The Alan Turing Institute

Turing Commons

ACTIVITY 4

Bias ~~Checklist~~ *Reflect*list

Responsible Research and Innovation in Data Science and AI

Summary

This activity is designed to help you (and your group) identify and understand how different biases can impact the design, development or deployment of a data-driven technology.

Using the project lifecycle model as a scaffold, you will consider how a variety of biases—categorised as instances of social, statistical, and cognitive biases—could enter into the project lifecycle, how they might impact downstream decisions, and which tailored actions could be undertaken to mitigate the impact of the bias.

Learning Objectives

- Become more familiar with the three types of bias (social, statistical, cognitive) and how they differentially impact the design, development and deployment of data-driven technologies.
- Learn to identify different types of mitigation techniques that are tailored to the different types of bias or the project lifecycle stage in which the bias is identified.
- Use the bias reflectlist to further embed your practical understanding of fainess (and related principles such as accountability or explainability).

Instructions

Pre-requisites

To carry out this activity, you will need the following:

- A copy of the project lifecycle model image
- A copy of the bias and mitigation technique cards
- A copy of the bias reflectlist worksheet
- (Optional) A copy of the bias reflectlist cheatsheet
- (Optional) A copy of the project lifecycle cheatsheet

Introduction

Bias impacts the design, development, and deployment of data-driven technologies in a variety of ways, which are often context-dependent. This presents a problem for a structured and generalisable approach to bias identification and mitigation activities.

Fortunately, we can use the project lifecycle model and the bias taxonomy to provide a structured approach to facilitate the process of bias mitigation. This doesn't provide us with a *checklist* of actions that we can follow to ensure that our project is free from bias, but it does provide us with a *reflectlist* that offers significant support for the process of bias mitigation.

Steps

The following steps apply regardless of whether this activity is carried out as a group or as an individual:

1. Review the list of biases (using either the cards or the cheatsheet) and select one bias from each category (social, statistical, and cognitive).

2. Identify one stage from the project lifecycle model where you think the bias poses a *significiant* risk to the project's goals or objectives.

3. Place the chosen bias card on the relevant stage and repeat the process for the other biases.

   → If the bias has no significant impact on any of the stages, or the risk is deemed very low, then discard the bias and select another one from the same category.

4 Once you have placed all of the bias cards, review the list of mitigation techniques, and select an appropriate mitigation technique for each of the biases. Place the mitigation technique card next to the bias card on the project lifecycle.

- If there is no appropriate mitigation technique, then use one of the blank cards to write down a mitigation technique that you think would be appropriate.

5 Record your first set of decisions on the bias reflectlist worksheet and repeat steps 1-4 until you run out of time or you have considered all of the biases.

Example

Here is a partial example of how some of the above steps could proceed:

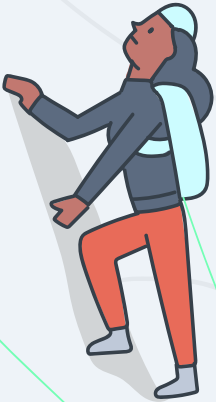
Group A select the following three biases:

- Historical Bias (Social)
- Missing Data Bias (Statistical)
- Decision-Automation Bias (Cognitive)

They place Missing Data Bias on the Data Analysis stage. Their justification for this placement is that although missing data may be caused by a variety of factors (e.g. barriers to participation in data collection activities), the impact is most likely to be identified and understood fully during exploratory data analysis.

The team recognise that a variety of techniques could apply to address the missingness in the dataset, and which technique will be best is conditional on the extent of the missing data (e.g. imputation, additional collection or augmentation of dataset). As such, they decide to attach the 'Stakeholder Engagement' mitigation technique to this bias to support the lower-level decision-making process.

They record this decision on the worksheet and repeat the process for the other two biases.



Goals and Objectives

The primary goal for this activity is to identify the practical impact of different types of bias on the design, development, and deployment of data-driven technologies, and to gain familiarity with the types of mitigation techniques that are available to address the impact of the bias.

By the end of the activity you should have the following:

1. A completed (or partially completed) bias reflectlist worksheet, which could be used to document the process of bias mitigation if the activity were carried out for real during a project.

Tips and Guidance

You will probably find that many biases could impact multiple stages of the project lifecycle. Therefore, you will need to decide which of the stages is most *significant*. By *"significant"* we mean the following:

1

The bias is likely to have a notable impact on the activities of the respective stage.

2

The chosen stage is likely to be the most appropriate time to mitigate the bias (e.g. most upstream, appropriate team skills and resources).

If you are struggling to identify a mitigation technique for a particular bias, then you can use the deliberative prompts from the cheatsheet to help. Each prompt is a question designed to support a process of reflection and deliberation.

As you progress through the activity, try to identify whether specific mitigation techniques are more common for certain types of bias or for certain stages of the project lifecycle.

Assets

- Project Lifecycle Model Graphic
- Cards (Biases and Mitigation Techniques)
- Bias Reflectlist Worksheet
- (Optional) Bias Reflectlist Cheatsheet
- (Optional) Project Lifecycle Cheatsheet

In-Person

If you are conducting this activity in-person, you will need to have a printed copy of all of the assets listed in the pre-requisites section (see above). The cards should be cut out to allow them to be placed upon the project lifecycle model.

You will also need a printed or digital copy of the worksheet to record the decisions made during the activity.

Online

If carrying out this activity online as a group you may wish to use an online whiteboard or collaborative note-taking tool such as Miro, Mural, HackMD. or Google Docs.

Our assets can be imported into some of these tools directly, or adapted to others.