



AI algorithms: defining and mitigating bias

Transparante Algoritmes voor Beter Beleid

Hinda Haned

November 28th, 2019

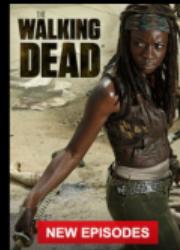
h.haned@uva.nl

How do AI algorithms help us?

NETFLIX



Popular on Netflix



TV Programmes based on Books



Voor jou in de Bonus

30 van je eerdere
aankopen zijn in de
Bonus

BONUS**1.49**

500 g

BONUS**0.59**

per stuk

BONUS**1.09**

per stuk

**2 VOOR 2.99****1.69**
per stuk**BONUS**

AH Mango

**2.29**
140 g**BONUS**

AH Sweet eve frambozen

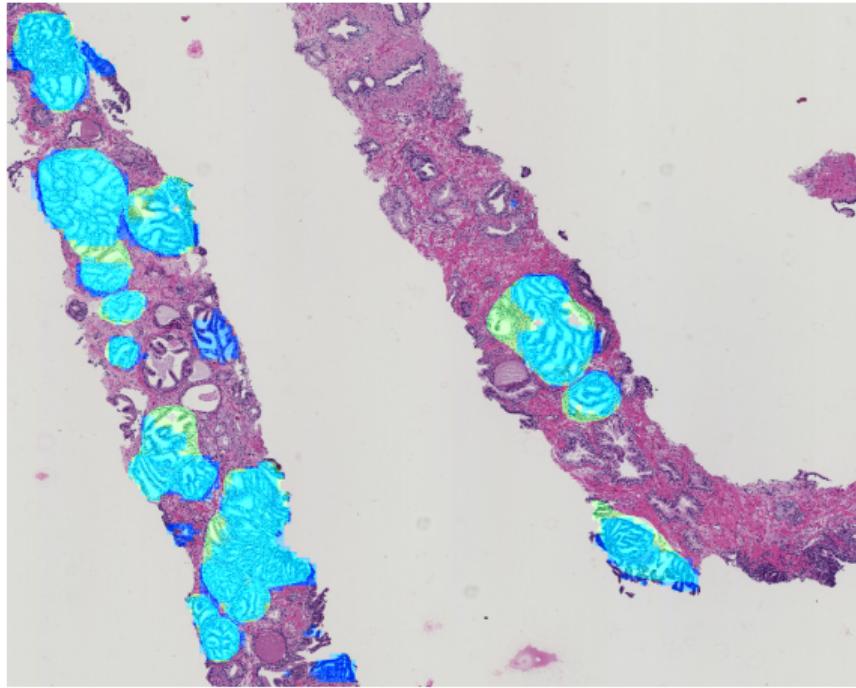
**BONUS**

AH Aubergine

0.59
per stuk**BONUS****4.08**
6 x 0,33 l**3.06**

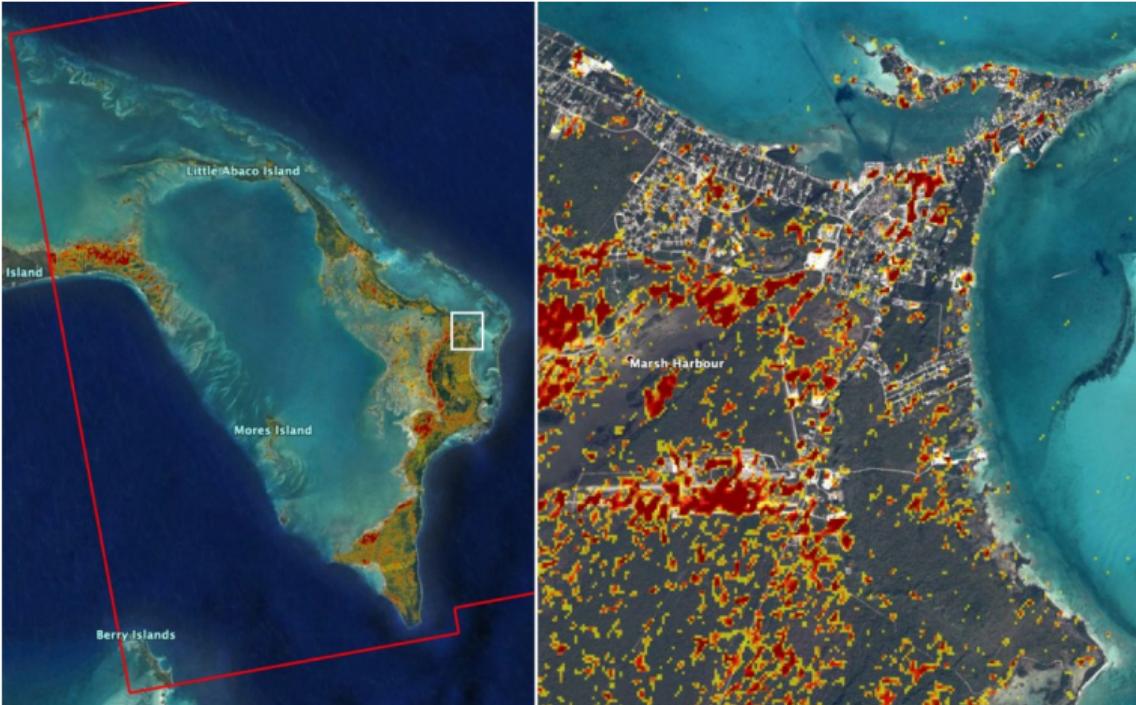
Coca-Cola Regular





Tumor growth pattern detection

Source: P. Ambrosini



Source: NASA

Can AI algorithms fail?

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's ([AMZN.O](#)) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

**nieuwsuur**

NIEUWSUUR • BINNENLAND • POLITIEK • MA 21 OKTOBER, 19:13

VN-rapporteur zeer bezorgd over Nederlands opsporingssysteem voor uitkeringsfraude

De VN-rapporteur voor de mensenrechten Philip Alston heeft ernstige zorgen over Nederland. De reden is een systeem dat uitkeringsfraude moet opsporen. In een brief aan de rechtbank in Den Haag schrijft Alston dat het systeem in strijd is met de mensenrechten omdat het mensen met weinig geld en mensen met een migratie-achtergrond discrimineert.



Sign in

Menu

NEWS

[Home](#) | [Video](#) | [World](#) | [UK](#) | [Business](#) | [Tech](#) | [Science](#) | [Stories](#) | [Entertainment & Arts](#)

[Business](#) | [Market Data](#) | [Global Trade](#) | [Companies](#) | [Entrepreneurship](#) | [Technology of Bus](#)

Apple's 'sexist' credit card investigated by US regulator

11 November 2019



Share

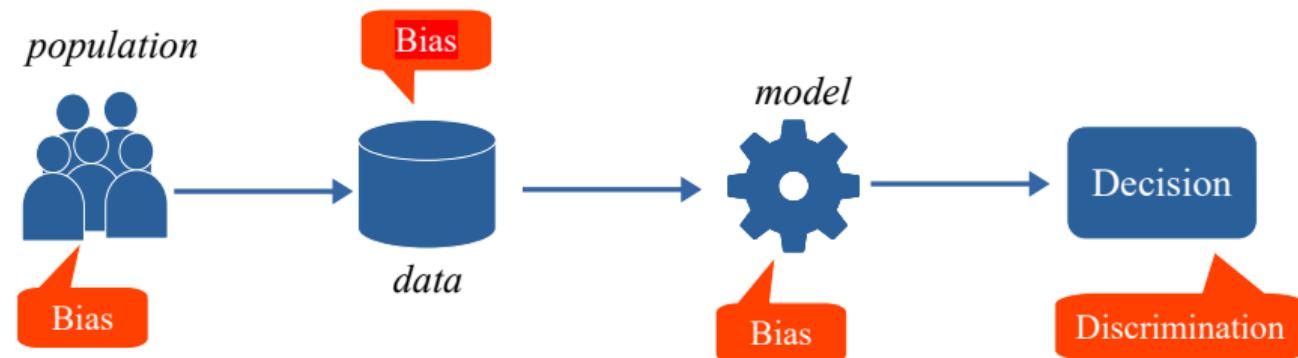
“Algorithms are written and maintained by people, and machine learning algorithms adjust what they do based on people’s behavior. As a result algorithms can reinforce human prejudices.”

C.C. Miller. When algorithms discriminate, NYT, 2019.

What is bias?

- Systematic errors that create unfair outcomes
- Sources: algorithm design, biased data collection or selection
- Algorithms learn and perpetuate bias

Bias occurs throughout the algorithmic pipeline



Types of bias

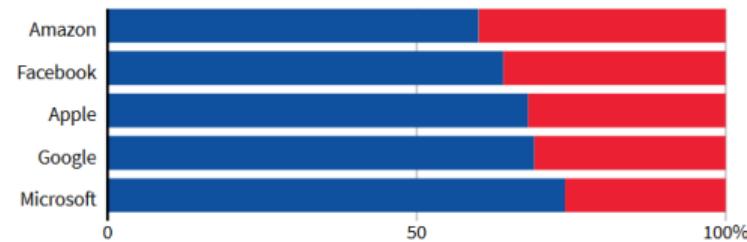
- **Historical bias** reflects structural societal issues
- **Representation bias** certain groups are under-represented in the training data
- **Measurement bias** training data are proxies for some ideal features and labels

simplified from Suresh & Guttag. A Framework for understanding unintended consequences of machine learning, 2019.

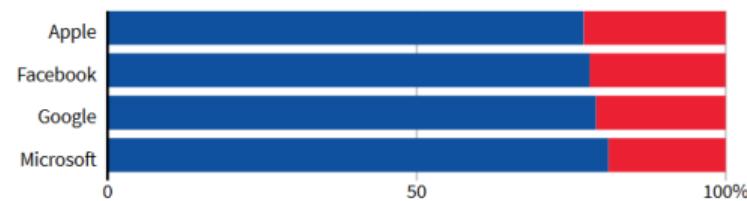
Historical bias

GLOBAL HEADCOUNT

■ Male ■ Female



EMPLOYEES IN TECHNICAL ROLES

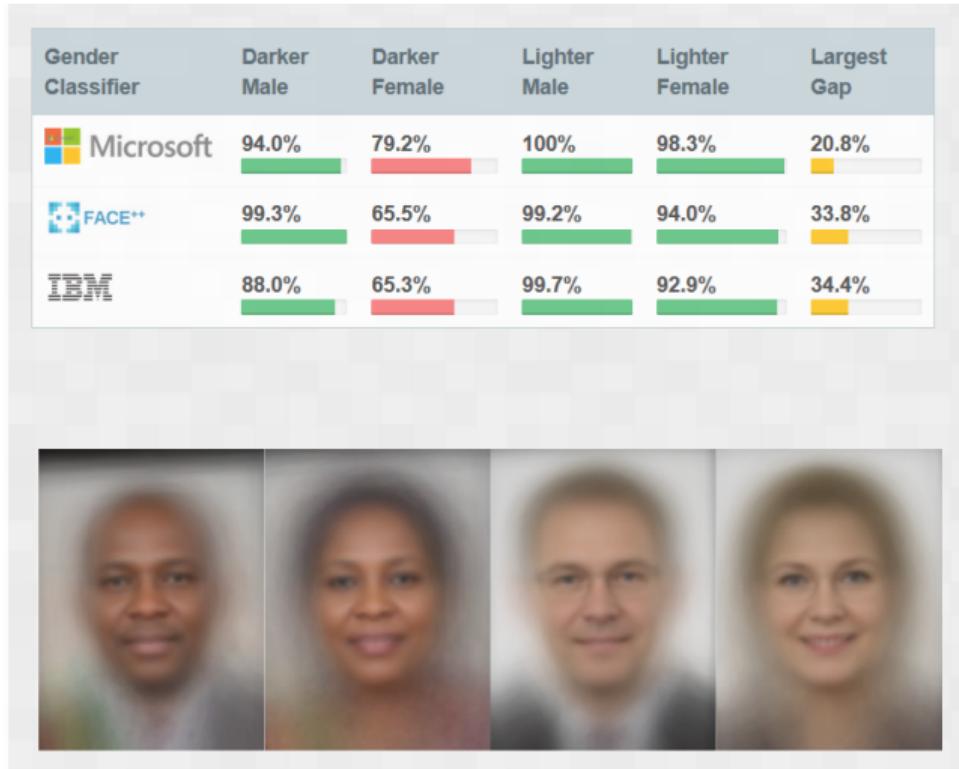


Note: Amazon does not disclose the gender breakdown of its technical workforce.

Source: Latest data available from the companies, since 2017.

By Han Huang | REUTERS GRAPHICS

Representation bias



<http://gendershades.org/overview.html>

Measurement bias

THE WALL STREET JOURNAL.

Subscribe Now | Sign In
SPECIAL OFFER: JOIN NOW

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate Search 

 Samsung Tries toAppease InvestorsBut Delays BigChanges

 Big Names TakeHit on Theranos

 HP EnterpriseUnveils Prototype ofNext-GenerationComputer

MARKETS

 Zenefits Hit With \$7 Million Fineby California Insurance Regulator

 HowUnilever Gets Hold Up to 80Days

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012

It was the same Swingline stapler, on the same [Staples.com](#) website. But for Kim Wamble, the price was \$15.79, while the price on Trude Frizzell's screen, just a few miles away, was \$14.29.

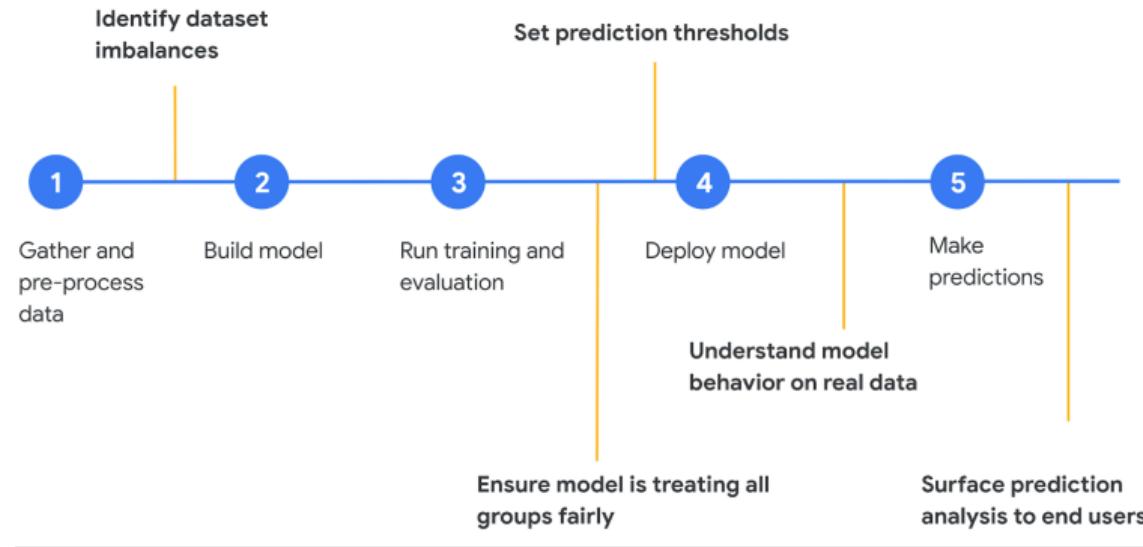
A key difference: where Staples seemed to think they were located.

Recommended Videos

- How Jill Stein's Election Recount Efforts Could Play Out
- Great Barrier Reef Suffers Largest Die-off

How can we avoid AI bias?

Best practice



Source: <https://ai.google/>

Regulation: GDPR

“Data subjects have a right to **meaningful information** about the **logic involved** and to the significance and the **envisaged consequence** of automated decision-making”

Ethics



FAT AI: Fairness Accountability & Transparency in AI



Bringing together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning

<https://www.fatml.org/>

Tools: bias testing and mitigation

The screenshot shows the homepage of the AI Fairness 360 Open Source Toolkit. At the top, there is a dark navigation bar with the text "IBM Research Trusted AI" on the left and a horizontal menu with the following items: Home (underlined in blue), Demo, Resources, Events, Videos, and Community. Below the navigation bar, the main content area has a light gray background. The title "AI Fairness 360 Open Source Toolkit" is centered at the top of this area. Below the title, there is a paragraph of text describing the toolkit's purpose and features. At the bottom of the main content area, there are two buttons: "API Docs" (in a dark gray box) and "Get Code" (in a blue box). The URL "https://ai-fairness360.com/" is visible at the very bottom of the page.

AI Fairness 360 Open Source Toolkit

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Containing over 70 fairness metrics and 10 state-of-the-art bias mitigation algorithms developed by the research community, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. We invite you to use it and improve it.

API Docs ↗ Get Code ↗

Open Data



Dataset

Werk en inkomen (Buurten)

Diverse datasets met statistieken van Onderzoek, Informatie en Statistiek.

Thema: Werk en inkomen,

Detailniveau: Buurten

Resources

Data

2.1 Bedrijfsvestigingen en werkzame personen 1) naar buurten, 1 januari 2016-2019

<https://api.data.amsterdam.nl/dcatt/datasets/-/fotdSpwjmSK9Q/purls/1>

2.2 Bedrijfsvestigingen buurten naar hoofdfunctie, 1 januari 2019

<https://api.data.amsterdam.nl/dcatt/datasets/-/fotdSpwjmSK9Q/purls/2>

2.3 Werkzame personen buurten naar hoofdfunctie, 1 januari 2019 1)

<https://api.data.amsterdam.nl/dcatt/datasets/-/fotdSpwjmSK9Q/purls/3>

2.4 Bedrijfsvestigingen en werkende personen buurten naar sectoren, 1 januari 2019

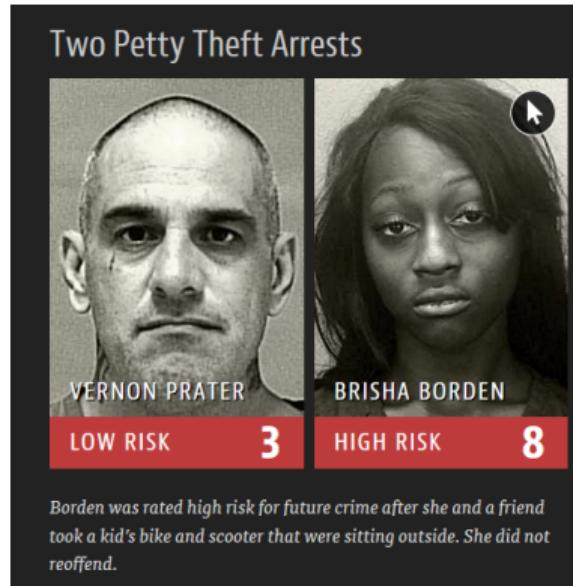
<https://api.data.amsterdam.nl/dcatt/datasets/-/fotdSpwjmSK9Q/purls/4>

2.5a Startende ondernemers naar buurten, 2013-2018 1)

<https://api.data.amsterdam.nl/dcatt/datasets/-/fotdSpwjmSK9Q/purls/5>

Open Data

“Through a public records request, ProPublica obtained two years worth of COMPAS scores from the Broward County Sheriff’s Office in Florida. We received data for all 18,610 people who were scored in 2013 and 2014.”



Ask fundamental questions

1. Why do you need AI for this task?
2. Is the system transparent?
3. When and how does the system fail?
4. Who is responsible for the errors?

Algorithmic bias: no quick fix

“Any real machine-learning system seeks to make some change in the world. To understand its effects, then, we have to consider it in the context of the larger socio-technical system in which it is embedded.”

Barocas et al. Fairness and machine learning, fairmlbook.org, 2019.

Diverse
Domains
e.g. aviation, justice

Diverse
Users
e.g. pilots, judges

Diverse
Criteria
e.g. fairness, privacy

Interdisciplinary Research

social science – mathematics – computer science – law – ethics

Aircraft safety

Adopt AI systems while ensuring transparency to stakeholders throughout the algorithmic pipeline.



Prof. Dr. Leon Gommans

Forensic evidence evaluation

Leveraging more performant models while ensuring transparency through explanations.



Dr. Corina Benschop

Socially aware AI

Empower and connect citizens and communities in a fair and inclusive manner, including those at the margins of society.



Dr. Sennay Ghebreab

Thank you

h.haned@uva.nl

github.com/hindantation