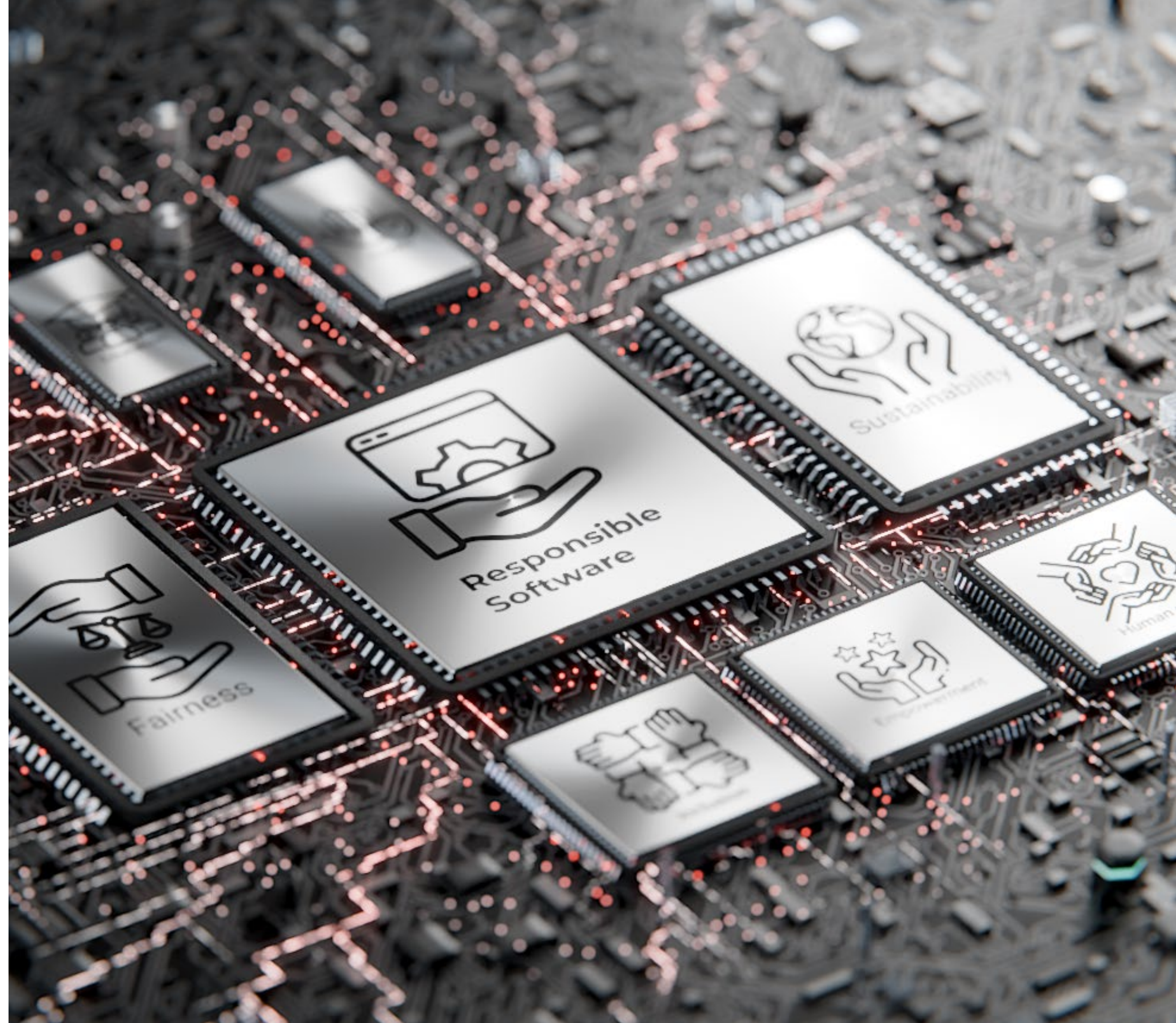


**EPFL**

**Empowerment 1  
Review & Case  
studies  
24 nov.**

Cécile Hardebolle

**Responsible  
Software**



# Agenda for today

---

1. Graded case
2. Final exam
3. Interactive review questions on Empowerment 1
4. Case studies:
  - a) Digital Ethics Canvas
  - b) Value analysis
  - c) “Dark” Patterns

# Graded Case

# Graded case

---

	Dates
<b>Release</b>	November 26
<b>Submission</b>	<b>December 9</b> at <b><u>23h59</u></b>

Grading
4% of total grade
Application of the Digital Ethics Canvas to a real product

- Reminder of the rules:
  - All documents allowed
  - Web search encouraged
  - No GenAI
  - In **groups of 3**, free choice

Choose your group on moodle **before November 26** (section "Graded Case")

# Final exam

# About the final exam

---

Select all the **correct** statements about the **final exam**:

- 20% a. It is in the winter exam session
- 1% b. It is on the last week of term
- 1% c. It includes programming
- 19% d. It includes case studies
- 15% e. It includes MCQs on the videos
- 0% f. All documents are allowed
- 21% g. Only one A4 paper of notes is allowed
- 5% h. The duration is 3 hours
- 16% i. The duration is 2 hours

URL: ttpoll.eu  
Session ID: cs290

# Format of the exam

---

## ■ In the Winter exam session

- Date: **21 January 2026, at 9h15**

- Duration: **2h**, except special arrangements

Note: this is a change compared to last year, duration was 1h30

## ■ Supervised in auditorium, individual

## ■ Pen and paper (QCM + case studies)

## ■ Closed book, **one A4 paper sheet** allowed (both sides, free format)

 The format of the mock test reflects exactly the format of the exam!

**Review questions**  
**Empowerment 1**

# Meditation app

URL: ttpoll.eu  
Session ID: cs290

ZenPath is an app dedicated to mental well-being that offers guided meditation sessions online. To reduce user dropout, they decide to display a popup after a user skips two sessions where the “Resume Today!” button is preselected.

What type of nudging technique is most likely used here?



# Use of data

URL: ttpoll.eu  
Session ID: cs290

← Back

## Data for Generative AI Improvement

Can LinkedIn and its affiliates use your personal data and content you create on LinkedIn to train generative AI models that create content?

Use my data for training content creation AI models On

This setting controls the training of generative AI models used to create content. When this setting is on LinkedIn and its affiliates may use your personal data and content you create on LinkedIn for that purpose. [Learn more.](#)

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

- 0% a. Most users will turn the setting off
- 0% b. Most users will turn the setting on
- 100% c. Most users will let the setting as is
- 0% d. Most users will change the setting

# Navigation app

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)

- 21% a. Takes advantage of System 1
  - 9% b. Takes advantage of System 2
  - 34% c. Transparent to the user
  - 2% d. Covert
  - 32% e. Ethically fine
  - 2% f. Ethically problematic
- 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 vs nudges

URL: ttpoll.eu  
Session ID: cs290

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

- 22% a. They modify the choice architecture
  - 16% b. They make users do things they didn't originally mean to
  - 26% c. They take advantage of how humans make decisions
  - 25% d. They intentionally bias user behavior
  - 0% e. They restrict choices
  - 0% f. They benefit users
  - 10% g. They benefit another party
  - 1% h. They make users lose track of time
- Shared characteristics (item b can be discussed...)
- Characteristics of either nudges or deceptive patterns. [Items g and f can lead to confusion and should be reframed]

# E-commerce platform

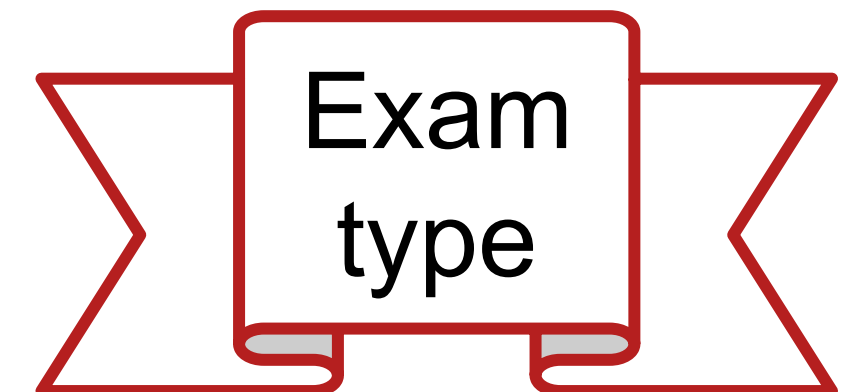
URL: ttpoll.eu

Session ID: cs290

The e-commerce platform Shine would like to implement new features to improve the experience of its various categories of users. Here is the list of envisaged features.

Which of them best matches the definition of a deceptive pattern?

- 17% a. Personalize style recommendations based on past browsing
- 8% b. Display user-provided past purchase data to recommend sizes
- 75% c. Register users to a ShineClub membership trial on checkout
- 0% d. Provide downloadable QR codes for the free return of items



# Translation

URL: ttpoll.eu  
Session ID: cs290

Consider the following translation. What is the issue here?

French ▾ ↔ English (American) ▾ Glossary

Dans un souci de durabilité, la recherche d'itinéraire dans Noodle Maps renvoie désormais 2 options d'itinéraire dans l'ordre suivant :

- 1) l'itinéraire consommant le moins de carburant mais le plus long
- 2) l'itinéraire le plus court **mais consommant plus de carburant**

×

In the interests of sustainability, the route search in Noodle Maps now returns 2 route options in the following order:

- 1) the most fuel-efficient but longest route
- 2) the shortest **but most fuel-efficient route**

- ✗ 0% a. Parity error
- ✗ 56% **b. Factuality error**
- ✗ 0% c. Measurement error
- ✓ 44% 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 **courseware**
  2. Find **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**

# Risks

---

## Which risks do you identify?

👉 1 post = 1 risk

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

The description needs to make explicit why the risk is classified under the ethical lens you have chosen

**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**

		Severity		
		low	mid	high
Probability	low	low	low	mid
	mid	low	mid	high
	high	mid	high	high

# 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)

4% a. Probability: low

18% b. Probability: medium

32% c. Probability: high

14% d. Severity: low

18% e. Severity: medium

14% 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](http://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)

34% a. Probability: low

17% b. Probability: medium

0% c. Probability: high

3% d. Severity: low

24% e. Severity: medium

21% f. Severity: high

# 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**

# Risk mitigation

---

Consider the following Sustainability risk:

**“Background processing may contribute to a higher footprint”.**

How could you mitigate this risk?

👉 1 post = 1 mitigation possibility

NOT DONE in class

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: **70741**



# 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

# Assessing software as we assess medicines

- **Expected benefits:**

treat mild to moderate pain and fever

- **Potential risks:**

- Hematological and lymphatic system disorders:  
*rare ( $\geq 1/10'000$ ,  $< 1/1'000$ )*
- Skin and subcutaneous tissue disorders:  
*occasional ( $\geq 1/1'000$ ,  $< 1/100$ )*

		Severity		
		low	mid	high
Probability	low	low	low	mid
	mid	low	mid	high
	high	mid	high	high



# Example: effects table

Effect	Short Description	Unit	Placebo N=131	Lenvatinib N=261	Uncertainties/ Strength of evidence	References
<b>Favourable Effects</b>						
PFS	Median time from randomization to progression or death	Months	3.6 (2.2, 3.7)	18.3 (15.1, NE)	Consistent and significant effect on PFS with a HR of 0.21 (0.14, 0.31)	See 'clinical efficacy' section
OS	Median time from randomization to death of any cause	Months	NE (20.3, NE)	NE (22.0, NE)		
<b>Unfavourable Effects</b>						
Hypertension	Incidence of grade 3 or 4 events	%	3.8	42.9	The association with these risks is further supported by the analysis in the extended safety population  The chosen dose of 24 mg is of special concern since it is associated with important levels of dose reductions and interruptions	Numbers presented were taken from the DTC Randomized Safety Set (see 'clinical safety' section)
Proteinuria	Incidence of grade 3 or 4 events	%	0	10.7		
Liver events	Incidence of grade 3 or 4 events	%	1	10.7		
Hypocalcaemia	Incidence of grade 3 and 4 events	%	0	4.9		
Diarrhoea	Incidence of grade 3 and 4 events	%	0	9.2		
Fatal AE	Incidence of treatment-related fatal AE	%	0	2.3	Uncertainties linked to low numbers	

Abbreviations: AE: adverse event; HR: hazard ratio; NE: not estimable; OS: overall survival; PFS: progression-free survival  
data cut-off dates : efficacy - PFS: 15 November 2013, OS:15 June 2014 ;safety: 25 March 2014.

# **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”

---

## Which values did you identify?

- 👉 1 post = 1 value
  - Full name of the value
  - **Manifestation** of the value in “For You”  
(excerpt from document)

Make sure to cite source

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: **57438**



# 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
  - Name / brief description
  - One value-based **benefit**
  - One value-based **harm**

NOT DONE in class

**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? If yes, what were they?
3. In **what kind of situation** might a software engineer find themselves implementing similar “dark” patterns in a software interface?

- Business pressure
- Doing the same as others
- Inattention
- Normalization in competitive industry

**What's next?**

# Next dates

---

	Monday (STCC Cloud C)	Tuesday (Computer Rooms)
24 Nov – 30 Nov	Empowerment 1 cases	Graded Assignment 2
1 Dec – 7 Dec	Debriefing Graded 2	Empowerment 2 notebook
8 Dec – 14 Dec	Empowerment 2 cases	Graded Case

## Debriefing” =

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

We will have a conclusion & Q&A session on the last week of the semester, I will ask you to post your questions in advance