Curating the future: the necessity of curation in open data and the threat of systemic algorithmic bias
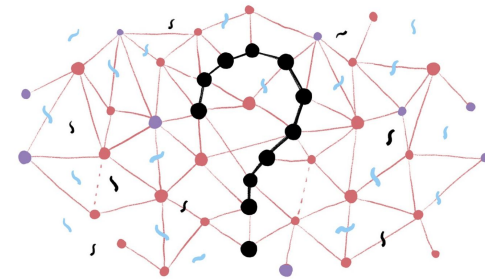
Ilias Koutsakis - CERN (RCS-SIS)

Algorithms? Why?

Machine Learning, Big Data, and other recent computational advancements, have drastically changed our lives and are at the core of many human processes nowadays (search engines, advertisements, etc).

However, those algorithms do not think. Instead, they merely exhibit the results of computations, implemented by humans and using human-generated data, which often leads to them unintentionally picking up human biases.
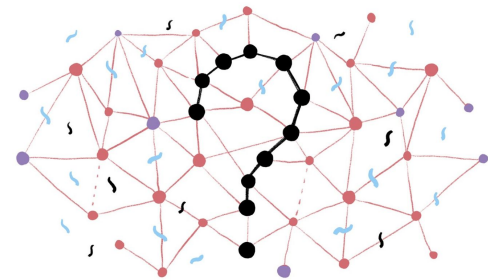
Algorithmic bias: what is it?

All learning algorithms work better, the more labeled data they have. However, it is inevitable that blind spots will be established, on the basis of the data they will be provided.

When a system displays the **inherent values** of the people (or society) that are involved in its setup, we are facing algorithmic bias.
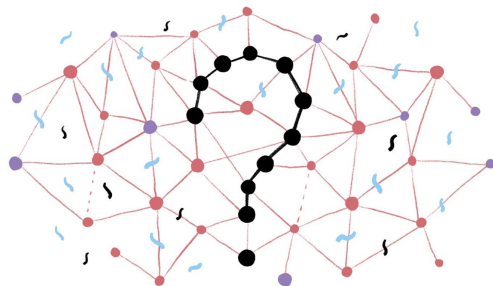
This issue has generated a lot of concerns among civil servants, academics, and industries, as results could be harmful, undesirable, and hard to comprehend (black box situation).

What we will do

We will examine notorious examples of algorithmic bias, explain why they happened, and try to provide solutions for the future.
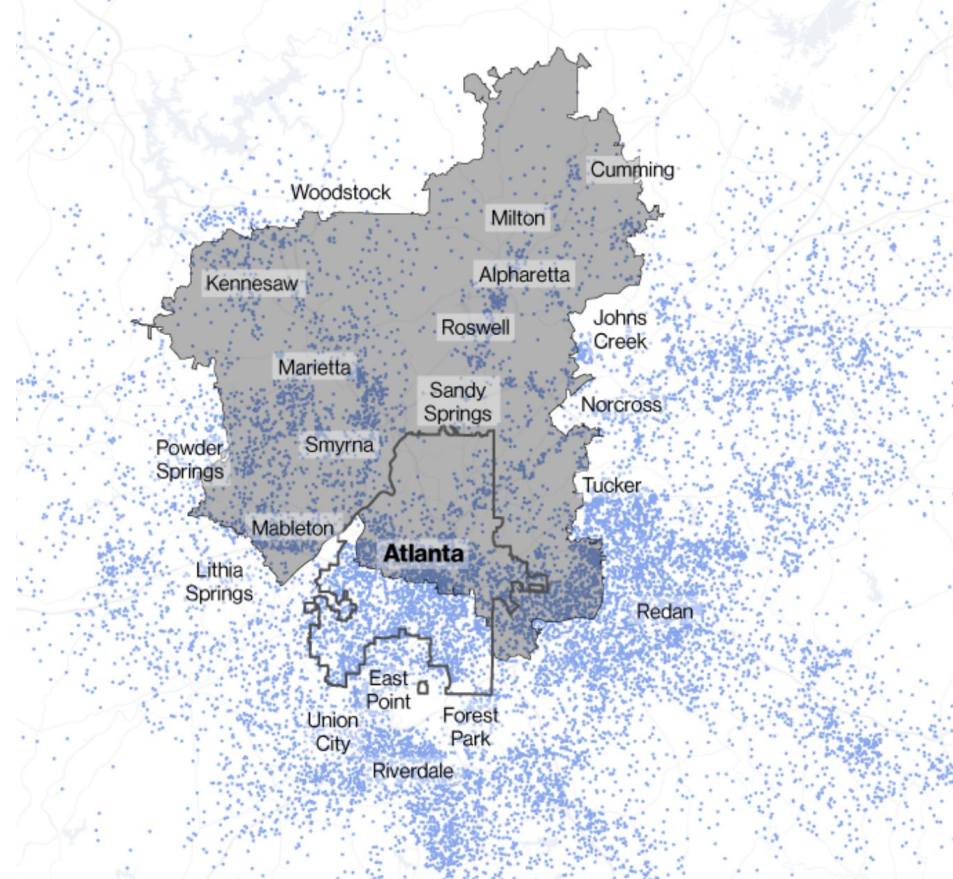
Examples of unintentional algorithmic bias

Amazon Prime incident:

Certain neighbourhoods excluded from same-day delivery.

Microsoft Tay:

While learning from Twitter users, within a few hours the Microsoft AI "Tay", started using hate speech.

Sentiment analysis:

ConceptNet researchers found and published examples of unintended bias, when models get trained on standard crowdsourced methods (news articles etc).

```
text_to_sentiment("Let's go get Italian food")
```
2.0429166109408983

```
text_to_sentiment("Let's go get Chinese food")
```
1.4094033658140972

```
text_to_sentiment("Let's go get Mexican food")
```
0.38801985560121732

```
text_to_sentiment("My name is Emily")
```
2.2286179364745311

```
text_to_sentiment("My name is Heather")
```
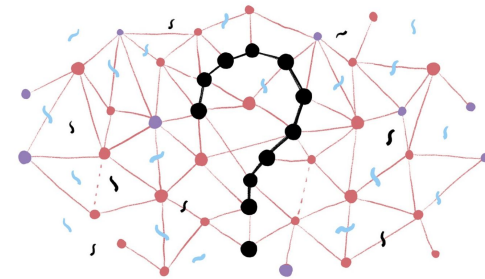1.3976291151079159

```
text_to_sentiment("My name is Yvette")
```
0.98463802132985556

```
text_to_sentiment("My name is Shaniqua")
```
−0.47048131775890656

And a lot more cases...

- LinkedIn and Amazon tried creating AI-assisted recruitment tools, revoked after being proven to be biased against women,
- Google Ads showed the highest paying jobs more to men than women,
- According to a 2012 study from Harvard University, ads for services that allowed people to search arrest records, would more often come up when researching a traditional African-American name,
- And a lot of other cases where societal discrimination and opinions manage to become "ground truth".

Solutions

Solutions need to be found on an international level, as those algorithms will become more and more prevailent the next few years

➜   **Better curation**

➜   **International cooperation**

➜   **Open datasets**

Thank you!

Sources:

- https://www.bloomberg.com/graphics/2016-amazon-same-day/

- https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

- http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/