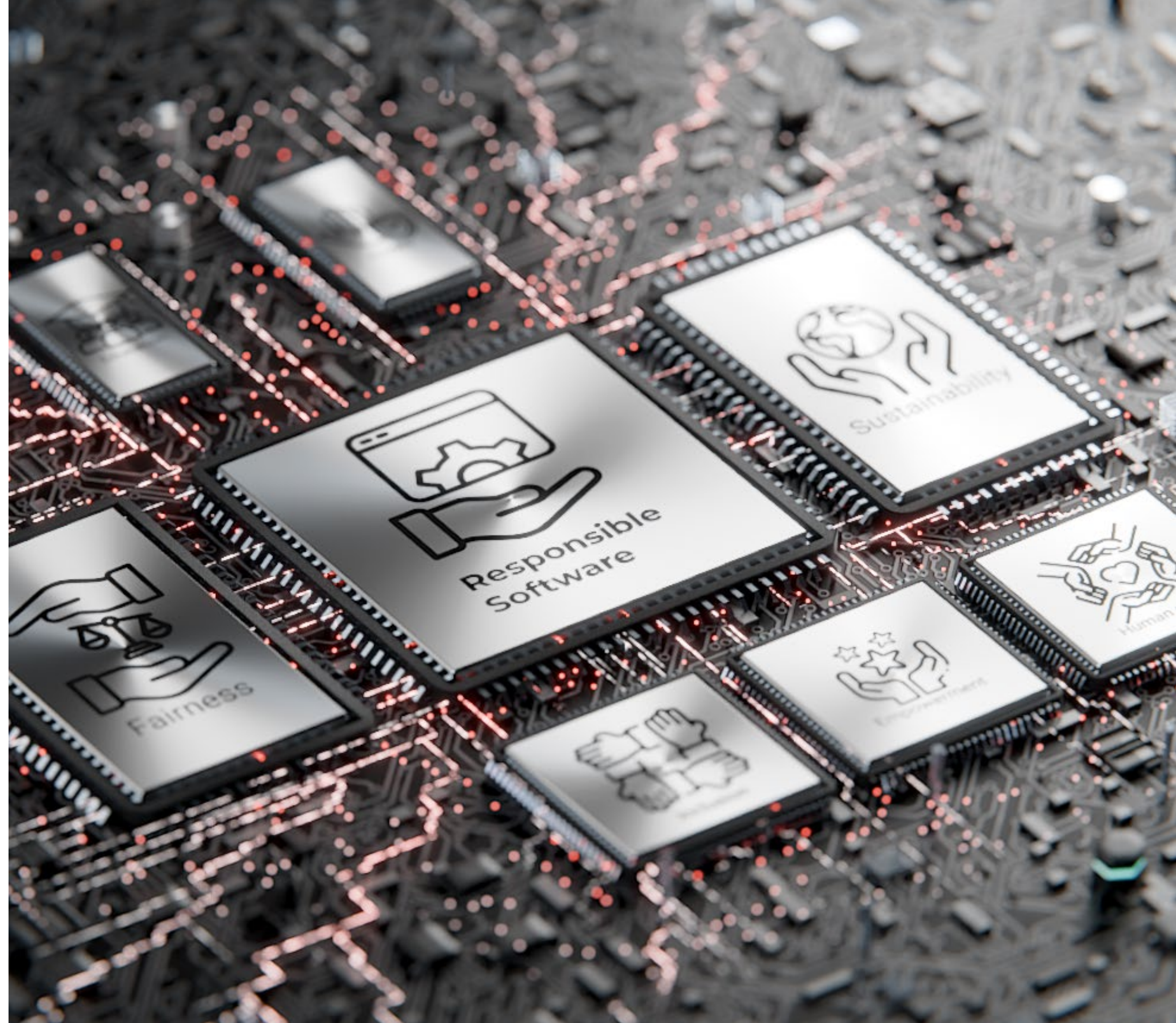


**EPFL**

**Sustainability 2  
Review & Case  
studies  
17 nov.**

Cécile Hardebolle

**Responsible  
Software**



# Agenda for today

---

1. Information about the Graded Case
2. Information about the Graded Assignment 2
3. Interactive review questions on Sustainability 2
4. Case studies:
  - a) Causal loop diagram
  - b) Ethical decision making

# 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")

# **Graded Assignment 2**

# Graded assignment 2

---

	Dates
Release	November 21
Submission	November 25 at <u>23h59</u>

Grading
8% of total grade
Coding questions + reflection questions

Topics
Safety 1 & 2
Fairness 1 & 2
Sustainability 1 & 2

## ■ Reminder of the rules:

- No GenAI
- No group work
- Use noto

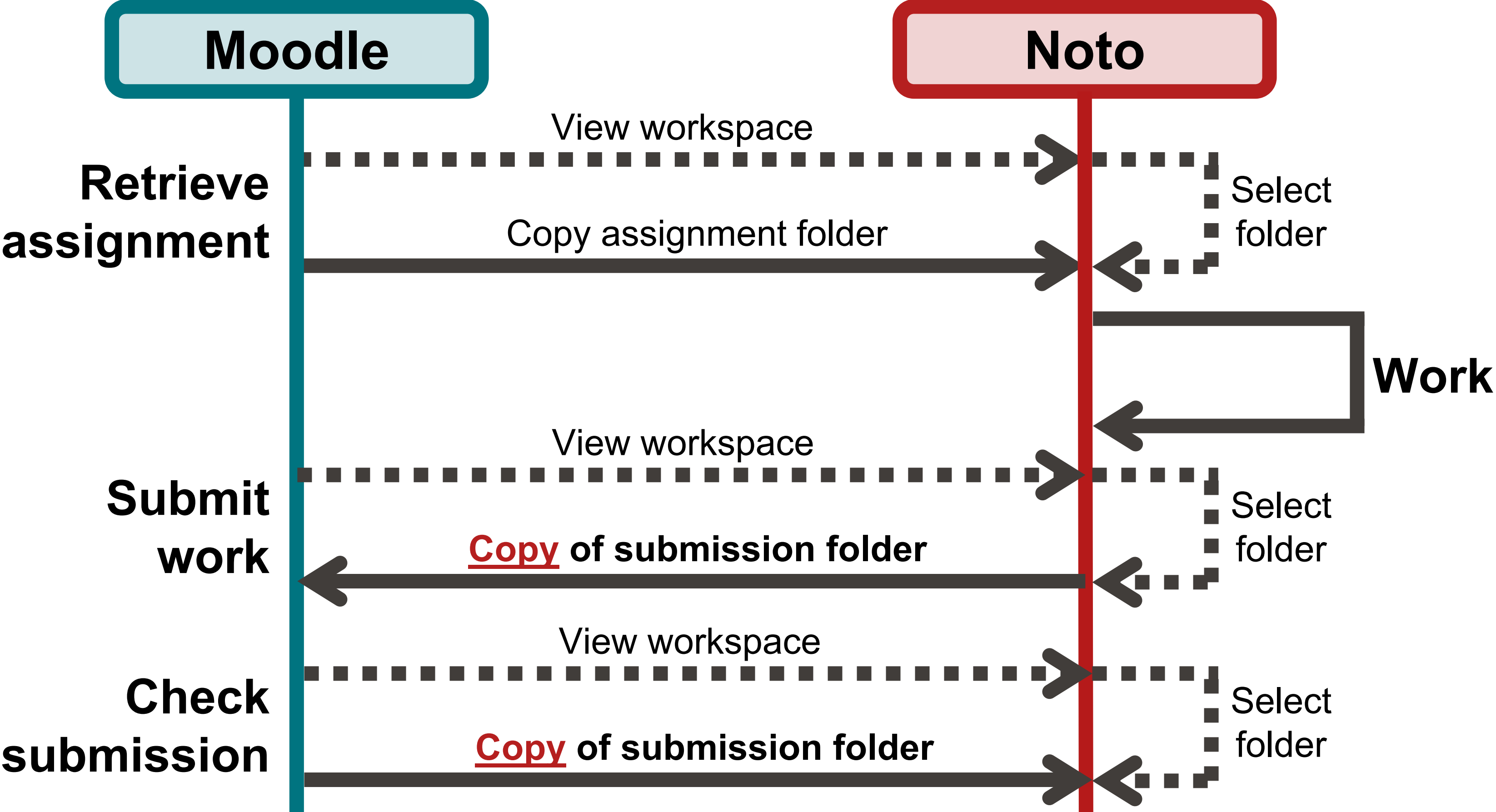
## ■ Support session with assistants on Tuesday, November 25, 10h-12h

👉 They do not have the solution!

They can help you debug or help with tech issues

They can help you submit your work on moodle

# Assignment: process on moodle & noto



# Checking what you have submitted

## 1. Go to moodle:

- Find the assignment “Test the moodle assignment” [**Fairness 1 section**]
- Click on it 🖱️ **View submission**
- Select **where to save this copy of your submission** on noto

### Submission status

Submission status	Submitted for grading
Grading status	Not graded
Time remaining	Assignment was submitted 6 days 15 hours early
Last modified	Friday, 14 November 2025, 08:32
Submission comments	▶ Comments (0)
Jupyter notebooks	<a href="#">View submission</a>

## 2. Go to noto:

- Find the folder where you have put the copy of your submission
- You should see a **new folder** with a long name:  
“test-assignment\_course18548\_student218401\_submission1276184\_Cécile\_Hardebolle”

# Important advice

---

## Logistics:

- Make sure to **submit your first work early**
  - We have been able to detect **submission issues** before the deadline
  - In case of **last-minute hitch**, there is still one submission we can grade, even if it is incomplete
  
- Make sure to **check your submission!**

## Assignment questions:

- **Read carefully the questions** and instructions
- Perform some sanity checks on your answers

**Review questions**  
**Sustainability 2**

# The footprint of training - 1

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

What are the 3 most important elements in the carbon footprint of ML training?

Rank them **by decreasing impact** (i.e. most impactful first) :

- 27% a. The training time
  - 15% b. The power consumption of the CPU
  - 16% c. The power consumption of the GPU
  - 22% d. The PUE of the datacenter
  - 20% e. The carbon intensity of the electricity
- The power consumed by CPUs is usually negligible compared to GPUs
- The multiplying factor from carbon intensity is usually higher than that of the PUE

# The footprint of training - 2

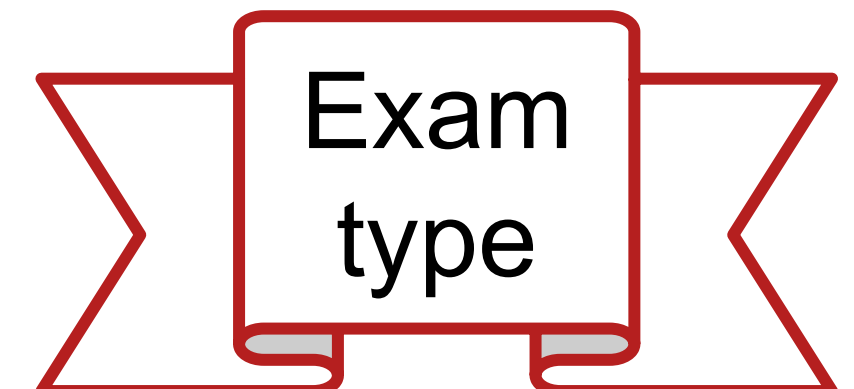
URL: ttpoll.eu  
Session ID: cs290

Let's consider the training of the model SupChat-7B. The computing node has 2 GPUs of the model Nvidia A100 80GB, which consume 400W each. Our datacenter, which has a PUE of 1.2, is located in Germany (carbon intensity: 381g CO<sub>2</sub>e / kWh).

The training time is 80 000 hours of total GPU computation time.

**What is the carbon footprint for the training of SUPMOD-7B?**


- 4% a. 14,63 tons CO<sub>2</sub>e
- 41% b. 29,26 tons CO<sub>2</sub>e
- 19% c. 14 630,4 tons CO<sub>2</sub>e
- 37% d. 29 260,8 tons CO<sub>2</sub>e



# Absolute time vs. GPU time

---

To compute energy consumption (kWh) we need:

- Instantaneous power consumption (kW)
- Time (h)  **there are two ways to report time:**
  - ◆ Absolute time = absolute duration, where N GPUs are run in parallel
  - ◆ GPU time = total GPU use time, as if only 1 GPU were used (i.e. the number of GPUs is already factored in)

$$time_{GPU} = time_{abs} \times n_{GPUs}$$

 Depending on how the time is reported you will use one of these formulas:

$$Energy = time_{GPU} \times power_{1GPU}$$

*or*

$$Energy = time_{abs} \times n_{GPUs} \times power_{1GPU}$$

# The footprint of training - 2

---

1. The computation time is provided as “GPU time” i.e. the number of GPUs is already factored in.
2. The power consumption is given in W we need to convert to kW to be able to use the carbon intensity (g/kWh)
3. Finally we need to convert from grams to tons

$$Footprint_{training} = time_{GPU} \times power_{1GPU} \times PUE \times CarbonIntensity$$

$$Footprint_{training} = 80\,000 \times 0,400 \times 1,2 \times 381 \times \frac{1}{1\,000\,000}$$

$$Footprint_{training} = 14,63 \text{ tons } CO_2e$$

# The footprint of inference - 1

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

What are the 3 most important elements in the carbon footprint of ML inference?

Rank them **by decreasing impact** (i.e. most impactful first) :



33%

a. The number of user queries



27%

b. The electricity consumed per query

15%

c. The PUE of the datacenter



24%

d. The carbon intensity of the electricity

# The footprint of inference - 2

---

The model SupChat-7B is now deployed in production. It is hosted on the same computing node with 2 GPUs of the model Nvidia A100 80GB, which consume 400W each. Our datacenter, which has a PUE of 1.2, is located in Germany (carbon intensity: 381g CO<sub>2</sub>e / kWh). Our model is able to serve 120 tokens per second ~~of computation time~~. It has an average of 2000 users daily and generates an average of 5000 tokens per user per day.

Confusing: speed is given in absolute time!

**What is the carbon footprint of 1 day of inference?**

1. What is the total GPU computation time used over 1 day (in h)?
2. What is the power consumed by the model for inference (in W)?
3. What is the total electricity consumed over 1 day (in kWh)?
4. What is the carbon footprint over 1 day (in kg CO<sub>2</sub>e)?

# The footprint of inference - 2

---

1. The **GPU computation time** used over 1 day can be obtained from the **speed** of the model (given in absolute time), the **total number of tokens served per day** and the **number of GPUs** + you need to convert from seconds to hours

$$time_{GPU} = \frac{nbusers_{perday} \times nbtokens_{peruser_{perday}}}{modelspeed} \times n_{GPUs} \times \frac{1}{3600}$$

$$time_{GPU} = \frac{2000 \times 5000}{120} \times 2 \times \frac{1}{3600}$$

$$time_{GPU} = 46,30 \text{ h}$$

Sanity check: with 2 GPUs in parallel, 46,30 h of GPU time is taking 23,15h of absolute time i.e. a bit less than 1 day.

# The footprint of inference - 2

---

2. Since we are working with GPU time, we need to use the instantaneous power consumption of 1 GPU (in kW):

$$Power_{inference} = 0,400 \text{ kW}$$

3. To get the electricity consumed we multiply GPU time with the instantaneous power consumption, we multiply by the PUE to account for the overhead electricity consumed by cooling:

$$Electricity_{inference} = time_{GPU} \times Power_{Inference} \times PUE$$

$$Electricity_{inference} = 46,30 \times 0,400 \times 1,2$$

$$Electricity_{inference} = 22,22 \text{ kWh}$$

# The footprint of inference - 2

---

4. To get the carbon footprint in kg CO<sub>2</sub>e we multiply the electricity consumed by the carbon intensity, then we scale to kg (i.e. divide by 1000)

$$Footprint_{inference} = Electricity_{inference} \times CarbonIntensity$$

$$Footprint_{inference} = \frac{22,22 \times 381}{1000}$$

$$Footprint_{inference} = 8,47 \text{ kg CO}_2\text{e}$$

# Total carbon footprint

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

We have obtained the carbon footprint of SupChat-7B at training and at inference time. What is its total carbon footprint?

- 4% a. Training
- 0% b. Inference
- 0% c. Training x Inference
- 0% d. Inference – Training
- 89% e. Training + Inference
- 7% f. Other We also need to add the footprint associated with embodied emissions (i.e. hardware manufacturing mainly)



# Hardware renewal

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

Session ID: cs290

We want to optimize the energy consumption of SupChat-7B at inference time. We decide to upgrade our hardware platform and to replace our A100 GPUS with H100 GPUs. The H100 are 4 times more performant than the A100 in terms of computation speed. Their power consumption is 700W at maximum use.

What effect(s) are we likely to observe (select all that apply)?

- 24% a. A decrease in the energy consumption
- 28% b. An increase in the energy consumption  Rebound effect (+ increased cooling needs)
- 21% c. A decrease in the overall carbon footprint
- 28% d. An increase in the overall carbon footprint  Embodied emissions!

# Water Usage Effectiveness

URL: ttpoll.eu  
Session ID: cs290

The datacenter hosting SupChat-7B consumes an average of 1 MW. This means annually a total of 8 760 MWh of electricity. It consumes approximately 15.8 million liters of water each year. What is the WUE of the datacenter (onsite only)?

0% a. 0,18

6% b. 0,55

67% c. 1,8

22% d. 18,03

6% e. 55,44

$$WUE = \frac{15\,800\,000}{8\,760\,000}$$
$$WUE = 1,8 \text{ L/kWh}$$

- The WUE is expressed in L/kWh  
-> Need to scale from MWh to kWh
- The reference value for the WUE is the closest possible to 0 (i.e. no water consumption)

Exam  
type

# **Case studies**

# Where to find the cases?

---

1. Go to moodle

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

3. Download:

- The **instruction sheet**
- 1 cheatsheet: Ethical Decision Making

**+ From previous chapters**, you will need:

- Causal Loop Diagram (2 - Safety 2)

# Where to find the cases?

---

1. Go to **courseware**
  2. Find **the case studies** for today: **Sustainability 2**
  3. Download:
    - The **instruction sheet**
    - 1 cheatsheet: Ethical Decision Making
- + From previous chapters**, you will need:
- Causal Loop Diagram (2 - Safety 2)

# **Causal Loop Diagram**

(review from Safety 2)

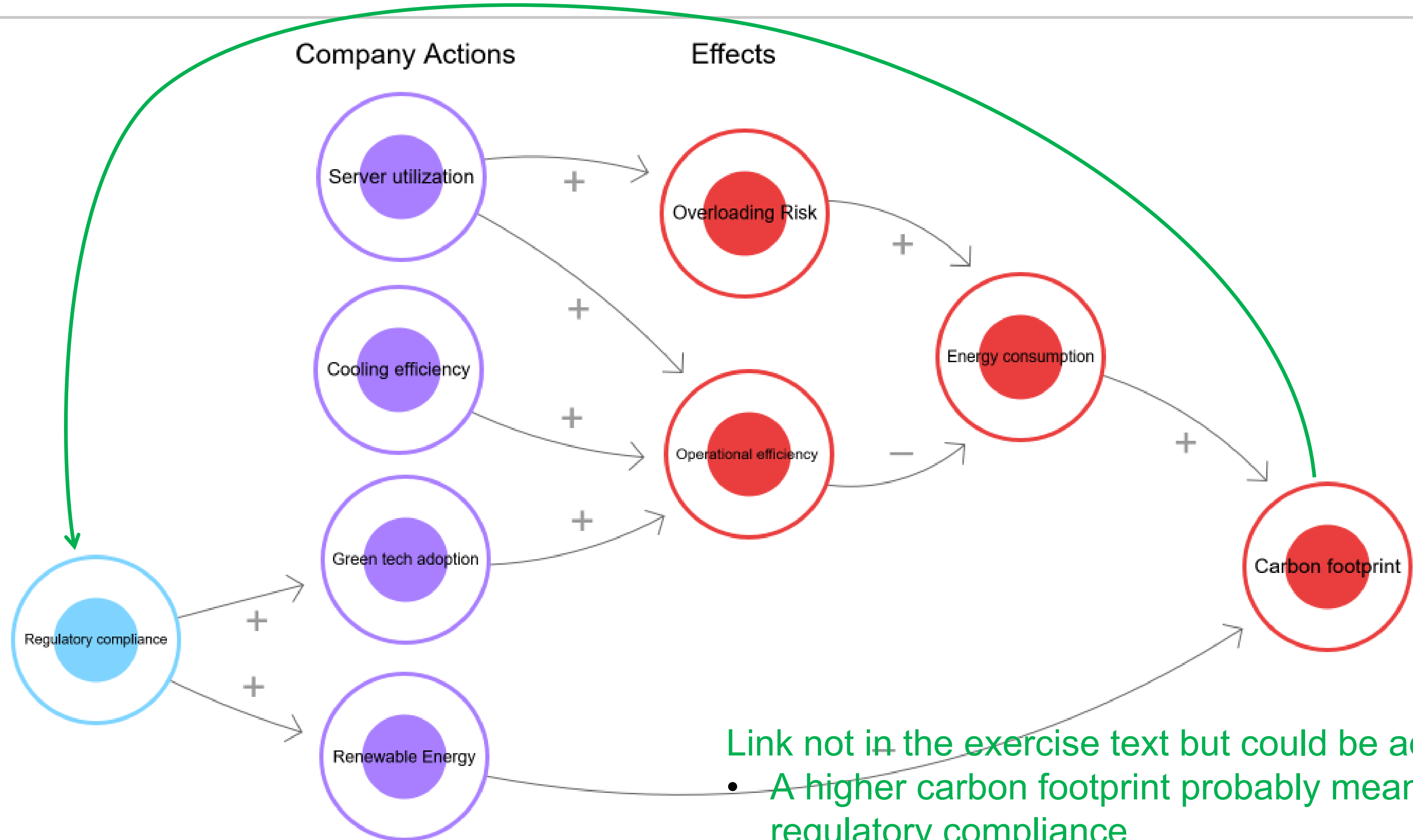
# Instructions – Creating the diagram

---

- Read the context description
- Build a causal loop diagram with the following variables (you need to add the causal links):
  1. energy consumption
  2. carbon footprint
  3. operational efficiency
  4. regulatory compliance
  5. renewable energy use
  6. adoption of green tech
  7. cooling efficiency
  8. server utilization
  9. risk of overloading servers

# Resulting diagram

<https://go.epfl.ch/su2-case1>



# Instructions – Using the diagram

---

Using the diagram, answer the following questions:

1. How would a significant improvement in the efficiency of existing cooling systems affect the energy consumption?
2. What happens to operational efficiency if the server utilization decreases?
3. What happens to the energy consumption if the carbon footprint increases?
4. Which type of actions from the company could decrease the overall carbon footprint?
5. What is the effect of stricter regulatory compliance on the adoption of greener technologies and renewable energy use?
6. What is an important downside of adopting greener technologies in a datacenter as a strategy to improve sustainability?

# Ethical Decision Making

# Instructions

---

- Read the context description
- Fill the table

<b>Ethical lens</b>	<b>Justification</b>	<b>Option chosen</b>
Rights		
Justice		
Utilitarian		
Common good		

# Ethical lens: Rights

---

Which option best respects the **rights** of all who have a stake?

👉 1 post =

- India / Switzerland
- Justification according to this lens

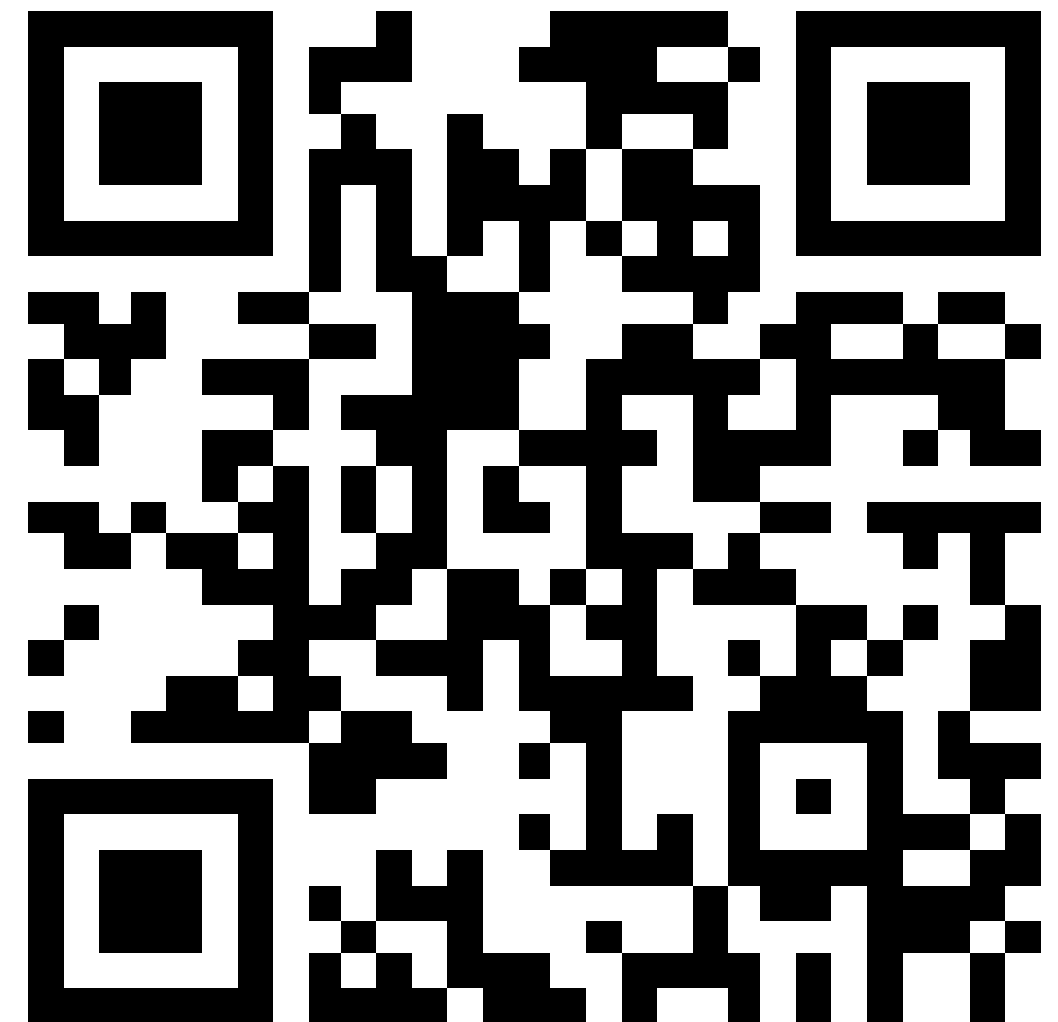
**To make clear in argument:**

- Which right(s) do you consider?
- For whom?

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: **68415**



# Ethical lens: Justice

---

Which option treats people **fairly**, giving them each what they are due?

👉 1 post =

- India / Switzerland
- Justification according to this lens

**To make clear in argument:**

- What do you make equal (e.g. benefits/burdens, punishment, reparations)?
- For whom?

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: **61431**



# Ethical lens: Utilitarian

---

Which option will produce the **most good** and do the **least harm** for as many stakeholders as possible?

👉 1 post =

- India / Switzerland
- Justification according to this lens

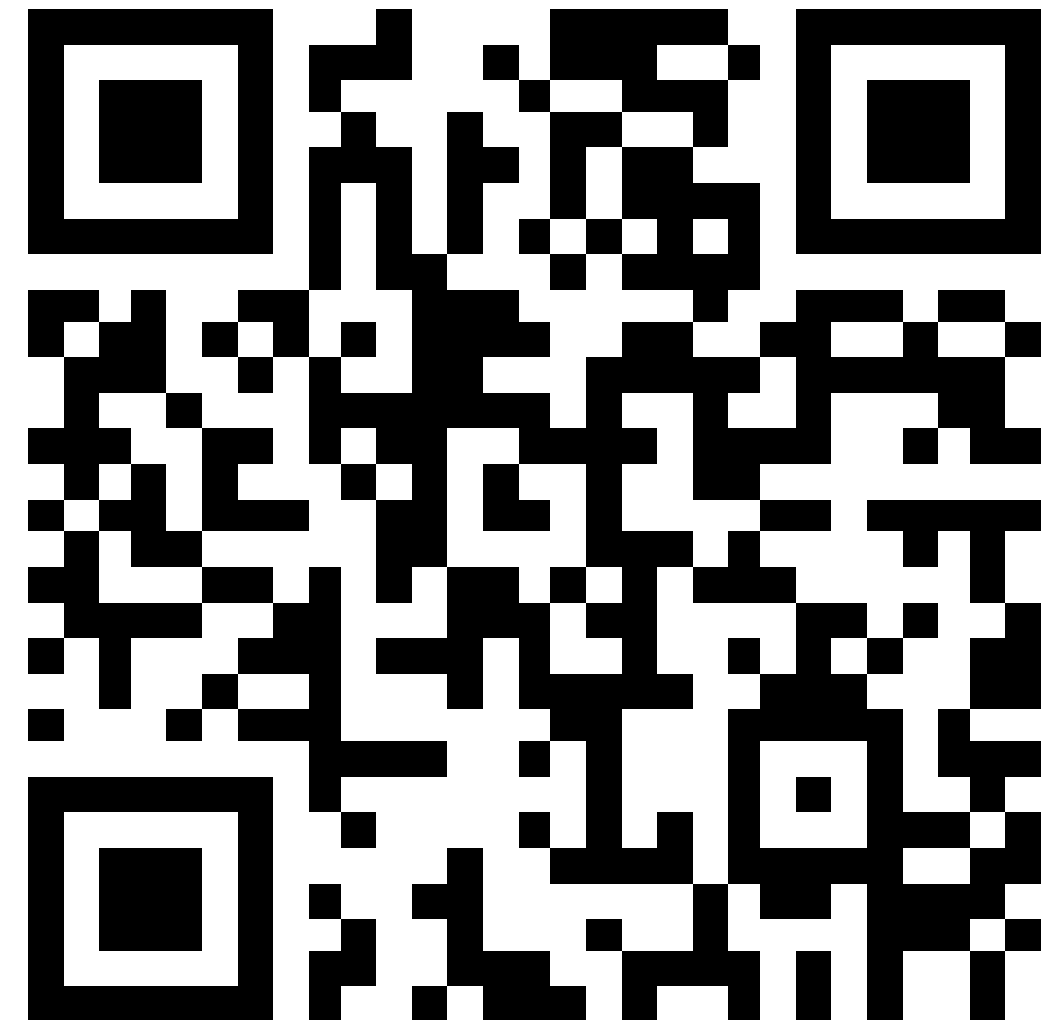
**To make clear in argument:**

- Which consequences do you count?
- How do you count them?

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: 44366



# Ethical lens: Common good

---

Which option best serves the **community** as a whole? And the most vulnerable?

👉 1 post =

- India / Switzerland
- Justification according to this lens

**To make clear in argument:**

- Which community(ies) do you consider?
- Which type of common good do you consider?

**Post your ideas:**

<https://speakup.epfl.ch>

Room key: **22633**



**What's next?**

# We start Empowerment 1!

---

Tomorrow, Tuesday 18: notebook on **automation bias**

By Monday 24:

- Watch **videos 7.1 to 7.4** + do the **quizzes**
- Finish the notebook  
(and any other leftover from previous weeks)

On Monday 24:

- Interactive questions on the theory
- Work on the **case studies together in class**