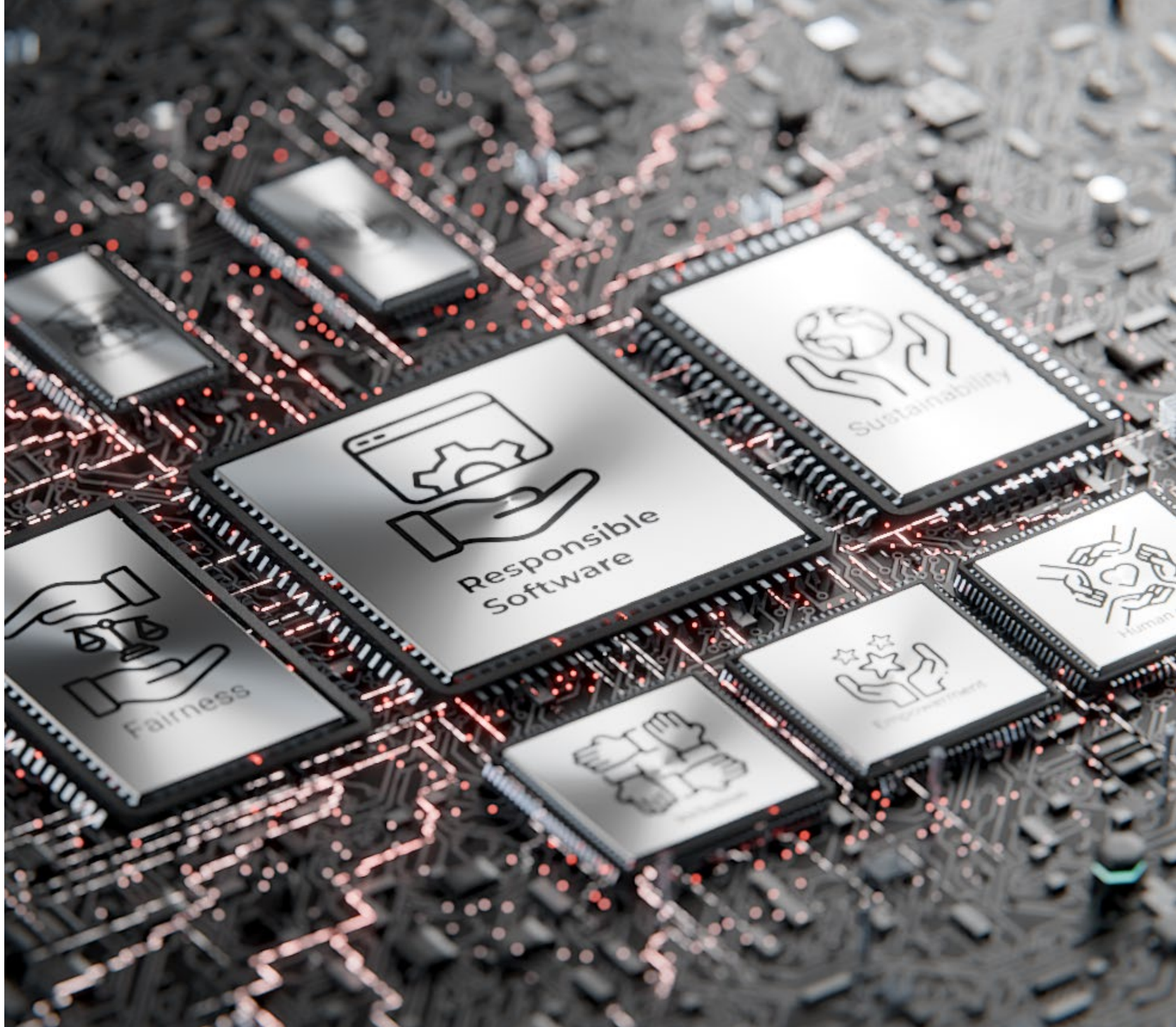
Empowerment 1 Review & Case studies

25 nov.

Cécile Hardebolle

Responsible Software



Agenda for today

1. Upcoming dates in the course

2. Interactive review questions on Empowerment 1

3. Case studies:
   a) Digital Ethics Canvas
   b) Value analysis
   c) “Dark” Patterns

Next dates

	Monday (SG1)	Tuesday (Computer Rooms)
25 Nov – 1 Dec	Empowerment 1 cases	Graded Assignment 2
2 Dec – 8 Dec	Debriefing Graded 2	Empowerment 2 notebook
9 Dec – 15 Dec	Empowerment 2 cases	Conclusion & Q&A in **SG1**
16 Dec – 20 Dec	Final exam	---

Debriefing" =

- Global **feedback** to the class + discuss your **questions**
- Work through **most difficult exercises**

"Conclusion & Q&A" =

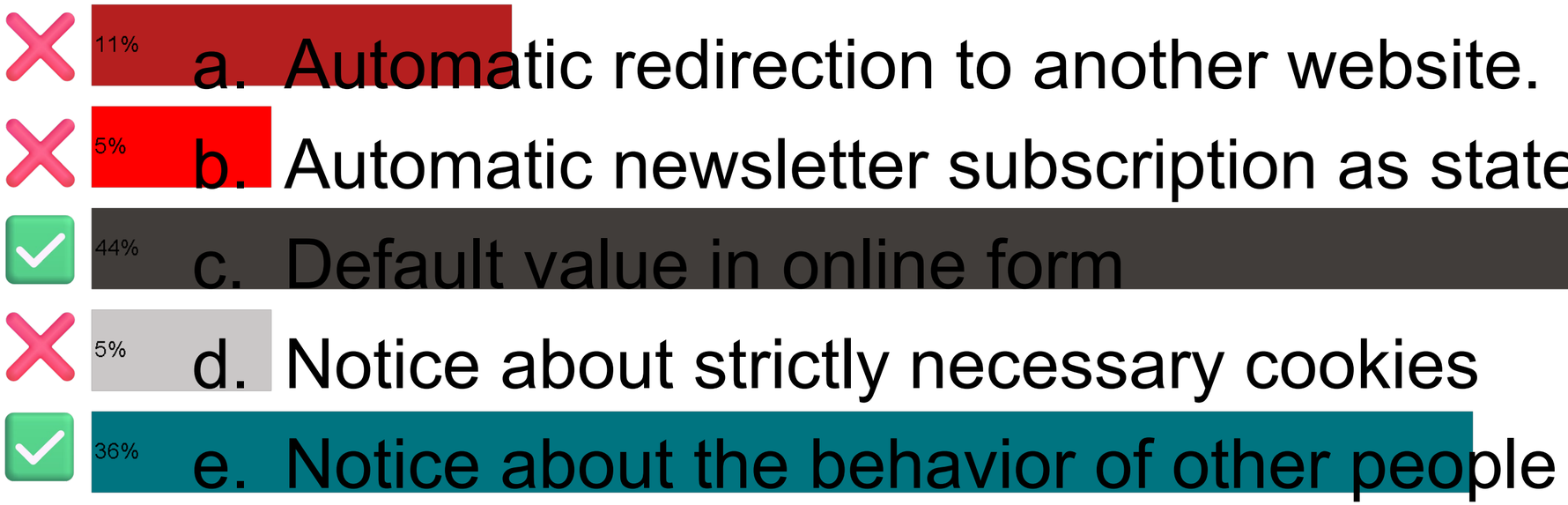
- Final **overview cases**
- Your **questions** (to post in advance - how is to be defined)

Review questions Empowerment 1

Nudges - 1

URL: ttpoll.eu
Session ID: cs290

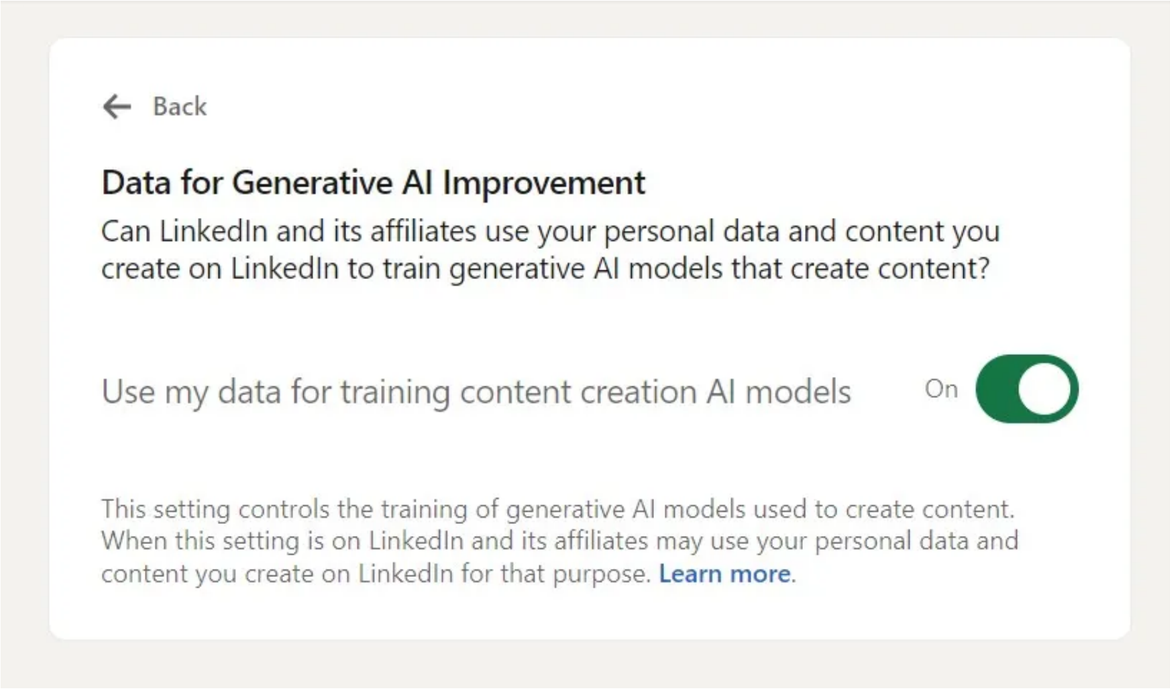
Which of the following are examples of digital (software) nudges? (select all that apply)



Nudge:
- Alter behavior
- Without: forbidding choices, changing incentives, taking away choice

Nudges - 2

URL: ttpoll.eu
Session ID: cs290



This is one of the settings on LinkedIn in the USA, set to its default value.
What is the most likely outcome?

- ❌ 3% a. Most users will turn the setting off
- ❌ 5% b. Most users will turn the setting on
- ✅ 92% c. Most users will let the setting as is
- ❌ 0% d. Other

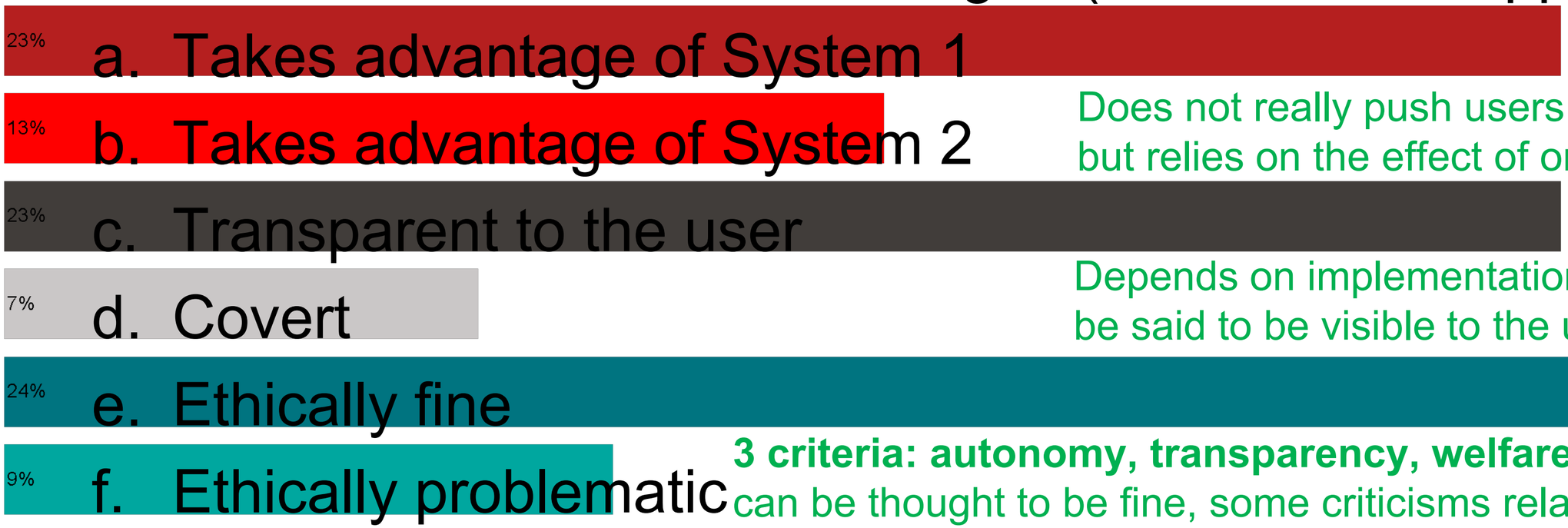
Nudges - 3

URL: ttpoll.eu
Session ID: cs290

In an effort towards more sustainability, the itinerary search in Noodle Maps now returns 2 itinerary options in the following order:

1) the most fuel-efficient but longest itinerary
2) the shortest but least fuel-efficient itinerary

What are the characteristics of this nudge? (select all that apply)



Does not really push users to reflect, but relies on the effect of order

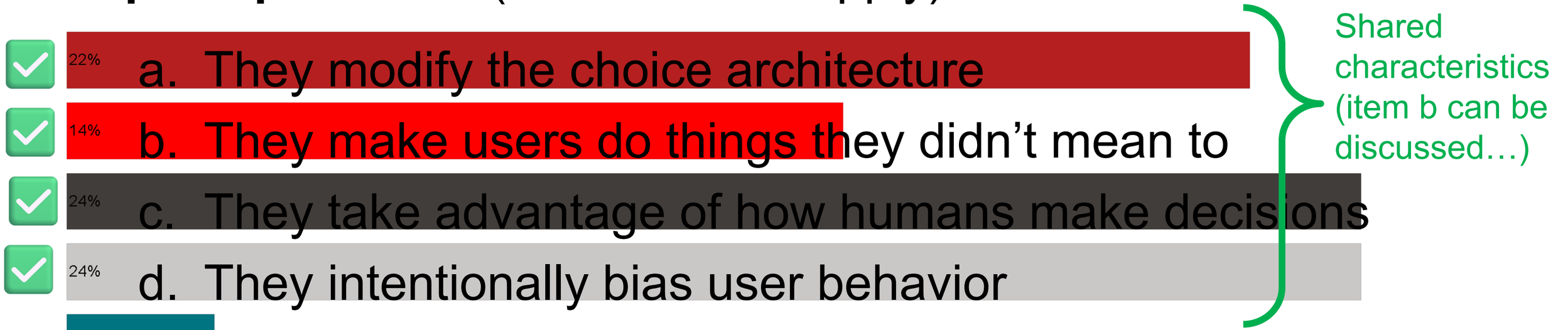
Depends on implementation, but can be said to be visible to the users

3 criteria: autonomy, transparency, welfare (this example can be thought to be fine, some criticisms relate to interfering with autonomy + benefit to community vs. individual user)

Deceptive patterns

URL: ttpoll.eu
Session ID: cs290

Which of the following are characteristics **shared** by nudges and **deceptive patterns**? (select all that apply)

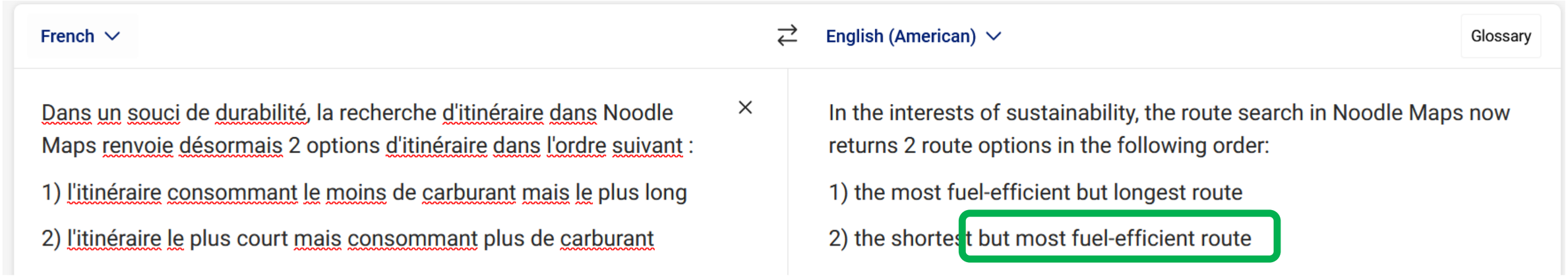


Characteristics of either nudges or deceptive patterns. [Apart from a few exceptions]

Translation

URL: ttpoll.eu
Session ID: cs290

Consider the following translation. What is the issue here?



❌ 8% a. Parity error

❌ 62% b. Factuality error

❌ 0% c. Measurement error

✅ 30% d. Faithfulness error

The response is erroneous compared to the input (prompt).
(Here since “Noodle Maps” does not exist, it cannot really be argued that the error relates to a known fact i.e. it is not a Factuality Error)

Case studies

Where to find the cases?

1. Go to **moodle**

2. Find the **link to the case studies** for today: **Empowerment 1**

3. Download:
   - The **instruction sheet**
   - 1 cheatsheet: Digital Ethics Canvas

+ From previous chapters, you will need:
- Value Analysis (3 - Fairness 1)

Digital Ethics Canvas

👉 Case 2

Instructions

- Read the context description
- Fill out the canvas:
  1. Evaluate the **benefits**
  2. Evaluate the **risks**
     a. Type of risk (i.e., description: what is the risk about?)
     b. Level of risk = Probability x Severity
  3. Reduce the risks: work on **mitigation**

Benefits

Which benefits do you identify for the app (think about different stakeholders)?

👉 1 post = 1 benefit

See posts on SpeakUp

Post your ideas:
https://speakup.epfl.ch
Room key: **41469**

Risks

Which risks do you identify?

👉 1 post = 1 risk

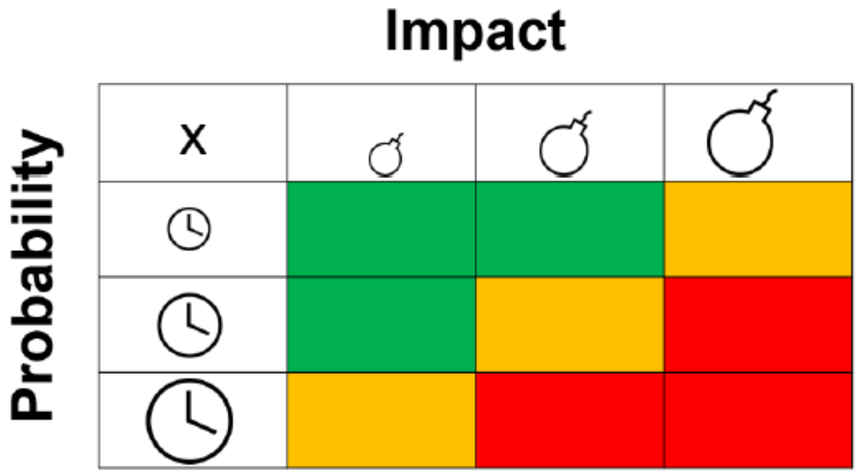
- **Name** of the ethical lens (welfare, fairness, autonomy, privacy, sustainability)
- **Description** of the risk

See posts on SpeakUp

Post your ideas:
https://speakup.epfl.ch
Room key: **14330**

Instructions

- Read the context description
- Fill out the canvas:
  1. Evaluate the **benefits**
  2. Evaluate the **risks**
     a. Type of risk (description: what is it about)
     b. Level of risk = Probability x Severity
  3. Reduce the risks: work on **mitigation**



Green: Low risk
Yellow: Medium risk
Red: High risk

Evaluating the level of risk - 1

URL: ttpoll.eu
Session ID: cs290

Consider the following Privacy risk: **“Tracks personal app usage”**
How would you evaluate the level of this risk in terms of probability and severity of impacts?
(select 2 options: 1 for probability, 1 for severity)

3% a. Probability: low

13% b. Probability: medium

35% c. Probability: high

3% d. Severity: low

34% e. Severity: medium

13% f. Severity: high

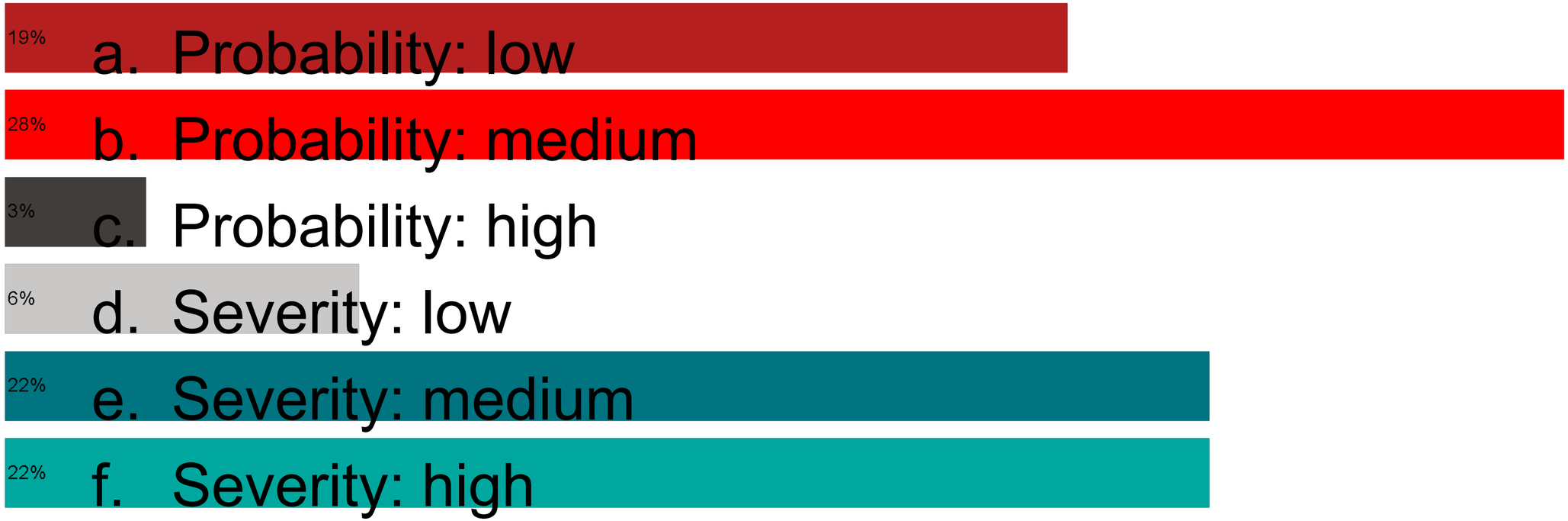
Qualitative evaluation: you need to provide a **justification** to support your evaluation of the probability/severity (including hypotheses you make on how the app is implemented), such as:

- Probability High: the app relies on tracking, so it necessarily is going to happen
- Severity High: tracking means collecting behavioral data over time, which can be considered sensitive (may disclose personal info)

Evaluating the level of risk - 2

URL: ttpoll.eu
Session ID: cs290

Consider the following Welfare risk: **“Excessive reminders could lead to stress or anxiety”**. How would you evaluate the level of this risk in terms of probability and severity of impacts?
(select 2 options: 1 for probability, 1 for severity)



Instructions

- Read the context description
- Fill out the canvas:
  1. Evaluate the **benefits**
  2. Evaluate the **risks**
     a. Type of risk (description: what is it about)
     b. Level of risk = Probability x Severity
  3. Reduce the risks: work on **mitigation**

Assessing software as we assess medicines

- **Expected benefits:**
  treat mild to moderate pain and fever

- **Potential risks:**
  - Hematological and lymphatic system disorders:
    rare (≥1/10'000, <1/1’000)
  - Skin and subcutaneous tissue disorders:
    occasional (≥1/1'000, <1/100)

Type, Severity, *Probability*

compendium.ch DAFALGAN Dolo cpr 500 mg (ec 12/23) UPSA Switzerland AG

Overall debriefing of the strategy

Designed to:

- Help software engineers think about a **range** **of ethical issues**
- Evaluate the **level** of ethical risks
  - Qualitative evaluation (what we have done)
  - Quantitative evaluation e.g., using metrics and experimental studies
- **Adapt** their designs based on the ethical risks

Value Analysis
(review from Fairness 1)

Values manifested in the product

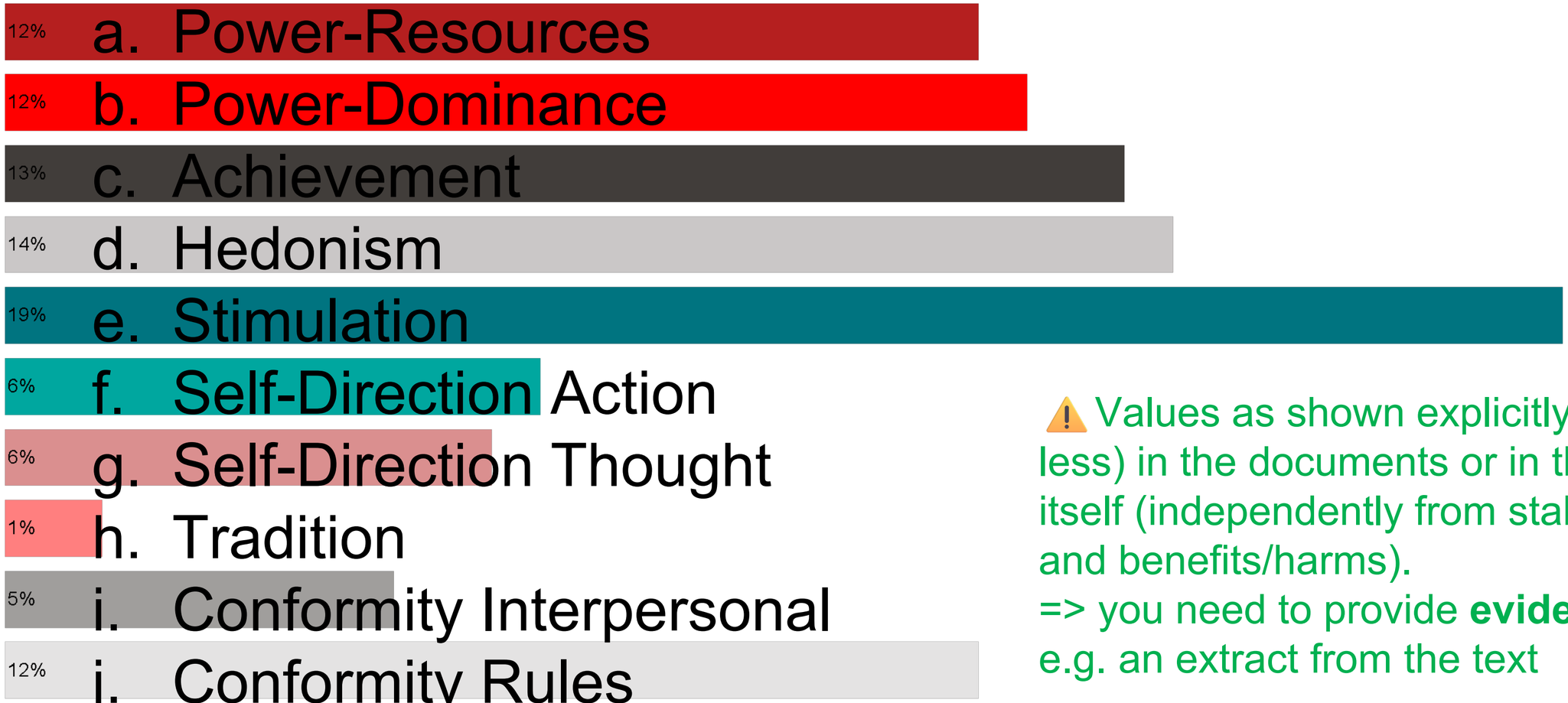
Individually,

- Read the information we have extracted from 4 sources
- Fill out the Artifact Values questionnaire in the appendix for the “For You” section of TikTok
  - Indicate **which values are visible**
  - Indicate how they **manifest**

Values in TikTok “For You”

URL: ttpoll.eu
Session ID: cs290

Select the values you have identified among these:



⚠️ Values as shown explicitly (more or less) in the documents or in the software itself (independently from stakeholders and benefits/harms).
=> you need to provide **evidence** for it, e.g. an extract from the text

Stakeholder values

- Identify a **stakeholder** of TikTok for whom there is:
  - One value-based benefit
  - One value-based harm
- Briefly describe the **profile** of this stakeholder (1 short paragraph)
- Fill out the **table:**

Stakeholder	Key values	Manifested	Benefits	Harms

Stakeholder

Which stakeholder(s) did you identify?

👉 1 post = 1 stakeholder (brief description)

See posts on SpeakUp

Post your ideas:
https://speakup.epfl.ch
Room key: **23109**

Value tensions

Draw the **value-based tension map** corresponding to the table:

1. Place the values
2. Add the stakeholder(s) concerned and indicate if it's a harm (red/"harm") or benefit (green/"benefit")

Do you identify **value-based tensions**?
Add lines to indicate the value tensions i.e., harm vs. benefit
They can be:

- Between different stakeholders or for the same stakeholder
- Between values or for the same value

“Dark” Patterns

Exploring “Dark” Patterns

Visit the following website: https://neal.fun/dark-patterns/
Engage with the various **examples** presented. Pay close attention to how these patterns affect your decision-making process.

1. What **emotions or reactions** did you experience when encountering the patterns presented on the website?
2. Have you **encountered similar or different types of patterns** in apps or websites you use? What were they, and can you think of other “dark” patterns that could be implemented to manipulate users?
3. How do “dark” patterns conflict with the idea of **user empowerment**?

What’s next?

Next dates

	Monday (SG1)	**Tuesday** (Computer Rooms)
25 Nov – 1 Dec	Empowerment 1 cases	Graded Assignment 2
2 Dec – 8 Dec	Debriefing Graded 2	Empowerment 2 notebook
9 Dec – 15 Dec	Empowerment 2 cases	Conclusion & Q&A in **SG1**
16 Dec – 20 Dec	Final exam	---