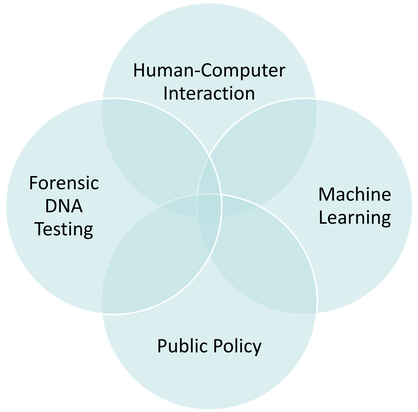
Adversarial Scrutiny of Evidentiary Statistical Software

Rediet Abebe[1], Moritz Hardt[2], **Angela Jin[1]**, John Miller[1], Ludwig Schmidt[3], and Rebecca Wexler[4]

[1] *EECS Dept., UC Berkeley*
[2] *Max Planck Institute for Intelligent Systems*
[3] *Paul G. Allen School of Computer Science & Engineering, University of Washington*
[4] *School of Law, UC Berkeley*



Convicted by Probabilistic Genotyping Software

In 2013, Mayer Herskovic, a Hasidic man living in Williamsburg, was convicted of a violent assault.

Photo by Demetrius Freedman for ProPublica.

Incriminating evidence: Output from probabilistic genotyping software (PGS)

But countless factors pointed towards his innocence. No other physical evidence linked him to the crime, and the PGS tool used in this case was never tested on a population as genetically insulated as the Hasidic Jews of Williamsburg.

Herskovic's conviction relied heavily on evidentiary statistical software his defense counsel could not scrutinize.

Story as reported in "Thousands of Criminal Cases in New York Relied on Disputed DNA Testing Techniques" by Lauren Kirchner in 2017.

Additional Concerns with PGS Validity

Different PGS tools analyzing the same data led to divergent conclusions [1, 2], and coding errors in one PGS tool affected dozens of cases in Queensland [2].

These cases have led to many calls for independent assessment of the validity of these software tools, but publicly available information about existing validation studies continues to lack details needed to perform this assessment [3].

Categories of Evidentiary Statistical Software

Category	Description	Scale of Use
Probabilistic genotyping	Analyze DNA mixtures	Used in a total of over 220,000 cases worldwide [4]
Breath testing	Estimate blood alcohol content from breath	In one year, used in over 30,000 cases [5]
Gunshot detection	Classify sounds as gunshots	Used in over 190 cases [6]
Toolmark analysis	Analyze surface of bullet and match to gun	In 2005, placed in over 230 state and local law enforcement agencies [7]

Prior Work: Designing an Audit Framework for Testing the Validity of Evidentiary Statistical Software (ESS)

How might defense counsel adversarially test ESS?

Adversarial Scrutiny in Law

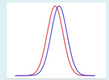
Judge & Jury

Prosecution: Prove guilt

Defense: Disprove the prosecution's case

Adversarial Robustness in ML

The study of how ML models fail under adversarial conditions such as changing subpopulations of adversarial input perturbations



Robust Adversarial Testing: Is the tool valid on cases similar to the defendant's?

Family of Distributions

Choose characteristics of the defendant's case to specify a set of distributions
e.g., Three distributions over DNA mixtures with 2 contributors
- D_1 ~ *2 contributors, DNA degraded as if exposed to outside conditions for 6 days*
- D_2 ~ *2 contributors, 95-105 pg of DNA*
- D_3 ~ *2 contributors, DNA from the Hasidic Jewish population of Williamsburg*

Quality Check Function

Choose an evaluation metric appropriate for the tool and case at hand.
e.g., Measure the PGS tool's inclusion error rate on a dataset and determine whether it exceeds 10%.

$$check\left(\ \ ,\ \ \right) \in \{Pass, Fail\}$$

An ESS A passes **(*F, check*)-robust adversarial testing** if, for every distribution D in family F, we have $check\ (A, D) = Pass$.

Operationalization

1. Compile a large database of inputs
2. Choose the family of distributions representing inputs similar to the defendant's case
3. Choose a quality check function appropriate for the tool and case at hand

Future Work

Case Study: Probabilistic Genotyping Software

Bridging expertise in forensic DNA testing, ML evaluation methods, and public policy, my goal is to:
- assess the technical feasibility of robust adversarial testing
- develop alternative methods for auditing probabilistic genotyping software
- explore governance and policy frameworks to support this evaluation

Understanding Defense Attorneys' Needs

We envision defense attorneys applying robust adversarial testing to evidentiary statistical software they face at trial. However, defense attorneys – especially those representing indigent clients – face a variety of barriers that pose challenges to implementing robust adversarial testing. There may also be other interventions that are more appropriate for ensuring rigorous evaluation of evidentiary statistical software.

My goal is to use need-finding and participatory design techniques to better understand defense attorneys' needs when confronted with evidentiary statistical software. I plan to work with defense attorneys, public defense offices, and organizations supporting them to better understand needs and co-design potential interventions.

Acknowledgements

Thank you to my collaborators Rediet Abebe, Moritz Hardt, John Miller, Ludwig Schmidt, and Rebecca Wexler. Thank you to Niloufar Salehi and Ludwig Schmidt for advising me on current research. Additional thank you to the AI Policy Hub and Nathan Adams for supporting the follow-up work and for many invaluable conversations.

References

[1] Jeanna Neefe Matthews, Graham Northup, Isabella Grasso, Stephen Lorenz, Marzieh Babaeianjelodar, Hunter Bashaw, Sumona Mondal, Abigail Matthews, Mariama Njie, and Jessica Goldthwaite. 2020. When trusted black boxes don't agree: Incentivizing iterative improvement and accountability in critical software systems. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 102–108.
[2] Katherine Kwong. 2017. The Algorithm says you did it: The use of Black Box Algorithms to analyze complex DNA evidence. Harv. JL & Tech. 31 (2017), 275.
[3] John M Butler, Hari Iyer, Rich Press, Melissa Taylor, Peter M Vallone, and Sheila Willis. 2021. DNA Mixture Interpretation: A NIST Scientific Foundation Review. National Institutes of Standards and Technology (2021).
[4] STRMix. Survey Shows STRmix Has Been Used in 220,000 Cases Worldwide. https://www.strmix.com/ news/survey-shows-strmix-has-been-used-in-220000-cases-worldwide/. Accessed: 2022-01-21.
[5] Stacy Cowley and Jessica Silver-Greenberg. "These Machines Can Put You in Jail. Don't Trust Them." The New York Times, November 3 (2019).
[6] ShotSpotter, Inc. "ShotSpotter Files Defamation Lawsuit Against Vice Media." ShotSpotter, Inc., October 12 (2021).
[7] Luke O. Brien. "3-D Imaging Goes Ballistic." WIRED, July 20 (2006).

The work described under "Prior Work" was produced in collaboration with Rediet Abebe, Moritz Hardt, John Miller, Ludwig Schmidt, and Rebecca Wexler. Our paper can be found on arXiv via the QR code to the left.

angelacjin@berkeley.edu

The work described under "Future Work" is supported by the UC Berkeley AI Policy Hub, an academic initiative advancing interdisciplinary research to anticipate and address AI policy opportunities.