Text, logo

Description automatically generatedLogo

Description automatically generated with medium confidence

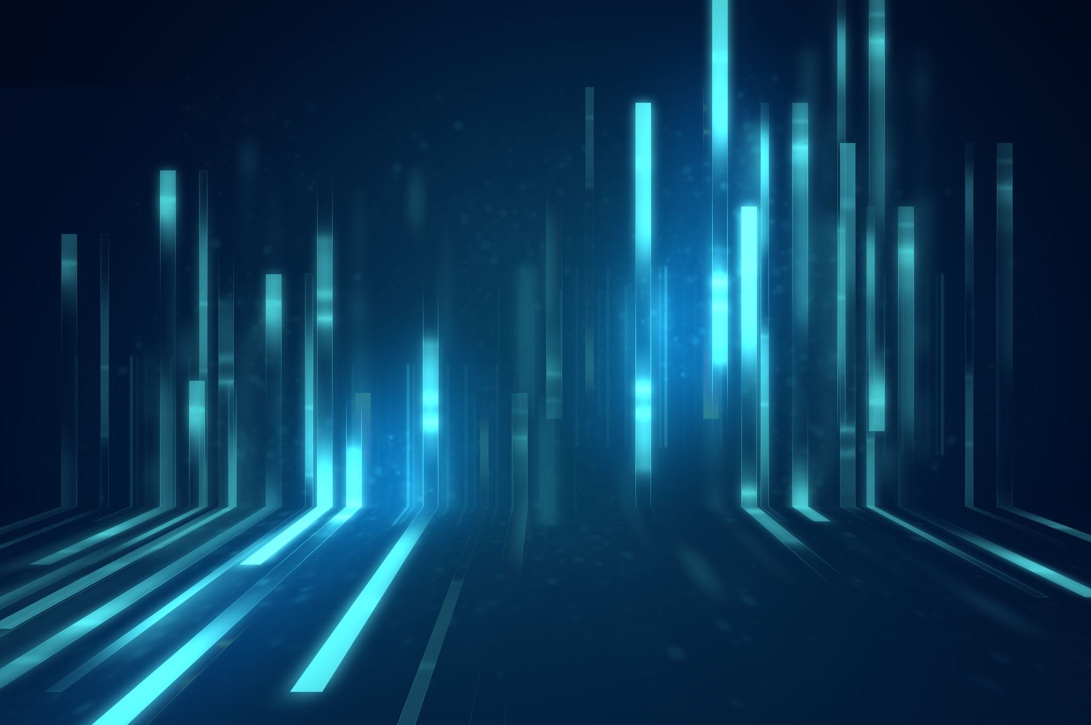
­­­

Fairness

Expert

Module 1

**Philosophical Concepts of Fairness**



# Introduction

This module will expand on the intermediate course on fairness. It will further link philosophical concepts of fairness to technical models that can be captured algorithmically.

­­­

# Pre-reading

* Justice and Fairness

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/justice-and-fairness>

* It's Not Fair! But What Is Fairness?

<https://www.psychologytoday.com/intl/blog/am-i-right/201205/its-not-fair-what-is-fairness>

* The Fairness Fixation

<https://www.philosophytalk.org/blog/fairness-fixation>

* Concepts of Justice and Fairness

<https://journal.viterbo.edu/index.php/at/article/view/342/136>

**Reading**

* COMPAS example

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

­­­



**Self-assessment pass/fail questions**

1) Substantive fairness means...

a. The rules are fair

b. The outcome is fair

c. Fairness doesn’t matter

2) Iris Marion Young’s theory focused on the importance of understanding the nature of...

a. Fairness

b. Transparency

c. Oppression

3) Can different conceptualizations of fairness be included in a single algorithmic model?

a. No

b. Yes

c. Maybe

4) How do we achieve algorithmic fairness?

a. Embed one of the definitions of fairness into the tool

b. Don’t worry about it because it cannot be achieved

c. Decide on which conceptualisation of fairness best fits the use case and continually revisit the model and its use to see if it is fair.

**Answers**

Qs 1) b, 2) c, 3) a, 4) c Qs

# Tasks

**Task 1**

Make notes to give a five-minute talk explaining to police why an AI tool you have developed for them is fair. Describe what the tool is for and how you have attended to and implemented fairness considerations into its design and use.

**Task 2**

Re-read the COMPAS case study and answer these questions:

How could this have been avoided?

Using the principles learned in this topic, evaluate why, or why not, the COMPAS system is unfair.

# Train the trainer

For each theory on fairness, compile its advantages and disadvantages into a table. Use this table to teach the content. Ask the students to engage in a discussion of whether perfect fairness in a society is ever achievable?

Trainers should keep up to date with developments in the EU AI Act regarding the notion of fairness. What is the definition of fairness? How is it recommended to be implemented in AI? Who is accountable for a lack of fairness according to this regulation?

You can also inform yourself also about the psychology of fairness:

<https://www.psychologytoday.com/us/blog/the-mindful-self-express/201408/the-neuroscience-fairness-and-injustice>