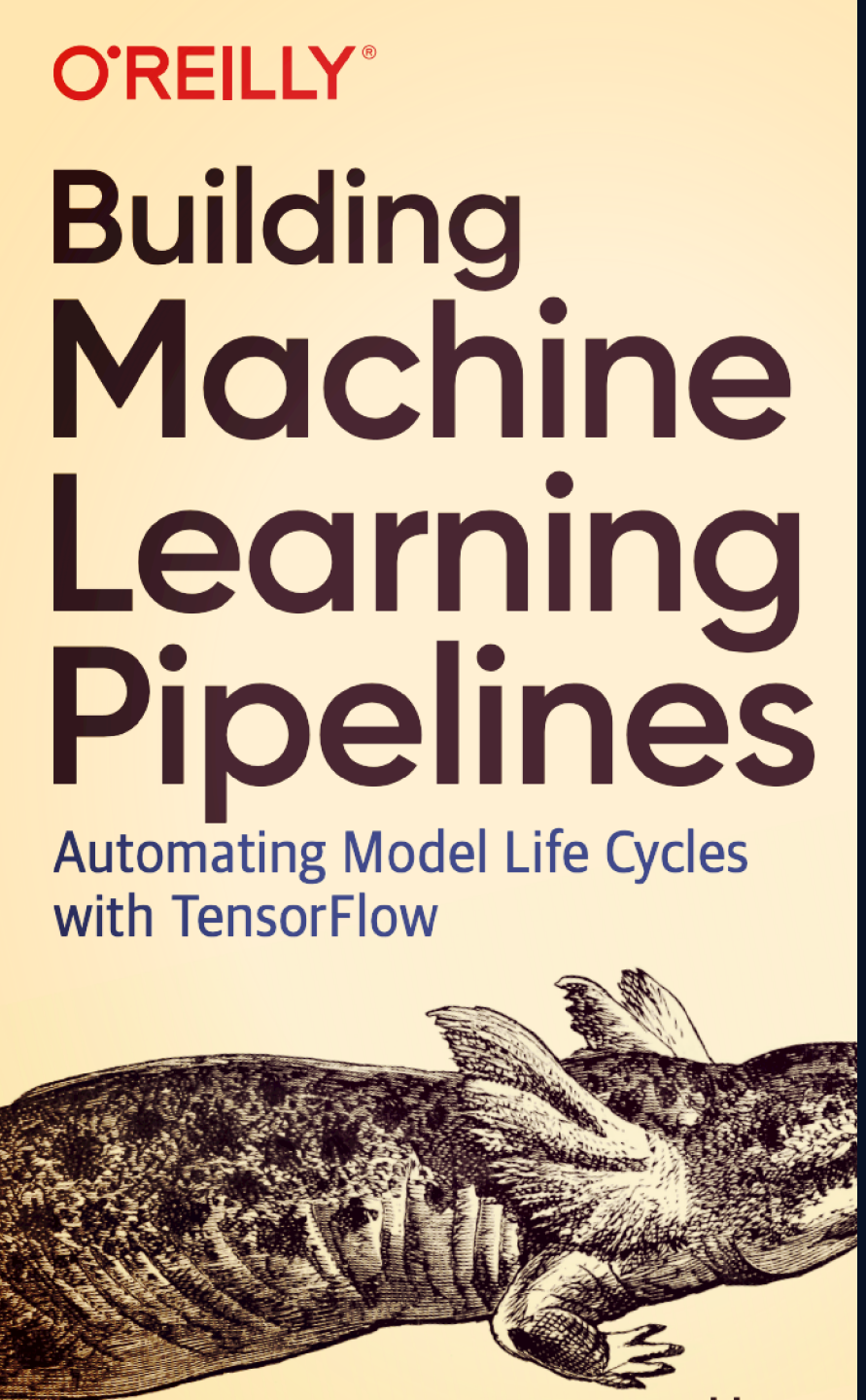
Fairness of data

Does our data work for everyone in a same way?



What we are talking about?

> To analyze whether our model is fair, we need to identify when some groups of people get a different experience than others in a problematic way. For example,a group of people could be people who don't pay back loans. If our model is trying to predict who should be extended credit, this group of people should have a different experience than others. An example of the type of problem we want to avoid is when the only people who are incorrectly turned down for loans are of a certain race.

Fairness

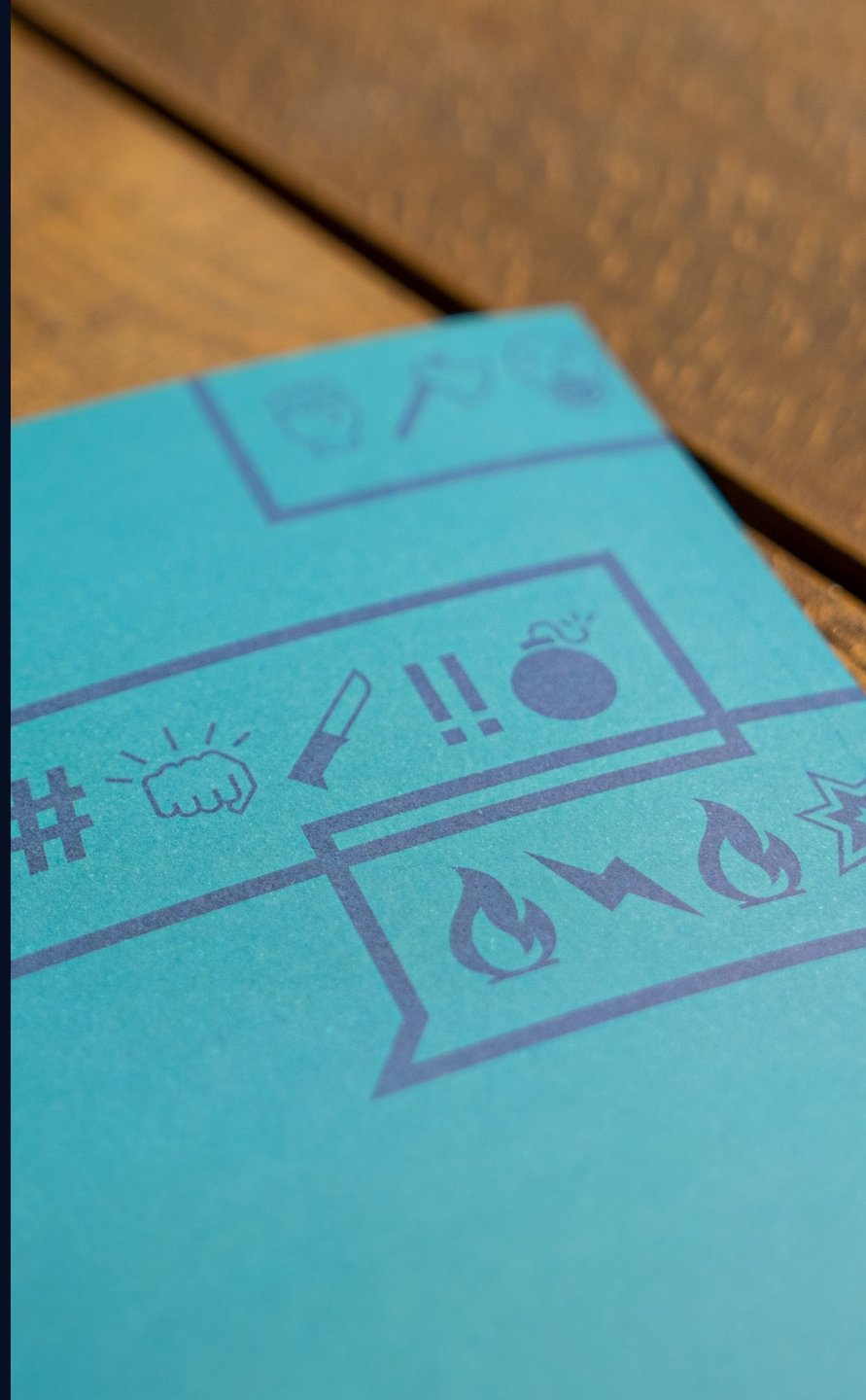
> We define bias here as data that is in some way not representative of the real world.

Biasing

Is your data is fair?

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.



Is data in cultural heritage is **FAIR**?

Hint: **NO**

Is it a problem with **my** dataset?

Hint: **No**

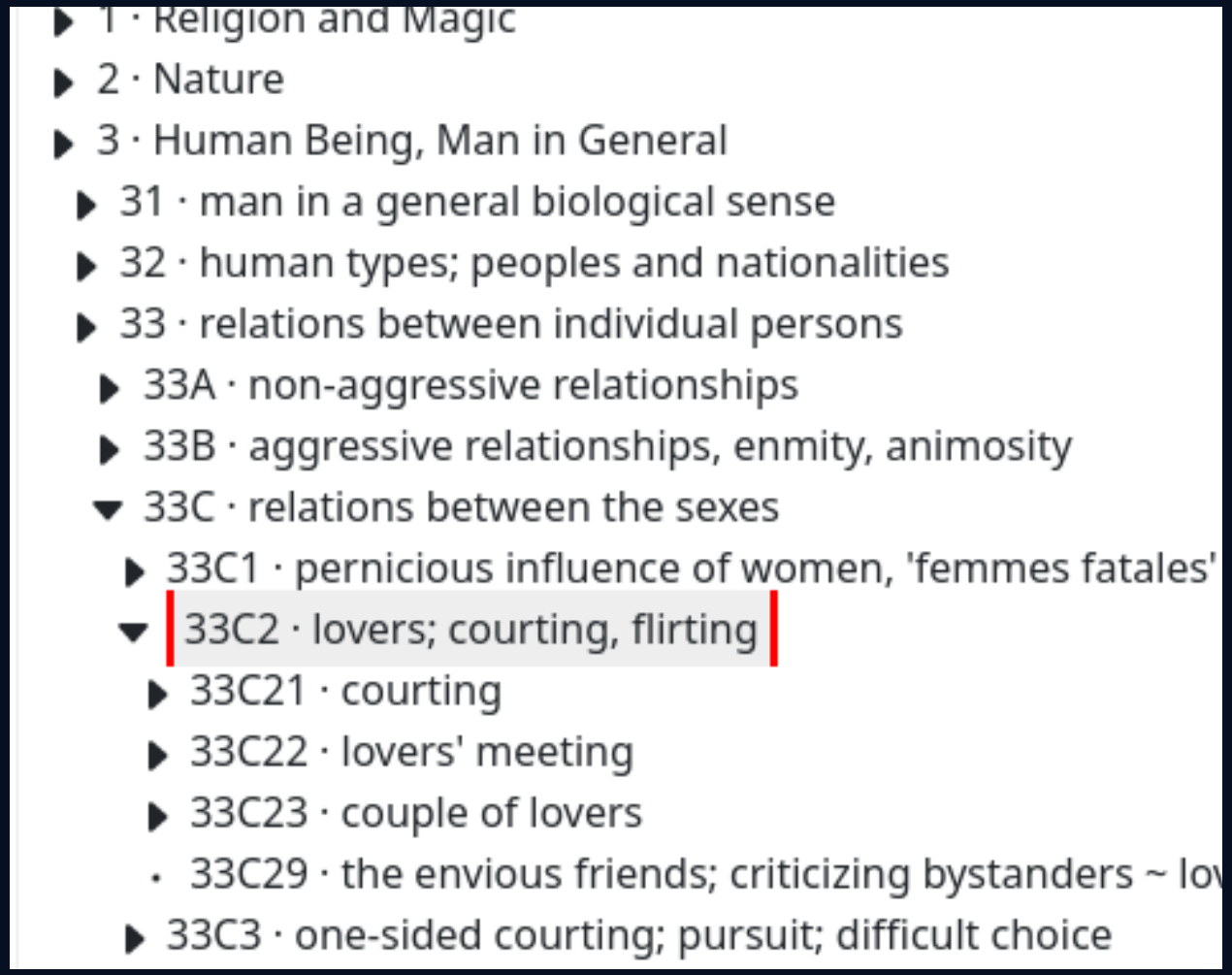
Facts against this statement

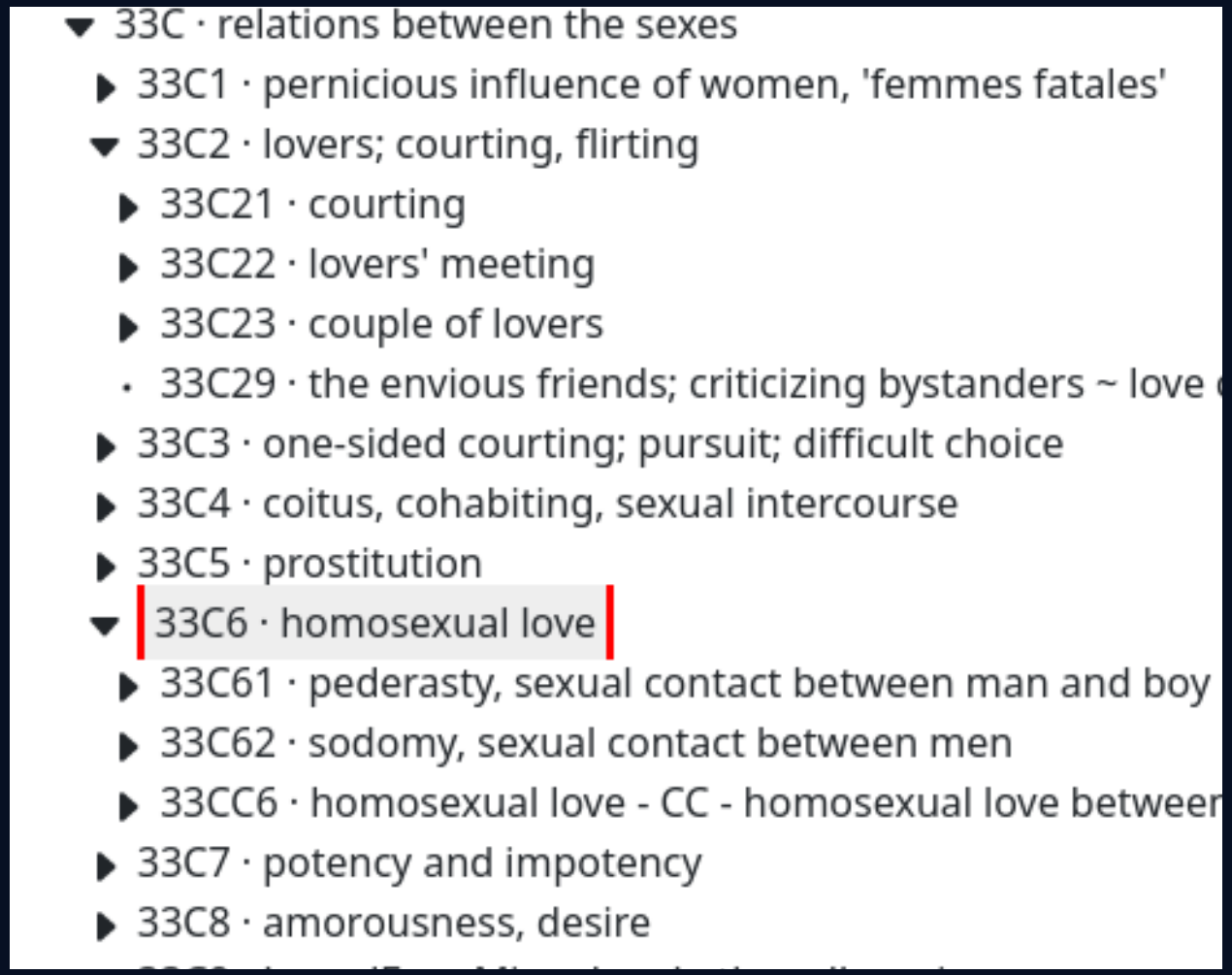
Model was trained using transfer leaning and it already has features for person detection

Data was collected without filtering by any visual scene context

Data follows the structure that predefined by experts before the epoch of fairness discussion

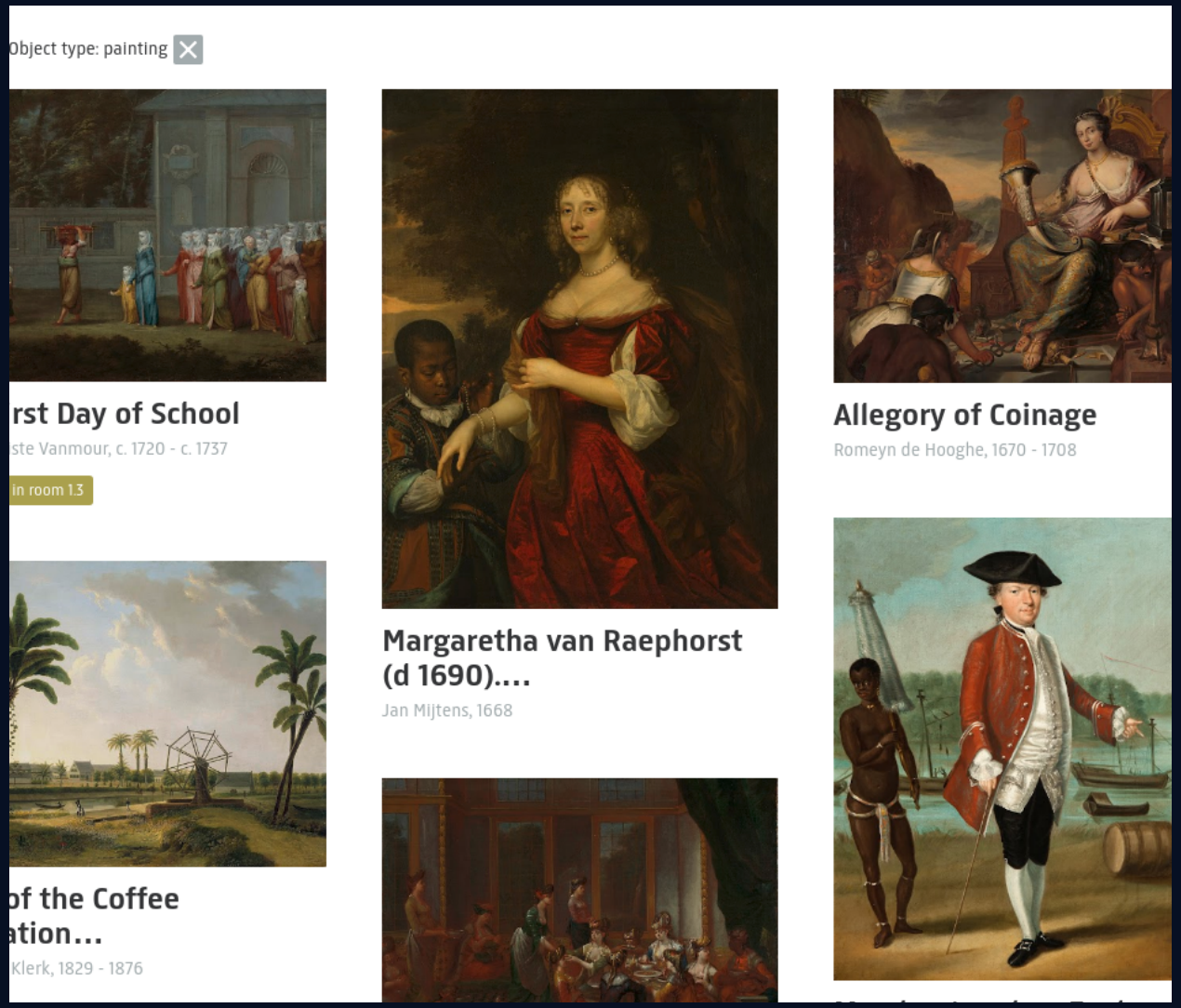
Discrimination by sexual orientation or representation of art?

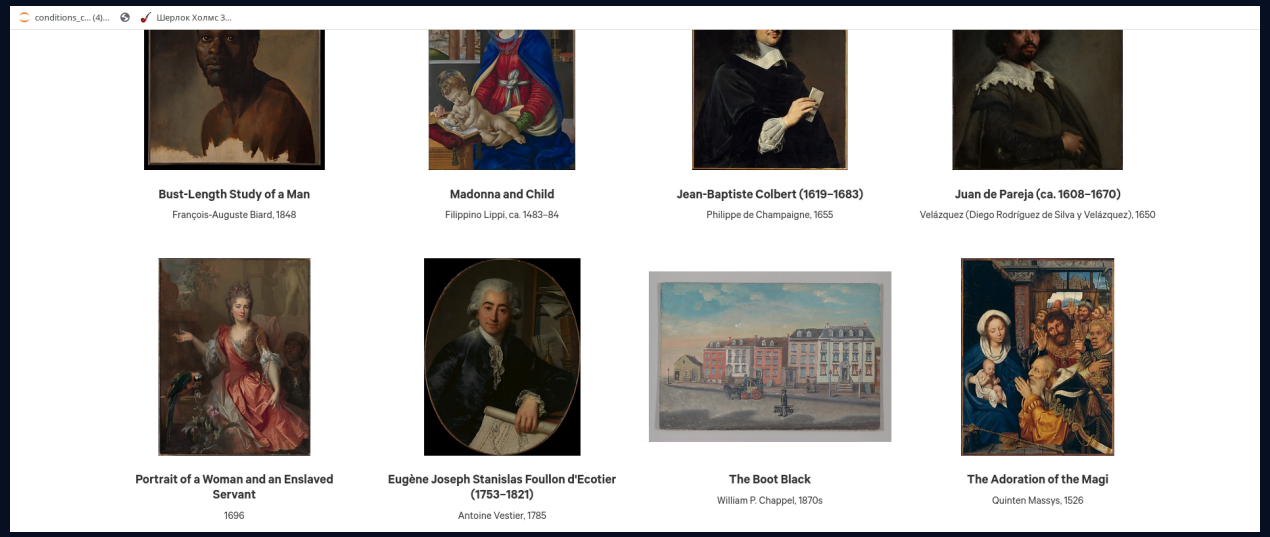


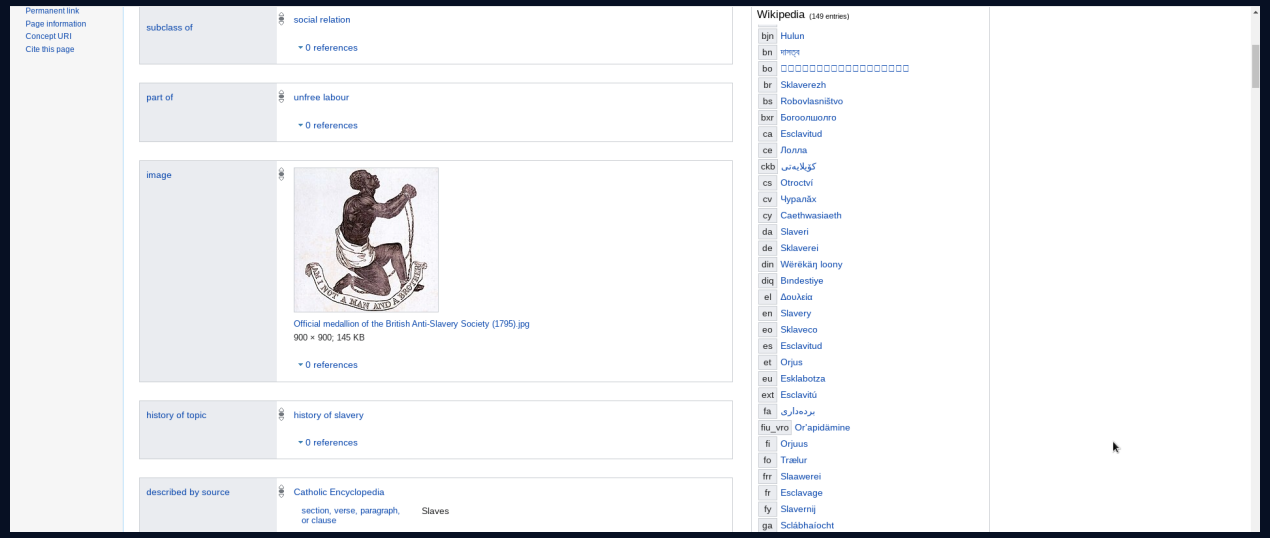


Does word “slavery” associate to specific ethnic group?

Hint: Data structure say “yes”

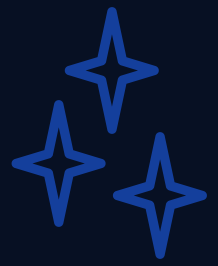






Algorithms don’t remember incidents of unfair bias. But customers do.

What we can do?



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.

What you know about your data?