

AI4ER CDT

Co-Designing Case Studies



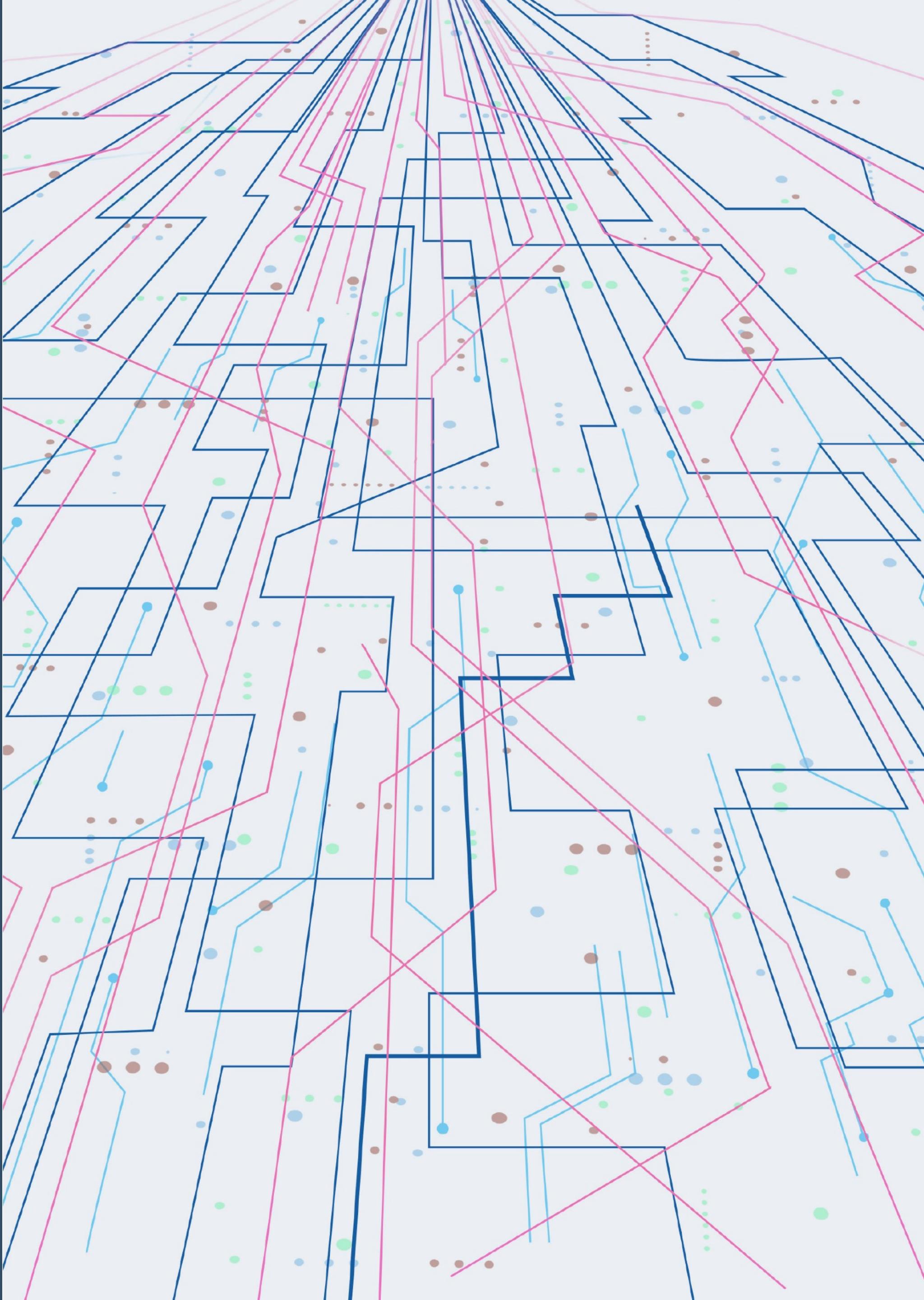
Dr Christopher Burr
Claudia Fischer
December 2022

- # Presentation Overview
- 1 Case Studies
 - 2 Illustrations
 - 3 Activities



Section 1

Co-Designing Case Studies

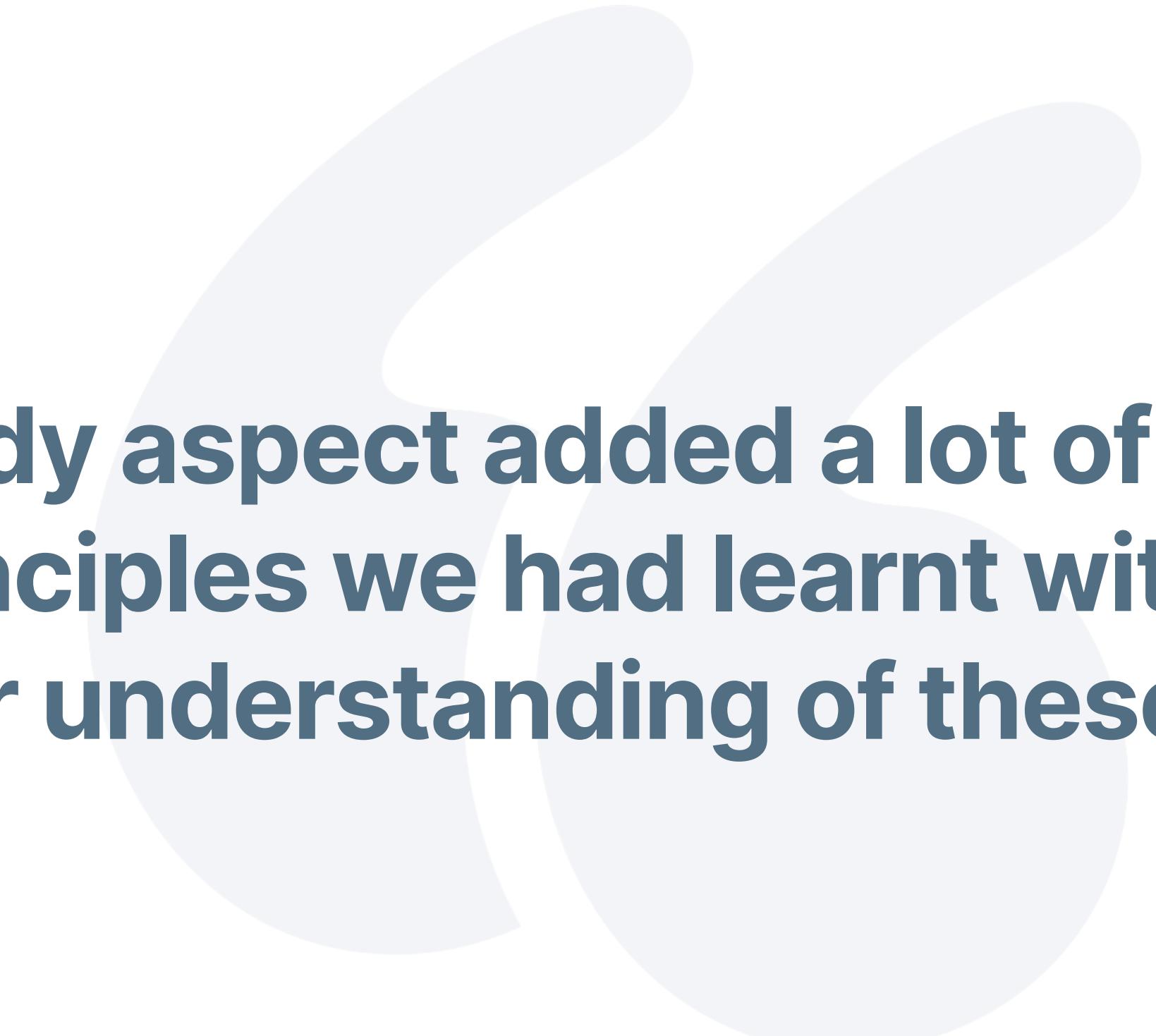


The screenshot shows the homepage of the Turing Commons website. The header features a dark teal bar with the "Turing Commons" logo, a search bar, and a GitHub link. Below the header, a navigation menu includes "Home", "Welcome", "Responsible Research and Innovation", "Public Engagement of Data Science and AI", and "Blog". The main content area has a white background with a large, central, circular illustration depicting a complex network of people, buildings, vehicles, and data storage units. To the left of the illustration, the text "Welcome to the Turing Commons" is displayed, followed by a description: "An online platform to support open dialogue and reflection about the responsible design, development, and deployment of data-driven technologies." A green "Get Started" button is located below this text. At the bottom of the page, a black footer bar contains the copyright notice "Copyright © 2022 Alan Turing Institute" and "Made with Material for MkDocs Insiders".

Online Platform

Turing Commons

An online platform to support open discussion and reflection about the responsible design, development, and deployment of data-driven technologies.



Feedback

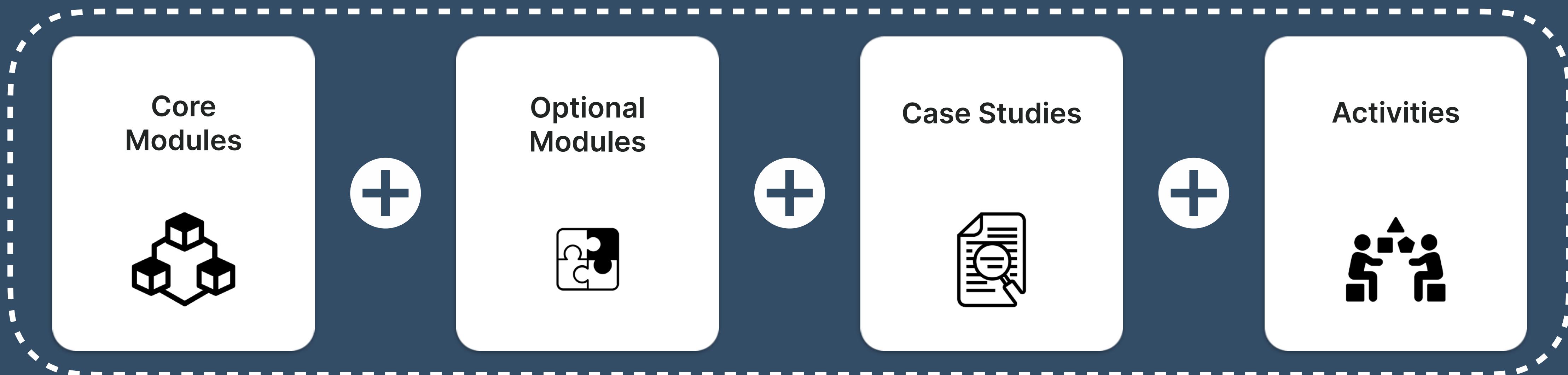
“I feel the case study aspect added a lot of value—it allowed us to apply the principles we had learnt within the course and strengthen our understanding of these principles.”

Challenge



How do we adapt our courses to make them more accessible to specific groups and communities?

Skills Tracks



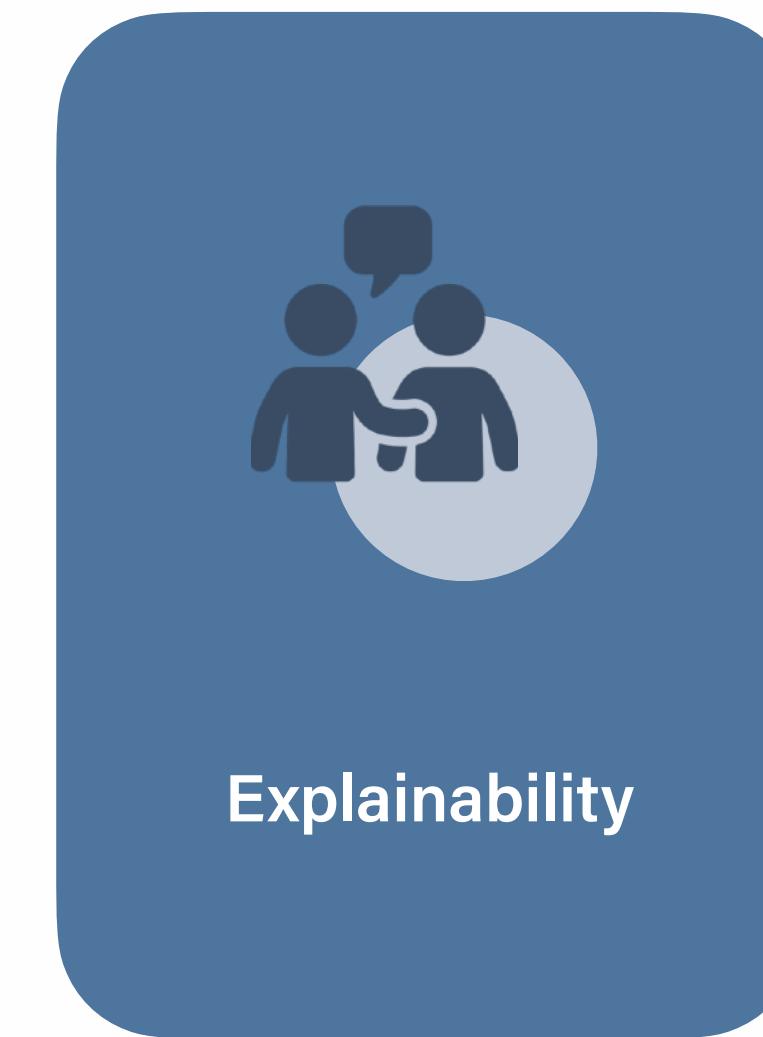
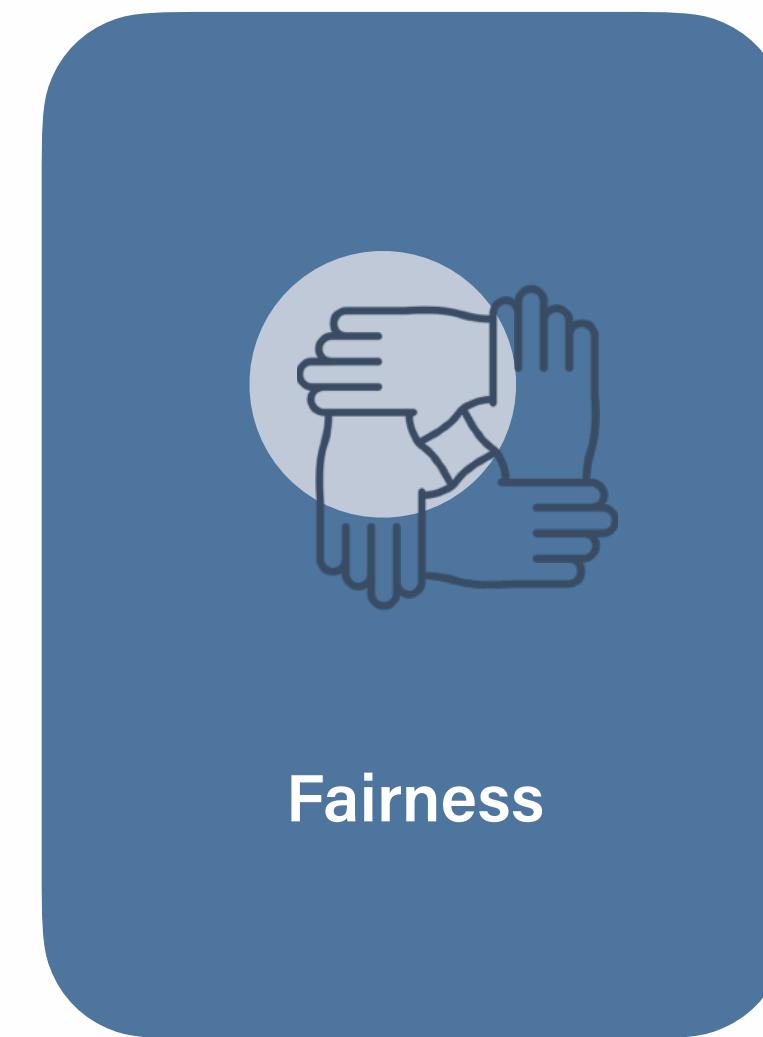
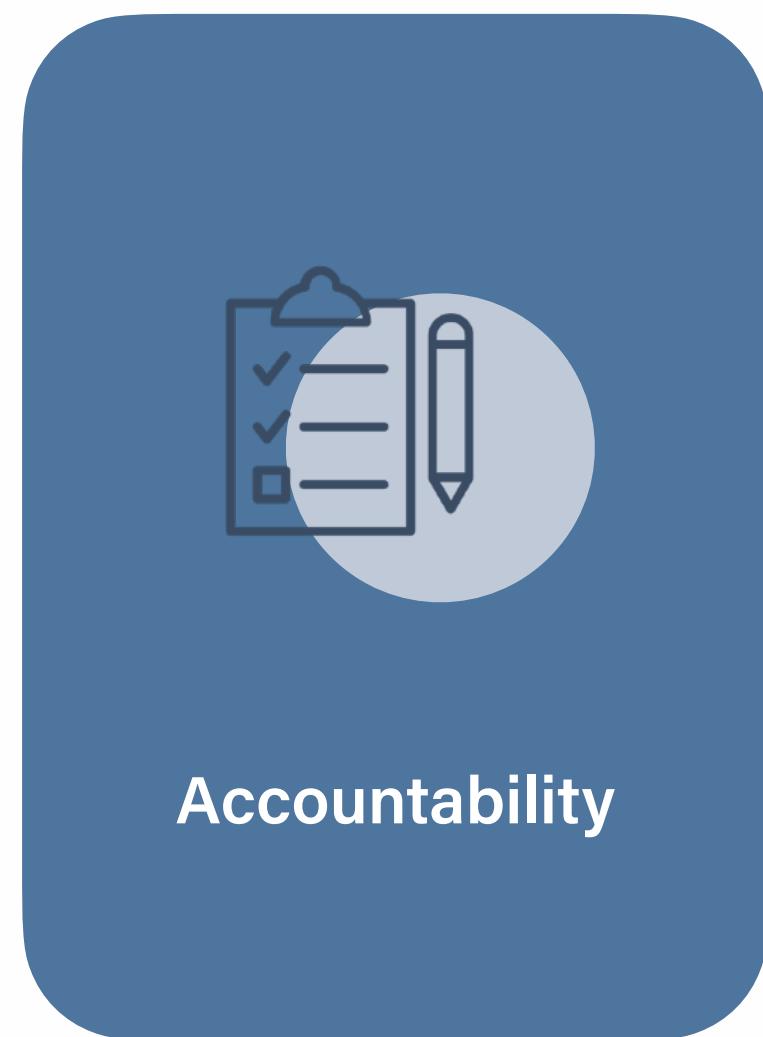
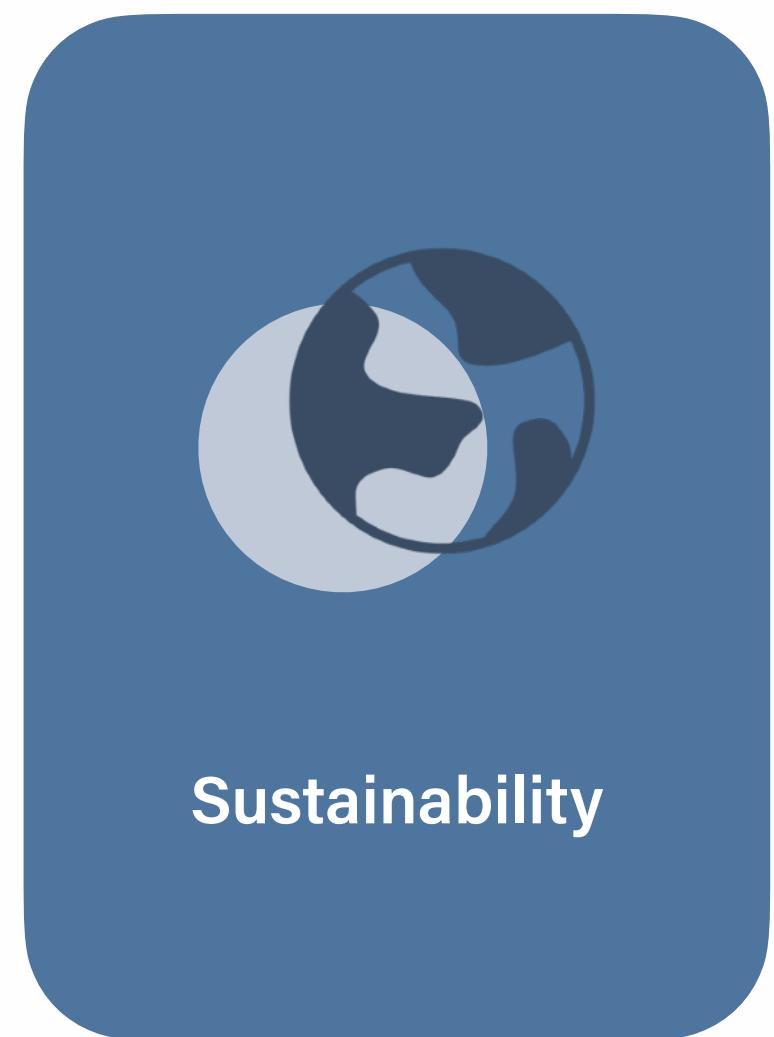


Modules

Responsible Research and Innovation

1. What is Responsible Research and Innovation?
2. The Project Lifecycle Model
3. (*SAFE-D Modules*)
4. Responsible Communication and Open Science

SAFE-D Principles



Supporting Reflection and Deliberation

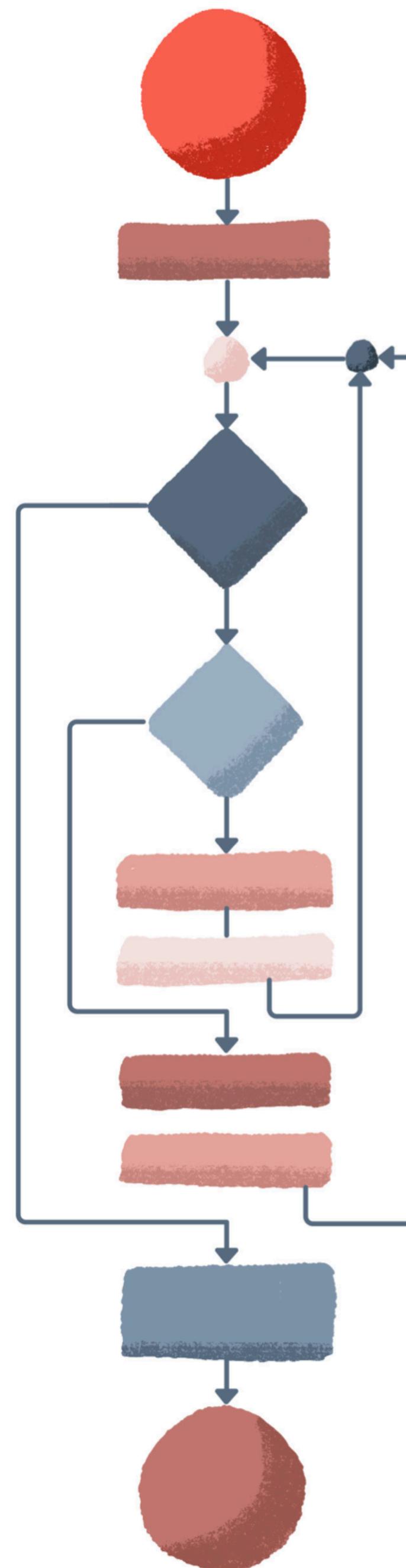
Hypothetical Case Studies



Case Study

Structure

- Title
- Authors
- Hero Image
- Summary Sentence
- Project Description
- Technology Description
- Key Issues
- Deliberative Prompts
- Stakeholders and Affected People
- Datasheet
 - Available Data
 - Algorithmic Techniques



```
{
  "title": "Decision Support System",
  "domain": "healthcare",
  "userGroup": 2,
  "personas": [
    "Psychiatrist",
    "Patient",
    "Developer"
  ],
  "isLive": true
}
```

Decision Support System

The Alan Turing Institute

Using natural language processing to help psychiatrists assess patients

The system uses a modern form of natural language processing that is reliant on neural networks to identify salient features of a patient's speech. This includes the words and phrases the patient states as well as extra-linguistic properties, such as intonation or pace.

The system operates in real-time to make recommendations to the psychiatrist. The recommendations can include suggestions about specific topics that were raised by the patient, which may require follow-up (e.g. reference to problematic relationships), as well as generic features that may be informative. Many recommendations require additional interpretation from the psychiatrist, and no automated decisions are made by the system.

The system's outputs can be shared with the patient where explanations are required. As such, the system has been designed to support patient-psychiatrist communication and participatory decision-making. For instance, phrases that were flagged as relevant are emphasised and specific recommendations are also accompanied by confidence ratings that can be explained by a trained psychiatrist.

© The Alan Turing Institute

Overview

Psychiatrists working for a national healthcare system have been provided access to an AI system that can support decisions made during initial patient assessment and diagnosis.

The system uses a modern form of natural language processing that is reliant on neural networks to identify salient features of a patient's speech. This includes the words and phrases the patient states as well as extra-linguistic properties, such as intonation or pace.

The system operates in real-time to make recommendations to the psychiatrist. The recommendations can include suggestions about specific topics that were raised by the patient, which may require follow-up (e.g. reference to problematic relationships), as well as generic features that may be informative. Many recommendations require additional interpretation from the psychiatrist, and no automated decisions are made by the system.

The system's outputs can be shared with the patient where explanations are required. As such, the system has been designed to support patient-psychiatrist communication and participatory decision-making. For instance, phrases that were flagged as relevant are emphasised and specific recommendations are also accompanied by confidence ratings that can be explained by a trained psychiatrist.

© The Alan Turing Institute

Key Consideration

The system has been designed for use as a *decision support system*. Are there properties of the system that could negatively impact the ability for the psychiatrist to perform their clinical duties effectively?

Deliberative prompts

- If the system was designed from scratch, how should psychiatrists and patients be involved in its design, development, and deployment?
- Why does it matter that the system functions as a *decision support tool* and not an automated decision-making system?
- Should patients have full access to the outputs of the *decision support system*?

© The Alan Turing Institute

Datasheet

Category	Details
Available Data	<ul style="list-style-type: none"> Automated transcription of the conversations between patient and psychiatrist Extracted extra-linguistic features from audio recording of the conversation Relevance feedback from the psychiatrist about salience of specific recommendations Electronic health record of patient, including any initial assessment data (e.g. PHQ-9) or prescriptions.
Analysis Techniques	<ul style="list-style-type: none"> Natural language processing (NLP): <ul style="list-style-type: none"> System uses convolutional neural networks for speech recognition and to identify features from audio recording, which are then classified according to whether they are informative (i.e. exceeding some relevance threshold) Visualisation techniques support explainability by highlighting salient features

© The Alan Turing Institute

Groups, Organisations and Affected Individuals

- Patients
- Psychiatrists

© The Alan Turing Institute

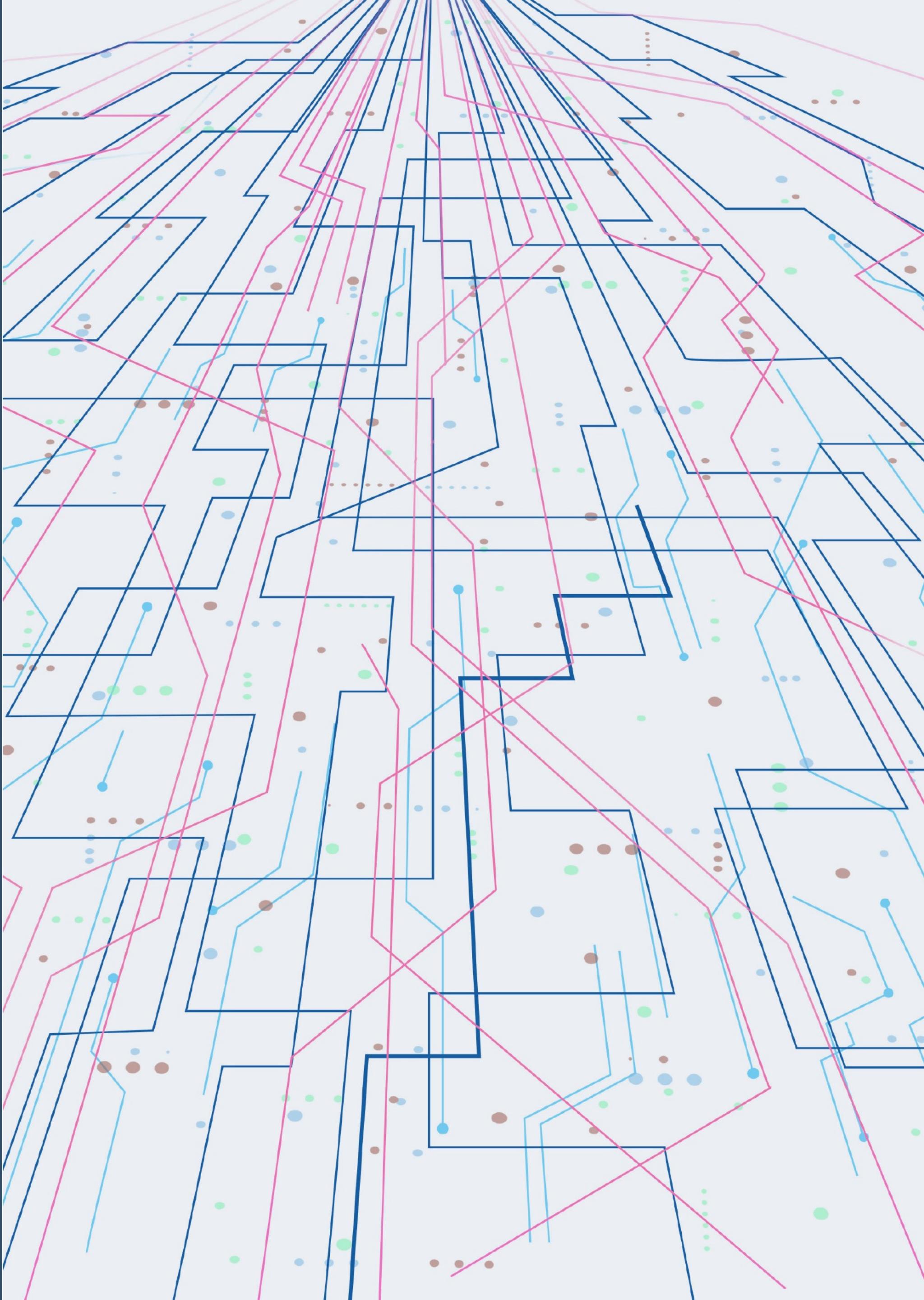
Task

Draft the *minimal* information for two case studies



Section 2

Illustrations



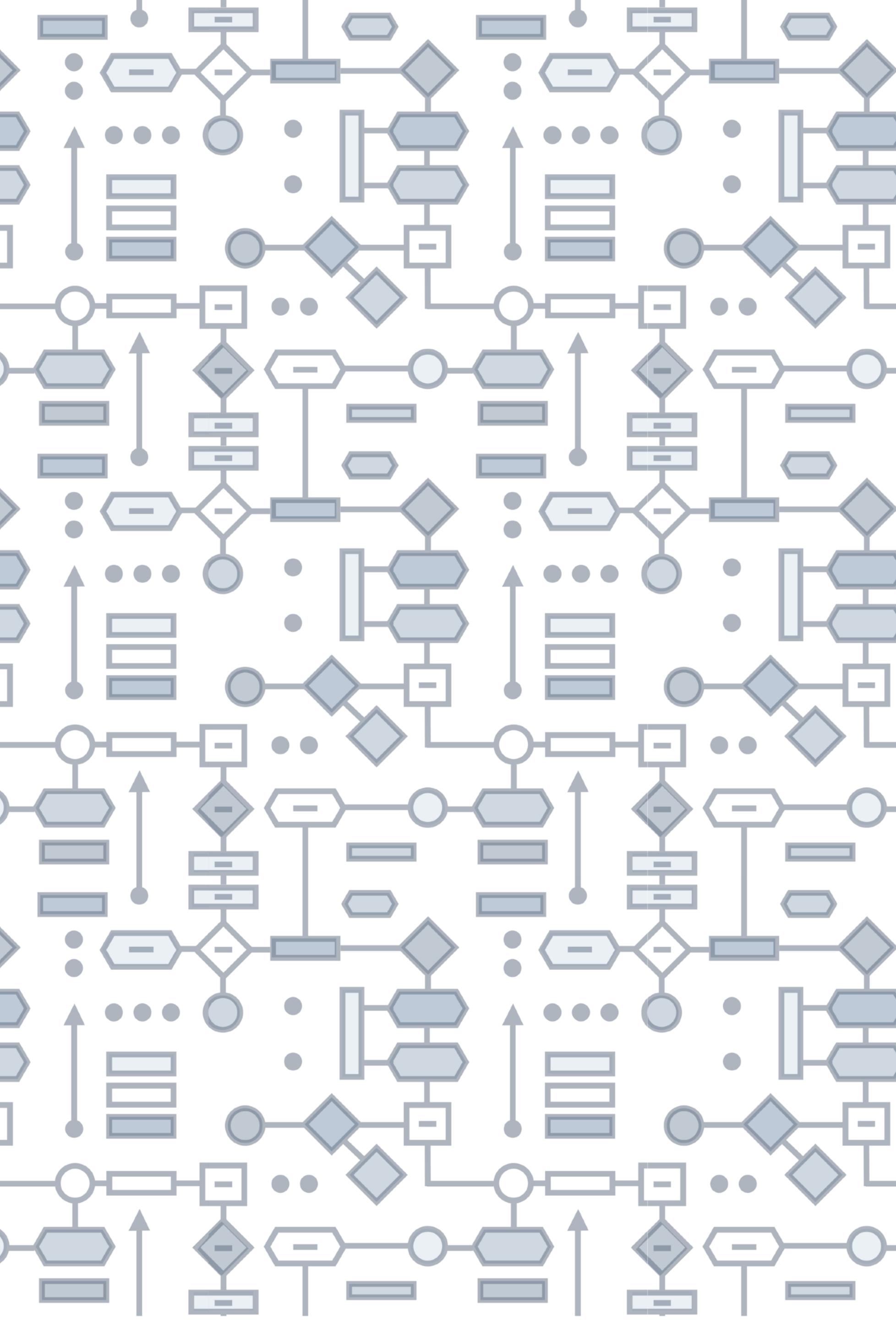
Task

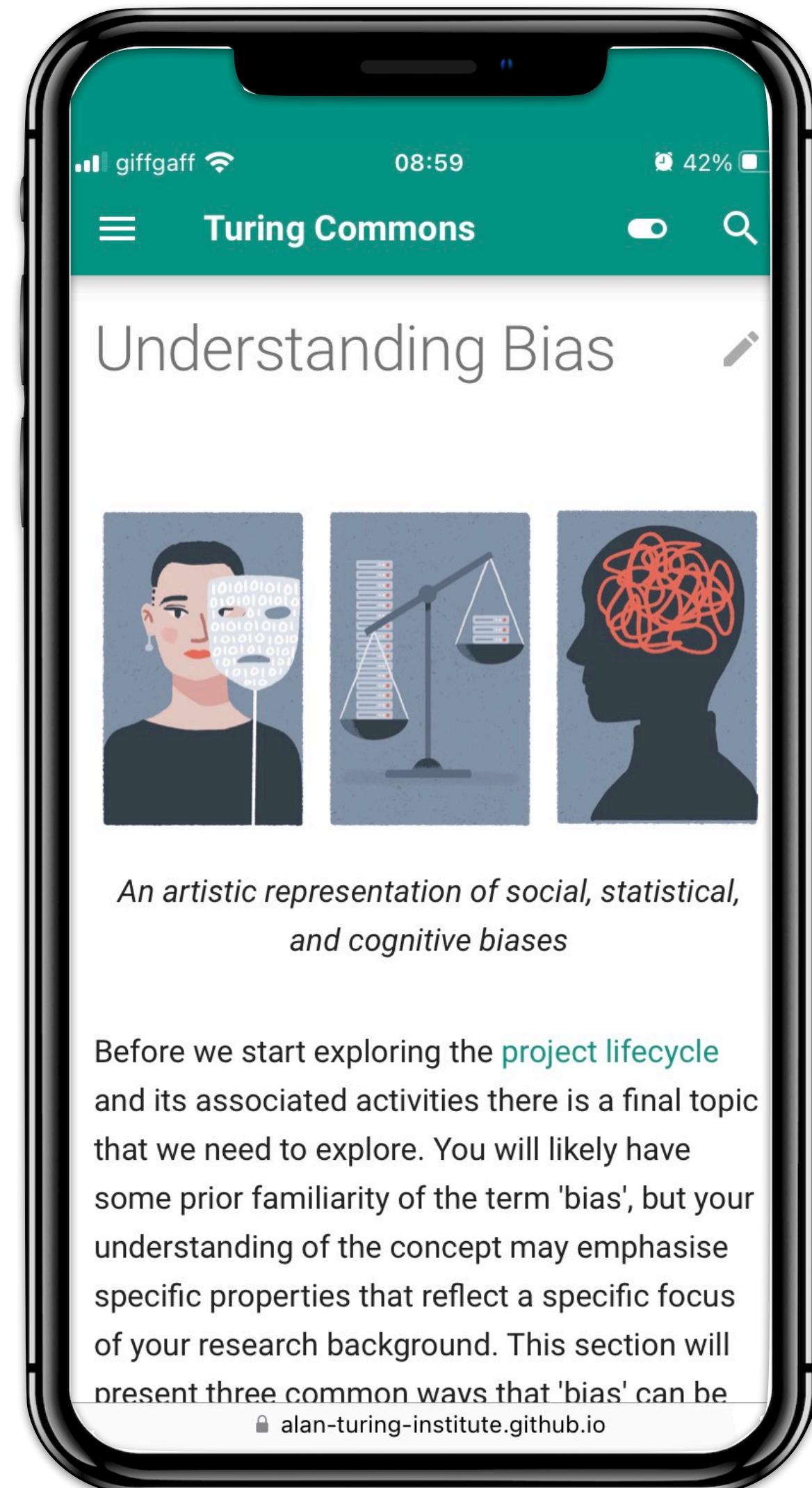
Co-design a hero image for your case study



Section 3

Activities





The browser window has a teal header bar with the 'Turing Commons' logo and navigation icons. The main content area has a sidebar with a tree menu:

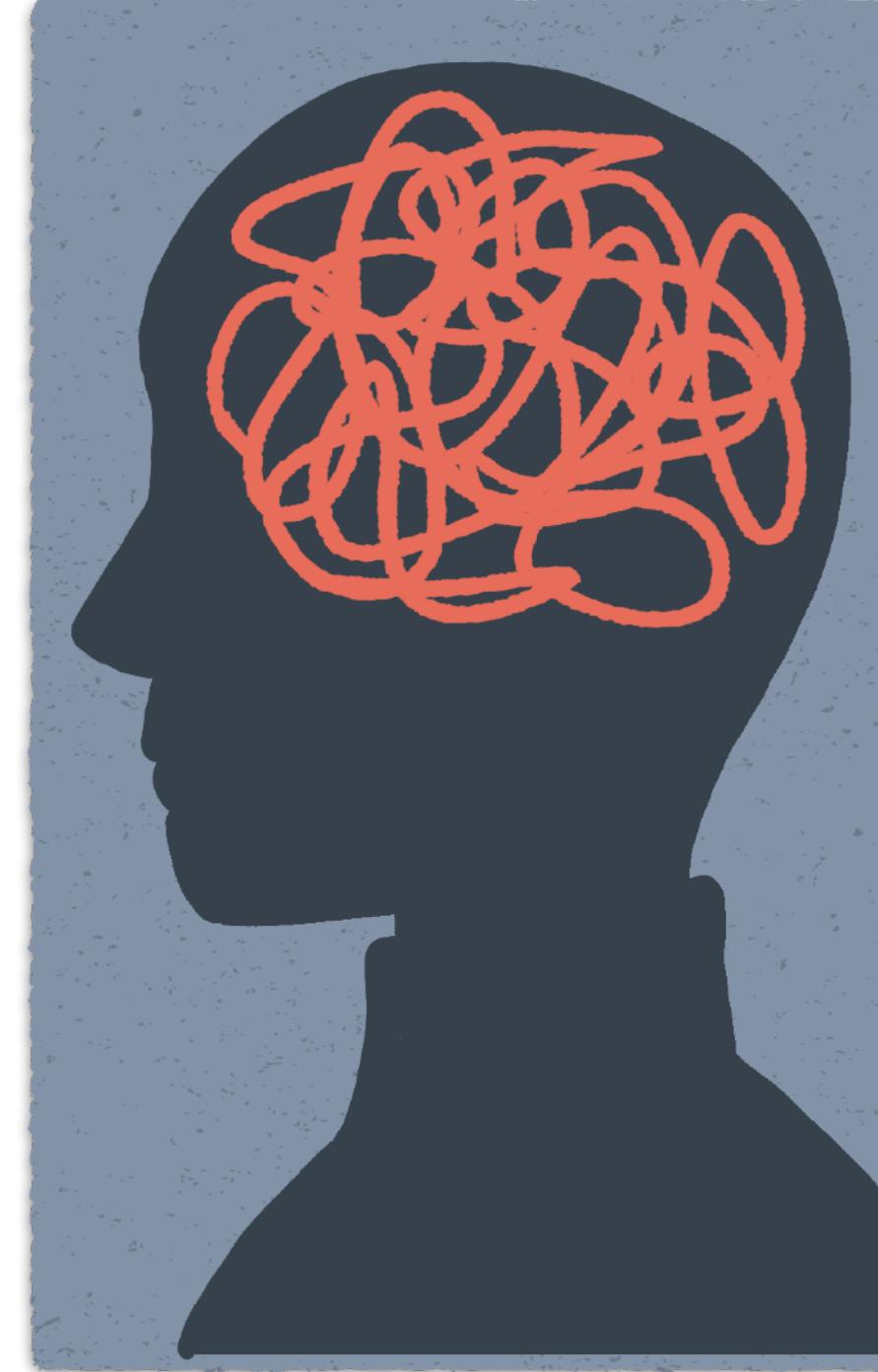
- Responsible Research and Innovation
 - Introduction
- What is Responsible Research and Innovation
 - What is Responsibility
 - A Short History of RRI
 - Science, Technology, and Society
 - Science and Technology Studies Timeline
- Responsible Data Science and AI
 - What is Responsible Data Science and AI
 - Introducing the Project Lifecycle
 - Roles and Responsibilities
 - Understanding Bias
- The Project Lifecycle
 - Case Studies
 - Project Planning
 - Problem Formulation
 - Data Extraction or Procurement
 - Data Analysis
 - Preprocessing and Feature Engineering
 - Model Selection and Training
 - Model Testing and Validation
 - Model Reporting
 - Model Productionalisation
 - User Training
 - System Use and Monitoring
 - Model Updating or Deprovisioning

The main content area starts with a section titled 'Data Analysis'. It includes a text block about the module's focus on data analysis, mentioning Jupyter notebooks. A note says you don't need Python or Jupyter Notebooks. A 'launch binder' button is present. To the right is a 'Table of contents' sidebar with a list of topics from 'What is Exploratory Data Analysis?' to 'Forms of Selection Bias (and more missing data)'.

Bias Mitigation

Self-Assessments

A *reflect-list* to support the identification of possible biases and actions that can be taken at various stages throughout the project lifecycle.

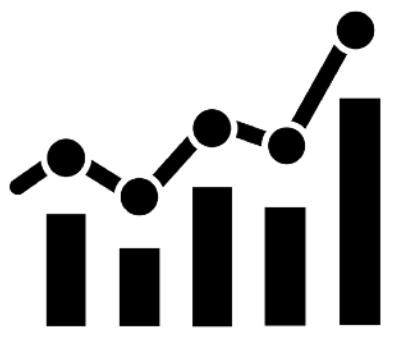


Question



What types of bias should be considered across a project's lifecycle?

Statistical



Cognitive



Social

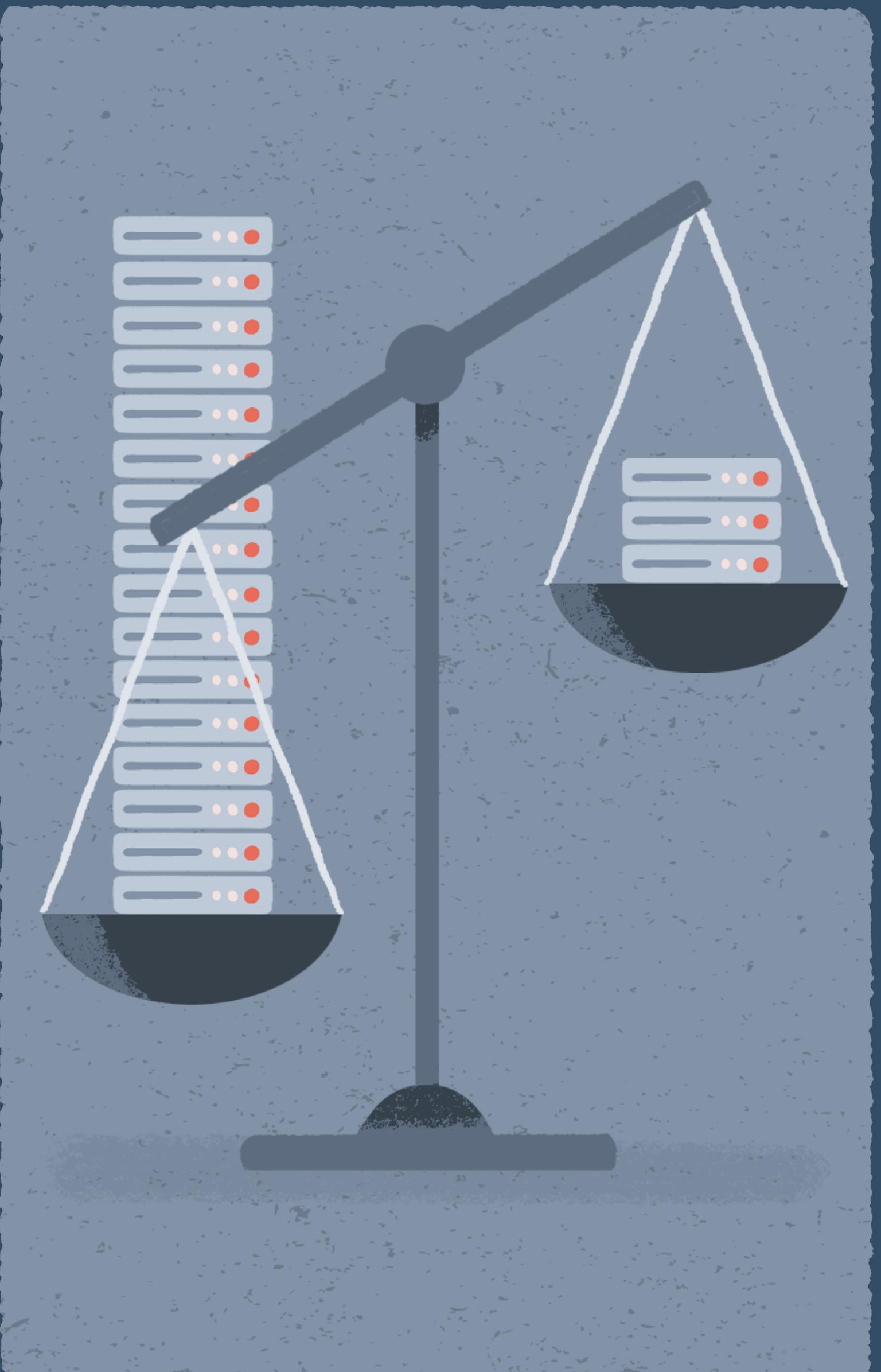


Understanding Bias

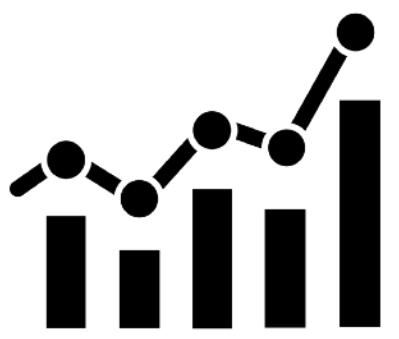
Statistical Bias

Statistical definitions of 'bias' will typically make reference to one of the following features (Aronson, 2018):

- **Systematicity:** bias arises from a systematic process, rather than a random or chance process.
- **Truth:** a realist assumption that the deviation is from a true state of the world
- **Error:** the bias reflects an error, perhaps due to sampling or measurement
- **Deviation (or Distortion):** a quantity in which the observed result is taken to differ from the actual result were there no bias.
- **Affected elements:** the study elements that may be affected by the bias include the conception, design, and conduct of the study, as well as the collection, analysis, interpretation, and representation of the data
- **Direction:** the deviation is directional, as it can be caused by both an under- or over-estimation



Statistical



Cognitive



Social

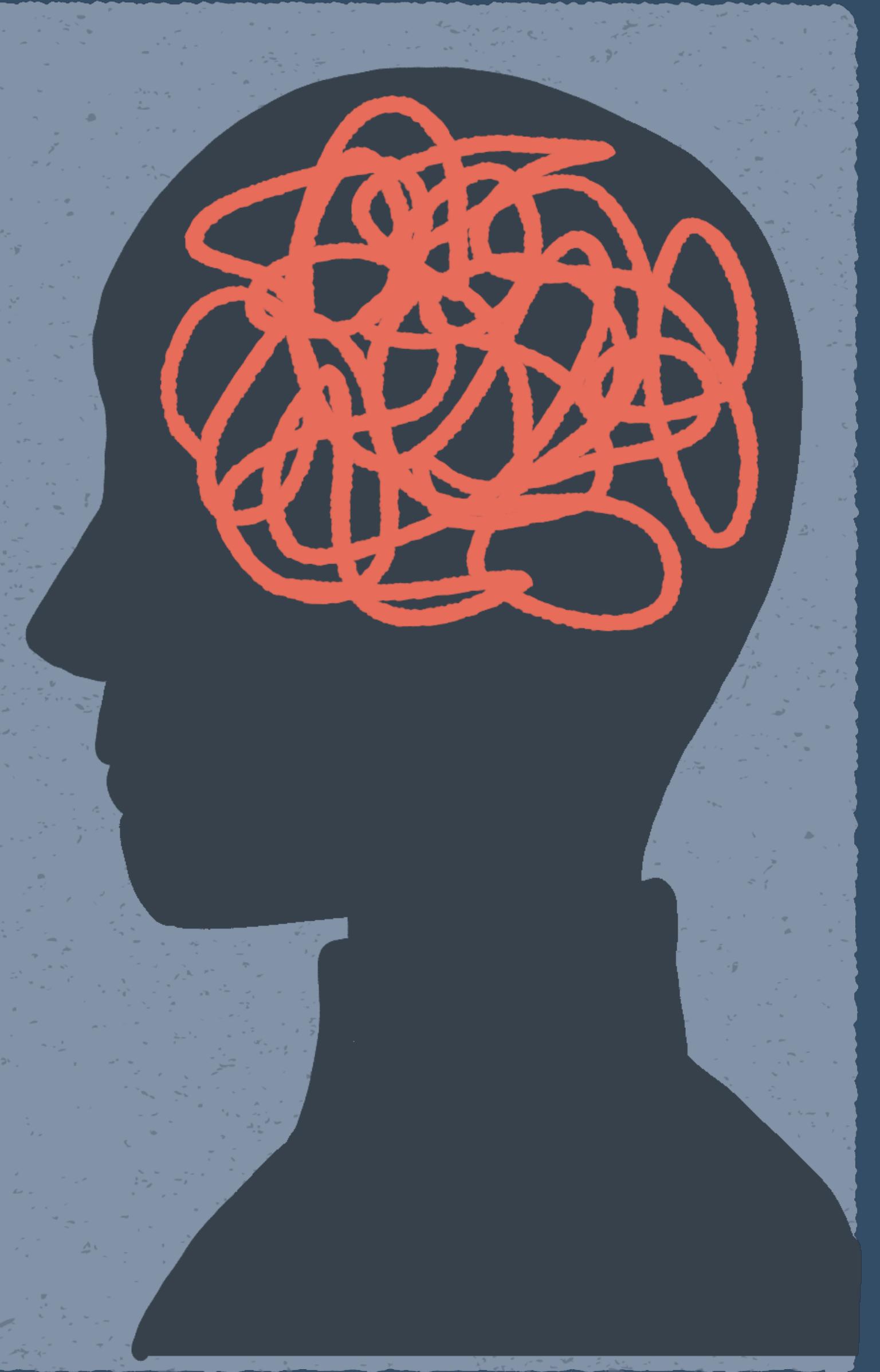


Understanding Bias

Cognitive Bias

Biases affect many facets of our cognition:

- Judgement
- Decision-Making
- Perception
- Information Retrieval
- Emotional Awareness
- Memory
- Logical Reasoning



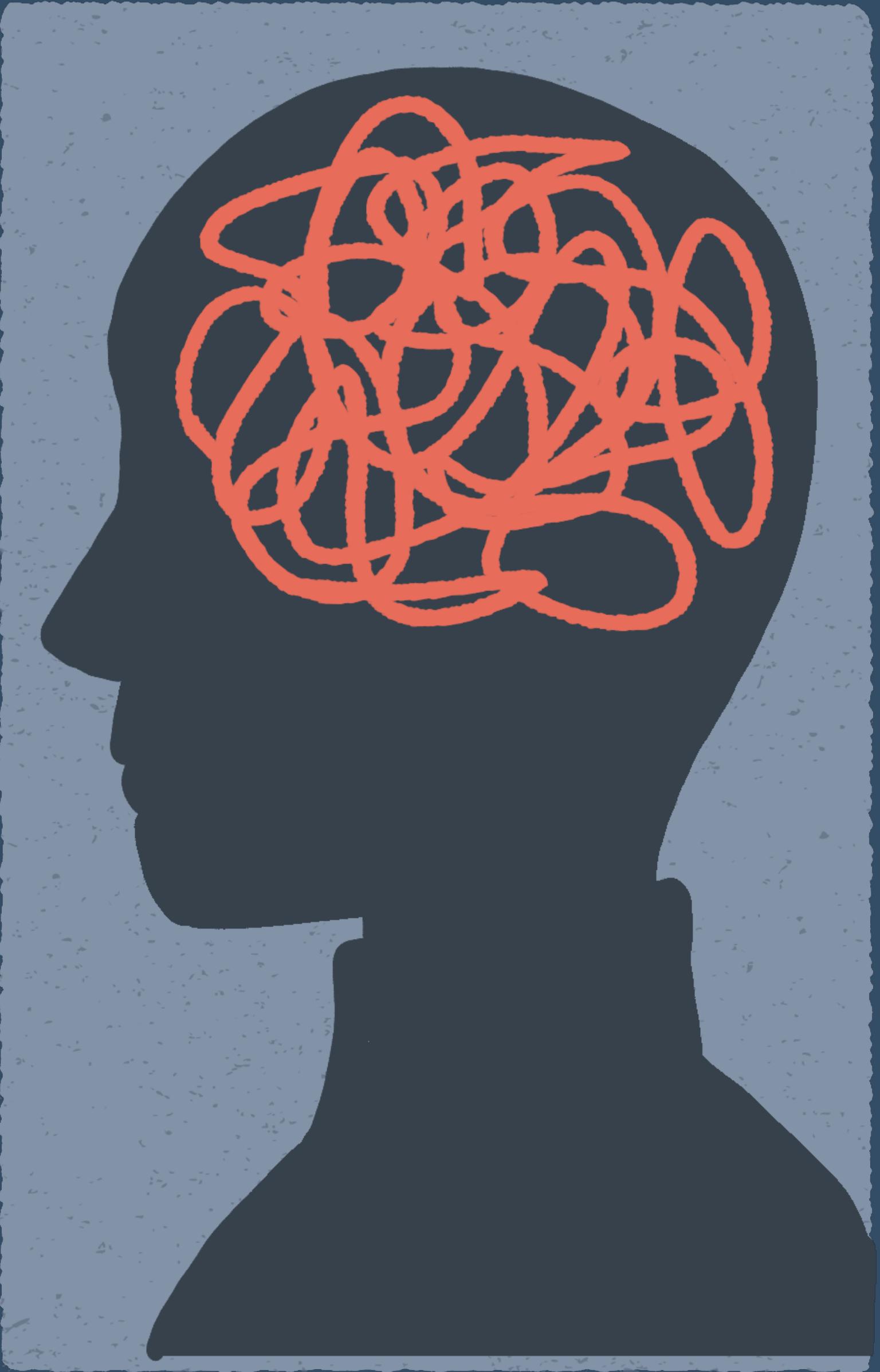
Understanding Bias

Cognitive Bias

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?

- A. Linda is a bank teller.
- B. Linda is a bank teller and is an active campaigner for women's rights.

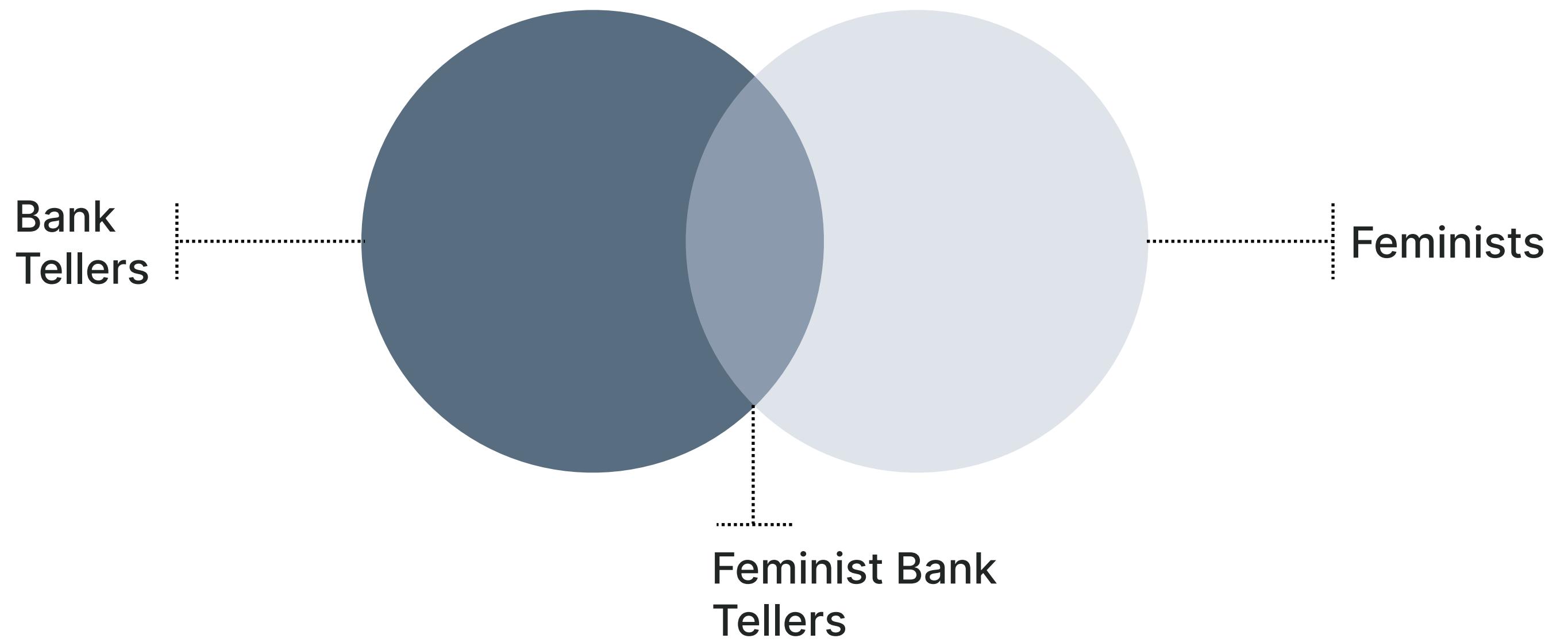




<https://www.menti.com/????>

Answer

Representativeness Heuristic



When Tversky and Kahneman posed this question to a group of 88 undergraduate students, only 15 got the correct answer (Tversky and Kahneman, 1983).

They attributed this to the representativeness heuristic.

The probability of two events occurring in conjunction, such as Linda being both a 'bank teller' and 'active in the feminist movement' must be less than or equal to the probability of either event occurring on its own.

Suppose we find 1,000 people who fit Linda's description and 10 of them work as bank tellers. How many of them are also feminists? At most, all 10 of them are; in that case, the two options are equally probable. If fewer than 10 are, the second option is less probable. But there is no way the second option can be more probable

Quotation

I am particularly fond of this example because I know that the [second] statement is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me, “but she can’t just be a bank teller; read the description.”

Statistical



Cognitive



Social



Understanding Bias

Social Bias

In ordinary usage, the term 'bias' is associated with myriad forms of prejudice or discrimination (i.e. a disposition to treat an individual or organisation in a way that is considered to be unfair).

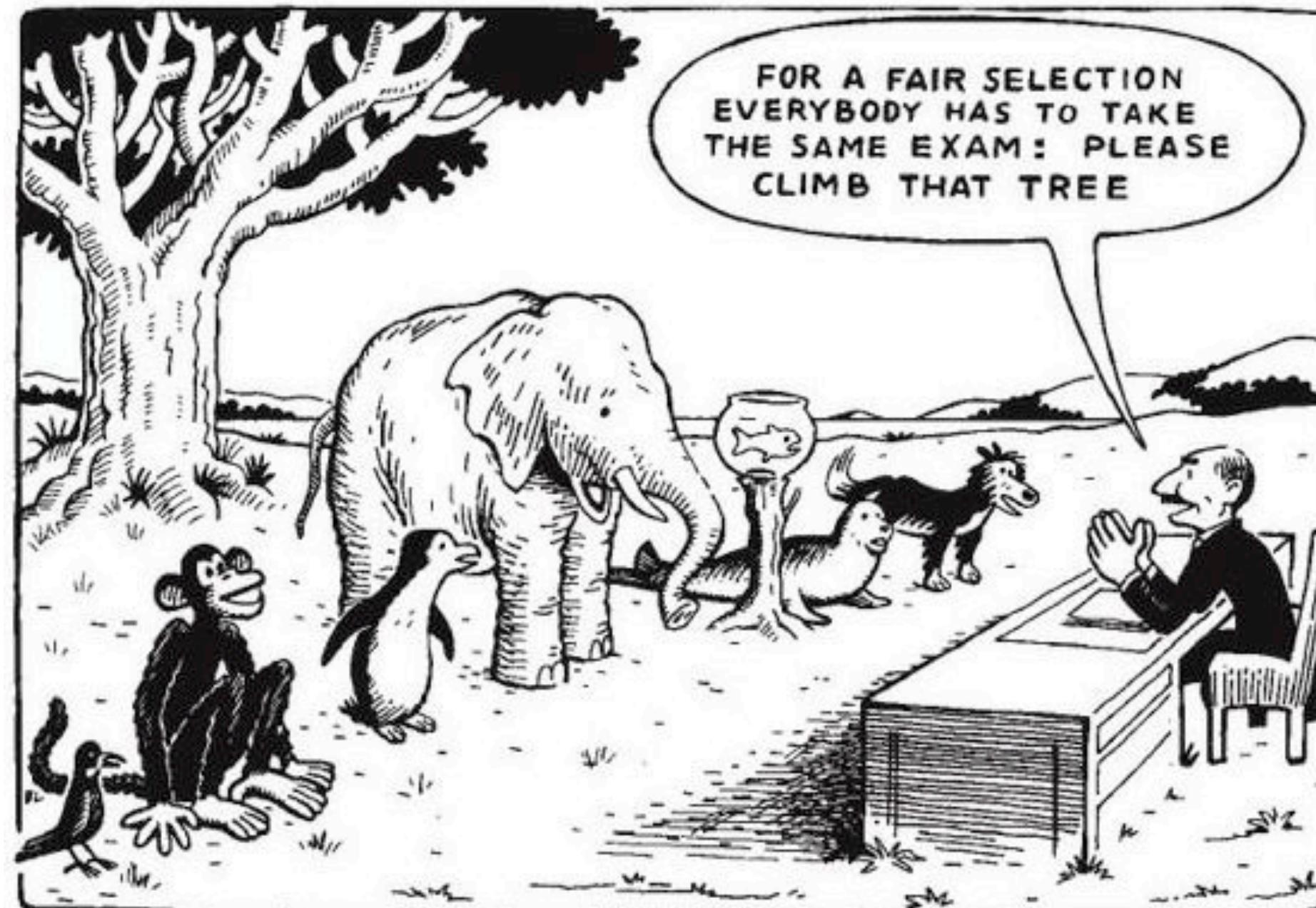
These biases may be embedded in social systems or organisational structures and practices (i.e., systemic or structural bias).

What is meant by this increasingly common phrase?



Social Bias

Structural Bias in Education

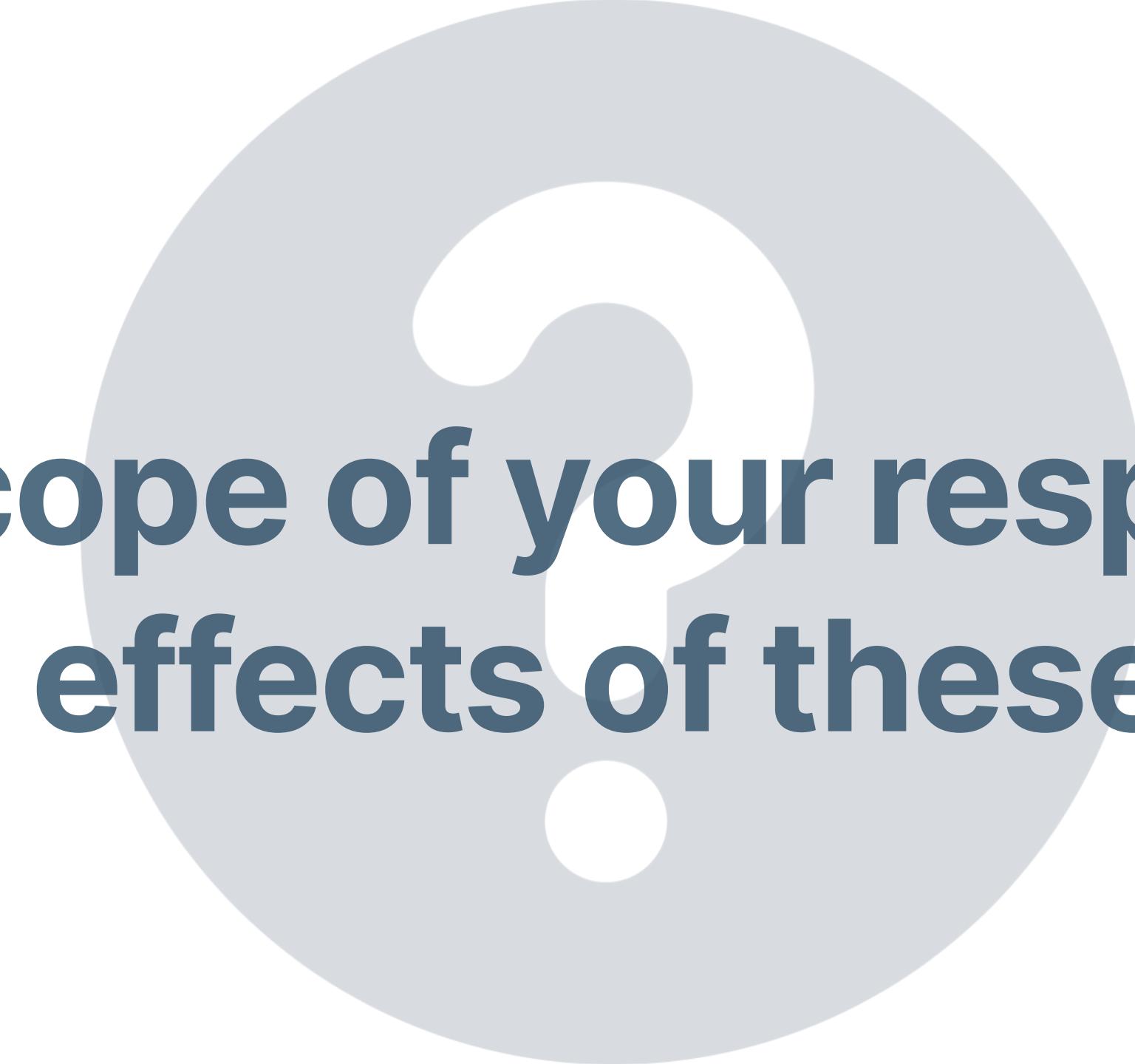


Amazon's recruitment tool perpetuated bias in hiring against women (Reuters, 2018)

Their algorithmic system learned to perpetuate a bias to prefer male candidates to female candidates because this reflected past hiring decisions.

This could have been an implicit (or cognitive) bias in their hiring procedure, or it could have been a result of structural bias in educational systems.

Deliberative Prompt



**What is the scope of your responsibility for
mitigating the effects of these biases?**

Bias Self-Assessment



Discussion

Are there other resources you think should be developed?



Thank you!

The
Alan Turing
Institute



Turing Commons