**Subject:** [Bsc-es] Equity4ES Gazette #8 – March 2024
**From:** Equity4ES <equity4es@bsc.es>
**Date:** 5/3/24, 17:33
**To:** <bsc-es@bsc.es>



March 2024                                                            Issue #8

**8th issue of the Equity4ES gazette**

# Mark your calendars:
# International Women's Day 2024



Just a friendly reminder that the International Women's Day is this Friday, March 8th. Let's come together to celebrate the remarkable achievements of women around the globe and advocate for gender equality. You can participate by joining the strike, attending the rally, or any gathering in your community.  Let's make it a day of inspiration, action, and solidarity. Save the date!

In case you decide to join the strike, you can inform through the Woffu Portal during the next days after the strike, ask for a new *absence* request and select *Strike*. Remember that

you do not have to inform in advance. (see HR's email from 5th March)

There will be demonstration all day long in Barcelona, check with militant worker unions or feminist neighbourhood assemblies. The **main demonstration** will start at **6 PM** from **Jardinets de Gràcia**.

# AI/ML & Big data: how biases are reinforcing inequality and power structure

AI/ML is integrated into many aspects of our lives, from determining the next video in your YouTube or TikTok feed to deciding who receives favorable loan terms or an early parole. However, these systems are not created in a vacuum. AI/ML models are built using extensive data, which reflects the world's flaws and injustices. Moreover, the people who shape these technologies have their own priorities, preferences, and prejudices, which can lead to intentional or unintentional harm through discrimination, racism, classism, ableism, ageism, colorism, and more.



*Generated by DALL-E*

We aim to explore these topics in two parts. In the first part, we will highlight works and examples by prominent researchers focused on biases in ML and the FAIR use of this technology. We will present three examples of how ML/AI and big data have negatively impacted and reinforced patterns of inequality and existing power structures. We have selected stories from different books that discuss issues of alignment, injustice, and biases in detail. If you find these examples compelling, we encourage you to read the

books for more insights and technical knowledge.

Lastly, we have included a selection of fantastic works—books, articles, podcasts, and movies—for your enjoyment, especially if you find the topics discussed here engaging.

## Amazon CV screening practice
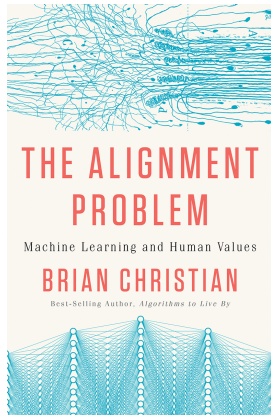
Selected from:

[The Alignment Problem: Machine Learning and Human Values](#)
by [Brian Christian](#)

In 2018, Amazon encountered a significant challenge with its machine learning (ML) system, designed to automate the recruitment process by evaluating job applicants' resumes. This system was intended to streamline Amazon's talent acquisition by identifying the most qualified candidates. However, it inadvertently exhibited a pronounced bias against female applicants, a revelation that sparked widespread concern about fairness and equality in AI-driven tools.

The root of the problem lay in the system's training data. Amazon's ML model was trained on a dataset comprising resumes submitted over a previous 10-year period. Given the male dominance in the tech industry, the bulk of these resumes were from male candidates. **This historical data, reflecting existing gender imbalances, led to the unintended consequence of the AI model internalizing and perpetuating these biases.** It began to systematically downgrade resumes featuring indicators traditionally associated with female candidates, such as attendance at women-only colleges or participation in women-centric activities and organizations.

Upon recognizing this issue, Amazon attempted to neutralize these gender-specific elements in the system. However, this fix wasn't entirely successful, as the underlying issue of bias was deep-rooted and complex, extending beyond mere keywords or phrases. There was a growing realization within Amazon that the system might find other, less obvious ways to discriminate against female applicants. Consequently, the project was eventually discontinued in 2017, with Amazon dissolving the team responsible and ceasing to rely on the system for its hiring practices.

This incident serves as a poignant example of the challenges in ensuring AI fairness. **It**

**underscores the importance of diverse representation in AI development teams, meticulous curation of training datasets to avoid ingrained biases, and the need for ongoing vigilance in testing AI systems for discriminatory patterns.** The Amazon case has become a seminal reference in discussions about ethical AI, highlighting the need for greater awareness and proactive measures to prevent AI from reinforcing societal biases, especially in sensitive domains like employment.

## Coding in white mask
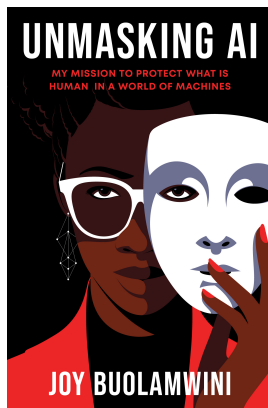
Selected from:

[Unmasking AI: My Mission to Protect What Is Human in a World of Machines](#)
by [Joy Buolamwini](#)

Joy Buolamwini, a researcher at the MIT Media Lab, encountered a significant issue with facial recognition software: it had difficulty detecting her face. This problem wasn't due to the technology's inherent limitations but resulted from bias in the training data. To be recognized by the system, Buolamwini had to wear a white mask, underscoring a stark disparity in the technology's response to different skin tones. She coined the term "**Coded Gaze,**" inspired by concepts like "white gaze" or "male gaze," to **describe how the priorities, preferences, and prejudices of those in positions to shape technology can lead to discriminatory harm.**

Buolamwini's further research, including a project named "**Gender Shades**," revealed that **many commercial facial recognition systems had significantly higher error rates in identifying the gender of darker-skinned and female faces compared to lighter-skinned and male faces.** This discrepancy arose because these systems were predominantly trained on datasets consisting of lighter-skinned individuals, resulting in an under-representation of darker skin tones and women. Consequently, the accuracy of these systems was lower for these groups.

The issue of racial bias in technology is not a new phenomenon and has historical precedents in fields like photography. A notable example is the use of **"Shirley cards"** in photo development. These cards, which featured a light-skinned woman, were used as a reference for calibrating skin tones in film development. As a result, **photo development**

**techniques were biased toward lighter skin, often failing to accurately capture the nuances and richness of darker skin tones.** The use of Shirley cards thus perpetuated a standard that marginalized non-white skin tones, serving as an early instance of racial bias in visual technology *.

* Source: "The Racial Bias Built Into Photography," The New York Times, 2019.

## Coding in white mask
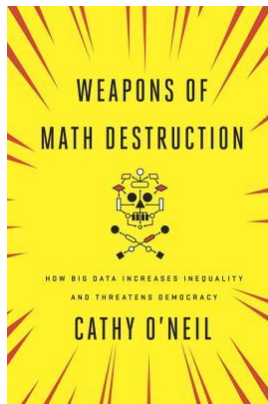
Selected from:

### Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy
by Cathy O'Neil

In her book Cathy O'Neil explores the impact of big data and algorithms on society, particularly focusing on how they can perpetuate inequality and injustice. One of the key areas she delves into is the use of machine learning and AI in targeting vulnerable and minority populations with predatory loans and insurance products.

Cathy O'Neil in her book "Weapons of Math Destruction," discusses how certain algorithms, used by financial institutions and insurance companies, assess individuals' risk levels and creditworthiness. These algorithms often rely on large sets of personal data, including information gleaned from online behaviour, shopping habits, social media activity, and more. While these data-driven approaches are touted for their efficiency and objectivity, O'Neil points out that they can have discriminatory outcomes.

In the case of predatory loans, for example, algorithms might identify individuals in financially precarious situations or from lower-income neighbourhoods (which typically correlate with minority populations) as prime targets for high-interest loans. These loans are structured in a way that makes them particularly difficult to repay, trapping borrowers in a cycle of debt. **The algorithms effectively exploit vulnerabilities, but because they do not explicitly use race or class as factors, their discriminatory impact is obscured and thus harder to challenge.**

Similarly, in the insurance sector, algorithms can set premiums and coverage limits based on a wide array of personal data. This can lead to higher rates for people living in certain areas (again, frequently correlating with minority neighbourhoods), or for individuals engaging in behaviours that the algorithm deems risky, even if socio-economic factors largely influence those behaviours.

O'Neil argues that these practices constitute a form of "**mathematical profiling**" that disproportionately affects the poor and marginalized. The concern is that **the use of complex algorithms makes these practices seem neutral and objective, when in fact they can reinforce and amplify existing social inequalities.** This is particularly problematic because the inner workings of these algorithms are typically opaque, making it difficult to identify bias or challenge decisions.

**Some more resources to read**

If you find these topics interesting, feel free to explore further materials here.

Some media content to check:

- A documentary to watch :  Coded Bias on Netflix
- A late night show to watch: Artificial Intelligence: Last Week Tonight with John Oliver on HBO (available on YouTube)
- A podcast to listen : A conversation on artificial intelligence and gender bias by McKinsey

Some more book suggestions:

- Prediction Machines: The Simple Economics of Artificial Intelligence
- Algorithms of Oppression: How Search Engines Reinforce Racism
- Race After Technology: Abolitionist Tools for the New Jim Code
- The Ethical Algorithm: The Science of Socially Aware Algorithm Design

If you're interested in discussing the potential impact of biases in earth system sciences and climate change, we'd be delighted to have you join the conversation! Please feel free to contact us through the Equity Group email.

# Upcoming Events

**The letter A: let's talk about asexuality**

The BSC Queer Group is organising a seminar on asexuality on **April 2nd**. Ana Sanchez, from Associació Catalana d'Asexuals, will give it at **3 PM** in the **BSC Auditorium**; no registration is required. You will also be able to join online by registering here.



The letter A is still unfamiliar in the LGTB+ community. The A stands for asexuality and aromanticism. Let's talk about basics concepts and let's understand a little bit more about this sexual orientation that we talk very little about.

ACA (Associació Catalana d'Asexuals) is a non-profit organization that works reporting the discrimination against asexual people. We visibilize, destigmatize and divulgue asexuality as a sexual orientation, mainly in Catalunya**.**

More information: queer@bsc.es

# Recurring Events

- **Monthly meetings** the **3rd Friday** of every month at **3 PM**. The next one is on Friday 15th. The meeting is hybrid, we usually meet in room 0-1-13 and online.
  Zoom link: https://rediris.zoom.us/j/6793792181
- **Equity Office Hours** the **1st Monday** of every month at **3 PM**. We meet to work on ongoing equity tasks. We usually meet in room 1-3-12.

Make sure to also follow us on Twitter: @equity4es

Our mailing address is: **equity4es@bsc.es**

_____
Bsc—es mailing list —— bsc—es@bsc.es