Rachel Longjohn

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

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 issues in safety, ethics, and responsibility.

EDUCATION

Ph.D. in Statistics Sep 2019 - Present
University of California, Irvine
Advisor: Padhraic Smyth

M.S. in Statistics Sep 2019 - Jun 2021
University of California, Irvine *(GPA: 3.99/4.00)*

B.S. in Applied and Computational Mathematics Aug 2015 - May 2019
University of Southern California *(GPA: 3.96/4.00)*
Specialization in Computer Programming

EXPERIENCE

Visiting Student in Statistical Sciences Jun 2023 - Present
Los Alamos National Laboratory

Graduate Student Researcher Sep 2019 - Present
University of California, Irvine

Data Science Intern May 2018 - Aug 2019
Obsidian Security

PUBLICATIONS

Peer-Reviewed

- Longjohn*, R., Kelly*, M., Singh, S., & Smyth, P. (2024). Benchmark data repositories for better benchmarking. *NeurIPS*. https://arxiv.org/abs/2410.24100
- 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
- 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

Workshop

- 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.*

Preprints / Under Review

- Longjohn, R., & Smyth, P. (2025). Score-based likelihood ratios using stylometric text embeddings. (Under Review).
- Longjohn, R., Gopalan, G., & Casleton, E. (2025). Statistical uncertainty quantification for aggregate task-performance metrics in ML benchmarks. (Under Review).

PRESENTATIONS

- 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.*
- 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.*
- Longjohn, R., & Smyth, P. (2023a). Bayes factors for the existence of changepoints in categorical sequences within digital forensics. *Joint Statistical Meetings.*
- 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.*
- Longjohn, R., & Smyth, P. (2022b). Tutorial on likelihood ratios with applications in digital forensics. *NIST Center for Statistics and Applications in Forensic Evidence Summer Webinar Series.*
- Longjohn, R., & Smyth, P. (2022a). Likelihood ratios for categorical evidence with applications to digital forensics. *Joint Statistical Meetings.*
- 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.*

INSTITUTIONAL SERVICE AND TEACHING

Workshop Organizer Apr 2025
The Future of Machine Learning Data Practices and Repositories @ ICLR

Reviewer Jun 2022, 2023, 2024
NeurIPS, Datasets and Benchmarks

Data Repository Curator and Librarian Jan 2020 - Present
UCI Machine Learning Repository

Teaching Assistant Fall 2019, 2024
Basic Inferential Statistics, University of California, Irvine

Ph.D. Student Mentor Sep 2021 - Jun 2023
Statistics Department, University of California, Irvine

AWARDS

University of California, Irvine

- Rose Hill Foundation Science and Engineering Fellowship Oct 2024
- Robert Newcomb Graduate Award in Statistics Honorable Mention Sep 2020

University of Southern California

- Summa Cum Laude Honors May 2019
- Phi Beta Kappa Honor Society Aug 2017
- Dean's List 2015-2019

SKILLS

- **Programming Languages**: Python, R, C++, MatLab
- **Libraries/Frameworks**: PyTorch, Stan, JAGS, R Tidyverse
- **Tools**: Git, Quarto