016

### Manuscripts Under Review/In Progress

3. Longjohn, R., Casleton, E., & Gopalan, G. (2024). Statistical uncertainty quantification for aggregate task-performance metrics in ML benchmarks. (Under Review).
4. Longjohn, R., Kelly, M., Singh, S., & Smyth, P. (2024). Benchmark repositories for better benchmarking. (Under Review).

### Conference Presentations

5. Gopalan, G., Casleton, E., Binette, O., & Longjohn, R. (2023). Statistical approaches for testing and evaluating foundation models. *Fall Technical Conference: Harmonizing Quality, Statistics, and Data Science.*
6. Longjohn, R., & Smyth, P. (2023a). Bayes factors for the existence of changepoints in categorical sequences within digital forensics. *Joint Statistical Meetings.*
7. Longjohn, R., & Smyth, P. (2023b). A likelihood ratio approach for detecting behavioral changes in device usage over time. *Annual Meeting of the American Academy of Forensic Sciences.*
8. Longjohn, R., & Smyth, P. (2022a). Likelihood ratios for categorical evidence with applications to digital forensics. *Joint Statistical Meetings.*
9. Longjohn, R., Smyth, P., & Stern, H. (2022). Likelihood ratios for categorical evidence with applications in digital evidence. *Annual Meeting of the American Academy of Forensic Sciences.*

### Webinars

10. Longjohn, R., & Smyth, P. (2022b). Tutorial on likelihood ratios with applications in digital forensics. *Center for Statistics and Applications in Forensic Evidence Summer Webinar Series.*

## RESEARCH AFFILIATIONS

**Los Alamos National Laboratory, CCS-6 Statistical Sciences Group** Jun 2023 - Present
Graduate Student Researcher

**NIST Center for Statistics and Applications in Forensic Evidence** Aug 2020 - Present
Graduate Student Researcher

**UCI Machine Learning Repository** Jan 2020 - Present
Dataset Curator and Librarian

**UCI Covid Awareness and Mobility Group** Jun 2020 - Jun 2022
Graduate Student Researcher

## WORK EXPERIENCE

**Los Alamos National Laboratory** Jun 2023 - Sep 2023
*Summer Student in Statistical Sciences Group*
Investigated statistical approaches for the testing and evaluation of foundation models in the context of nuclear nonproliferation, with a particular focus on uncertainty quantification in performance metrics.

**Obsidian Security** May 2018 - Aug 2019
*Data Science Intern*
Constructed generalizable data representations in Python used for modeling cybersecurity insights into enterprise cloud environments. Interfaced with data from over 10 APIs including those from Box, DropBox, GSuite, AWS, Office 365, Slack, and Salesforce.

## TEACHING EXPERIENCE

**STATS 7: Basic Statistics** Sep 2019 - Dec 2019
*Statistics Department, University of CA, Irvine*
Led two weekly discussions of 50 students each, wrote homework and exams, and held six office hours a week. Course covers basic inferential statistics, including confidence intervals and hypothesis testing on means and proportions, t-distributions, Chi-squared distributions, and introductory regression concepts.

## INSTITUTIONAL SERVICE

**Reviewer** Jul 2022 - Jul 2024
*NeurIPS Datasets and Benchmarks*

**Data Repository Curator and Librarian** Jan 2020 - Present
*UCI Machine Learning Repository*
Curate and review datasets that are to be listed in the UCI Machine Learning Repository operated by the UCI Computer Science department. Edit and improve metadata for datasets in the repository to better support reproducible research.

**Ph.D. Student Mentor** Sep 2021 - Present
*Statistics Department, University of California, Irvine*
Mentor first-year Ph.D. students in the Statistics department by meeting weekly to provide guidance on topics such as identifying research interests, finding an advisor, preparing for qualifying exams, and managing time as a graduate student.

**Editor-in-Chief, Viterbi Conversations in Ethics** Jan 2018 - May 2019
*Viterbi School of Engineering, University of Southern California*
Developed the USC Viterbi School of Engineering's ethics magazine and website for publishing student-authored papers on ethics topics in science. Led a team of five editors through the editorial process and collaborated with website designer to publish the website in January 2019.

## AWARDS

**Robert Newcomb Graduate Award in Statistics Honorable Mention** Sep 2020
Awarded by UCI Statistics Department for strong performance on qualifying exams

**Summa Cum Laude** May 2019
University of Southern California

**Phi Beta Kappa** Aug 2017
University of Southern California