Fairness & Bias: Definitions

This week, we will dive into the subfield of Responsible AI, which aims to ensure AI is safe, trustworthy, and unbiased. It is a relatively new subfield, which has seen a stark increase in popularity over the last couple of years. Unfortunately, many responsible AI tools currently do not support image data. So, this week we may need to push ourselves to think outside the box.

Let's start with our trailblazer journey:

Today, we will focus on the concepts of 'fairness' and 'bias'. How are they defined? Are there different types of 'fairness' and/or 'bias'? Lastly, in the workshop, we will familiarize ourselves with two principles to achieve fairness; equality and equity.

Learning objectives

1. Define the terms 'individual fairness', 'group fairness', 'bias', 'equity', and 'equality'.
2. Describe different types of bias, and connect them to the CRISP-DM cycle phases by providing concrete examples.
3. Explain the difference between the terms 'individual fairness' and 'group fairness'.
4. Explain the difference between the terms 'equity' and 'equality'.

Table of contents

1. Introduction: 3 hours
2. Workshop: 2 hours
3. Preparation DataLab: 2 hours

Questions or issues?

If you have any questions or issues regarding the course material, please first ask your peers or ask us in the Q&A in Datalab!

> Tip: Note down any important questions you might have!

Good luck!

1) Introduction

Artificial Intelligence, often in the form of machine learning, underpins decisions that profoundly affect our society. For example, it may decide if you get car insurance or if you are eligible for parole.

Machine learning uses data by learning through the generalization of examples, defining rules that apply for past cases, and predicting future unseen cases (See Figure 1).
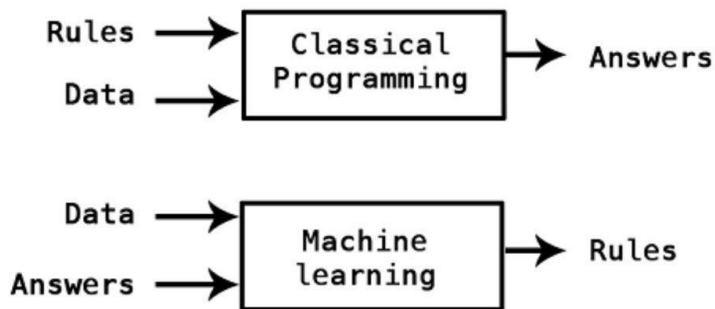

Figure 1: Traditional vs. machine learning.

Machine learning can come in handy when making data-driven decisions, uncovering relevant factors that humans might overlook. However, it does not ensure fair decisions 😕.

Before we can start with our quest to make AI more 'fair', we need to define the concept a bit further:

1a Watch the video Why monkeys (and humans) are wired for fairness by Sarah Brosnan.

Video 1. Why monkeys (and humans) are wired for fairness.

1b Explain how the terms 'fairness' and 'cooperation' relate to each other by giving a real-life example. Write your answer down.

1c Read the comic book Fairness & Friends by Khan et. al., which you can find, here. This comic book is completely awesome, because it will introduce you to all the main concepts related to fairness and AI in a fun,

and comprehensible way.

> Note: There is no real consensus on defining bias in machine learning; some scholars claim there are three distinct types, while others claim there are six specific types of bias.

1d List, and describe at least five types of bias presented in either Fairness & Friends or A survey on bias and fairness in machine learning.
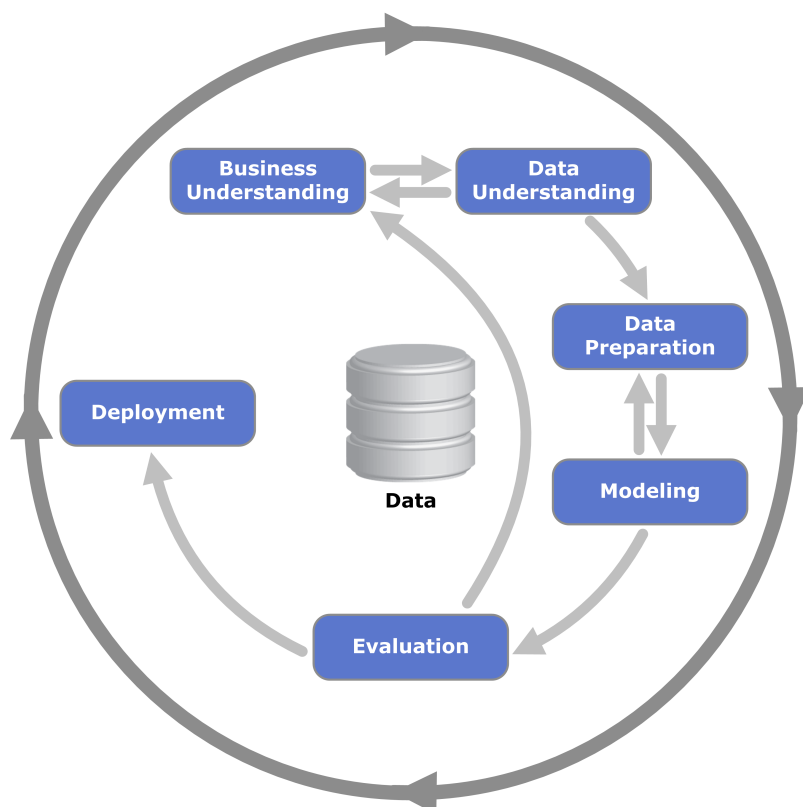

Figure 2. CRISP-DM cycle.

1e Give an example for each type of bias listed in exercise 1d, and connect it to one of the CRISP-DM phases. Write your answer down, and limit your word count to a maximum of 150 words.

1f Explain the difference between 'individual fairness', and 'group fairness'. Write your answer down.

2) Workshop: The difference between equality and equity

Now, we are introduced to the terms 'fairness' and 'bias', it is time to deepen our knowledge. In the workshop, you will learn about how two principles, equality, and equity, can help to achieve fairness:

> Though often used interchangeably, equality and equity are quite different. If fairness is the goal, equality and equity are two processes through which we can achieve it. Equality simply means everyone is treated the same exact way, regardless of need or any other individual difference. Equity, on the other hand, means everyone is provided with what they need to succeed. In an equality model, a coach gives all of his players the exact same shoes. In an equity model, the coach gives all of his players shoes that are their size (Source).

The current fairness metrics, which will be the topic of the coming independent study days, are closely connected to one of these principles, namely equality.

2.1 Individual exercises

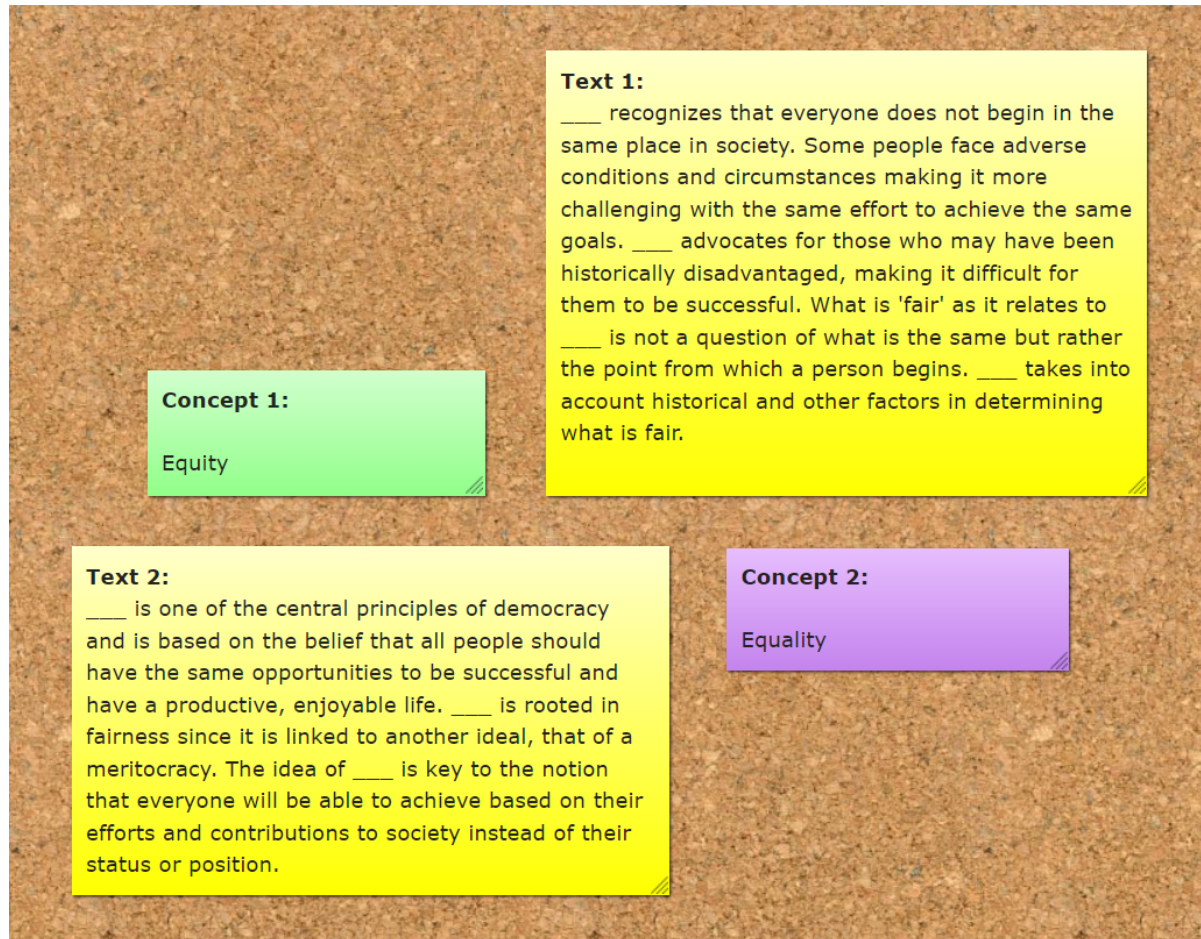**2.1a** Read the Post-Its in Figure 3.

Figure 3. Post-Its on a bulletin board.

2.1b Fill the gaps in the 'text' Post-Its by connecting it to the correct 'concept' Post-It. Write your answer down.

2.1c Provide clear, and concise answers to the following questions:
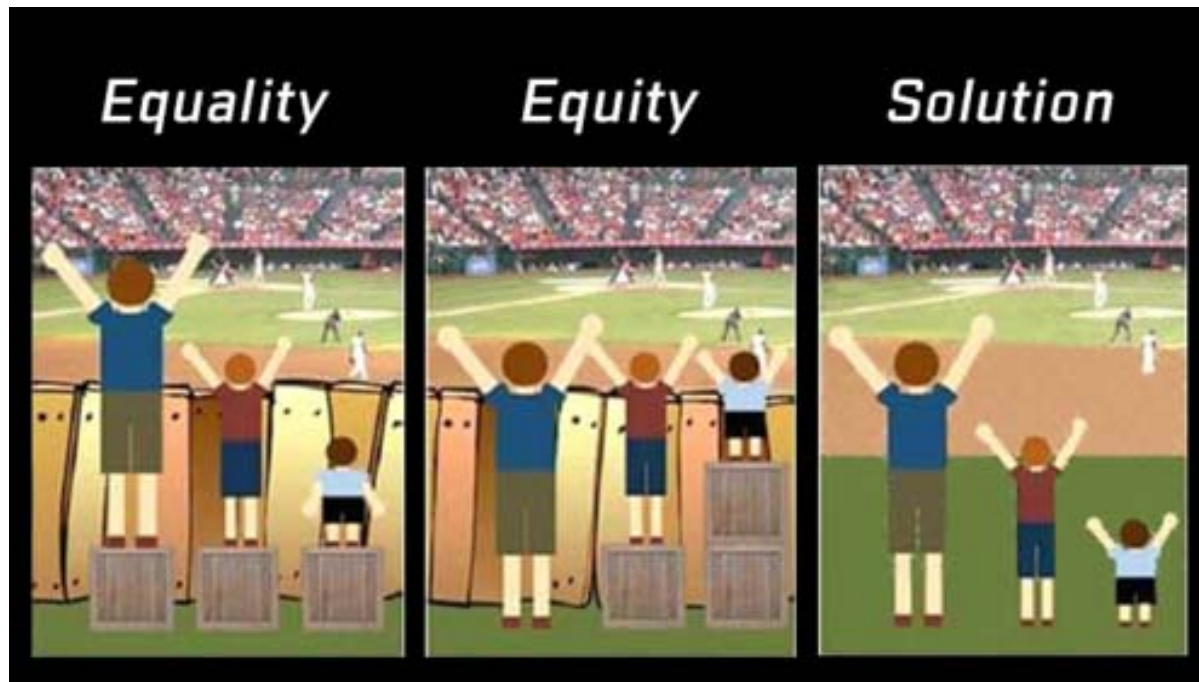
Figure 4. Equality vs. equity.

1. Looking at the images in Figure 4, why does the difference between equity and equality matter?
2. How does the baseball game metaphor relate to the guiding quote about shoes?
3. What are some other real-world examples of equity and equality?
4. How are equality and equity related to fairness? In school, sports, society, etc.?
5. Are there other ways to achieve fairness besides equity and equality? (For example, the removal of the wooden fence eliminates the structural barrier to access.)

2.2 Group exercises

2.2a Team up with at least one fellow student, and discuss your answers to the individual exercises. Write down the name(s) of your teammate(s), and the differences/commonalities in your provided answers.

3) Preparation for DataLab 1

- ☐ Watch the video 'What's your worldview?' by Impact 360 institute:

Video 2. What's your worldview?

- ☐ Read the article Eurocentrism by Keena Hays to find out how an Eurocentric worldview reinforces western beliefs of superiority.

- ☐ Become familiar with the Open Images dataset. Answer the following questions:

1. What is the Open Images dataset?
2. Who created it?
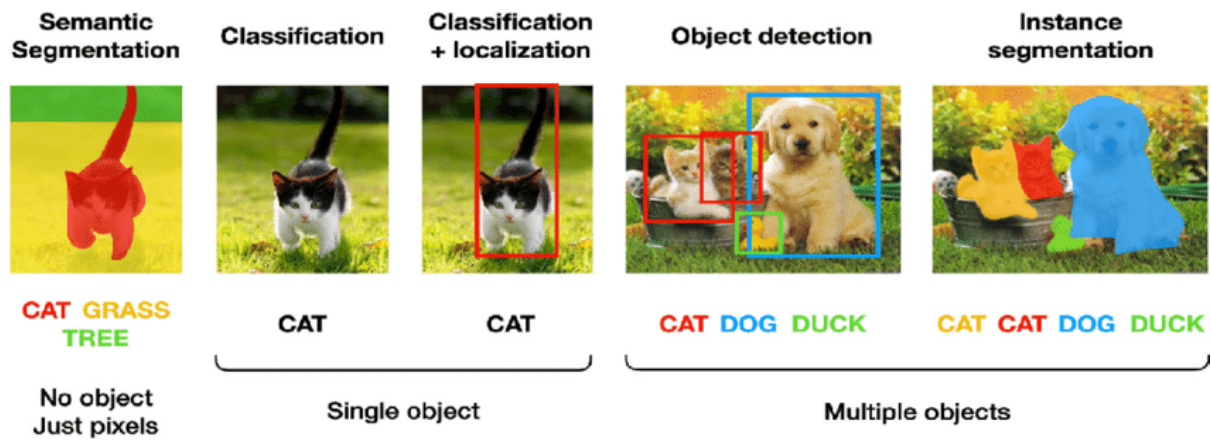3. For what kind of computer vision tasks can it be used?

Figure 5. Types of computer vision tasks.

Write your answers down, and upload them to GitHub.

- ☐ Become familiar with A Designer's Critical Alphabet.

Resources

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35.

Khan, F. A., Manis, E., & Stoyanovich, J. (2021, March). Fairness and Friends. In Beyond static papers: Rethinking how we share scientific understanding in ML-ICLR 2021 workshop.