

# Rachel Longjohn

Website: [rlongjohn.github.io](https://rlongjohn.github.io)

Email: [rlongjohn@gmail.com](mailto:rlongjohn@gmail.com)

## EDUCATION

---

### Ph.D. in Statistics

University of California, Irvine

Advisor: Padhraic Smyth (joint faculty in Statistics and Computer Science)

Sep 2019 - Present

### M.S. in Statistics

University of California, Irvine

Sep 2019 - Jun 2021

GPA: 4.00

### B.S. in Applied and Computational Mathematics

University of Southern California

Specialization in Computer Programming

Aug 2015 - May 2019

GPA: 3.96

## RESEARCH INTERESTS

---

I am broadly interested in problems at the intersection of statistics and machine learning, particularly in testing and evaluation for scientific applications. I am also passionate about ethics and data curation, provenance, and documentation issues.

## RESEARCH OUTPUT

---

### Publications

1. Longjohn\*, R., Kelly\*, M., Singh, S., & Smyth, P. (2024). Benchmark data repositories for better benchmarking. *NeurIPS Datasets and Benchmarks*, 37.
2. 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>
3. 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>

### Conference Presentations

4. Longjohn\*, R., Gopalan\*, G., & Casleton, E. (2024). Statistical uncertainty quantification for aggregate task-performance metrics in ML benchmarks. *NeurIPS Workshop on Statistical Frontiers in LLMs and Foundation Models*.
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

Graduate Student Researcher

Jun 2023 - Present

### NIST Center for Statistics and Applications in Forensic Evidence

Graduate Student Researcher

Aug 2020 - Present

### UCI Machine Learning Repository

Dataset Curator and Librarian

Jan 2020 - Present

### UCI Covid Awareness and Mobility Group

Graduate Student Researcher

Jun 2020 - Jun 2022

## 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, with a particular focus on uncertainty quantification in aggregate 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 >10 APIs including those from DropBox, GSuite, AWS, Office 365, and Salesforce.

## TEACHING EXPERIENCE

---

### STATS 7: Basic Statistics

Fall 2019, Fall 2024

*Statistics Department, University of California, Irvine*

Led weekly discussions of 50 students each. The course covers basic inferential statistics, including confidence intervals, hypothesis testing, and introductory regression concepts.

## INSTITUTIONAL SERVICE

---

### Reviewer

2022 - 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 and led the editorial team for the USC Viterbi School of Engineering's online magazine on ethics topics in science.

## AWARDS

---

### Rose Hill Foundation Science and Engineering Fellowship

Oct 2024

Awarded by UCI Graduate Division for demonstrated academic excellence and leadership qualities.

### 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

Awarded by University of Southern California for strong academic performance.

### Phi Beta Kappa

Aug 2017

Awarded by University of Southern California for strong academic performance.