Saint George on a Bike

How are data collections and vocabularies teaching AI systems human stereotypes?

What are we talking about?

Idea and concepts

Digitalization of CH objects for machine ingestion may have started for **preservation** reasons, but it allowed **applying AI technology** in order to extract knowledge that can improve user experience and be a **valuable** resource for GLAM. Some use case are

- Automatic metadata annotation
- Interaction with minorities such as visually impaired citizens
- New forms of interaction with users through web-pages and app
- Improving search engine

Idea and concepts

But **risks** and **benefits** of using **AI** are two sides of the same coin
Why?
The terms “fairness” and ”bias” are often used interchangeably to refer to whether different groups of people experience different performance from a AI model.

“Fairness”

“The terms ”fairness” and “bias” are often used interchangeably to refer to whether different groups of people experience different performance from a machine learning model.”

“Bias”

We define bias here as data that is in some way not representative of the real world.

“Building Machine Learning Pipelines Automating Model Life Cycles with TensorFlow” Hannes Hapke and Catherine Nelson

Is your data fair enough?

Why is fairness important?

Just like **humans**, artificial intelligence can be **sexist and racist**

Princeton University study finds machine learning **copies human prejudices** when learning language

Using the popular GloVe algorithm, trained on around 840 billion words from the internet, three Princeton University academics have shown AI applications replicate the stereotypes shown in the human-generated data.

These prejudices related to both race and gender.

Fairness in GLAM

Case 1

Saint George on a Bike Project

Two people were classified by two different classes: “Person” and “Monkey”

Two people were classified by two different classes: “Person” and “Monkey”

Fairness in GLAM

Case 2
IconClass

33C2 lovers; courting, flirting

Search with these related keywords:

human being, lover, relationship

Add more detail:

- 33C21 · courting
- 33C22 · lovers' meeting
- 33C23 · couple of lovers
- 33C29 · the envious friends; criticizing bystanders ~ love couple
- 33C2(+0) · lovers; courting, flirting (+ variant)

Classification of “Love” representation in IconClass

33C6 homosexual love

Search with these related keywords:

homosexual, human being, lesbianism, relationship

Add more detail:

- 33C61 · pederasty, sexual contact between man and boy
- 33C62 · sodomy, sexual contact between men
- 33CC6 · homosexual love - CC - homosexual love between women: lesbianism
- 33C6(+0) · homosexual love (+ variant)

Potentially Offensive grouping of different art content

Homosexuality in Art

Kyiv art museum, an icon from St. Catherine's Monastery on Mt. Sinai in Israel. It shows two robed Christian saints. Between them is a traditional Roman 'pronubus' (a best man), overseeing a wedding. The pronubus is Christ. The married couple are both men. Image source: medievaltumblr.com

How can it happen?

Data Leakage

Data leakage is a concept in AI when information from outside the training dataset is used to create the model. This additional information can allow the model to learn or know something that it otherwise would not know and in turn invalidate the estimated performance of the mode being constructed.

Pipeline of ML Project

Good intention

Good idea

We decided to train a recommendation system for a suggestion of Art objects based on previous searches of a user

Data collection

We scraped data and used tags of a publicly available vocabulary or classification system.
Date leakage

Data preparation

Manual checking is a highly resource demanding process. EDA was done by one person

Production

We trained a model and put it into production

Data is spoiled

Even if model shows good performance, without checking data for fairness any type of ML will be biased towards some groups

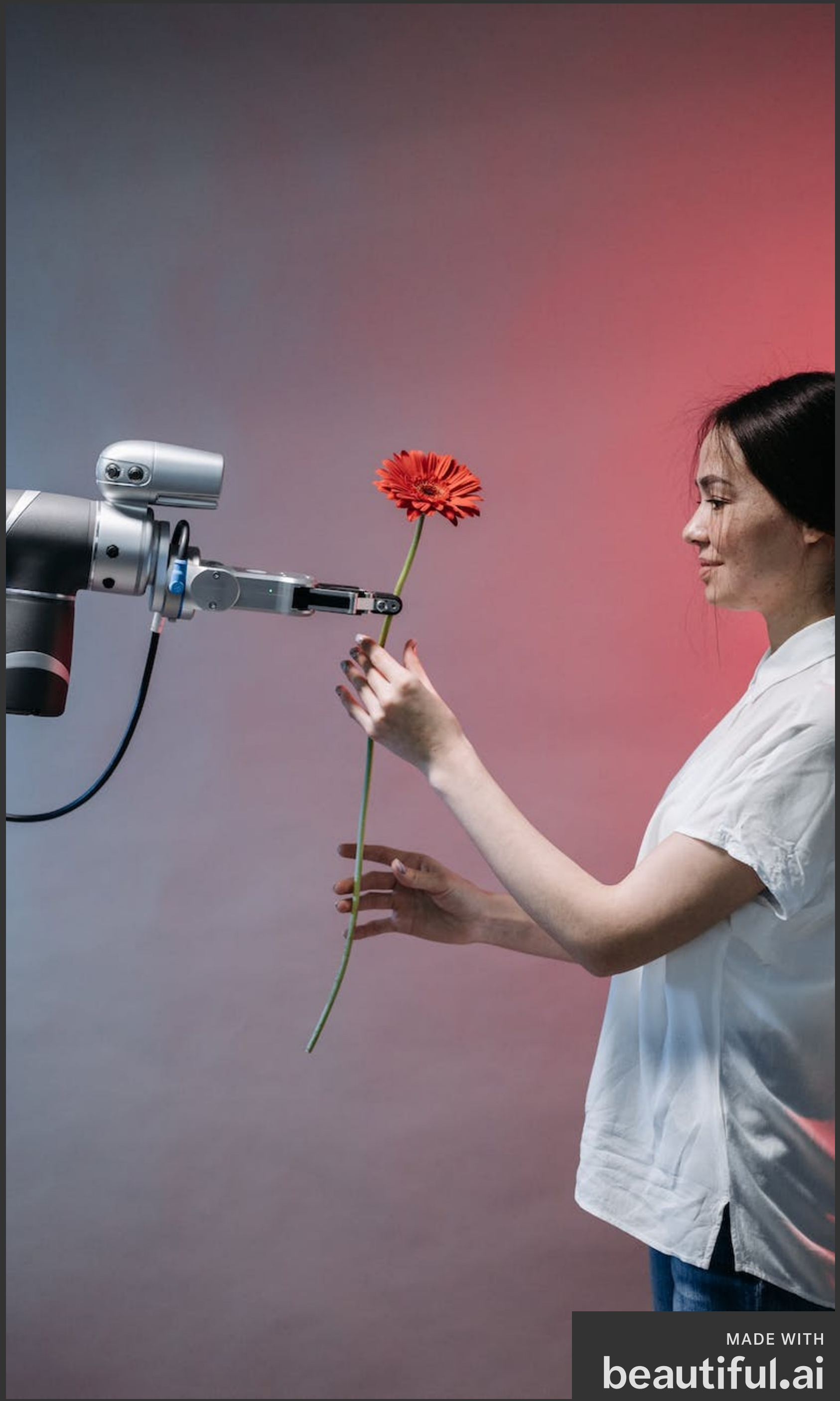
User experience

Without awareness of fairness, results will be biased.

Bad user experience

A guide to machine learning (ML) fairness - Think with Google

Algorithms don't remember incidents of unfair bias. But customers do.



What can we do?

Several steps which helps you to avoid bad user experience

Be transparent.

Tell people how your algorithm makes decisions. Knowing how your product works — and how well it works across groups — will make people more comfortable using it.

Test, tune, and test again.

Inspect training datasets for bias using a fairness indicator, visualizer, or other tool. Even a widely used dataset might have flaws, so it's important to review it carefully. Teams should also continue monitoring algorithms after they are released.

Seek different points of view.

Hire people with diverse backgrounds and areas of expertise. Invite the public to share local knowledge. Collaborate with community groups and advocates. A wide range of input makes data more robust.

Ask questions.

Does your product use data such as race, skin color, religion, sexual orientation, socioeconomic status, income, country, location, health, language, or dialect?

A guide to machine learning (ML) fairness - Think with Google

Questions?