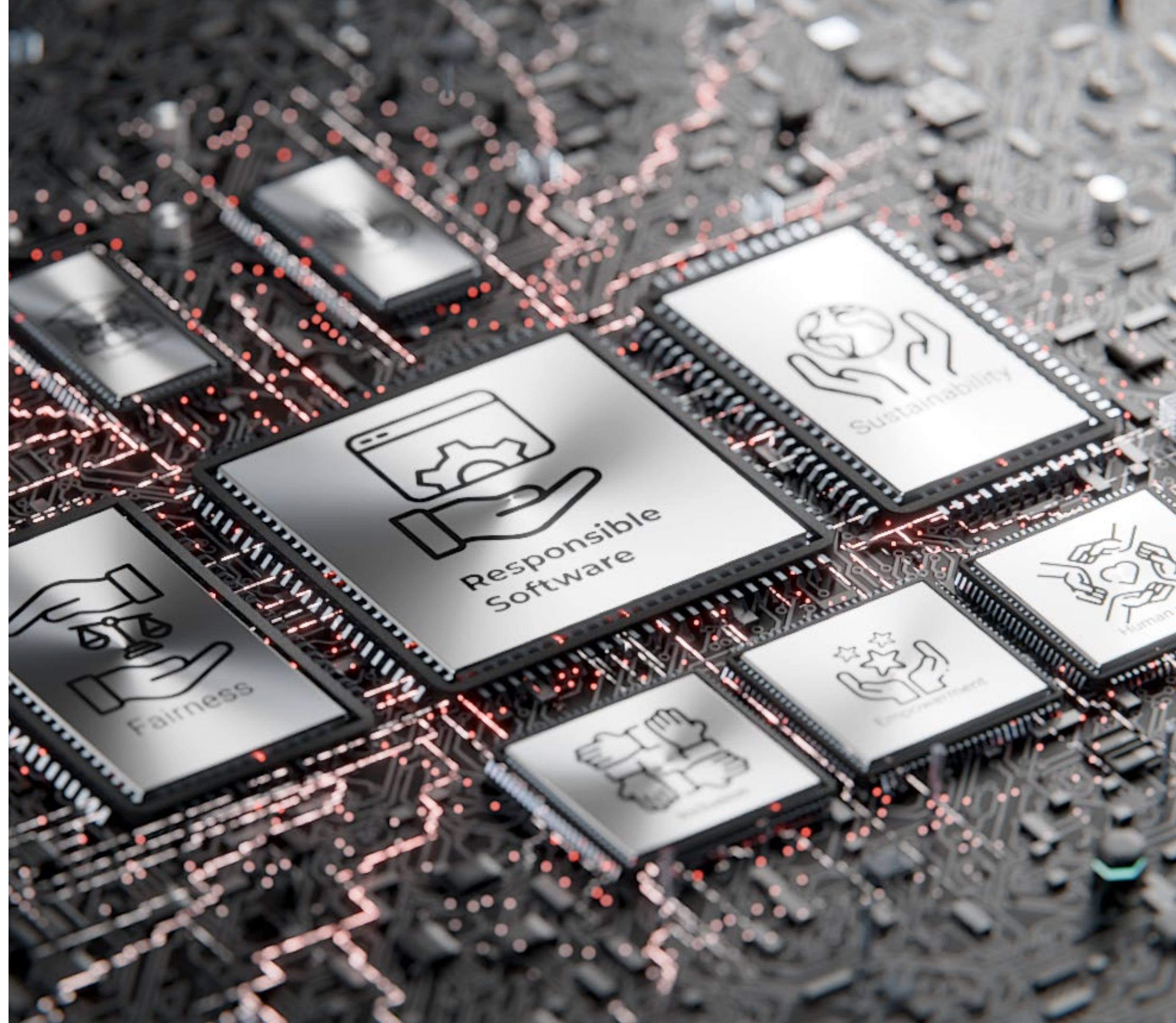


# **Blank class quizzes - Chapters 0 to 4**

Cécile Hardebolle

**Responsible  
Software**





# To keep in mind

---

Upon request of some students, I am providing you with blank versions of the interactive quizzes we have done in class for chapters 0 to 4.

They may help you review content and deepen your understanding but please keep in mind that these quizzes have been designed for **generating discussion in class** i.e. they are **NOT designed like exam questions** (e.g., they may have several correct answers).

# Autonomous car software - 1

---

The software of an autonomous car fails to recognize traffic signs correctly.

We are in the presence of (select all that apply):

- a. A safety threat
- b. A security threat
- c. A safety hazard
- d. A security hazard

URL: ttpoll.eu

Session ID: cs290

# Autonomous car software - 2

---

When weather conditions include rain, the software of an autonomous car fails to recognize traffic signs correctly.

We are in the presence of (select all that apply):

- a. A safety threat
- b. A security threat
- c. A safety hazard
- d. A security hazard

URL: ttpoll.eu

Session ID: cs290

# Worldwide “CrowdStrike” outage in 2024

This event is an example of:

- a. Malfunction
- b. Misuse, abuse
- c. Unintended use
- d. Intended use

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290

## CrowdStrike IT outage affected 8.5 million Windows devices, Microsoft says

20 July 2024

Share ↵ Save +

Joe Tidy  
Cyber correspondent, BBC News



The New York Times

## Stranded in the CrowdStrike Meltdown: ‘No Hotel, No Food, No Assistance’

Airlines pledged assistance, refunds and reimbursements to passengers whose travel had been disrupted by this summer’s software outage. Instead, passengers told us, they were on their own.

# Bad actors, safety and security

---

- a. Bad actors generate safety issues only
- b. Bad actors generate security issues only
- c. Bad actors generate both security and safety issues

URL: ttpoll.eu

Session ID: cs290



# Bad actors and the 4 scenarios

---

Bad actors can be involved in (select all that apply):

- a. Malfunction
- b. Misuse, abuse
- c. Unintended use
- d. Intended use

URL: ttpoll.eu

Session ID: cs290

# The “confusing” matrix - 1

---

We use software to detect fissures in concrete walls before they become visible to the naked eye.

A positive result means presence of fissure.

URL: ttpoll.eu  
Session ID: cs290

Select all the correct statements:

- a. TN = actual absence of fissure, correct prediction
- b. TP = actual absence of fissure, correct prediction
- c. FN = actual presence of fissure, incorrect prediction
- d. FP = actual presence of fissure, incorrect prediction



# The “confusing” matrix - 2

---

We use software to detect fissures in concrete walls before they become visible to the naked eye.

A positive result means presence of fissure.

From a safety perspective, the indicator we should pay most attention to is:

- a. TN
- b. TP
- c. FN
- d. FP

URL: ttpoll.eu  
Session ID: cs290

# The “confusing” matrix - 3

We use software to detect fissures in concrete walls before they become visible to the naked eye.

A positive result means presence of fissure.

Here is the confusion matrix you get 🙌

What is the False Negative Rate (FNR)?

- a. 13%
- b. 17%
- c. 20%
- d. 25%

|        |            | Predicted |            |
|--------|------------|-----------|------------|
|        |            | Fissure   | No Fissure |
| Actual | Fissure    | 60        | 15         |
|        | No Fissure | 20        | 100        |

# Which among these are bad actors in Catter?

---

Select all that apply:

- a. Sassy posts a fake picture of Dogs invading Purrville
- b. KitKat re-posts covert Dog propaganda
- c. Felix promotes cat-biscuits that his cousin cooks
- d. Tuna posts a series of angry replies to Catnip's post

URL: ttpoll.eu

Session ID: cs290



# Harm categories - 1

---

A user sees their post unfairly censored.  
This harm is in the category (select one):

- a. Physical injury
- b. Emotional or psychological injury
- c. Opportunity loss
- d. Economic loss
- e. Dignity loss
- f. Liberty loss
- g. Privacy loss
- h. Environmental impact
- i. Manipulation
- j. Social detriment

URL: [ttpoll.eu](https://ttpoll.eu)

Session ID: cs290

# Harm categories - 2

---

A fitness app leaks GPS location data on social media.

This harm is in the category (select one):

- a. Physical injury
- b. Emotional or psychological injury
- c. Opportunity loss
- d. Economic loss
- e. Dignity loss
- f. Liberty loss
- g. Privacy loss
- h. Environmental impact
- i. Manipulation
- j. Social detriment

URL: [ttpoll.eu](https://ttpoll.eu)

Session ID: cs290

# Harm categories - 3

---

Online ads lead a compulsive shopper to additional purchases.

This harm is in the category (select one):

- a. Physical injury
- b. Emotional or psychological injury
- c. Opportunity loss
- d. Economic loss
- e. Dignity loss
- f. Liberty loss
- g. Privacy loss
- h. Environmental impact
- i. Manipulation
- j. Social detriment

URL: ttpoll.eu

Session ID: cs290

# Harm categories - 4

---

A recruitment software indirectly discriminates based on people's name.

This harm is in the category (select one):

- a. Physical injury
- b. Emotional or psychological injury
- c. Opportunity loss
- d. Economic loss
- e. Dignity loss
- f. Liberty loss
- g. Privacy loss
- h. Environmental impact
- i. Manipulation
- j. Social detriment

URL: ttpoll.eu

Session ID: cs290



# Harm categories - 5

---

The results of an image search engine for “Nurse” show only women.  
This harm is in the category (select one):

- a. Physical injury
- b. Emotional or psychological injury
- c. Opportunity loss
- d. Economic loss
- e. Dignity loss
- f. Liberty loss
- g. Privacy loss
- h. Environmental impact
- i. Manipulation
- j. Social detriment

URL: [ttpoll.eu](https://ttpoll.eu)

Session ID: cs290

# Macro-level perspective

URL: ttpoll.eu  
Session ID: cs290

A macro-level perspective is useful (select all correct statements):

- a. When software is under design
- b. After software is deployed
- c. After an analysis with a meso-level perspective
- d. When considering expanding to new countries
- e. When software is used by public institutions

# Disinformation

---

URL: ttpoll.eu  
Session ID: cs290

A piece of information is false but created without intention to harm.  
It is (select all that apply):

- a. Misinformation
- b. Disinformation
- c. Malinformation
- d. Fake news

# False beliefs

---

URL: ttpoll.eu  
Session ID: cs290

If you are exposed to a dis/mis-information post by Melon Husk, you are more likely to believe it because of:

- a. System 2
- b. False consensus
- c. Source cues
- d. Illusory truth

# Software & disinformation

URL: ttpoll.eu

Session ID: cs290

Software playing a role in disinformation can be (select all that apply):

- a. Generative AI
- b. Bots
- c. Content moderation systems
- d. Content recommendation systems



# Humans & disinformation

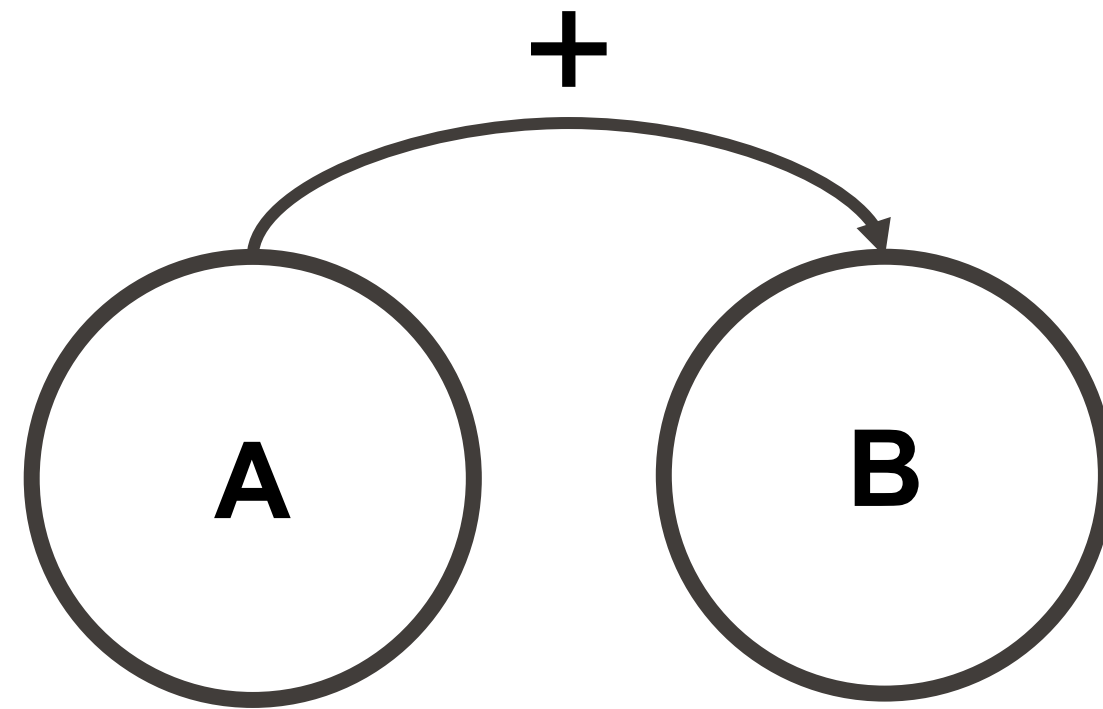
URL: ttpoll.eu  
Session ID: cs290

Humans playing a role in disinformation do it (select all that apply):

- a. Because it's their work
- b. By inattention
- c. For political reasons
- d. To please other people
- e. To spark emotions

# Causal Loop Diagrams

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290



The arrow with label “+” means:

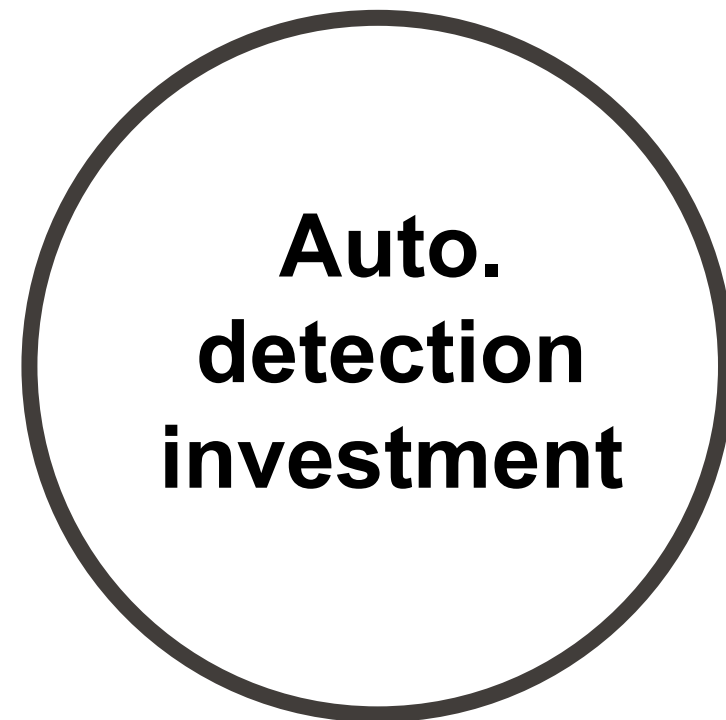
- a. There's a transition from state A to state B on token “+”
- b. The quantity in A is added to the quantity in B
- c. A and B change in a positive direction
- d. B changes in the same direction as A

# Causal Loop Diagrams - 1

---

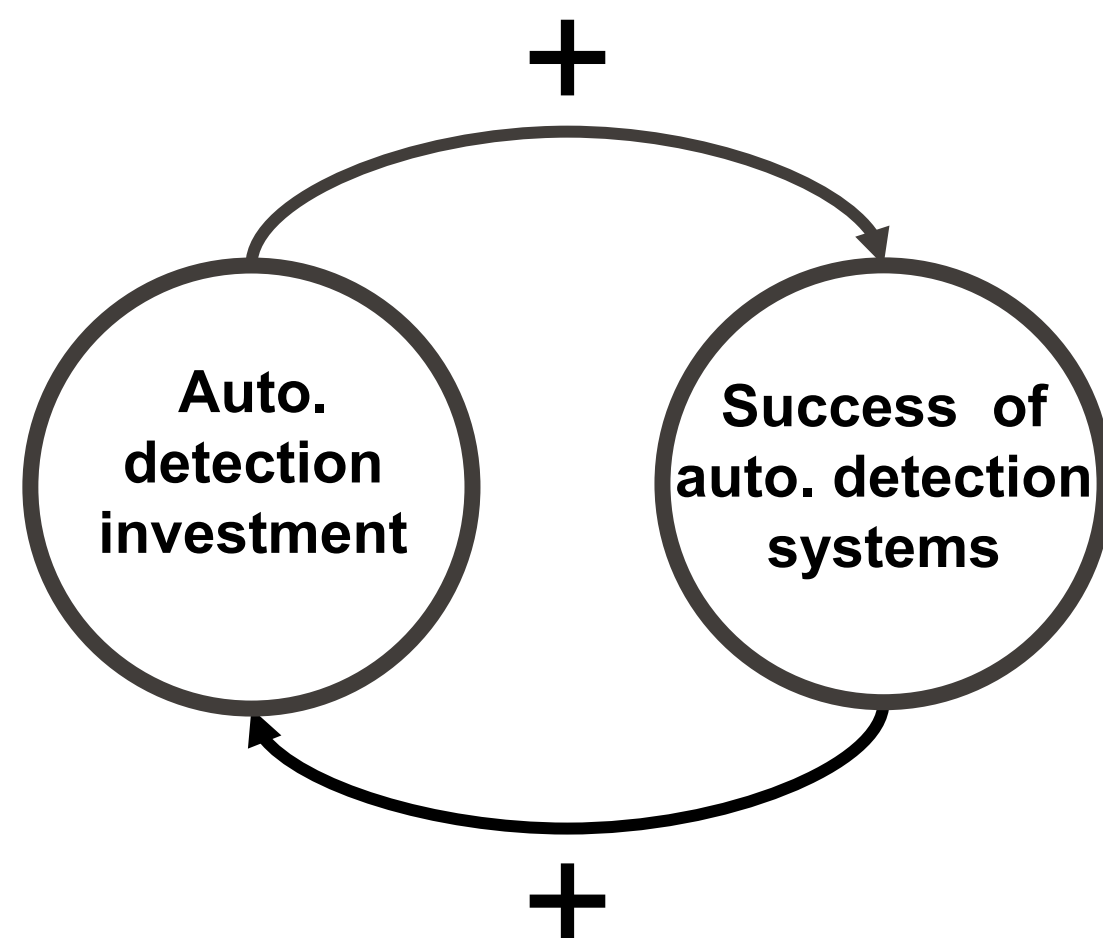
Key variables:

- Investment of resources in automatic detection systems to detect/filter fake news
- Success of automatic detection systems to detect/filter fake news



# Causal Loop Diagrams – 1a

URL: ttpoll.eu  
Session ID: cs290



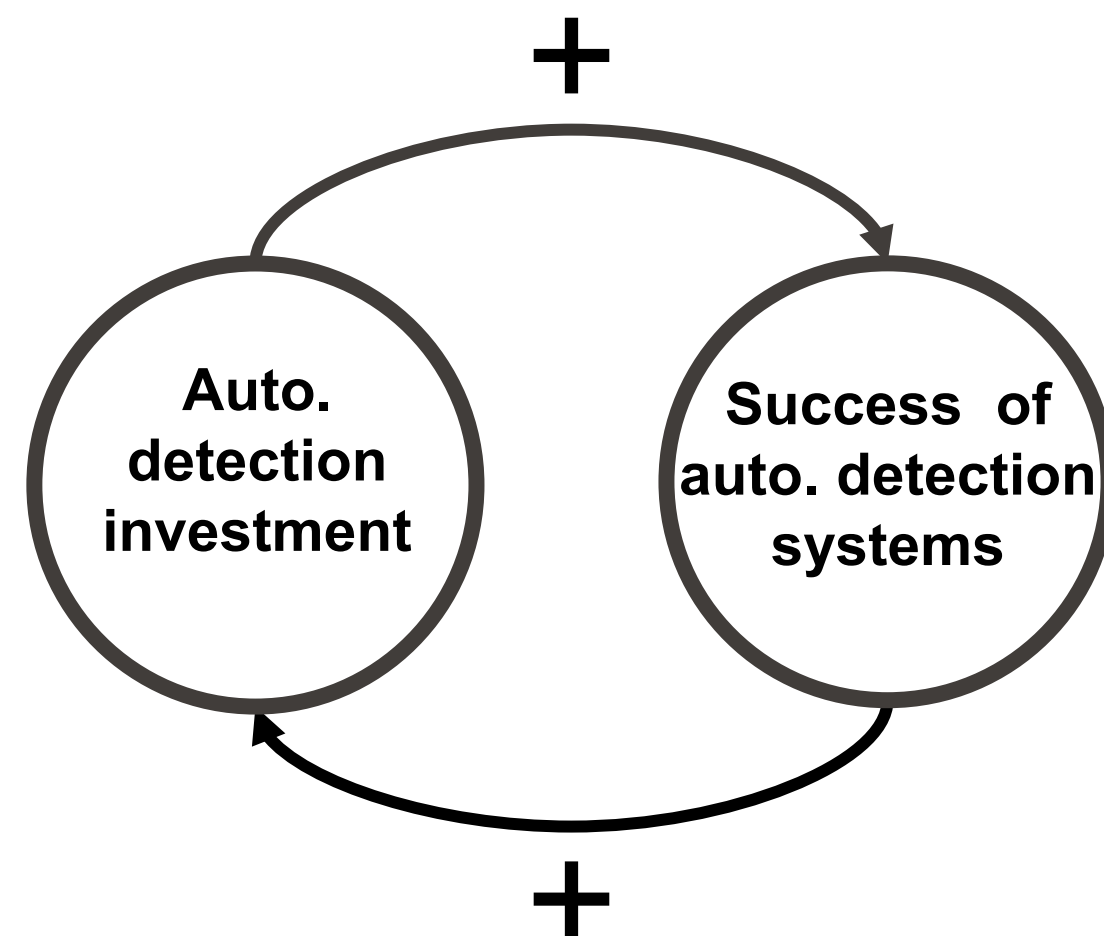
Over time, the quantities in this system will:

- a. Stabilize
- b. Increase
- c. Decrease
- d. It depends

Simulation: <https://go.epfl.ch/cs290-cld-1>

# Causal Loop Diagrams – 1b

URL: ttpoll.eu  
Session ID: cs290



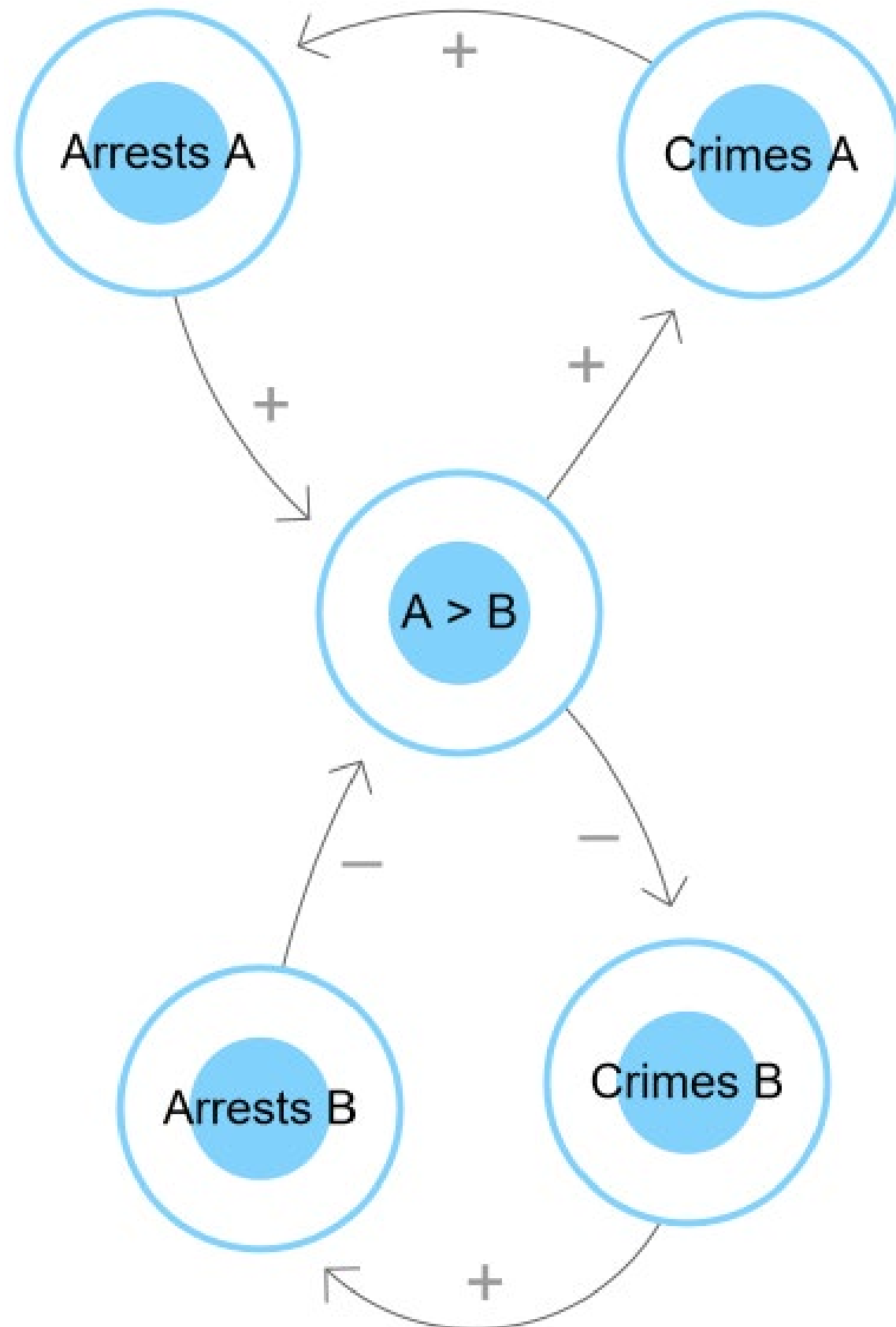
The loop in this diagram is:

- a. Balancing
- b. Reinforcing



# Predictive policing

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290



Over time, the quantities in this system will:

- a. Stabilize
- b. Increase
- c. Decrease
- d. It depends



# Attributes - 1

URL: ttpoll.eu  
Session ID: cs290

**Hair color as an attribute to represent people is:**  
(select all that apply)

- a. A sensitive attribute
- b. A protected attribute
- c. An observed variable
- d. A latent variable
- e. An objective representation of people
- f. A subjective representation of people

# Attributes - 2

URL: ttpoll.eu  
Session ID: cs290

Let's imagine a software that relies on SAT scores (standardized test for university admission in the US) to make recommendations of when to approve study loans.

**The SAT score is a \_\_\_\_\_ attribute.**

- a. Sensitive
- b. Protected
- c. Private
- d. Public
- e. Proxy
- f. System

# Bias - 1

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290

The city of Lozhann decides to deploy a smartphone app that allows residents to report potholes throughout the city to help with the identification of repair needs.

**The data collected by the app will probably exhibit:**  
(select all that apply)

- a. Preexisting bias
- b. Confirmation bias
- c. Representation bias
- d. Measurement bias
- e. Automation bias



# Biases in the ML lifecycle - 1

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290

Simpson's paradox is when the patterns observed at the level of the full sample and at the level of subgroups are opposed.

**When training a ML model, Simpson's paradox can lead to**  
(select 1 answer):

- a. Evaluation bias
- b. Aggregation bias
- c. Optimization choices
- d. Deployment bias

# Biases in the ML lifecycle - 2

URL: ttpoll.eu  
Session ID: cs290

The society RetailProtect develops a ML model to identify instances of shoplifting in retail shops. They evaluate their model on a benchmark in which actors from diverse ethnicities simulate a range of shoplifting actions.

**This can lead to (select 1 answer):**

- a. Evaluation bias
- b. Aggregation bias
- c. Optimization choices
- d. Deployment bias



# Fairness metrics - 1

URL: [ttpoll.eu](http://ttpoll.eu)

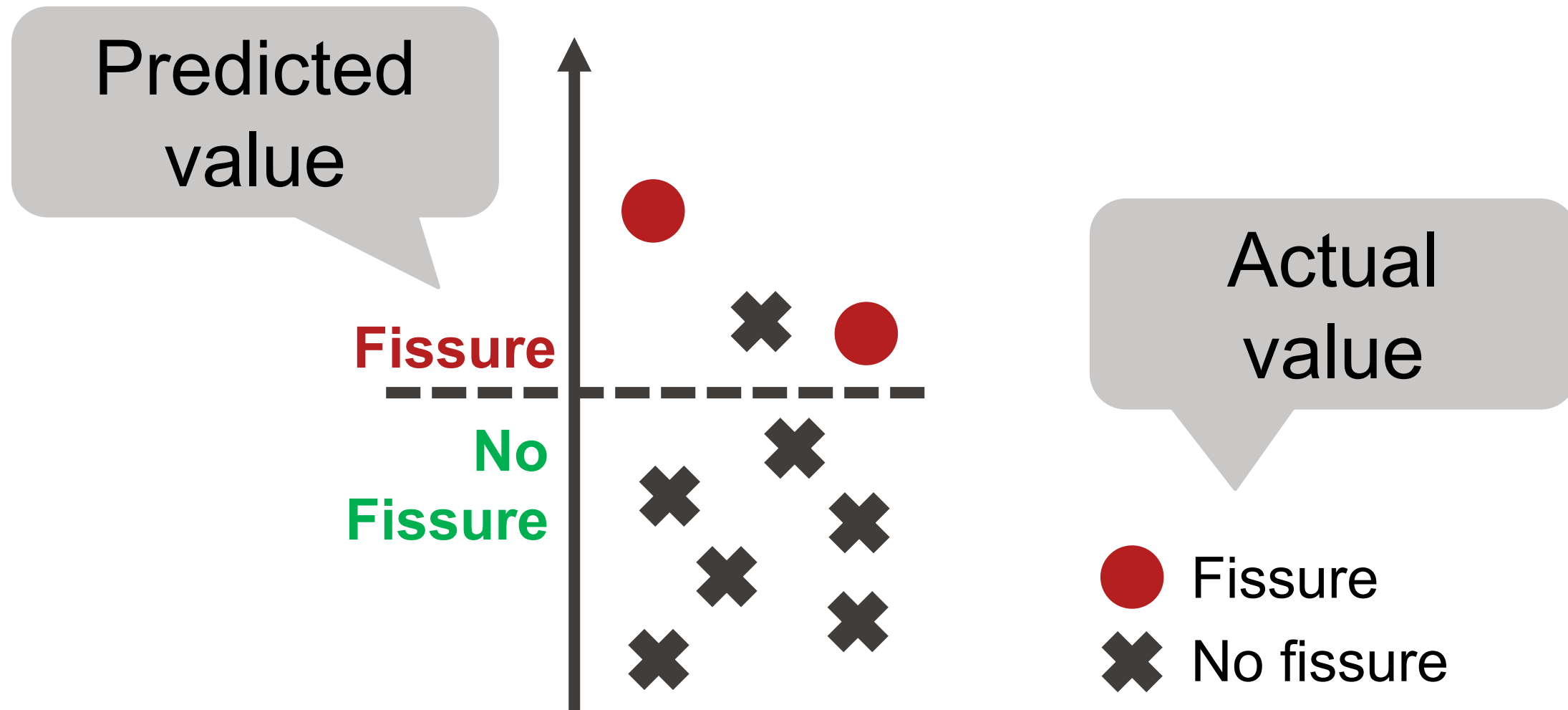
Session ID: cs290

Among the metrics below, **which can be used to assess the fairness** of a piece of software? (select all that apply)

- a. Accuracy
- b. False Positive Rate
- c. False Negative Rate
- d. False Discovery Rate
- e. False Omission Rate
- f. Positive Predictive Value
- g. Negative Predictive Value
- h. Positive prediction rate (also called acceptance rate)

# Fairness metrics – 2

The company SuperCrack has developed a model to detect fissures in concrete before they become visible. They evaluate their model against a benchmark. The results look like this:



|        |            | Predicted |            |
|--------|------------|-----------|------------|
|        |            | Fissure   | No Fissure |
| Actual | Fissure    | 2         | 0          |
|        | No Fissure | 1         | 6          |



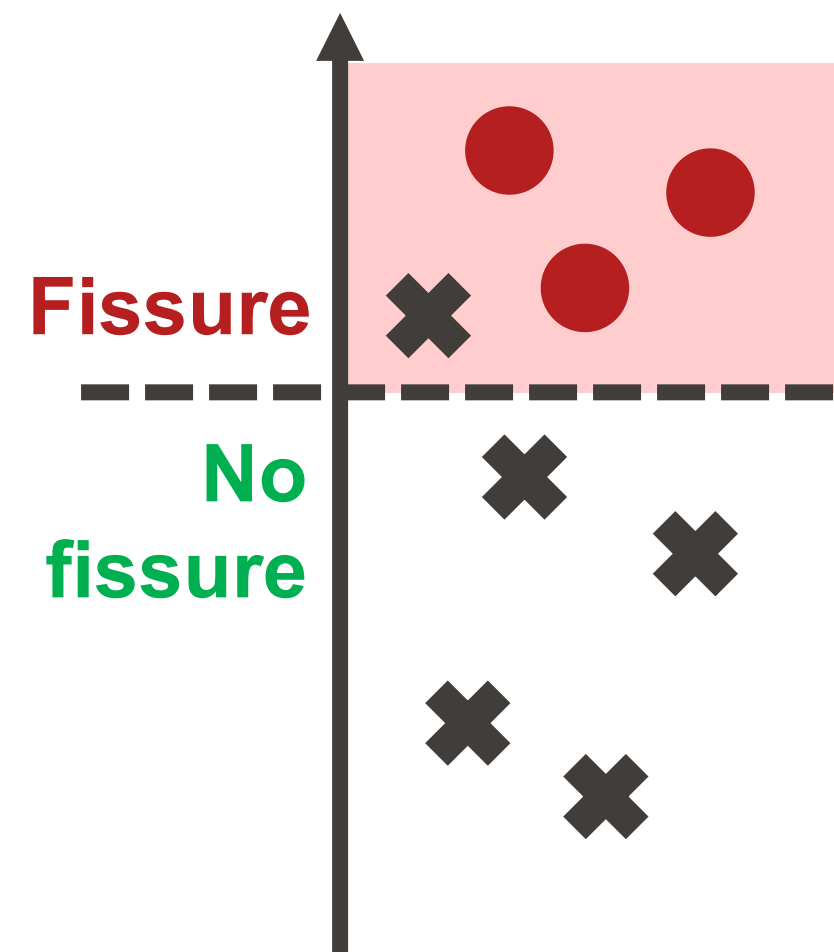
# Fairness metrics – 2a

URL: [ttpoll.eu](http://ttpoll.eu)  
Session ID: cs290

They want to know whether their model performs equally well for plain concrete and for reinforced concrete. Here are the results:

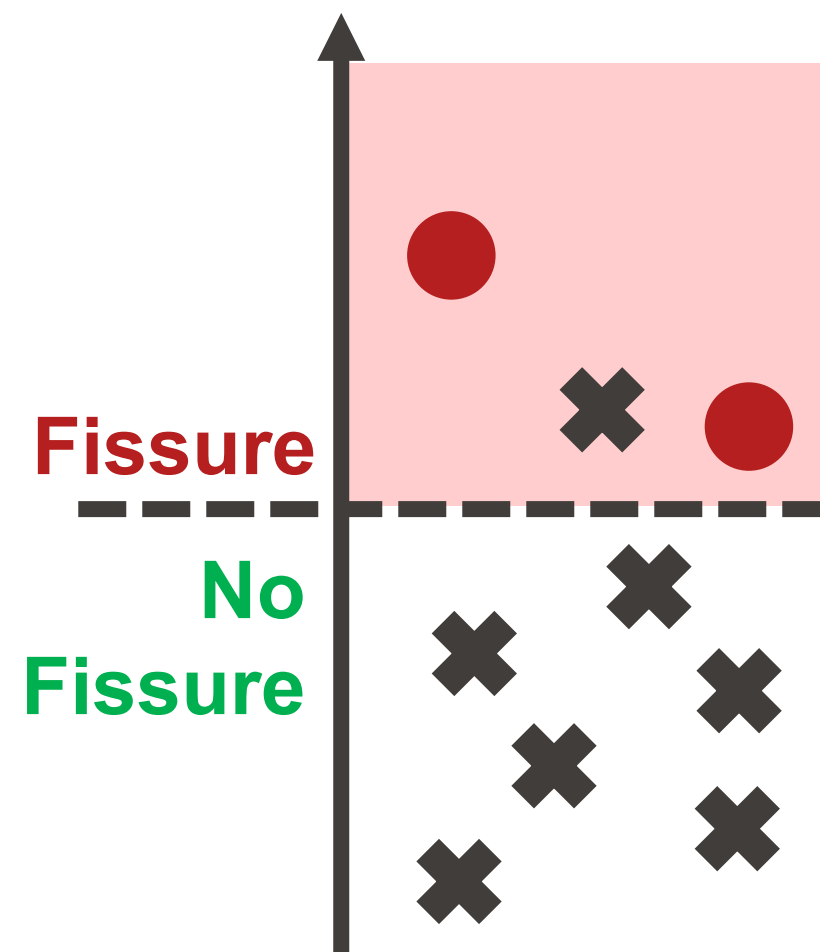
Metric = **4** / 8

Plain  
Concrete



Metric = **3** / 9

Reinforced  
Concrete



**Which metric are they using?** (select 1 answer)

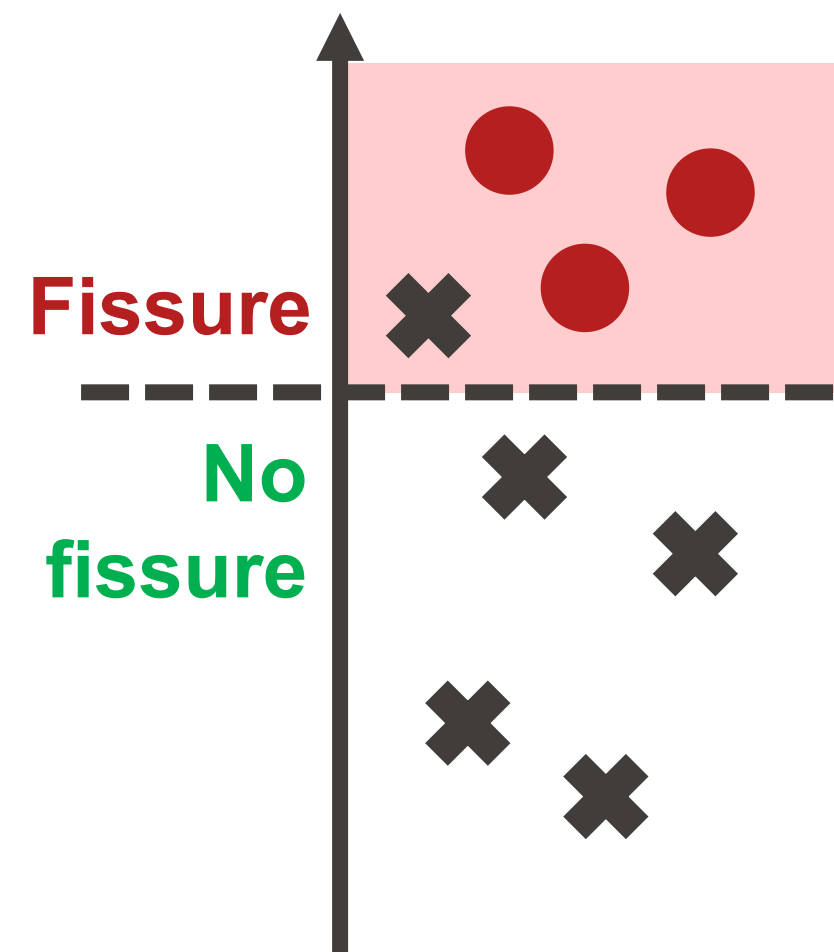
- a. Equal accuracy
- b. Error rate balance
- c. Error parity
- d. Demographic parity

# Fairness metrics – 2b

URL: ttpoll.eu  
Session ID: cs290

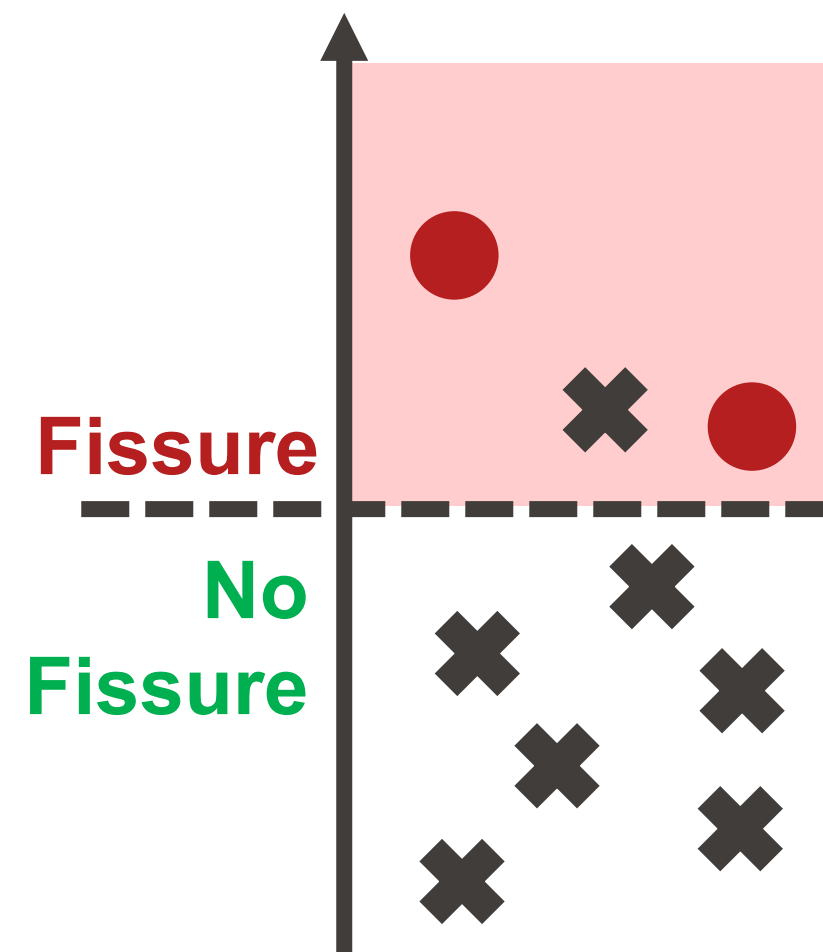
Metric = 4 / 8

Plain  
Concrete



Metric = 3 / 9

Reinforced  
Concrete



**According to this metric,  
is their model fair?**  
(select 1 answer)

- a. Yes
- b. No
- c. Other option