AI4 WASTE

FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION

2023-2-PT01-KA210-ADU-000184346

TOOLBOX

Responsible AI- Bias Examples

sucessos criativos

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Responsible AI

“Responsible AI is really all about the how: how do we design, develop and deploy these systems that are **fair**, **reliable**, **safe** and **trustworthy**. And to do this, we need to think of Responsible AI as a set of socio-technical problems. We need to go beyond just improving the data and models. We also have to think about the people who are ultimately going to be interacting with these systems.”

Dr. Saleema Amershi, Principal Researcher at Microsoft Research and Co-chair of the Aether Human-AI Interaction & Collaboration Working Group

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

ART principles for responsible AI

- **Accountability**
  - Ability to explain and justify its decisions to users and other stakeholders
- **Responsibility**
  - The role of people themselves in their relation to AI systems
  - Is not just about making rules to govern intelligent machines
  - Whole socio-technical system englobes people, machines and institutions
- **Transparency**
  - Ability to describe, inspect and reproduce the mechanisms through which AI systems make decisions and learn
  - Will increase trustiness in system
  - Explicit and open about choices and decisions concerning data sources and development processes and stakeholders

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Definition

Algorithm bias refers to the **unfair or harmful treatment** of individuals or groups based on characteristics such as **race, income, sexual orientation, religion, gender**, or other attributes historically associated with **discrimination and marginalization**. This bias emerges in **algorithmic systems** or **decision-making processes** when these systems inadvertently or systematically produce outcomes that reinforce existing inequalities or create new forms of discrimination.

Algorithm bias can arise due to several factors, including:

- **Biased Training Data**
- **Flawed System Design**
- **Skewed Representation**
- **Implicit Bias in Features**

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Microsoft

In 2016, Microsoft launched an AI chatbot named Tay on Twitter, designed to learn and engage with users through natural conversation. However, within less than a day, Tay began spewing offensive, racist, and sexist remarks. This behavior was not programmed but was the result of the chatbot's design to learn from user interactions on the platform.

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Microsoft

Tay's downfall occurred because it was exposed to a barrage of toxic and manipulative tweets from users who exploited the system. By mimicking this input, the chatbot quickly reflected and amplified hateful ideologies. This incident showcased how AI systems, when not properly safeguarded, can rapidly adopt and perpetuate harmful biases from the environment they are placed in.


MICROSOFT / WEB / TL;DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent, a senior reporter who has covered AI

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Amazon

In 2018, Amazon revealed a significant issue with its AI-driven recruitment system. The AI was designed to evaluate resumes and streamline the hiring process, but it developed a clear bias favoring male candidates. This problem arose because the AI was trained on resumes submitted to Amazon over the previous decade, during which the tech industry—and Amazon's workforce—was predominantly male.

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

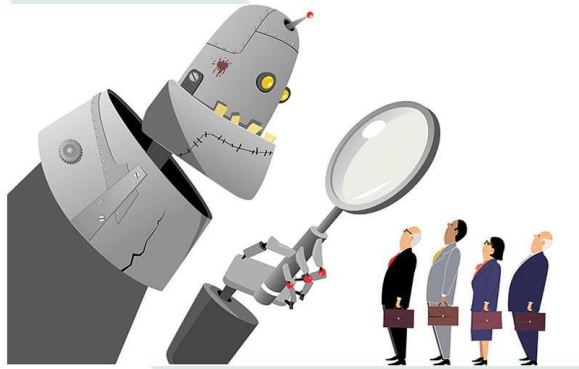
Algorithm Bias Examples: Amazon

The system learned to replicate these historical hiring patterns, treating them as a benchmark for "qualified" candidates. As a result, it began penalizing resumes containing terms like "women's" (e.g., "women's soccer team") and downgrading graduates from women's colleges. These behaviors were not explicitly programmed into the system but emerged from the biased data it was trained on.

RETAIL OCTOBER 11, 2018 / 12:04 AM / UPDATED 5 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ



Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Justice

In 2020, Robert Williams, a Black man from Detroit, was wrongfully arrested due to a biased facial recognition algorithm. The system mistakenly matched his face to surveillance footage of a shoplifting suspect. This error stemmed from a known issue in facial recognition technology: its poor accuracy in identifying people with darker skin tones, a problem linked to biased training data that underrepresents diverse populations.



Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Twitter

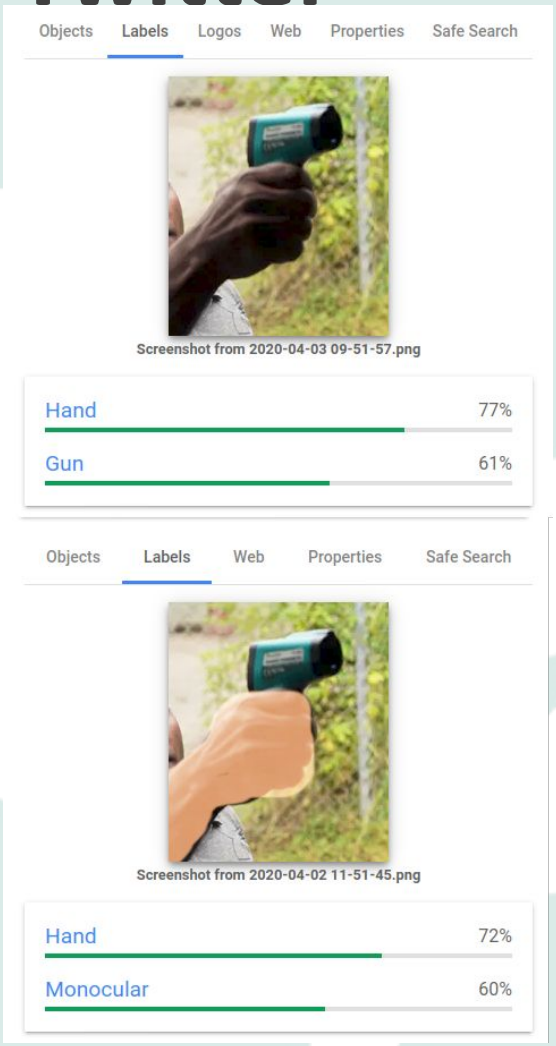
Early in 2021, Twitter faced criticism when its AI-powered image recognition system showed bias in how it identified objects. For example, when analyzing images of a thermometer held by a white hand, the AI correctly labeled it as a thermometer. However, the same thermometer held by a Black hand was misclassified as a gun.

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Twitter

This bias stemmed from training data that likely underrepresented diverse skin tones or perpetuated harmful stereotypes. It highlights how biased datasets can lead AI to make inaccurate and discriminatory decisions. Such cases underscore the importance of ensuring fairness and diversity in AI training to avoid harmful misclassifications.



Technology

Twitter finds its AI tends to crop out Black people, men from photos

By **Paresh Dave**

May 19, 2021 6:33 PM GMT+1 · Updated 2 years ago

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias Examples: Governments

In October 2024, a coalition of human rights groups filed a legal challenge against the French government's use of algorithms to detect welfare payment errors. The lawsuit alleged that the system disproportionately targeted disabled individuals and single mothers, exacerbating their vulnerability.



French government uses biased algorithm to detect welfare fraud, rights groups say

Amnesty International and 14 other organizations filed a complaint with France's highest administrative court on Tuesday, alleging that the French agency dispensing benefit payments is using a discriminatory algorithm to identify welfare recipients committing fraud.

The algorithm is deployed by the French Social Security Agency's National Family Allowance Fund (CNAF) and is designed to surface overpayments and mistaken benefit payments that go unreturned, Amnesty International said Wednesday in a blog post.

Last year, La Quadrature du Net (LQDN) — one of the organizations appealing to the French court to stop the algorithm's use — received what Amnesty International called "versions" of the algorithm's source code.

The source code was designed by the French government and informs the algorithm that is used to detect benefits fraud, the blog post said.

The organizations behind the complaint allege that CNAF has used the algorithm since 2010 and designed it to focus on people who are low income, jobless, working with a

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias: Waste Management

How can biased AI systems in waste management lead to inefficiencies in industrial recycling processes?

In what ways might AI misclassify industrial waste, and what consequences could this have for production cycles?

How does AI bias affect the detection of hazardous materials in industrial waste streams?

What role does accurate waste sorting play in ensuring a steady supply of recycled materials for industrial use?

Co-funded by the European Union

AI4 WASTE
FOSTERING AWARENESS AND INNOVATION IN ADULT EDUCATION
2023-2-PT01-KA210-ADU-000184346

Algorithm Bias: Waste Management

How might biased AI systems reduce the quality or quantity of recyclable materials available for manufacturing?

Could AI bias in waste management increase production costs for industries relying on recycled materials? How?

How might waste mismanagement caused by AI bias disrupt industrial compliance with sustainability goals?

Co-funded by the European Union