Causes & impacts of data bias

Worksheet section	Contents
1	Types of data bias
2	Causes & mitigation of bias

Version: 1.0

If you require this document in an alternative format, such as large print or a coloured background, please contact

hello@effini.com

or

4th Floor, The Bayes Centre
47 Potterrow
Edinburgh
EH8 9BT

1. Types of data bias

1) Which of these definitions of data ethics is correct?

   a) "When data is used to prejudice against a person"

   b) "When data is used to prejudice for or against one person or group"

   c) "When data is used to prejudice for a group of people"

2) Many websites such as Amazon ask customers to provide reviews of the products. Not all customers fill in reviews for the products they buy. How could this cause bias in the reviews of the products?

kkay	taimafan
★★★★★ **The best pencil**	★☆☆☆☆ **They are school pencils, not what they described/pictured as sketching pencils**
Reviewed in the United Kingdom on 26 July 2022	Reviewed in the United Kingdom on 11 June 2022
Style Name: Noris 2B | Size Name: 12 Count (Pack of 1) | **Verified Purchase**	Style Name: Noris 2B | Size Name: 12 Count (Pack of 1) | **Verified Purchase**
I had to order the 2b from amazon couldn't find it anywhere else, soft very esy to draw and write with and does not scratch the paper, the durability is ok but for softer tip is shorter 5han other overall great product	Don't buy if you want sketching pencils, because they are not. You will get school pencils. I have returned mine.
Helpful | Report abuse	Helpful | Report abuse

3) Search engines such as Google can show examples of algorithmic bias.

> **Reminder:** Algorithmic bias is a computer system that makes decisions that are unfair to certain individuals or groups

Use a search engine to find images for these phrases. Describe what the images look like and whether they appear to be biased in any way.

Teacher style

Plumber person

Nurse person

Scottish person

1. Types of data bias

If you would like any further reading on how search engines can show algorithmic bias, please read this article from July 2022, "Gender bias in search algorithms has effect on users"

https://www.nyu.edu/about/news-publications/news/2022/july/gender-bias-in-search-algorithms-has-effect-on-users--new-study-.html

Now you are going to play a game called **Survival of the best fit*** that will show you more about algorithmic bias. Click on the link below to start the game.

https://www.survivalofthebestfit.com/game/

4) Were you able to create an unbiased automated solution for hiring new employees?

SURVIVAL OF THE BEST FIT

5) If a company wanted to create an automated recruitment process like the one in the game, how could they try to reduce the impact of bias on the system,

a) When the data is collected?

b) When the data is being analysed/system is being created?

c) When the system is being used to hire new employees?

2. Causes and mitigation of data bias

1) Fill in the missing words in these causes of data bias.

Sample bias   The dataset used for data analysis does not represent the [_____]

Exclusion bias   Predictive variables are [_____] prior to the data analysis.

Measurement bias   Arises from a systematic issue with the [_____] of the data.

Confirmation bias   Arises from [_____] data that confirms a human's existing beliefs.

Stereotype bias   Cultural or [_____] stereotypes may led to an uneven distribution of data.

Survivorship bias   Arises from concentrating on the successful output of the process and [_____] those who don't survive.

2) Explain why this data bias problem has been caused by **confirmation bias.**

> A business is looking to launch a new product. They conduct a survey of potential customers and most of the people don't like the new product. However, when the new product is shown to a friend of the product developer their friend really likes it. The product developer places more importance on the views of their friend than the survey of potential customers.

[_____]

3) Explain why this data bias problem has been caused by **survivorship bias.**

> A TV company is trying to understand what makes a successful TV show. They look at variables related to successful shows such as length of the show, type of show, what time it is broadcast etc to try and predict whether future shows will be successful.

[_____]

4) Explain what are the best ways to mitigate against data bias.

[_____]

2. Causes and mitigation of data bias

5) Correlation bias is when there appears to be a relationship between two data items but they are not actually linked.

The website below contains examples of data items that show the same patterns but are not related.
https://www.tylervigen.com/spurious-correlations

Review the website and copy and paste an interesting example below.