Understanding Decision Subjects' Fairness Perceptions and Retention in Repeated Interactions with AI-Based Decision Systems

Meric Altug Gemalmaz, Ming Yin

Purdue University

AIES-22, Oxford, UK, August 2022

AI-based Decision Systems

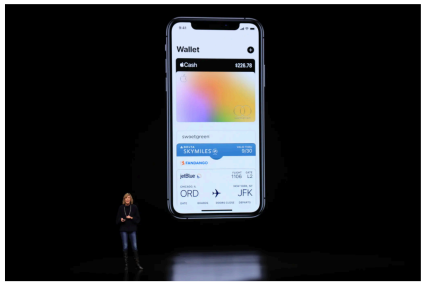
AI can be BIASED!

The New York Times

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sexist" against women applying for credit.

Give this article



Jennifer Bailey, vice president of Apple Pay. Regulators are investigating Apple Card's algorithm, which is used to determine applicants' creditworthiness. Jim Wilson/The New York Times

nature

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 24 October 2019 | Update 26 October 2019

Millions of black people affected by racial bias in health-care algorithms

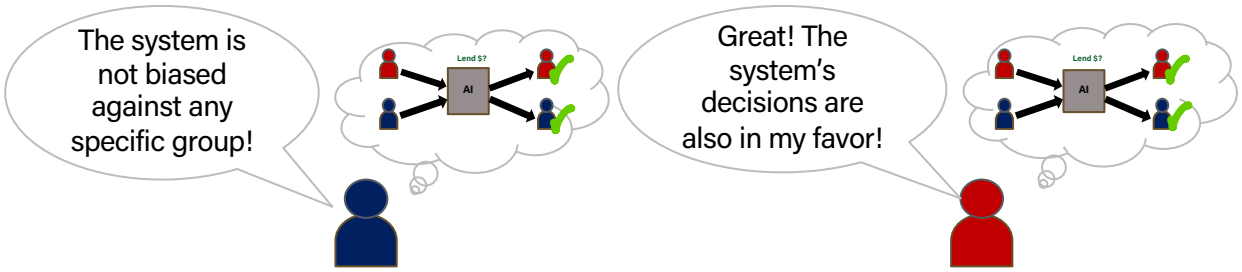
Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford

PURDUE UNIVERSITY | Department of Computer Science

Fairness Perceptions of AI-based Decision Systems

What factors affect the fairness perceptions of decision subjects (i.e., people who are subject to the AI system's decisions), and how?



Decision Subject 1 Decision Subject 2

[Wang et al. 2020]

Decision subjects can often repeatedly interact with an AI-based decision system!

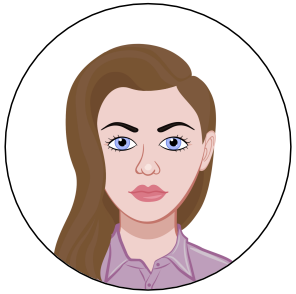
Our Research Questions

When decision subjects interact with the AI-based decision systems repeatedly…

- **RQ1:** How are decision subjects' fairness perceptions and their retention in the AI system affected by the AI system's fairness level across groups, and its tendency to favor the subject's own group?

- **RQ2:** What role does a decision subject's qualification level play in influencing her fairness perceptions and retention in the AI system?

- **RQ3:** What role does a decision subject's sensitivity level to fairness play in influencing her fairness perceptions and retention in the AI system?



Name: Jane

Group: Red

Credit score range: 720-750

Credit history (years): 11

Home ownership: Own

Small business industry: Food and accommodation service

Available Balance: 600 coins

Round 1 out of 10

Would you like to apply for a loan or not?

Apply for a loan! ☑

Leave the game and claim my bonus! ☐

AI system's decision on your application for this round:

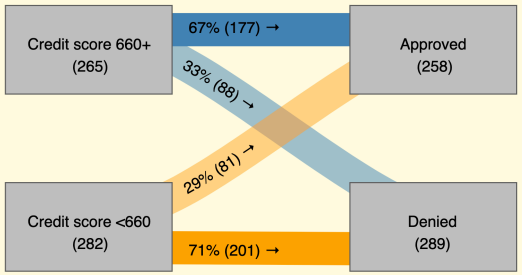
✓

Congratulations!

Your application is approved!

AI system's decisions on all 1104 applicants in this round:

Group Red



Among 547 red group applicants in this round, 258 applicants' (47%) applications have been approved. More specifically:

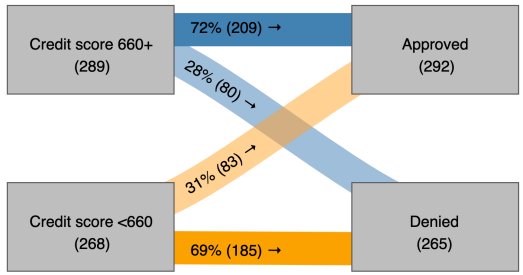
For the 265 (48%) applicants whose credit score is 660 or higher:

- 177 of them (67%) get their applications approved.
- 88 of them (33%) get their applications denied.

For the 282 (52%) applicants whose credit score is below 660:

- 81 of them (29%) get their applications approved.
- 201 of them (71%) get their applications denied.

Group Blue



Among 557 blue group applicants in this round, 292 applicants' (52%) applications have been approved. More specifically:

For the 289 (52%) applicants whose credit score is 660 or higher:

- 209 of them (72%) get their applications approved.
- 80 of them (28%) get their applications denied.

For the 268 (48%) applicants whose credit score is below 660:

- 83 of them (31%) get their applications approved.
- 185 of them (69%) get their applications denied.

Study 1

Two treatments by varying properties of the bank's AI model:

In the fair model treatment, the bank's AI model treats applicants from different groups **equally** and uses **same** decision matrix!

In the unfair model treatment, the bank's AI model is unfair as it is **in favor of applicants from the red group** thus uses **different** decision matrix for each group!

Same Approve and Deny chance respectively!

Credit/Decision	Approve	Deny
≥ 660	70%	30%
< 660	30%	70%

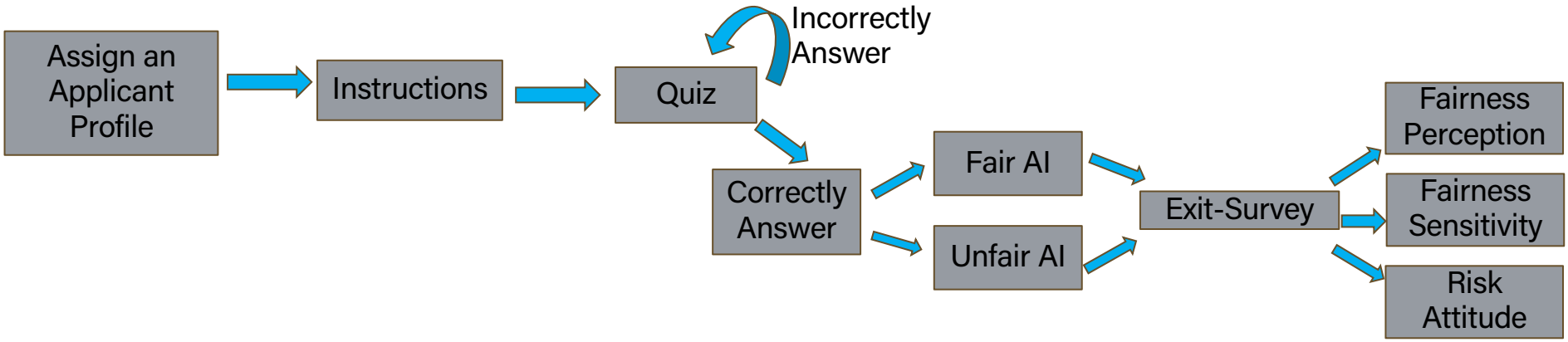
Higher for Red!

Credit/Decision	Approve	Deny
≥ 660	90%	10%
< 660	40%	60%

Lower for Red!

Credit/Decision	Approve	Deny
≥ 660	50%	50%
< 660	20%	80%

Experimental Procedure



809 Subjects Participated!

What are the Results?

RQ1: The impacts of the AI system's decision outcomes

	Perceived Fairness		**Retention**	
	Model 1	*Model 2*	*Model 3*	*Model 4*
Biased treatment	-0.12 (0.28)		-0.12 (0.26)	
Advantaged		0.57 (0.35)		0.58† (0.32)
Disadvantaged		-0.68* (0.33)		-0.69* (0.30)
Risk attitude	0.33*** (0.03)	0.34*** (0.03)	0.18*** (0.03)	0.18*** (0.03)
Constant	9.15*** (0.50)	9.09*** (0.51)	4.03*** (0.46)	3.97*** (0.46)

What are the Results?

RQ2: The role of decision subjects' qualification levels

	Perceived Fairness				Retention			
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*	*Model 8*
Biased treatment	-0.12 (0.28)	0.86† (0.52)			-0.14 (0.25)	0.37 (0.48)		
Advantaged			0.57 (0.36)	1.98** (0.67)			0.58† (0.32)	0.64 (0.60)
Disadvantaged			-0.68* (0.33)	-0.08 (0.62)			-0.73* (0.30)	0.05 (0.56)
Qualification	0.00 (0.04)	0.09 (0.06)	0.01 (0.04)	0.09 (0.06)	0.19*** (0.04)	0.23*** (0.05)	0.19*** (0.04)	0.23*** (0.05)
Qualification × Biased treatment		-0.18* (0.08)				-0.09 (0.07)		
Qualification × Advantaged				-0.26* (0.11)				-0.01 (0.10)
Qualification × Disadvantaged				-0.11 (0.09)				-0.14† (0.09)
Risk attitude	0.34*** (0.03)	0.34*** (0.03)	0.32*** (0.03)	0.35*** (0.03)	0.18*** (0.03)	0.18*** (0.03)	0.18*** (0.03)	0.18*** (0.03)
Constant	9.13*** (0.55)	8.64*** (0.59)	9.06*** (0.54)	8.54*** (0.59)	3.04*** (0.49)	2.78*** (0.54)	2.96*** (0.49)	2.76*** (0.53)

What are the Results?

RQ3: The role of decision subjects' sensitivity to fairness

	Perceived Fairness		**Retention**	
	Model 1	*Model 2*	*Model 3*	*Model 4*
Biased treatment	-0.16 (0.30)		-0.14 (0.26)	
Advantaged		0.43 (0.38)		0.51 (0.33)
Disadvantaged		-0.65† (0.35)		-0.66* (0.31)
Fairness sensitivity	-0.18* (0.07)	-0.18* (0.07)	-0.13* (0.06)	-0.13* (0.06)
Constant	16.50*** (0.75)	16.52*** (0.75)	8.25*** (0.66)	8.28*** (0.65)

Study 2

Fairness perceptions and retention are mainly influenced by the system's tendency to favor the subject's own group.

- **RQ4:** Are the changes in subjects' fairness perceptions and retention when the system favors or disfavors their own group caused by the subjects' own prospects of receiving the favorable decision, or the relative advantage or disadvantage they have towards the other group?

Unbiased Red Advantaged Red Disadvantaged

What are the Results?

Subjects' retention in the AI system is mainly driven by subjects' own prospects of receiving the favorable decision.

Subjects in both the red advantaged treatment and the red disadvantaged treatment increased their perceived fairness!

Fairness perceptions may be affected by many factors including...

- One's own prospects of receiving the favorable decision
- One's relative advantage against others
- The AI system's *overall* likelihood of granting favorable decisions

Summary

- Decision subjects' fairness perceptions and retention in repeated interactions with AI-based decision systems is significantly affected by whether the AI system tends to favor them.
- Different characteristics of the decision subjects might have their own influences on the fairness perception and the retention.
- Decision subjects' fairness perceptions of an AI system may be influenced by the system's treatment on themselves and on others in a complex way, meanwhile their retention in the system seems to be mostly driven by their own prospects of receiving the favorable decision from the system.

Thank You!