Lessons from Computer Science for the Future of Predictive Policing

WITH A FOCUS ON PERSON-BASED PREDICTIVE POLICING

ELISSA M. REDMILES

GEORGETOWN UNIVERSITY

PRESENTATION TO THE
NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE

WITH SIGNIFICANT CONTRIBUTIONS FROM NINA GRGIĆ-HLAČA (MAX PLANCK INSTITUTE FOR SOFTWARE SYSTEMS, GERMANY)

AGENDA

1. **Trust**
2. **AI Futures**
3. **Recommendations**

Slides: https://georgetown.box.com/v/Redmiles-NASEM-PredPolicing

1 Trust

- Performance: Does the technology work?
- Fairness: Is the technology fair?

Does the technology work?

what seems like the simple answer is

Predictive accuracy

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

PREDICTIVE ACCURACY

EXAMPLE: COMPAS [1]

Scenario

A judge making a decision about pretrial detention

Goal

Predict who will commit a crime in the future.

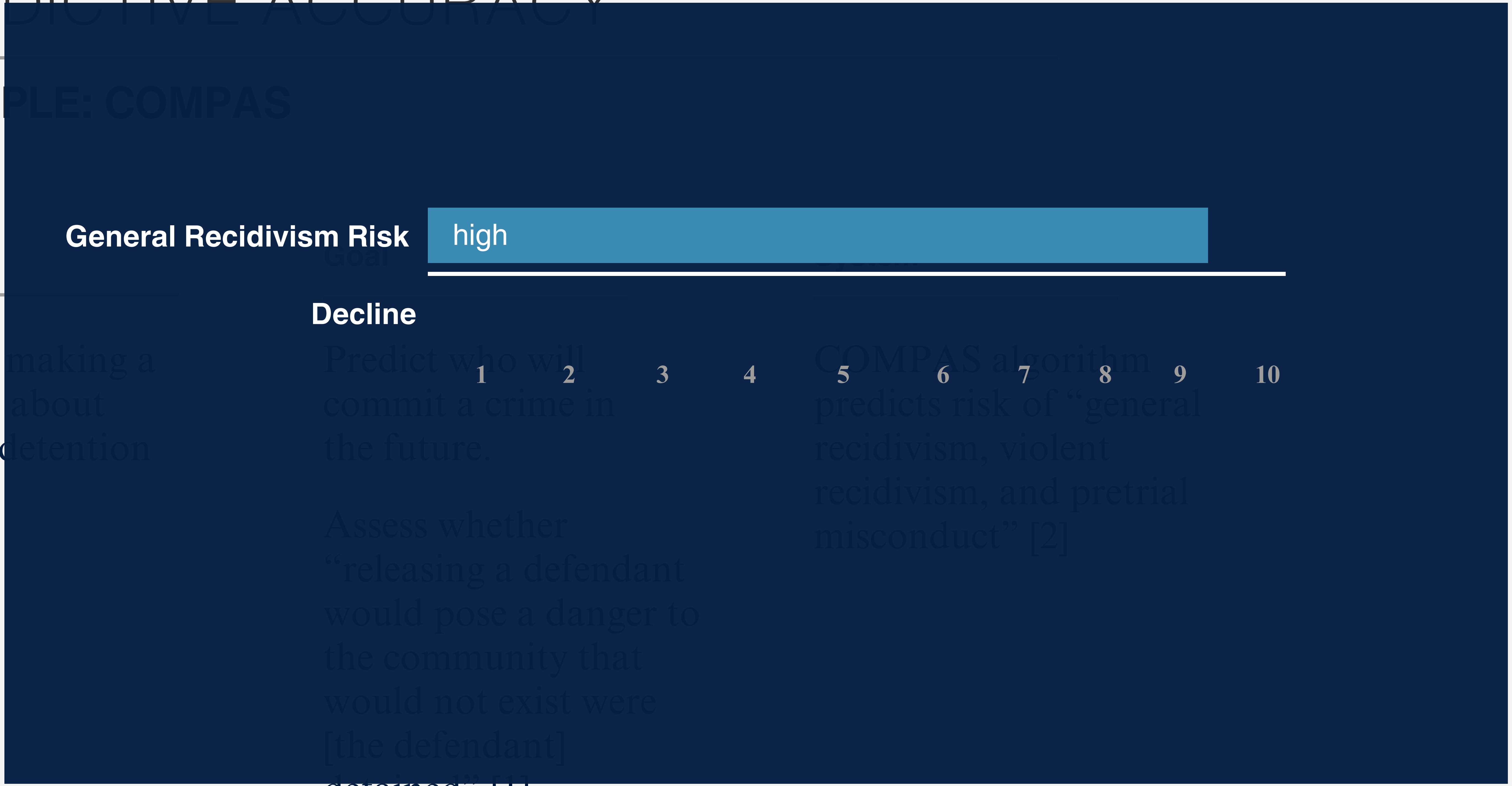
Assess whether “releasing a defendant would pose a danger to the community that would not exist were [the defendant] detained” [2]

System

COMPAS algorithm predicts risk of “general recidivism, violent recidivism, and pretrial misconduct” [3]

PRE

EXAM



Scenario

A judge
decision
pretrial

detained" [1]

PREDICTIVE ACCURACY **CHALLENGES**

PREDICTIVE ACCURACY **CHALLENGES**

#1 Measuring the outcome

Example **COMPAS**

Predicted outcome: risk score of how likely defendant is to commit a crime

Measure: Observations of arrests

Takeaway

All we can validate is whether we've predicted re-arrests. We cannot actually know if we're any good at person-based crime prediction.

PREDICTIVE ACCURACY **CHALLENGES**

#2 Accurately predicting social outcomes [4]

Example **ProPublica**

ProPublica obtained the risk scores of 7k+ people in Broward County, FL 2013-2014 and checked how many were charged with new crimes over the next 2 years.

Takeaway

“Only 20 percent of the people predicted to commit violent crimes actually went on to do so…Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years”

ANGWIN, J., LARSON, J., MATTU, S., & KIRCHNER, L. (2016, MAY 23). MACHINE BIAS. RETRIEVED FROM PROPUBLICA WEBSITE: HTTPS://WWW.PROPUBLICA.ORG/ARTICLE/MACHINE-BIAS-RISK-ASSESSMENTS-IN-CRIMINAL-SENTENCING

Similar accuracy to lay people predicting recidivism [5].

PREDICTIVE ACCURACY **CHALLENGES**

#3 Achieving our goal

Example **COMPAS**

Goal: Keep the community safe (assess whether "releasing a defendant would pose a danger to the community...")

Predicted outcome: risk score of how likely defendant is to be re-arrested ~~commit a crime~~

Takeaway

"Accurately predicting the occurrence of future crimes is not the same thing as helping to reduce crime…

accurate predictions of crime might simply cause the police to observe more crimes and generate more arrests rather than preventing those crimes from happening in the first place...

The police might be better off estimating the deterrent effect of police intervention" [6]

Does the technology work?

we must consider the

(Un)intended Consequences

Predictive Accuracy

(UN)INTENDED **CONSEQUENCES**

Example **Anti-trafficking SMS systems**

Law enforcement or NGOs are working to identify and/or help victims of sex trafficking.

They scrape advertisements for sexual services, compile contact information into a database, and/or sells data to other organizations so they can contact those in the database.

(UN)INTENDED **CONSEQUENCES**

#4 Leaking **dangerous private information**

Example **Anti-trafficking SMS systems**

Law enforcement or NGOs aim to help victims of sex trafficking.

Data collection is one approach:
We studied NGOs that scrape advertisements for sexual services, compile contact information into a database, and sell that data to other NGOs who text (SMS) those in the database.

→

Takeaway

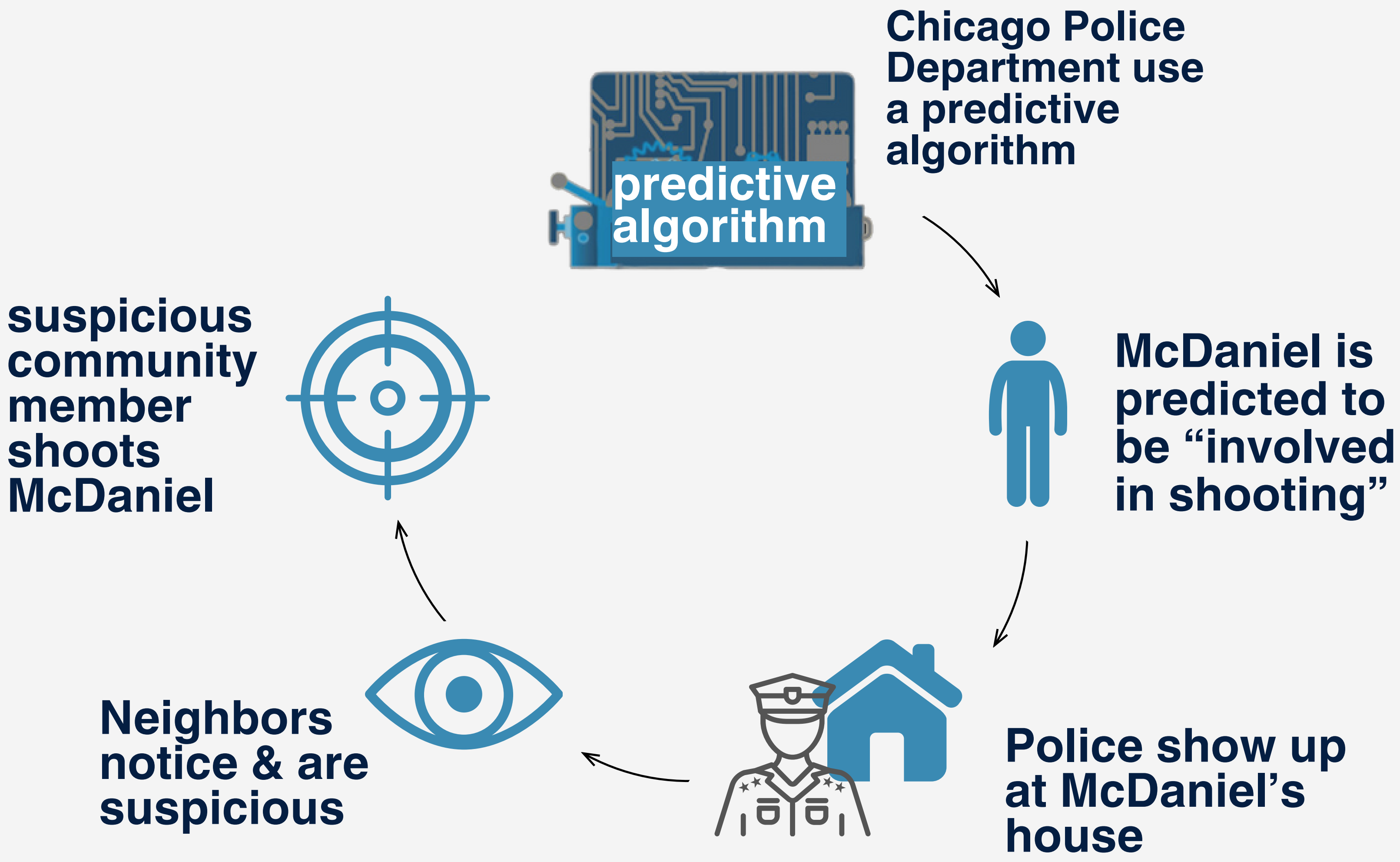
“Girls who I’ve worked with up in the massage parlors say things like ‘If my boyfriend knew I did this, he would kill me.’ ” [7]

System users are blind to this harm:
Organizations recognized that scaling outreach increased spam to those they contacted, but did not see how that could cause harm

(UN)INTENDED **CONSEQUENCES**

#4 Leaking **dangerous private information**

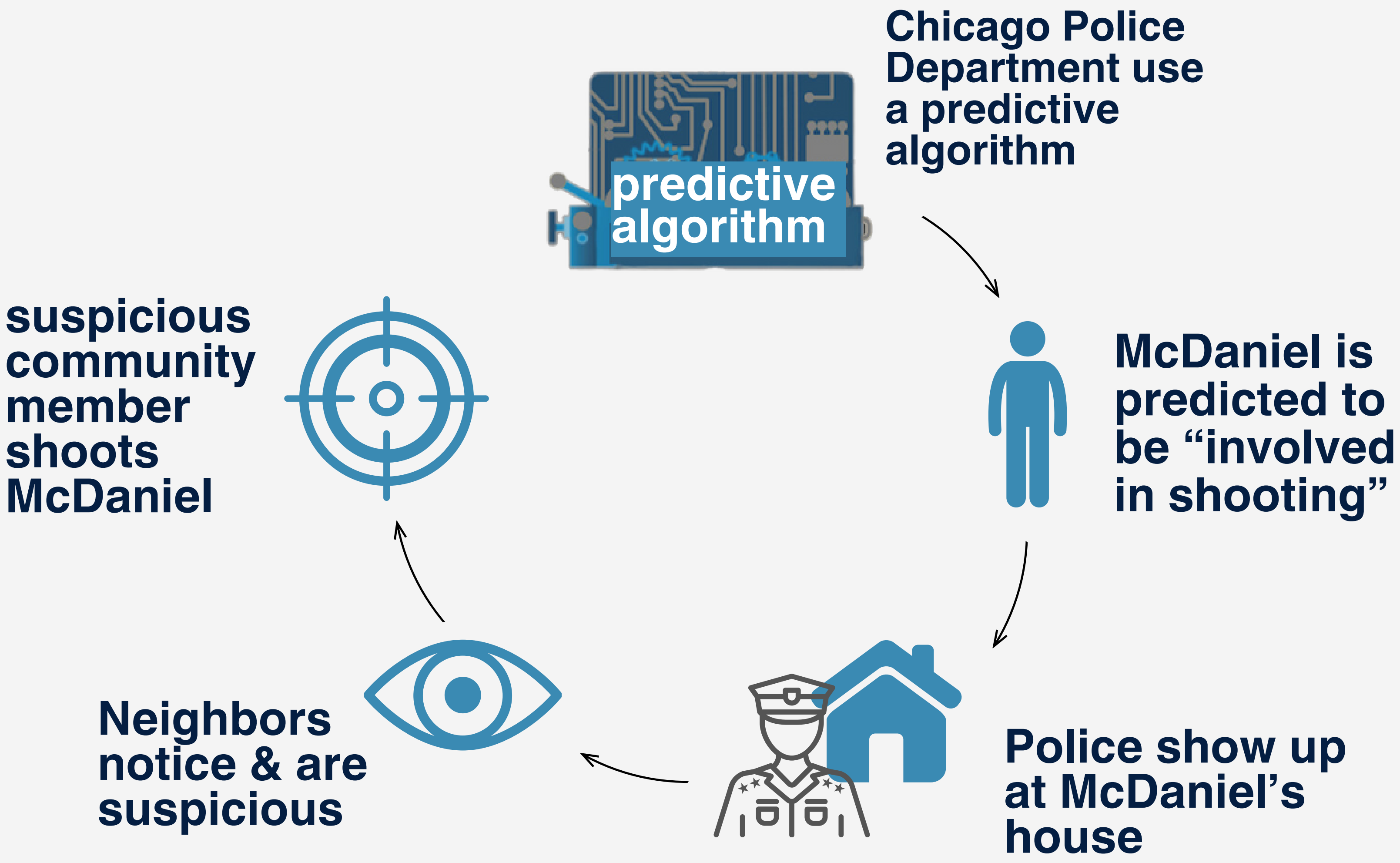
Example **The Shooting of Robert McDaniel [8]**



(UN)INTENDED **CONSEQUENCES**

#5 Creating **self-fulfilling prophecies**

Example **The Shooting of Robert McDaniel [8]**



(UN)INTENDED **CONSEQUENCES**

#5 Creating **self-fulfilling prophecies**

Example **The Shooting of Robert McDaniel [8]**

Police started following McDaniel, visiting his place of work and questioning his associates

Takeaway

Eventually, McDaniel was arrested for marijuana possession.

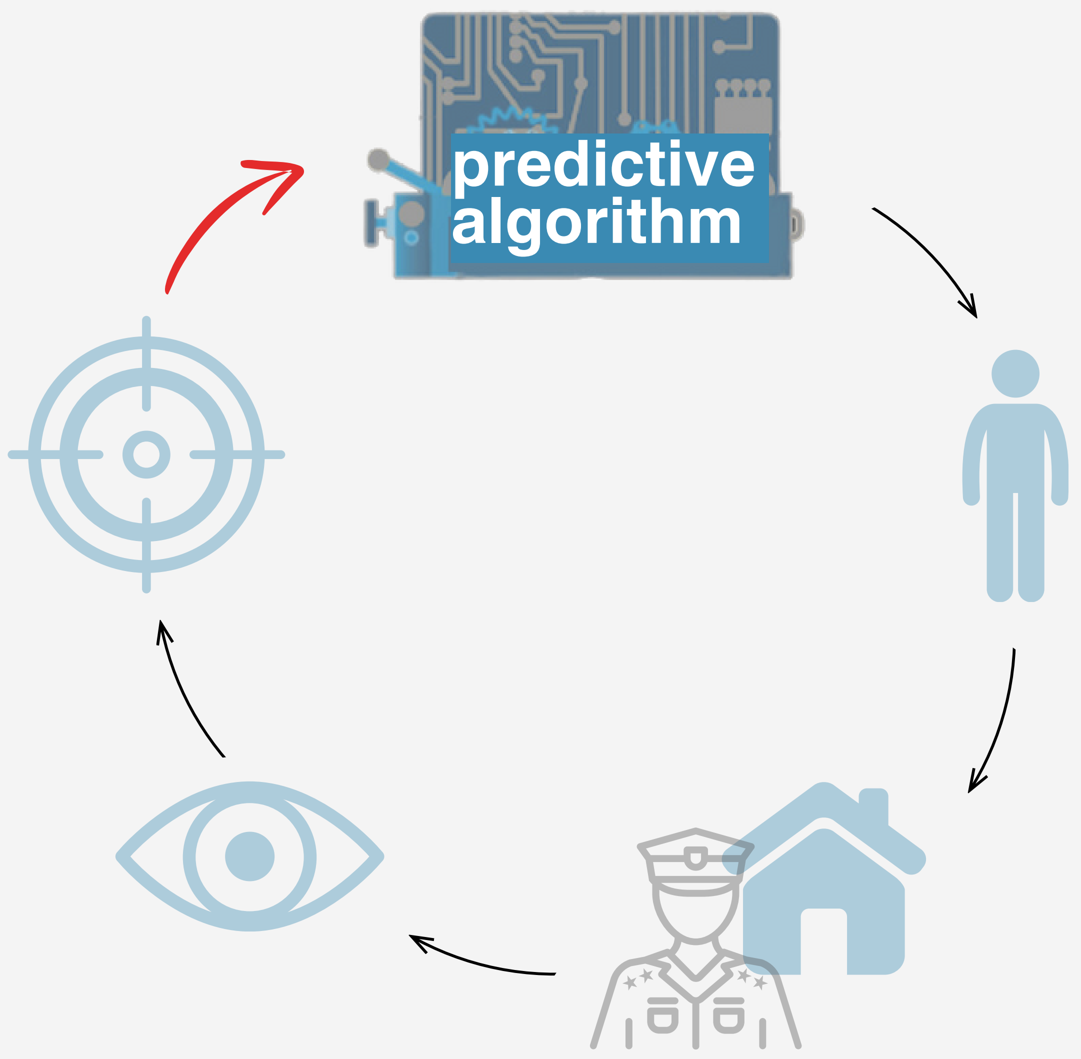
“a higher bail amount—based on predicted recidivism—can increase the likelihood of recidivism. In credit, a loan premium decided using a predictive model can negatively affect the probability of repayment” [9]

(UN)INTENDED **CONSEQUENCES**

#6 Creating **data feedback loops**[10,11]

Example **The Shooting of Robert McDaniel [8]**

Each self-fulfilling (or self-negating) prediction is recorded as a data point



(UN)INTENDED **CONSEQUENCES**

#6 Creating **data feedback loops**[10,11]

Example **The Shooting of Robert McDaniel [8]**

Each self-fulfilling (or self-negating) prediction is recorded as a data point

Takeaway

Performative Prediction [12,13]

> “The newly observed criminal acts that police document as a result of these targeted patrols then feed into the predictive policing algorithm on subsequent days, generating increasingly biased predictions. This creates a feedback loop where the model becomes increasingly confident that the locations most likely to experience further criminal activity are exactly the locations they had previously believed to be high in crime: selection bias meets confirmation bias” [10]

1

Analyze whether the outcome used to measure success matches the prediction

2

Justify predicting negative behavior vs. positive intervention efficacy

3

Articulate how prediction will turn into action, with what consequences

- privacy
- self-fulfilling prophecies
- feedback loops (performativity)

1 Trust

- Performance: Does the technology work?
- Fairness: Is the technology fair?

DEEP DIVE INTO **ACCURACY**

Example **COMPAS**

Predictive Accuracy

67.0% vs. **63.8%**

Overall predictive accuracy for **white defendants**

Overall predictive accuracy for **Black defendants**

DEEP DIVE INTO **ACCURACY**

Example **COMPAS**

Predictive Accuracy

67.0% vs. **63.8%**

Overall predictive accuracy for **white defendants**

Overall predictive accuracy for **Black defendants**

Takeaway

Systematic overestimation of recidivism for Black defendants [5]

False Positives [5]

"Black defendants who did not recidivate were incorrectly predicted to reoffend at a rate of 44.9%, nearly **2x as high as their white counterparts** at 23.5%"

False Negatives [5]

"White defendants who did recidivate were incorrectly predicted to not reoffend at a rate of 47.7%, **nearly twice as high as their black counterparts** at 28.0%.

Is the technology fair?

what seems like the simple answer is

Computational fairness metric

THERE ARE AT LEAST
21 COMPUTATIONAL DEFINITIONS OF FAIRNESS [14]

Name	Criterion	Note	Reference
Independence	Indep.	Equiv.	Calders et al. (2009)'
Group fairness	Indep.	Equiv.	
Demographic parity	Indep.	Equiv.	
Conditional statistical parity	Indep.	Relax.	Corbett-Davies et al. (2017)
Darlington criterion (4)	Indep.	Relax.	Darlington (1971)
Equal opportunity	Separ.	Relax.	Hardt, Price, Srebro (2016)
Equalized odds	Separ.	Equiv.	Hardt, Price, Srebro (2016)
Conditional procedure accuracy	Separ.	Equiv.	Berk et al. (2017)
Avoiding disparate mistreatment	Separ.	Equiv.	Zafar et al. (2017)
Balance for the negative class	Separ.	Relax.	Kleinberg et al. (2016)
Balance for the positive class	Separ.	Relax.	Kleinberg et al. (2016)
Predictive equality	Separ.	Relax.	Corbett-Davies et al. (2017),
Equalized correlations	Separ.	Relax.	Woodworth (2017)
Darlington criterion (3)	Separ.	Relax.	Darlington (1971)
Cleary model	Suff.	Relax.	Cleary (1966)
Conditional use accuracy	Suff.	Equiv.	Berk et al. (2017)
Predictive parity	Suff.	Relax.	Chouldechova (2016)
Calibration within groups	Suff.	Equiv.	Chouldechova (2016)
Darlington criterion (11), (2)	Suff.	Relax	Darlington (1971) -

BAROCAS, S., HARDT, M., & NARAYANAN, A. (2023). FAIRNESS AND MACHINE LEARNING: LIMITATIONS AND OPPORTUNITIES. PAGES 74-75. MIT PRESS

Concerns about equality of mistakes
(e.g., equality of false positive and/or false negative rates for different demographic groups)

Name	Criterion	Note	Reference
Independence	Indep.	Equiv.	Calders et al. (2009)'
Group fairness	Indep.	Equiv.	
Demographic parity	Indep.	Equiv.	
Conditional statistical parity	Indep.	Relax.	Corbett-Davies et al. (2017)
Darlington criterion (4)	Indep.	Relax.	Darlington (1971)
Equal opportunity	Separ.	Relax.	Hardt, Price, Srebro (2016)
Equalized odds	Separ.	Equiv.	Hardt, Price, Srebro (2016)
Conditional procedure accuracy	Separ.	Equiv.	Berk et al. (2017)
Avoiding disparate mistreatment	Separ.	Equiv.	Zafar et al. (2017)
Balance for the negative class	Separ.	Relax.	Kleinberg et al. (2016)
Balance for the positive class	Separ.	Relax.	Kleinberg et al. (2016)
Predictive equality	Separ.	Relax.	Corbett-Davies et al. (2017),
Equalized correlations	Separ.	Relax.	Woodworth (2017)
Darlington criterion (3)	Separ.	Relax.	Darlington (1971)
Cleary model	Suff.	Relax.	Cleary (1966)
Conditional use accuracy	Suff.	Equiv.	Berk et al. (2017)
Predictive parity	Suff.	Relax.	Chouldechova (2016)
Calibration within groups	Suff.	Equiv.	Chouldechova (2016)
Darlington criterion (11), (2)	Suff.	Relax	Darlington (1971) -

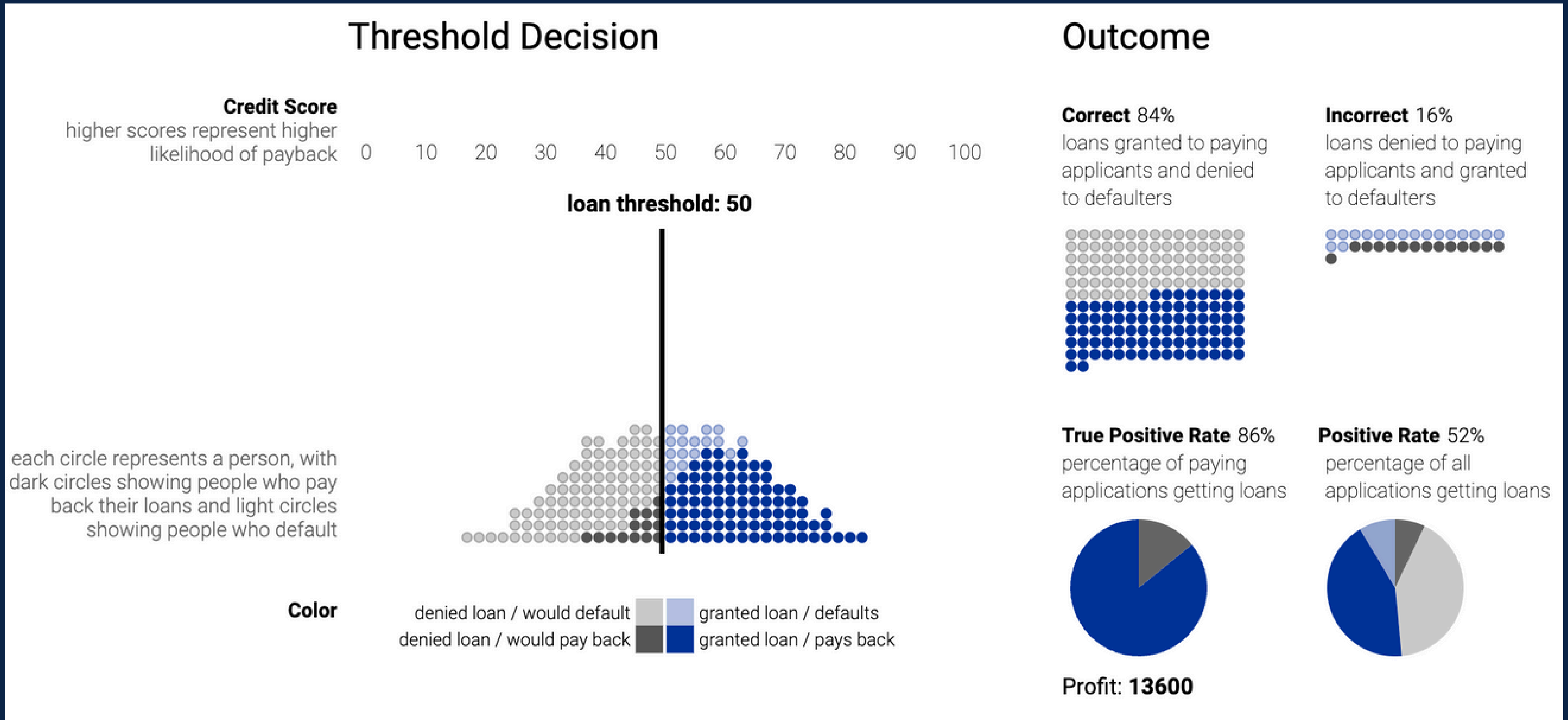
Focused on *outcomes* not *accuracy*.
Demographic parity is achieved when the rate of outcomes is the same between groups. E.g., 5% of Black defendants and 5% of white defendants are predicted to recidivate, respectively.

Name	Criterion	Note	Reference
Independence	Indep.	Equiv.	Calders et al. (2009)'
Group fairness	Indep.	Equiv.	
Demographic parity	Indep.	Equiv.	
Conditional statistical parity	Indep.	Relax.	Corbett-Davies et al. (2017)
Darlington criterion (4)	Indep.	Relax.	Darlington (1971)
Equal opportunity	Separ.	Relax.	Hardt, Price, Srebro (2016)
Equalized odds	Separ.	Equiv.	Hardt, Price, Srebro (2016)
Conditional procedure accuracy	Separ.	Equiv.	Berk et al. (2017)
Avoiding disparate mistreatment	Separ.	Equiv.	Zafar et al. (2017)
Balance for the negative class	Separ.	Relax.	Kleinberg et al. (2016)
Balance for the positive class	Separ.	Relax.	Kleinberg et al. (2016)
Predictive equality	Separ.	Relax.	Corbett-Davies et al. (2017),
Equalized correlations	Separ.	Relax.	Woodworth (2017)
Darlington criterion (3)	Separ.	Relax.	Darlington (1971)
Cleary model	Suff.	Relax.	Cleary (1966)
Conditional use accuracy	Suff.	Equiv.	Berk et al. (2017)
Predictive parity	Suff.	Relax.	Chouldechova (2016)
Calibration within groups	Suff.	Equiv.	Chouldechova (2016)
Darlington criterion (11), (2)	Suff.	Relax	Darlington (1971) -

BUT, WE **CANNOT ACHIEVE ALL** COMPUTATIONAL **FORMS OF FAIRNESS** SIMULTANEOUSLY[15]

BUT, WE **CANNOT ACHIEVE ALL** COMPUTATIONAL **FORMS OF FAIRNESS** SIMULTANEOUSLY

Interactive tool from Google Research that explains this further:



https://research.google.com/bigpicture/attacking-discrimination-in-ml

Scan here!

IN SUM,

- **Predictive mistakes are not distributed evenly across people**
- **There is no universal computational definition of fairness**
- **All computational forms of fairness cannot be achieved simultanously**

PUBLIC BELIEF IN FAIRNESS MATTERS

One approach to deciding what's fair is to ask the public.

Research in "empirical jurisprudence" [16] finds that people are more likely to comply with the law when it aligns with their moral views [16-18]

PUBLIC BELIEF IN FAIRNESS MATTERS

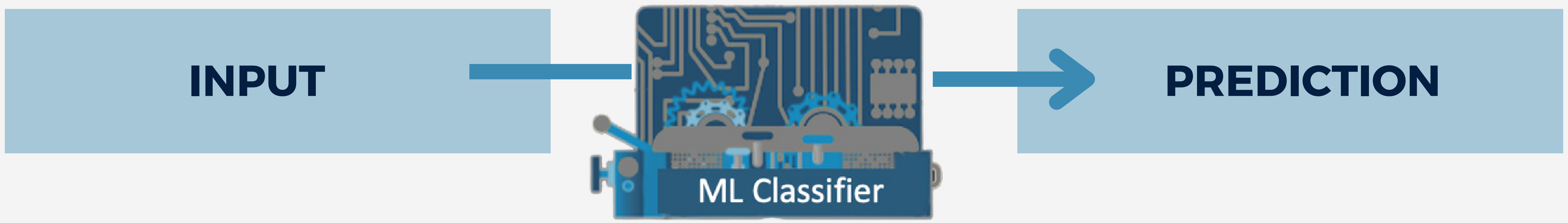
Which form(s) of computational fairness align with public perceptions?

Depends on the context [19-21] & application of the prediction [21]

PUBLIC BELIEF IN FAIRNESS MATTERS

Public perception of fairness isn't just about outcomes [22, 23]

IS THIS ALGORITHM FAIR?



Procedural Fairness

Is it fair to use certain data to make this decision?

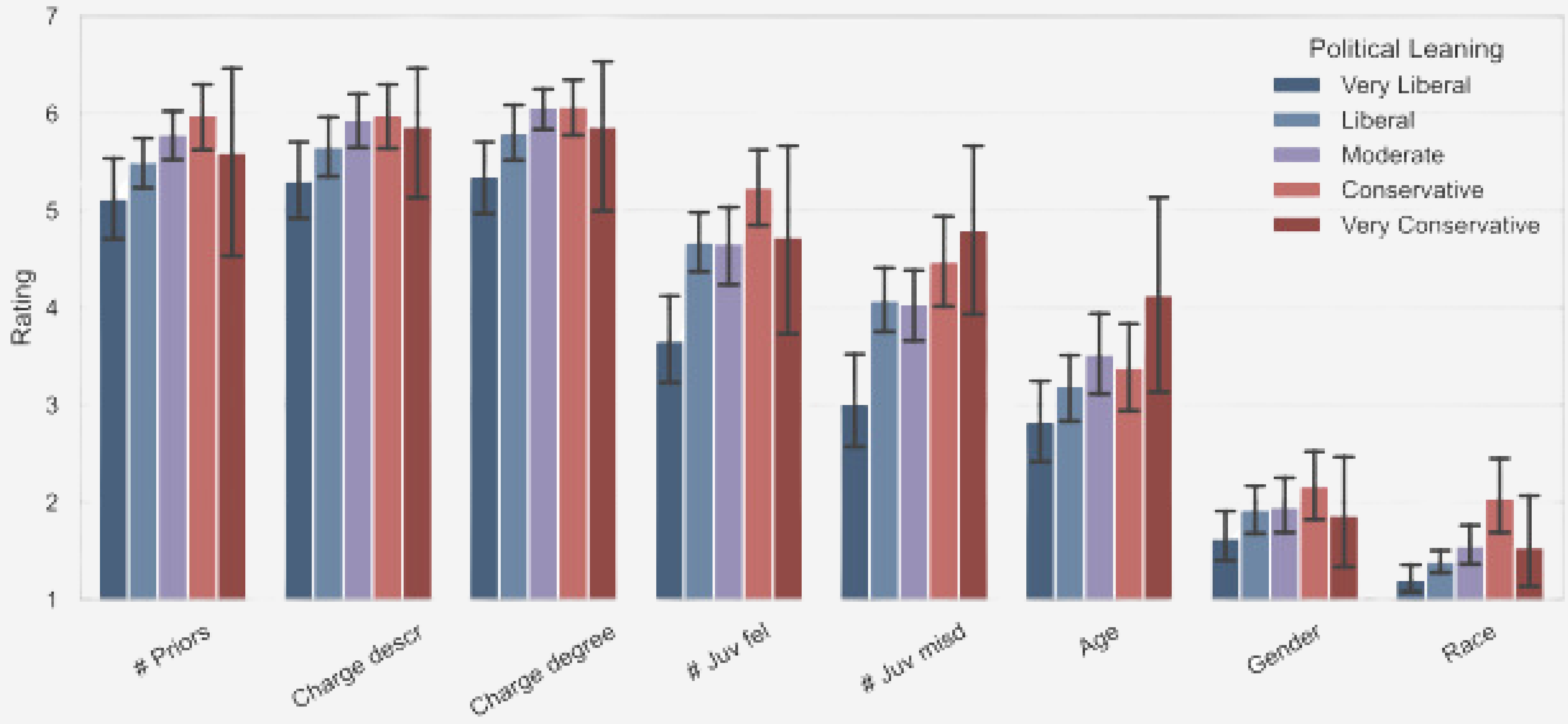
Distributed Fairness

Is the distribution of outcomes fair?

FAIRNESS PERCEPTIONS DIFFER AMONG PEOPLE

Example **COMPAS**

- Relevant **experiences influence fairness perception**: People who attended a bail hearing rate some features as less fair. [24]
- **Ideology influences fairness perception**: liberal participants rate features as less fair than conservatives, consistent with moral foundations theory. [24]



GRGIĆ-HLAČA, N., LIMA, G., WELLER, A., & REDMILES, E. M. (2022, OCTOBER). DIMENSIONS OF DIVERSITY IN HUMAN PERCEPTIONS OF ALGORITHMIC FAIRNESS. IN PROCEEDINGS OF THE 2ND ACM CONFERENCE ON EQUITY AND ACCESS IN ALGORITHMS, MECHANISMS, AND OPTIMIZATION (PP. 1-12)

GET FEEDBACK FROM A DIVERSITY OF STAKEHOLDERS

Research Shows

- **Research shows public belief in fairness matters**

- **Research shows people's ideology & experiences influence their beliefs.**

Takeaway

- **Get feedback:**
  - Example methodology: ORCAA Ethical Matrix "Fix a use case, elicit concerns from stakeholders, validate & prioritize concerns" [25]
  - For additional methodological considerations and guidance, see [e.g., 22, 26]

- **Only as good as the diversity of stakeholders**
  - Avoid Token Stakeholders: Anti-trafficking SMS system developers & users claimed what they were doing was not harmful because they had one survivor as part of their organization. [7]

1

Analyze whether the outcome used to measure success matches the prediction

2

Justify predicting negative behavior vs. positive intervention efficacy

3

Articulate how prediction will turn into action, with what consequences

4

Select fairness metrics:

- per use case
- with input from public & experts with a diversity of ideologies, experiences & demographics

5

Consider both procedural fairness and distributive fairness

AGENDA

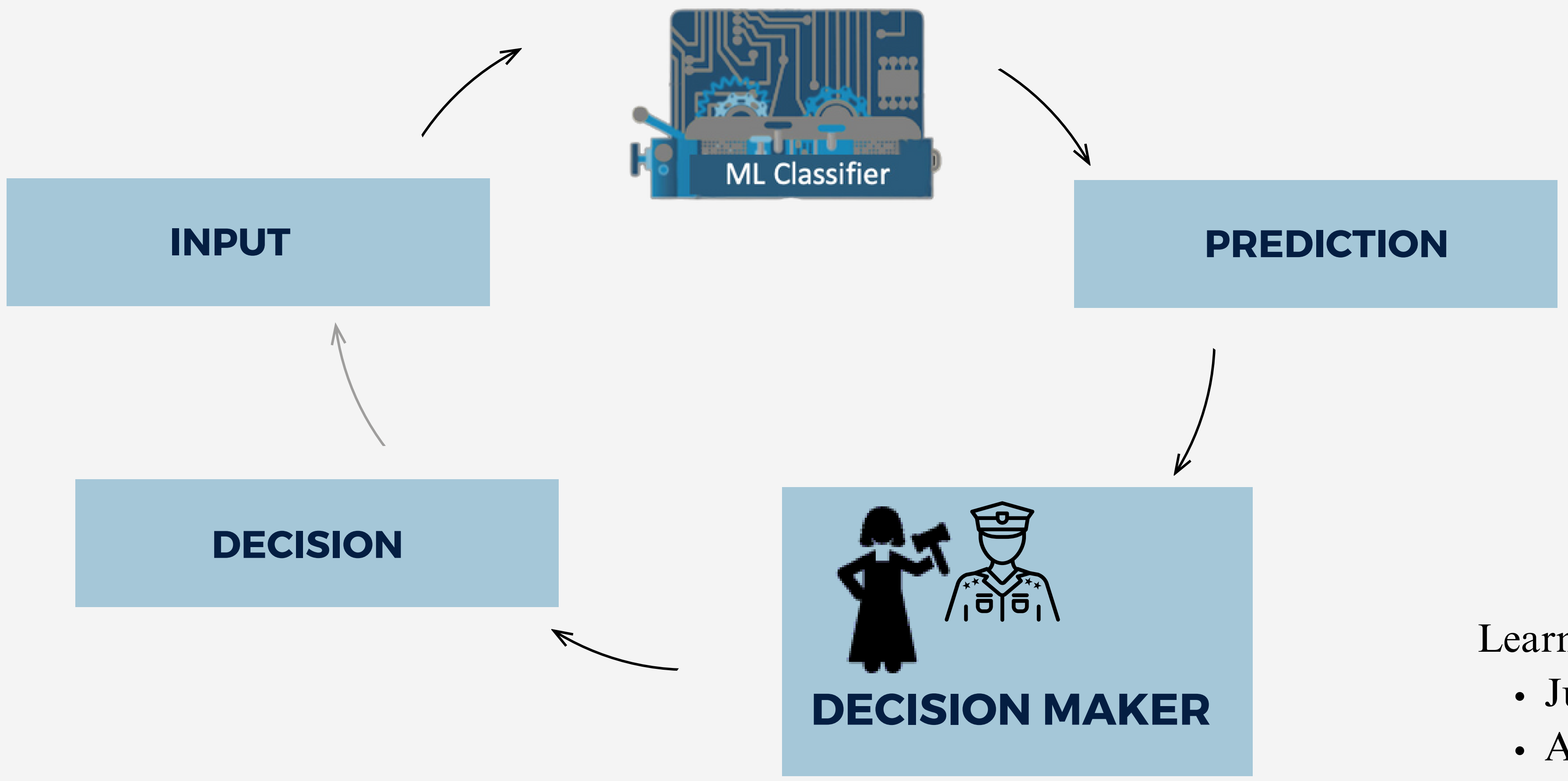
1 Trust

2 **AI Futures**

3 Recommendations

MEASURE THE FULL DECISION SYSTEM NOT JUST THE AI

A prediction is not a decision. Research finds many factors influence how the prediction influences the final decision.



Learn more:

- Judge-Advisor System framework [27]
- AI Assisted Decision Making [e.g., 28]

TRUSTED AI NEEDS TRANSPARENT BENCHMARKS & AUDITS

Closed systems breed mistrust

Benchmarks

Example: Wisconsin Supreme Court requires warnings that COMPAS is a proprietary system without public evaluation [29]

> "The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined."

Example: NIST unable to evaluate forensic probabilistic genotyping software [30]

> “There is not enough publicly available data to independently assess the reliability of these methods”

TRUSTED AI NEEDS TRANSPARENT BENCHMARKS & AUDITS

Closed systems breed mistrust

Benchmarks

Example: Wisconsin Supreme Court requires warnings that COMPAS is a proprietary system without public evaluation [29]

Example: NIST unable to evaluate forensic probabilistic genotyping software [30]

Audits

- Internal audits
- External audits
  - using both computational metrics & stakeholder feedback
- Unsolicited independent audits
  - require open benchmarks and system access

1

Analyze whether the outcome used to measure success matches the prediction

2

Justify predicting negative behavior vs. positive intervention efficacy

3

Articulate how prediction will turn into action, with what consequences

4

Select fairness metrics per use case & with human input

5

Consider both procedural fairness and distributive fairness

6

Measure accuracy & fairness of the full decision system

7

Require external audits that use computational metrics & stakeholder input

8

Require ongoing access for unsolicited independent evaluation

9

Publicly report incidents, evaluation outcomes, and design decisions

AGENDA

1. Trust
2. AI Futures
3. Recommendations

Leading Computing Researchers Argue:

DO NOT DO PERSON-BASED "PREDICTIVE OPTIMIZATION" [3]

Against Predictive Optimization:

On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy

Angelina Wang, Sayash Kapoor, Solon Barocas, Arvind Narayanan.

FAccT 2023 (earlier draft)
Journal of Responsible Computing 2023

Our argument

Predictive optimization is a distinct type of automated decision making that has **proliferated widely**. It is sold as accurate, fair, and efficient.

We identify a **recurring set of flaws** that apply broadly to predictive optimization, are hard to fix technologically, and negate its claimed benefits.

Any application of predictive optimization should be considered **illegitimate** by default unless the developer justifies how it avoids these flaws.

LESSONS FROM COMPUTER SCIENCE FOR THE FUTURE OF PREDICTIVE POLICING

0

Consider whether to do person-based prediction AT ALL

1

Analyze whether the outcome used to measure success matches the prediction

2

Justify predicting negative behavior vs. positive intervention efficacy

3

Articulate how prediction will turn into action, with what consequences

4

Select fairness metrics per use case & with human input

5

Consider both procedural fairness and distributive fairness

6

Measure accuracy & fairness of the full decision system

7

Require external audits that use computational metrics & stakeholder input

8

Require ongoing access for independent evaluations by third-parties

9

Publicly report incidents, evaluation outcomes, and design decisions

Slides

Bibliography

[1] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias. Retrieved from ProPublica website: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] U.S. Department of Justice. (n.d.). Criminal resource manual 26: Release and detention pending judicial proceedings (18 U.S.C. 3141 et seq.). Retrieved from https://www.justice.gov/archives/jm/criminal-resource-manual-26-release-and-detention-pending-judicial-proceedings-18-usc-3141-et

[3] Northpointe. (2015). Practitioner's Guide to COMPAS Core. Retrieved from https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

[4] Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. (2023). Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy. ACM Journal on Responsible Computing. https://doi.org/10.1145/3636509

[5] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science advances, 4(1), eaao5580.

[6] Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. MIT Press.

[7] Bhalerao, R., McDonald, N., Barakat, H., Hamilton, V., McCoy, D., & Redmiles, E. (2022). Ethics and Efficacy of Unsolicited Anti-Trafficking SMS Outreach. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 1–39. https://doi.org/10.1145/3555083

[8] Stroud, M. (2021, May 24). An automated policing program got this man shot twice. Retrieved from The Verge website: https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list

[9] Wang, Angelina, et al. "Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy." ACM Journal on Responsible Computing (2023

[10] Lum, K., & Isaac, W. (2016). To predict and serve?. Significance, 13(5), 14-19.

[11] Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018, January). Runaway feedback loops in predictive policing. In Conference on fairness, accountability and transparency (pp. 160-171). PMLR

Bibliography

[12] Hardt, M., & Mendler-Dünner, C. (2023). Performative prediction: Past and future. arXiv preprint arXiv:2310.16608.

[13] Perdomo, J., Zrnic, T., Mendler-Dünner, C., & Hardt, M. (2020, November). Performative prediction. In International Conference on Machine Learning (pp. 7599-7609). PMLR.

[14] Narayanan, A. (2018). Tutorial: 21 definitions of fairness and their politics. In Conference on Fairness, Accountability, and Transparency, NYC Feb (Vol. 23).

[15] Jon M. Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores 8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9--11, 2017, Berkeley, CA, USA. 43:1--43:23.

[16] Tobia, K. (2022). Experimental jurisprudence. The University of Chicago law review, 89(3), 735-802.

[17] Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., & Tyler, T. R. (2012). Why do people comply with the law? Legitimacy and the influence of legal institutions. British journal of criminology, 52(6), 1051-1071.

[18] Tyler, T. R., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. Psychology, public policy, and law, 20(1), 78

[19] Srivastava, M., Heidari, H., & Krause, A. (2019, July). Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2459-2468).

[20] Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2020). How do fairness definitions fare? Testing public attitudes towards three algorithmic definitions of fairness in loan allocations. Artificial Intelligence, 283, 103238.

[21] Harrison, G., Hanson, J., Jacinto, C., Ramirez, J., & Ur, B. (2020, January). An empirical study on the perceived fairness of realistic, imperfect machine learning models. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 392-402).

Bibliography

[22] Grgic-Hlaca, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018, April). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In Proceedings of the 2018 world wide web conference (pp. 903-912).

[23] Plane, A. C., Redmiles, E. M., Mazurek, M. L., & Tschantz, M. C. (2017). Exploring user perceptions of discrimination in online targeted advertising. In 26th USENIX Security Symposium (USENIX Security 17) (pp. 935-951).

[24] Grgić-Hlača, N., Lima, G., Weller, A., & Redmiles, E. M. (2022, October). Dimensions of diversity in human perceptions of algorithmic fairness. In Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-12).

[25] Cathy O'Neil. Dealing with High Impact AI. Statement to the Fourth Bipartisan Senate Forum On Artificial Intelligence.

[26] Feffer, M., Skirpan, M., Lipton, Z., & Heidari, H. (2023, August). From preference elicitation to participatory ml: A critical survey & guidelines for future research. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 38-48).

[27] Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. Organizational behavior and human decision processes, 101(2), 127-151.

[28] Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technological Forecasting and Social Change, 175, 121390. https://doi.org/10.1016/j.techfore.2021.121390

[29] Engel, C., & Grgić-Hlača, N. (2021). Machine advice with a warning about machine limitations: experimentally testing the solution mandated by the Wisconsin Supreme Court. Journal of Legal Analysis, 13(1), 284-340.

[30] Butler, J. M., Iyer, H., Press, R., Taylor, M. K., Vallone, P. M., & Willis, S. (2021). DNA Mixture Interpretation: A NIST scientific foundation review. Natl. Inst. Stand. Technol. NISTIR, 8351.

Appendix

PERFORMANCE RECOMMENDATIONS

Recommendation 1

Require system developers to **analyze whether the outcome used to measure success matches the prediction**

- Example: predicting crime, observing arrests

Recommendation 2

Require developing & procuring organizations to **justify predicting negative behavior vs. positive intervention efficacy**

- Example: predicting shooting involvement vs. efficacy of strategic lighting or suicide prevention sign placement

Recommendation 3

Require procuring organizations to articulate:

- the **actions that will be taken on the basis of predictions**
- community & expert feedback on **privacy leaks & consequences of actions**
- an evaluated **plan to manage self-fulfilling prophecies & feedback loops**

FAIRNESS RECOMMENDATIONS

Recommendation 1

Select fairness metrics **per use case**

Recommendation 2

Select fairness metrics **using both public & expert input** from people diverse in ideology, experiences related to the use case & demographics

Recommendation 3

Consider both procedural fairness and distributive fairness (i.e., assess computational and perceived fairness of system inputs & outputs)

AI FUTURES RECOMMENDATIONS

Recommendation 1	Require system developers to **measure accuracy & fairness of the full decision system**
Recommendation 2	Require system developers to **contract periodic external audits** from firms that are reputable and use both computational metrics and stakeholder input/
Recommendation 3	Require system developers to **enable independent evaluations** by making benchmarks & system functionality public or at minimium accessible to credentialed independent auditors (academics, journalists, NIST, etc.)
Recommendation 4	Require procuring & developing organizations to **publicly report incidents, evaluation outcomes, and design decisions**

PREDICTIVE ACCURACY **CHALLENGES**

#1 We **can’t measure the outcome** we’re trying to predict.

#2 We **struggle to predict** social outcomes

#3 We **can’t achieve our goal** with what we're trying to predict.

(UN)INTENDED **CONSEQUENCES**

#3 We **leak dangerous private information** when acting on predictions

#4 We create **self-fulfilling prophecies** when acting on predictions

#5 We create **data feedback loops** when acting on predictions