In the past century, computation and machine learning have become ubiquitous and invaluable aspects of our everyday existence. Automation improves our quality of life in countless ways. With this boon, however, comes the sobering observation that automated decisions, though seemingly objective, often violate the fundamental rights of individuals and reinforce an unjust status quo. My research focuses on the intersection of ethical algorithm design and socially aware machine learning. In particular, I develop theoretically rigorous and empirically verified algorithms to protect individual privacy and mitigate automated bias.

Machine learning tools can compromise rights by leaking private information, an issue that differential privacy provides strong tools and mathematical guarantees to address. However, these tools have not been widely applied in finance. One resulting problem is that clients in securities lending scenarios are often incentivized to lie about their demands to protect their privacy. With joint differential privacy, we devised a resource allocation algorithm that is simultaneously private, approximately optimal, and incentivizes truth-telling [4]. Subsequently, we again considered ways to formally balance utility and privacy in financial markets, proposing a differentially private mechanism for call auctions [3]. My contributions in the former were theoretical, deriving the incentive and optimality properties of our algorithm, and my work in the latter consisted of performing simulations and proving a key composition property of our method. These papers provided theoretically sound algorithms to fill a gap in the literature and practical methods for industry.

Even when privacy leaks are addressed, however, entrenched biases can linger. There have been many documented cases of marginalized groups being discriminated against financially. In [5], we utilized minimax group fairness – fairness measured by worst-case outcomes across groups – to create a portfolio design method balancing utility and risk tolerances among socio-economic groups. Noting that minimax fairness was relatively unexplored in the algorithm literature, we then developed and tested an oracle-efficient and convergent algorithm to provably achieve minimax group fairness in general settings [7], for which I developed theory, prototyped the algorithms, and lead the experimental design. We contributed a rich, publicly available code base and accompanying experimental analysis to the literature, which provided the foundation for extensions such as [6, 9, 1]. In evaluating the performance of these algorithms, however, we observed that the fairness literature lacks user-friendly, comprehensive baselining tools. In response, I am collaborating with Amazon to build a rich yet easy-to-use framework allowing users to visualize trade-offs between metrics of interest across different bias mitigation techniques and data sets. In addition to developing software, I adapted an existing pre-processing technique designed to equalize error rates between groups to handle other types of fairness metrics [2]. With open source software and a working paper soon to be released, our goal is to provide a comparison of state-of-the-art techniques along with an online platform where researchers can contribute their own algorithms and data sets to form a dynamic benchmarking toolkit.

It is important to note, however, that sensitive features such as race, sex, and age are often not permitted as explicit factors in legal decision making. This leads to a conundrum; how can we train a model to be fair with respect to race, sex, or age without these features? In [8], we recently showed how to produce a proxy that allows one to train a fair model even when the original sensitive features are not available. We proved that obeying multiaccuracy constraints with respect to a given model class suffices for this purpose and provided algorithms for learning such proxies. With a thorough empirical analysis, we published one of the first positive results in this domain, showing that it is possible to efficiently train proxies that can stand in for missing sensitive features to effectively train downstream classifiers subject to a variety of demographic fairness constraints. My contributions to this paper include designing the algorithms, proving their convergence properties, protoyping their implementations, and guiding the empirical investigation.

One concern with this approach, however, is that there is no guarantee that individual privacy is protected when such a proxy is used. Moving forward, I am exploring techniques that generate a *non-disclosive* proxy for mitigating bias. We are investigating this problem in the setting of *data filtering* – using a small sample of data with sensitive attributes, we train a proxy that allows us to filter a data set with unobserved sensitive attributes into a sub-sample that is balanced with respect to these attributes. While ensuring that sensitive groups are evenly represented in a data set does not eliminate the potential for algorithmic bias, we hope that it does mitigate issues purely arising do to relative group sizes. Similarly, on the empirical side, I am beginning to investigate whether several popular "race blind" credit scoring algorithms may in fact be poorly calibrated on racial minority groups. In addition to exposing research questions of why and how this trend may hold, this investigation brings into focus important policy questions about disparate treatment legislation. As I progress in my career, I aim to explore such interdisciplinary questions with a long-term research program focused on both theoretical and empirical questions in this dynamic field.

References

[1] Martin Bertran, Natalia Martinez, Alex Oesterling, and Guillermo Sapiro. Distributionally robust group backwards compatibility. *arXiv preprint arXiv:2112.10290*, 2021.

[2] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, June 2002. ISSN 1076-9757.

[3] 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. ISBN 9781450379755. doi: 10.1145/3391403.3399500. URL `https://doi.org/10.1145/3391403.3399500`.

[4] 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. ISBN 9781450375849. doi: 10.1145/3383455.3422541. URL `https://doi.org/10.1145/3383455.3422541`.

[5] 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*, page 371–389. Association for Computing Machinery, New York, NY, USA, 2021. ISBN 9781450385541. URL `https://doi.org/10.1145/3465456.3467646`.

[6] 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. ISBN 978-3-95977-187-0. doi: 10.4230/LIPIcs.FORC.2021.6. URL `https://drops.dagstuhl.de/opus/volltexte/2021/13874`.

[7] Emily Diana, Wesley Gill, Michael Kearns, Krishnaram Kenthapadi, and Aaron Roth. *Minimax Group Fairness: Algorithms and Experiments*, page 66–76. Association for Computing Machinery, New York, NY, USA, 2021. ISBN 9781450384735. URL `https://doi.org/10.1145/3461702.3462523`.

[8] Emily Diana, Wesley Gill, Michael Kearns, Krishnaram Kenthapadi, Aaron Roth, and Saeed Sharifi-Malvajerdi. Multiaccurate proxies for downstream fairness. *ACM Conference on Fairness, Accountability, and Transparency*, 2022. To Appear.

[9] Vijay Keswani, Matthew Lease, and Krishnaram Kenthapadi. Towards unbiased and accurate deferral to multiple experts. *arXiv preprint arXiv:2102.13004*, 2021.