

Emily R. Diana

Office 265 South 37th Street, Room 434A
Academic Research Building,
Philadelphia, PA 19104

Phone +1 (860) 389 7309
Email ediana@wharton.upenn.edu
Website www.emilydiana.com

Education

- 2018-** Ph.D. Student in Statistics and Data Science - The Wharton School, University of Pennsylvania
Advisors: Michael Kearns and Aaron Roth
Committee: Michael Kearns, Aaron Roth, Shane Jensen, Edgar Dobriban
Thesis: *Approaches Addressing Algorithmic Bias and Disclosiveness*
Expected Graduation Date: May 2023
- 2017-2018** M.S. in Statistics - Stanford University
- 2011-2015** B.A. in Applied Mathematics, cum laude - Yale College
Thesis: *Maintaining Bipartite Structure with a Modified Louvain Algorithm*
Supervisor: Daniel Spielman

Selected Awards

- The Wharton School's J. Parker Memorial Bursk Prize for Excellence in Research (2022)
- Michigan Institute for Data Science Future Leader (2022)
- MIT EECS Rising Star (2021)
- Brown Best Student Paper Award (2021)
- Wellcome Data Re-Use Prize: Malaria (£15000, 2019)
- Weapons Simulation and Computing Code Development Silver Star Award (2017)

Research Interests

Ethical algorithm design, fairness and privacy in machine learning, financial applications of data science, statistical learning, bias mitigation under distribution shift, proxy attributes, tool development

Peer-Reviewed Publications

- Emily Diana**, Wesley Gill, Michael Kearns, Krishnaram Kenthapadi, Aaron Roth, and Saeed Sharifi-Malvajerdi. Multiaccurate proxies for downstream fairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 1207–1239, New York, NY, USA, 2022. Association for Computing Machinery.
- Emily Diana**, Wesley Gill, Ira Globus-Harris, Michael Kearns, Aaron Roth, and Saeed Sharifi-Malvajerdi. Lexicographically Fair Learning: Algorithms and Generalization. In *2nd Symposium on Foundations of Responsible Computing (FORC 2021)*, volume 192 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:23, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- Emily Diana**, Wesley Gill, Michael Kearns, Krishnaram Kenthapadi, and Aaron Roth. Minimax group fairness: Algorithms and experiments. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 66–76, New York, NY, USA, 2021. Association for Computing Machinery.
- Emily Diana**, Travis Dick, Hadi Elzayn, Michael Kearns, Aaron Roth, Zachary Schutzman, Saeed Sharifi-Malvajerdi, and Juba Ziani. Algorithms and learning for fair portfolio design. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, EC '21, page 371–389, New York, NY, USA, 2021. Association for Computing Machinery.

5. **Emily Diana**, Hadi Elzayn, Michael Kearns, Aaron Roth, Saeed Sharifi-Malvajerdi, and Juba Ziani. Differentially private call auctions and market impact. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, page 541–583, New York, NY, USA, 2020. Association for Computing Machinery.
6. **Emily Diana**, Michael Kearns, Seth Neel, and Aaron Roth. Optimal, truthful, and private securities lending. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2020. Association for Computing Machinery.
7. Charles R. Noble, Andrew T. Anderson, Nathan R. Barton, Jamie A. Bramwell, Arlie Capps, Michael H. Chang, Jin J. Chou, David M. Dawson, **Emily Diana**, Timothy A. Dunn, Douglas R. Faux, Aaron C. Fisher, Patrick T. Greene, Ines Heinz, Yuliya Kanarska, Saad A. Khairallah, Benjamin T. Liu, Jon D. Margraf, Albert L. Nichols, Robert N. Nourgaliev, Michael A. Puso, James F. Reus, Peter B. Robinson, Alek I. Shestakov, Jerome M. Solberg, Daniel Taller, Paul H. Tsuji, Christopher A. White, and Jeremy L. White. Ale3d: An arbitrary lagrangian-eulerian multi-physics code. Technical Report <https://www.osti.gov/biblio/1361589LLNL-TR-732040>, Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States), May 2017.

Contributed Articles

1. **Emily Diana**, Mingzi Niu, and Georgy Noarov. Sigecom winter meeting 2022 highlights. *ACM SIGecom Exchanges*, 20(1):3–23, July 2022.
2. Shuxiao Chen, **Emily Diana**, Sheng Gao, Siyu Heng, Hongming Pu, Hua Wang, and Dylan Small. Rethinking the causal relationship between malaria and anemia for african children: A community-level perspective via two-step matching adjustment, 2019. Report for Wellcome Data Re-Use Competition: Malaria (First Place Winners).

Talks and Presentations

1. *University of Illinois Urbana-Champaign*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies.” 2023. (Upcoming Seminar)
2. *University of Chicago*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies.” 2023. (Upcoming Seminar)
3. *Baruch College*. “Addressing Algorithmic Bias and Dislosiveness: Multiaccurate Proxies for Downstream Fairness.” 2023. (Seminar)
4. *University of Washington*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies.” 2023. (Seminar)
5. *University of Virginia*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies.” 2023. (Seminar)
6. *University of Florida*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies.” 2023. (Seminar)
7. *University of California, Irvine*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies for Redacted Features.” 2022. (Seminar)
8. *Gies College*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies for Redacted Features.” 2022. (Seminar)
9. *University of Michigan*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies for Redacted Features.” 2022. (Seminar)
10. *University of South Carolina*. “Addressing Algorithmic Bias and Dislosiveness: Minimax Group Fairness and Multiaccurate Proxies for Redacted Features.” 2022. (Seminar)
11. *Foster School of Business*. “Addressing Algorithmic Bias and Dislosiveness: Multiaccurate Proxies for Downstream Fairness.” 2022. (Seminar)
12. *Tippie College of Business*. “Addressing Algorithmic Bias and Dislosiveness: Multiaccurate Proxies for Downstream Fairness.” 2022. (Seminar)
13. *Michigan AI Symposium*. “Balanced Filtering via Non-Disclosive Proxies.” 2022. (Lightning Talk)

14. *IPAM Summer School on Algorithmic Fairness 2022*. “Multiaccurate Proxies for Downstream Fairness.” 2022. (Poster)
15. *ACM FAccT Conference 2022*. “Multiaccurate Proxies for Downstream Fairness.” 2022. (Short Talk, Long Talk, Poster)
16. *Lawrence Livermore National Laboratory Applied Statistics Group Seminar*. “Approaches Assessing Automated Bias: Minimax Fairness and Bias-Aware Proxies for Redacted Features.” 2022. (Invited Seminar)
17. *3rd Symposium on Foundations of Responsible Computing (FORC 2022)*. “Multiaccurate Proxies for Downstream Fairness.” 2022. (Talk)
18. *Wharton Statistics PhD Student Seminar Day*. “Bias Mitigation Benchmarking.” 2022. (Short Talk)
19. *MIDAS Future Leaders Summit*. “Multiaccurate Proxies for Downstream Fairness.” 2022. (Lightning Talk)
20. *Rutgers Conference on Advances in Bayesian and Frequentist Statistics*. “Multiaccurate Proxies for Downstream Fairness.” 2022. (Poster)
21. *EECS Rising Stars 2021*. “Multiaccurate Proxies for Downstream Fairness.” 2021. (Poster)
22. *AMLC 2021 Workshop on Trustworthy AI*. “Multiaccurate Proxies for Downstream Fairness.” 2021. (Talk)
23. *Lawrence D. Brown Distinguished Lecture and Student Workshop*. “Multiaccurate Proxies for Downstream Fairness.” 2021. (Talk)
24. *2nd Symposium on Foundations of Responsible Computing (FORC 2021)*. “Lexicographically Fair Learning: Algorithms and Generalization.” 2021. (Long Talk)
25. *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society*. “Minimax Group Fairness: Algorithms and Experiments.” 2021. (Short Talk, Long Talk, Poster)
26. *ALT Mentorship Workshop*. “Minimax and Lexicographically Fair Learning: Algorithms, Experiments, and Generalization.” 2021. (Talk Dissection)
27. *AMLC Workshop on Fairness and Bias in AI*. “Convergent Algorithms for (Relaxed) Minimax Fairness.” 2020. (Talk)
28. *6th Annual Bloomberg-Columbia Machine Learning in Finance*. “Optimal, Truthful, and Private Securities Lending.” 2020. (Invited Talk)
29. *Joint Statistical Meetings, Philadelphia, PA*. “Is Anemia Prevalence a Good Proxy for Malaria Prevalence for Children? A Community-Level Perspective via Matched Logistic Regression.” 2020. (Talk)
30. *NeuRIPS Workshop on Robust AI in Financial Services: Data, Fairness, Explainability, Trustworthiness, and Privacy, Vancouver, CA*. “Optimal, Truthful, and Private Securities Lending.” 2019. (Spotlight Talk)
31. *Grace Hopper Celebration of Women in Computing, Houston, TX*. “Domain Decomposition with Recursive Inertial Bisection.” 2016. (Poster)

Teaching Assistantships

The Wharton School, University of Pennsylvania

CIS 399: Science of Data Ethics (Spring 2020, Undergraduate)

STAT 613: Regression Analysis for Business (Fall 2019, Graduate)

STAT 102: Introduction to Business Statistics (Spring 2019, Undergraduate)

Stanford University

CS 161: Design and Analysis of Algorithms (Winter 2018-2019, Undergraduate)

CS 106A: Programming Methodologies (Fall 2018, Undergraduate)

Selected Experience

Jun 2021 - Amazon Web Services (Remote)

April 2022 *Applied Scientist Intern*

Supervisors: Michele Donini and Bilal Zafar

Research Topic: Bias Mitigation Benchmarking for Amazon SageMaker Clarify (Working Paper)

Language: Python

Jun 2020 - Amazon Web Services (Remote)

May 2021 *Applied Scientist Intern*

Supervisors: Michael Kearns, Krishnaram Kenthapadi, Aaron Roth

Research Topics: Minimax Group Fairness, Multiaccurate Proxies for Downstream Fairness.

Mar 2018 - Center on Poverty and Inequality, Stanford University, Stanford, CA

Aug 2018 *Research Assistant*

Supervisors: David Grusky and Adrian Raftery

Research Topic: Developing methodologies to analyze trends in contemporary social mobility based on contingency tables of longitudinally-linked Census data.

Aug 2015- Lawrence Livermore National Laboratory, Livermore, CA

Sep 2017 *Scientific Software Developer*

Parallelized and integrated a domain decomposer, Recursive Inertial Bisection, into the mesh generation step of ALE3D, a multi-physics “Arbitrary Lagrangian-Eulerian 3D” numerical simulation code. Primary developer for LLNL’s ParticlePack code. Member of team integrating a GPU portability abstraction into ALE3D’s advection package. Presented research internally on implications of strided memory access patterns on GPU-accelerated computing.

Service and Leadership

TheWebConf External Reviewer (2023)

FAccT Program Committee Member (2022, 2023)

UpML Workshop at ICML Program Committee Member (2022)

SEAS Center on Safe, Explainable, and Trustworthy AI-enabled Systems Member (2022)

Invited Contributor: Exchanges Cover of the SIGecom Winter Meeting on Algorithmic Fairness (2022)

AMLC Workshop on Trustworthy AI External Reviewer (2021)

FORC External Reviewer (2021)

Wharton Doctoral Program Student Representative (2021)

Wharton Doctoral Program Peer Mentor (2020-present)

Stanford Women in Mathematics Mentoring (2017)

LLNL Division Representative for Girls Who Code (2016-2017)

Other

Two Patents Filed (2021)

University of Pennsylvania Center for Teaching and Learning Teaching Certification (2022)

References

Name Prof. Michael Kearns
Reference Type Research, Teaching (Thesis Advisor)
Email mkearns@cis.upenn.edu
Website <https://www.cis.upenn.edu/~mkearns/>
Phone (215) 898 7888
Work Address 509 Levine Hall,
3330 Walnut Street,
Philadelphia, PA 19104

Name Prof. Aaron Roth
Reference Type Research (Thesis Advisor)
Email aaroth@cis.upenn.edu
Website <https://www.cis.upenn.edu/~aaroth/>
Phone (215) 746 6171
Work Address 3401 Walnut Street,
Room 406B,
Philadelphia, PA 19104

Name Dr. Krishnaram Kenthapadi
Reference Type Research (Industry Collaborator)
Email krishnaram@fiddler.ai
Website <https://cs.stanford.edu/people/kngk/>
Phone (650) 804 2502

Name Prof. Abraham Wyner
Reference Type Teaching
Email ajw@wharton.upenn.edu
Website <https://statistics.wharton.upenn.edu/profile/ajw/research>
Phone (215) 898 2439
Work Address 309 Academic Research Building,
265 South 37th Street,
Philadelphia, PA 19104

Dissertation Abstract

While data science enables rapid societal advancement, deferring decisions to machines does not automatically avoid egregious equity or privacy violations. Without safeguards in the scientific process — from data collection to algorithm design to model deployment — machine learning models can easily inherit or amplify existing biases and vulnerabilities present in society. In this dissertation, I focus on techniques to encode algorithms with ethical norms and construct frameworks ensuring that statistics and machine learning methods are deployed in a socially responsible manner. In particular, I develop theoretically rigorous and empirically verified algorithms to mitigate automated bias and protect individual privacy.

In the first chapter, I discuss two definitional contributions to the algorithmic fairness literature: minimax group fairness and lexicographic fairness. In the second chapter, I introduce a pair of algorithms for the strategic development of proxy attributes under equity constraints. Finally, I consider the application of differential privacy in financial applications where we aim to maximize a variety of social desiderata, such as privacy, truthfulness, and welfare. When available, I accompany each discussion with an open-source software package and thorough experimental analysis. All of the algorithms presented are oracle-efficient, provably convergent, and enjoy strong out-of-sample generalization guarantees.