

## SA2001 Individual assignment: Gender bias in algorithms

One of the cases on the website [Gendered Innovations](#) is about *gender bias in machine learning* and the challenge is explained as:

*"Machine learning algorithms can contain significant gender and ethnic bias. Where in the machine learning pipeline does bias reside: The input data? the algorithm itself? the types of deployment? More importantly, how can humans intervene in automated processes to enhance and, at least, not harm social equalities? And who should make these decisions?"*

*Importantly, AI is creating the future (technology, i.e., our devices, programs, and processes shape human attitudes, behaviors, and culture). In other words, AI may unintentionally perpetuate past bias into the future, even when governments, universities, and companies such as Google and Facebook have implemented policies to foster equality. So, the big question is: how can we humans best ensure that AI supports social justice?"*

The bias cases on [Gendered Innovations](#) are well described, with both *gender bias* and *ethnic bias*, as well as *gender intersecting with ethnicity*. Mapping and systemic solutions to the bias problems are briefly presented and discussed. However, the references may be a bit outdated in a field that moves as fast as machine learning (none of the 30+ references are published after 2018). Thus, we have for your assignment selected a *review article* published as late as 27 August 2022. This article is judged to give a better introduction and more thorough overview of the problem and state-of-the-art solutions and recommendations to algorithmic bias problems. The article's main sections are:

- *Introduction and The Algorithm Concept*
- *Data-Algorithmic Bias: Definitions and Classifications*
- *Examples of Gender Bias*
- *Datasets with Gender Bias*
- *Initiatives to Address Gender Bias*
- *An Illustrative Numerical Example*
- *Recommendations to Prevent, Identify, and Mitigate Gender Bias and Conclusions.*

### Study material for this individual assignment

As a quick introduction to the topic, it is recommended to first watch the 8-minute TED Talk:

[How I'm fighting bias in algorithms](#) by Joy Buolamwini of the Algorithmic Justice League.

before reading the 16-page review article:

[Castaneda, et al., \*Dealing with Gender Bias Issues in Data-Algorithmic Processes: A Social-Statistical Perspective\*, Algorithms 2022.](#)

## Individual reflection questions

The three questions in this individual assignment are reflections on different sections of the review article by (Castaneda, et al., 2022). You should read the whole article but read certain sections more carefully.

### **Reflection question 1**

From your own prior knowledge, skills, and approach, describe your main take-aways from section **2.2. Algorithm Concept in Social Sciences**

My take-away is that Algorithms and the Mathematics behind them have an enormous impact on society, since they are implemented in multiple sectors and contexts within society. Because algorithms are created and implemented by many different sectors it is hard to narrow the study of their social impacts into only a few societal contexts. Therefore, it is important to obtain feedback as holistic as possible which can be done by letting users all over the world and in different contexts get the chance to provide it. Mathematicians, Engineers and Computer Scientists creating the algorithms might not have sufficient insight to the area of usage of the algorithm, which is why collaboration between them and Social Scientists and others implementing the algorithm is a must, due to the consequences the algorithm might have. Although, the algorithm itself might be biased, it seems like the training sets for it might have the largest impact on whether the implementation is biased or to a lesser extent.

### **Reflection question 2**

Describe your own experience with all or any of the four types of gender bias described in section **4. Examples of Gender Bias** and also indicate in the text box below which of the four types you would like to have a case for the group work with the [Intersectional Design Cards](#) at the next two seminars. However, you are not bound to selected from those four cases, but please indicate which case you would like to work with in the upcoming seminars.

For instance, when I and a female friend went to the grocery store we could see in the entrance camera video there was a square around my (male) face and not around hers. Although this was some time ago, it is obviously still the case. At that time I do not recall we reflected too much on the reasons. I have also heard about, as indicated in the provided video, that suspected criminals can be picked out in a higher when the training sets are not sufficiently complete. I guess I would like decision management to be the group's case, since I reckon that might be one of the areas where gender bias in algorithms has the biggest societal impact.

### **Reflection question 3**

Reflect upon another section of the article which you select yourself. What would you like to highlight from your selected article section?

Regarding section 5, and the data sets. Even though there seems to be imaginative ideas to come around the problem of algorithms biases, it is obvious that the data sets must be more objective. Then the data collection procedures probably need to be more restricted and unbiased. But as the article indicates it seems as a hard task since there are a kind of catch 22 or a loop because people are the ground of the data sets. I guess there is a need for politicians and lawyers, too. It seems like the best way to avoid bias is to collect as diverse data as possible.