

Algorithmic fairness and GDPR transparency practice

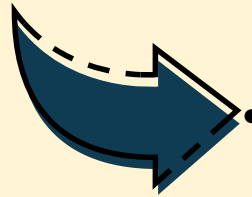
Ulrik Franke

RISE Research Institutes of Sweden
Adjunct professor, KTH

RI
SE

Some ethical concerns with AI and algorithms

algorithms \supset AI \supset ML



- **Transparency.** Difficult to explain why any particular classification or decision was made—systems become ‘black boxes’
- **Non-maleficence.** Concerns about safety and security; some mundane, some more far-fetched
- **Responsibility.** Who, if anybody, is to blame if a highly autonomous machine does harm?
- **Privacy.** Many successful algorithms feed on personal data



- **Fairness.** Prevention, monitoring or mitigation of unwanted bias and discrimination

Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9):389–399, doi: 10.1038/s42256-019-0088-2

Theory: Algorithmic fairness

Why is algorithmic fairness a concern?

Business Research (2020) 13:795–848
<https://doi.org/10.1007/s40685-020-00134-w>



ORIGINAL RESEARCH


Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development

Alina Köchling¹  · Marius Claus Wehner¹ 

IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE, VOL. 3, NO. 1, JANUARY 2021

101

Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?

Jacqueline G. Cavazos¹ , P. Jonathon Phillips, *Fellow, IEEE*,
Carlos D. Castillo, *Member, IEEE*, and Alice J. O'Toole

RESEARCH

RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}

Classification problems and bias (I)

Classify samples into the categories square and circle



Classification problems and bias (II)

Classification Sample	●	■
● (P)	TP	FN
■ (N)	FP	TN

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}$$

(miss rate)

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

(false alarm rate)

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(precision)

An intuitive family of fairness measures: Classification is not biased against any group Impossible!

(In practice)



Classification	●	■
Sample	●	■
●	TP_r	FN_r
■	FP_r	TN_r

Classification	●	■
Sample	●	■
●	TP_b	FN_b
■	FP_b	TN_b

$$PPV_r = PPV_b \text{ (equal precisions)}$$

$$FNR_r = FNR_b \text{ (equal miss rates)}$$

$$FPR_r = FPR_b \text{ (equal false alarm rates)}$$

Impossibility theorems

There are many intuitively compelling statistical measures of fairness, and most of them are not jointly satisfiable except in marginal cases (such as perfect predictors)

- Chouldechova A. (2017) Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5(2):153–163, doi: 10.1089/big.2016.0047
- Kleinberg J., Mullainathan S., Raghavan M. (2017) Inherent trade-offs in the fair determination of risk scores. In: 8th Innovations in Theoretical Computer Science Conference (ITCS 2017), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, vol 67, p 43, doi: 10.4230/LIPIcs.ITCS.2017.43
- Miconi, T. (2017). The impossibility of “fairness”: a generalized impossibility result for decisions. *arXiv preprint arXiv:1707.01195*.

Illustration. Classify samples into the categories square and circle, but treat the red and blue subsets equally!

Each of the linear classifiers x , y , and z satisfies two parities, but fails a third.

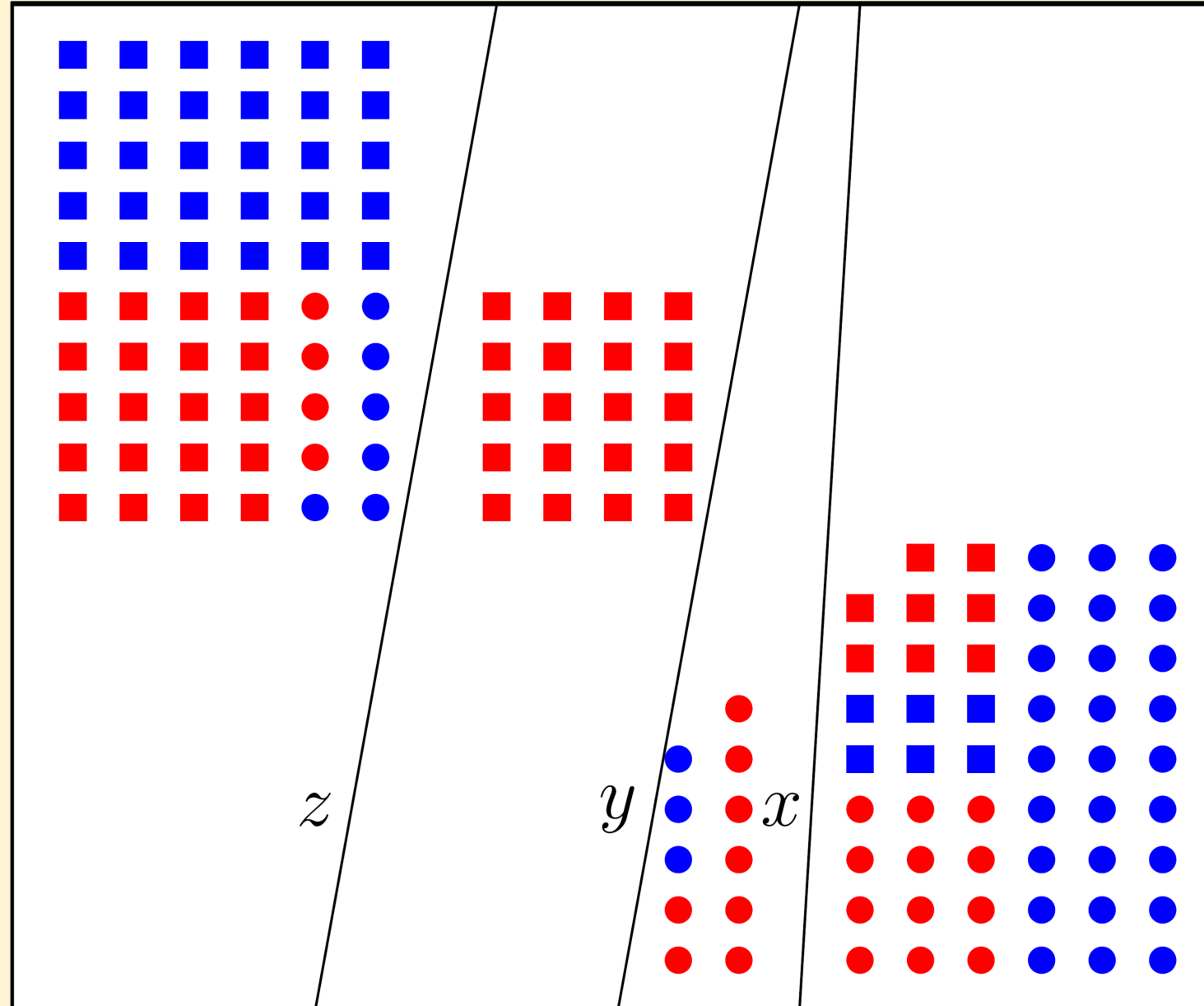
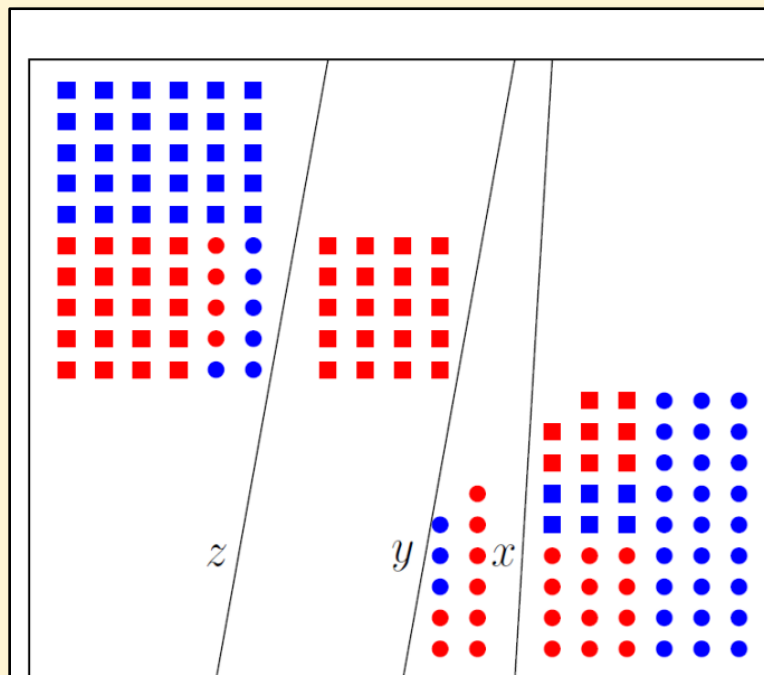


Illustration. Classify samples into the categories square and circle, but treat the red and blue subsets equally!

Each of the linear classifiers x , y , and z satisfies two parities, but fails a third.



x satisfies PPV-parity and FNR-parity, but not FPR-parity:

$$\begin{aligned} \text{PPV}_r &= \frac{40}{52} = \frac{10}{13} = \text{PPV}_b = \frac{30}{39} = \frac{10}{13} \\ \text{FNR}_r &= \frac{8}{48} = \frac{1}{6} = \text{FNR}_b = \frac{6}{36} = \frac{1}{6} \\ \text{FPR}_r &= \frac{12}{24} = \frac{1}{2} \neq \text{FPR}_b = \frac{9}{36} = \frac{1}{4} \end{aligned}$$

y satisfies FPR-parity and FNR-parity, but not PPV-parity:

$$\begin{aligned} \text{FPR}_r &= \frac{4}{24} = \frac{1}{6} = \text{FPR}_b = \frac{6}{36} = \frac{1}{6} \\ \text{FNR}_r &= \frac{8}{48} = \frac{1}{6} = \text{FNR}_b = \frac{6}{36} = \frac{1}{6} \\ \text{PPV}_r &= \frac{40}{44} = \frac{10}{11} \neq \text{PPV}_b = \frac{30}{36} = \frac{5}{6} \end{aligned}$$

z satisfies PPV-parity and FPR-parity, but not FNR-parity:

$$\begin{aligned} \text{PPV}_r &= \frac{20}{24} = \frac{5}{6} = \text{PPV}_b = \frac{30}{36} = \frac{5}{6} \\ \text{FPR}_r &= \frac{4}{24} = \frac{1}{6} = \text{FPR}_b = \frac{6}{36} = \frac{1}{6} \\ \text{FNR}_r &= \frac{28}{48} = \frac{7}{12} \neq \text{FNR}_b = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

Confusion matrices
of classifier x

$$\begin{array}{ll} \text{TP}_r = 40 & \text{FN}_r = 8 \\ \text{FP}_r = 12 & \text{TN}_r = 12 \end{array}$$

$$\begin{array}{ll} \text{TP}_b = 30 & \text{FN}_b = 6 \\ \text{FP}_b = 9 & \text{TN}_b = 27 \end{array}$$

Confusion matrices
of classifier y

$$\begin{array}{ll} \text{TP}_r = 40 & \text{FN}_r = 8 \\ \text{FP}_r = 4 & \text{TN}_r = 20 \end{array}$$

$$\begin{array}{ll} \text{TP}_b = 30 & \text{FN}_b = 6 \\ \text{FP}_b = 6 & \text{TN}_b = 30 \end{array}$$

Confusion matrices
of classifier z

$$\begin{array}{ll} \text{TP}_r = 20 & \text{FN}_r = 28 \\ \text{FP}_r = 4 & \text{TN}_r = 20 \end{array}$$

$$\begin{array}{ll} \text{TP}_b = 30 & \text{FN}_b = 6 \\ \text{FP}_b = 6 & \text{TN}_b = 30 \end{array}$$

Responses to the impossibility theorems in the literature

- Reject statistical measures in favor of individual measures

Dwork C., Hardt M., Pitassi T., Reingold O., Zemel R. (2012) Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, Association for Computing Machinery, New York, NY, USA, ITCS '12, p 214–226, doi: 10.1145/2090236.2090255

- Even though we cannot have parity of all measures at the same time, we can decide which measures are the most important in a given situation

Holm, S. (2022) The Fairness in Algorithmic Fairness. *Res Publica*, 1-17. doi: 10.1007/s11158-022-09546-3

Baumann, J., & Loi, M. (2023). Fairness and Risk: An Ethical Argument for a Group Fairness Definition Insurers Can Use. *Philosophy & Technology*, 36(3), 45. doi: 10.1007/s13347-023-00624-9

- Only one measure is really necessary for fairness

Hedden, B. (2021) On statistical criteria of algorithmic fairness. *Philosophy and Public Affairs*, 49(2). doi: 10.1111/papa.12189

- We cannot choose measures in a completely non-biased way, but it is still meaningful to try

Franke, U. (2022) First- and second-level bias in automated decision-making. *Philosophy & Technology* 35:21 doi: 10.1007/s13347-022-00500-y

...

Further reading

A more technical review:

- Chouldechova A., Roth A. (2020) A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM* 63(5):82–89, doi: 10.1145/3376898

A more philosophical review:

- Fazelpour, S., & Danks, D. (2021) Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), e12760. doi: 10.1111/phc3.12760

Practice: GDPR transparency in insurance

A right to explanation

vicious circle. Algorithms using crime and other data are also susceptible to self-fulfilling prophecies that discriminate against poorer or minority areas. A big problem is that people usually have no way of knowing what their profiles are based on — or that they exist at all.

There is an asymmetry in algorithmic power and accountability that lawmakers should correct. At the very least, there should be broader discussion of the principle that personal data belongs to an individual.

“A simplistic over-reliance

People should have the right to see their own data, how profiles are derived and have the right to challenge them. Some researchers

The GDPR right to explanation

Article 15

Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

[(a)–(g) omitted]

(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

h) Förekomsten av automatiserat beslutsfattande, inbegripet profilering enligt artikel 22.1 och 22.4, varvid det åtminstone i dessa fall ska lämnas meningsfull information om logiken bakom samt betydelsen och de förutsedda följderna av sådan behandling för den registrerade.

Testing the GDPR in practice

- Requests for information about how home insurance premiums are set were sent to 26 insurance companies in Denmark, Finland, The Netherlands, Poland and Sweden.
- Volunteers who were actual customers were recruited.

Country	Market share covered (%)
Sweden	90–95
Denmark	68
Finland	72
The Netherlands	45
Poland	40–45

Dexe, Jacob, et al. "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information." *The Geneva Papers on Risk and Insurance-Issues and Practice* 47.3 (2022): 669-697. <https://doi.org/10.1057/s41288-022-00271-9>

Testing the GDPR in practice (cont'd)

Hi!

In accordance with article 15, section 1h, of the General Data Protection Regulation 2016/679 I would like information on how the premium of my home insurance is determined. This article in the regulation should be applicable if pricing (i) is automated and (ii) is based on personal data (both collected from me and collected by other means).

I would be pleased to receive this information in suitable form (e.g., mathematical formulæ or descriptive text) that meets the requirements of the regulation on meaningful information about the logic involved in automated decision-making. Thanks a lot for your help!

Best regards etc.

Dexe, Jacob, et al. "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information." *The Geneva Papers on Risk and Insurance-Issues and Practice* 47.3 (2022): 669-697.
<https://doi.org/10.1057/s41288-022-00271-9>

Testing the GDPR in practice (cont'd)

- Considerable variation in responses
- No clear systematic differences between countries
- No clear systematic differences between companies with different sizes, ages or ownership structures

Dexe, Jacob, et al. "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information." *The Geneva Papers on Risk and Insurance-Issues and Practice* 47.3 (2022): 669-697. <https://doi.org/10.1057/s41288-022-00271-9>

	Living area (m ²)	Family status (e.g., number of people)	Address	Real estate data	Age	Deductible	Indemnity limit	Income and financial data	Security measures (e.g., locks)	Age of policy (loyalty)	Claims history	No response
DK1	L		L	X	X	X	L			X	X	
DK2	L		L	L	X		L	X			X	
DK3		X	X	X	X	X	X			X		
DK4			X	X	X							
DK5			X				X					
FI1	X	X	X		X	X						
FI2	X		X		X							
FI3	X		X		X	X	X			X		
FI4												X
NL1				X						X		
NL2												
NL3	X	X	X	X	X				X	X		
NL4												
PL1			X	X	X			X	X		X	
PL2												
PL3	X		X	X							X	
PL4			X	X			X		X	X	X	
PL5	X		X	X			X		X	X	X	
PL6												X
SE1	X	X	X		X							
SE2	X	X	L	X	L	X	X	X	X	L	X	
SE3			L		L							
SE4		X	X	X				X	X		X	
SE5			X		X							
SE6	X			X								
SE7		X	X		X					X	X	

Testing the GDPR in practice (cont'd)

- Some companies refer to business secrecy, but it is not necessarily the case that these companies are less forthcoming than those that do not.

Dexe, Jacob, et al. "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information." *The Geneva Papers on Risk and Insurance-Issues and Practice* 47.3 (2022): 669-697. <https://doi.org/10.1057/s41288-022-00271-9>

Process description	Fairness	Legal basis	General logic	Contact details	Information on other customers	Business confidentiality	How a profile is created
DK1	X	X	X	X			
DK2	X	X	X	X			
DK3	X			X			
DK4	X	X		X		X	X
DK5	X			X			
FI1	X		X	X			
FI2						X	
FI3	X		X	X			X
FI4							
NL1	X	X	X	X		X	
NL2	X			X			
NL3	X		X	X		X	
NL4	X	X	X			X	
PL1						X	
PL2		X	X			X	
PL3		X				X	
PL4		X	X	X		X	
PL5							
PL6							
SE1							
SE2	X	X			X		X
SE3		X					
SE4	X			X			
SE5	X			X			
SE6	X	X	X	X		X	X
SE7		X	X	X	X	X	

Testing the GDPR in practice (cont'd)

Dexe, Jacob, et al. "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information." *The Geneva Papers on Risk and Insurance-Issues and Practice* 47.3 (2022): 669-697. <https://doi.org/10.1057/s41288-022-00271-9>

Hi!

You have requested that SE7 inform you about how we calculate the premium for your home insurance.

SE7 uses all the data we have access to. It is about, e.g., information about you and your household. Where do you live,¹ how many are there in your household,² how old are you,³ how long have you had insurance in SE7,⁴ how many claims do you have.⁵ But also about other information about other customers, e.g., how many claims come from a certain residential area.

Information on other customers

The purpose is to calculate a premium that is as fair to each customer as possible in relation to the risk. It is not possible to set completely individual premiums because the idea of insurance is to spread the risks over a collective.

Fairness

Not all people suffer injuries, but if an injury occurs, it can be costly if you have to pay for everything yourself. When a collective bears the overall risk, the cost for each individual in the collective is lower. It will be a win-win situation for policyholders and the insurance company.

General logic

We do not provide information on exactly how e.g. what the actuarial formulæ look like. This is a business strategic and critical information that each insurance company keeps to itself. We have no obligation to disclose that information.

Business confidentiality

If you have any questions/comments on the above, you're welcome to contact us. You have all my contact details below.

Contact details

