

EPFL

Fairness 1

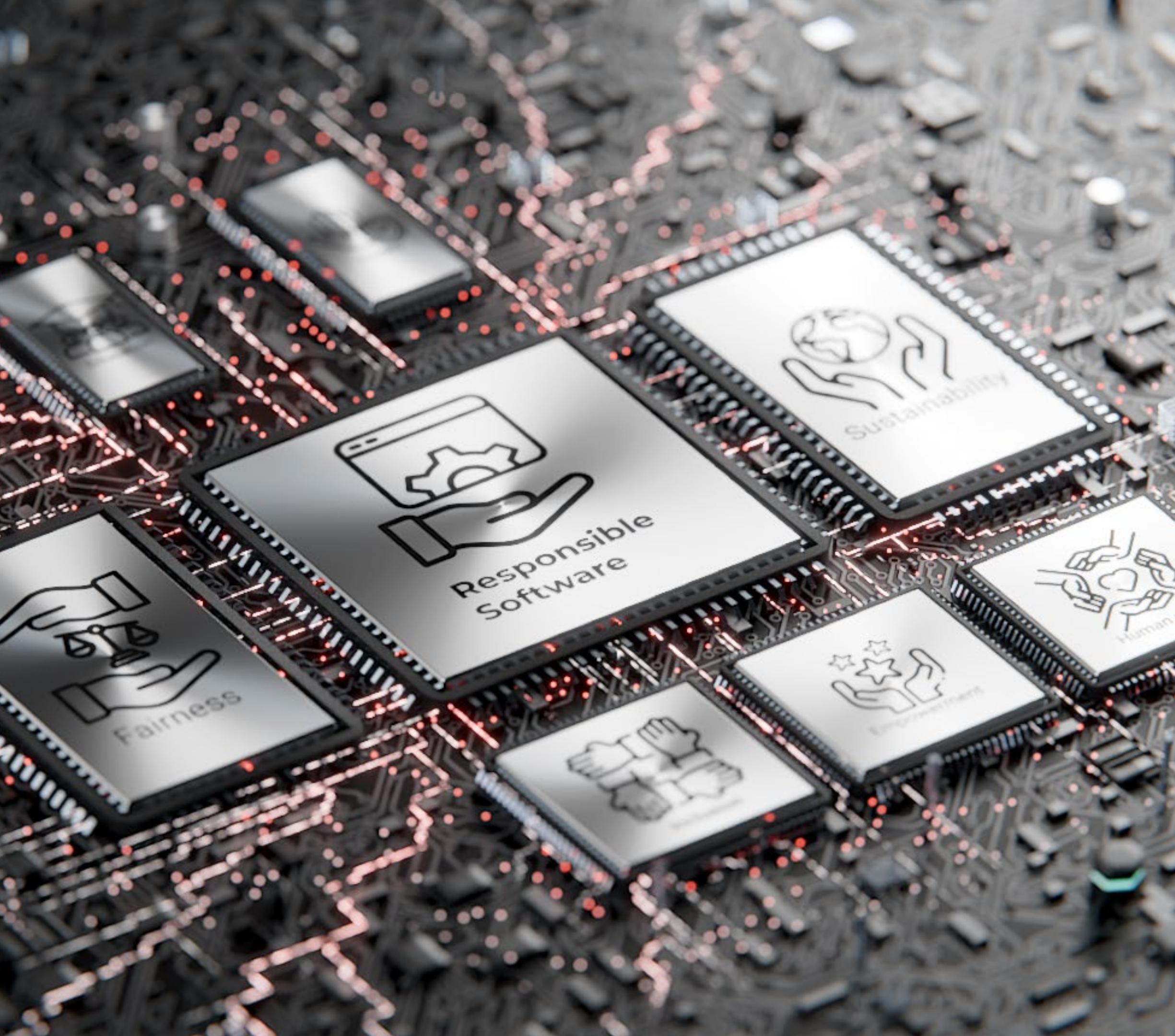
Review &

Case studies

7 oct.

Cécile Hardebolle

Responsible Software



Agenda for today

1. Graded assignment 1:
 - a) Logistics
 - b) Test the submission process
2. Interactive review questions on Fairness 1
3. Case studies:
 - a) Analyzing values
 - b) Inclusive design

Graded assignment 1: logistics

You will be **assigned a seat** in a computer room, communicated this week on moodle

- If you see an issue with the seat you are assigned, please contact me!
- Make sure to display your camipro card on your table

The assignment will start at **8h15** and you have **1h30** in total, except special arrangements

■ **Distribution of the assignment:**

you will find it directly into your noto workspace

■ **Submission of the assignment:**

you will use moodle  let's test that now!

Assignment submission test

1. Go to noto:

make sure you have your CS-290 notebooks in your workspace!

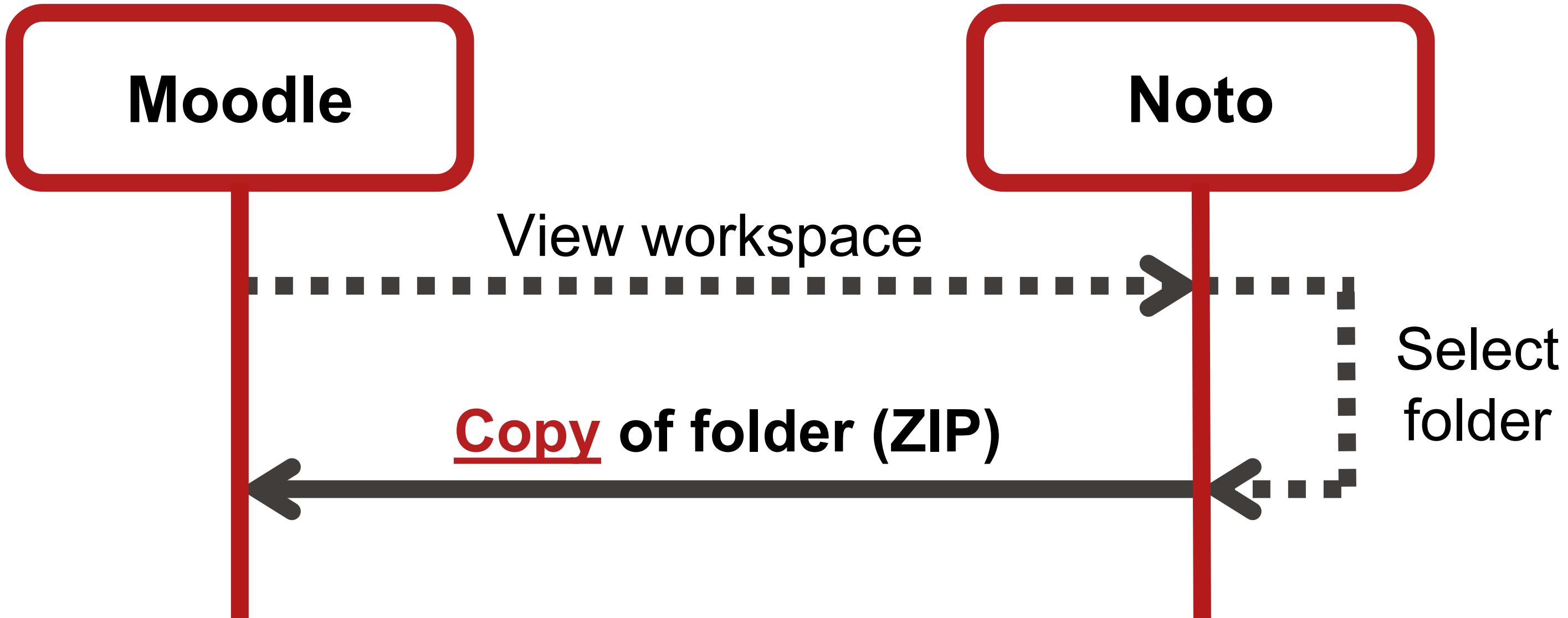
- If necessary, find the moodle link for Fairness 1

2. Go to moodle:

- Find the assignment “Test the submission process” in the Fairness 1 section
- Click on it, add a submission
- Select the folder Fairness 1 from noto
- Save changes

For those who have technical issues to access their notebooks on noto: a fix is under way, we will keep you informed with a post on Ed.

How does it work?



- The copy on moodle is the reference for grading
- After submission, changes on noto are NOT reflected on moodle **unless you resubmit**

Review questions

Fairness 1

Attributes - 1

URL: ttpoll.eu
Session ID: cs290

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

- 24% a. A sensitive attribute
- 10% b. A protected attribute
- 32% c. An observed variable
- 5% d. A latent variable
- 17% e. An objective representation of people
- 13% f. A subjective representation of people

What someone sees as “red” can be described as “auburn” by someone else, we would need to use a set of predefined categories, and some cases would be difficult to fit in (there is the same issue with skin color by the way)
-> the only way to make it “objective” would be to measure the color with colorimetry

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.



Article on SAT scores' correlation with race:
<https://www.brookings.edu/articles/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility/>

Bias - 1

URL: 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.

Article on biases in pothole reporting:
<https://www.autovision-news.com/industry/ai-bias-pothole-detection/>

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

- 23% a. Preexisting bias
- 10% b. Confirmation bias
- 27% c. Representation bias
- 37% d. Measurement bias
- 3% e. Automation bias



Case studies

Where to find the cases?

1. Go to moodle
2. Find the **link to the case studies** for today: **Fairness 1**
👉 this link will send you to courseware
(where you can find all the course material)
3. Download:
 - The **instruction sheet**
 - 2 cheatsheets: Value Analysis, Inclusive Design
 - 1 template

Value Analysis

Part 1 – Your own values

- 1. Fill the “own values questionnaire” (template / appendix 1.1)**
- 2. Discuss with your neighbor: what are your conclusions about the nature of values?**

Part 2 – Values manifested in products

Individually, fill out the Artifact Values Questionnaire for Google Maps (template / appendix 1.2):

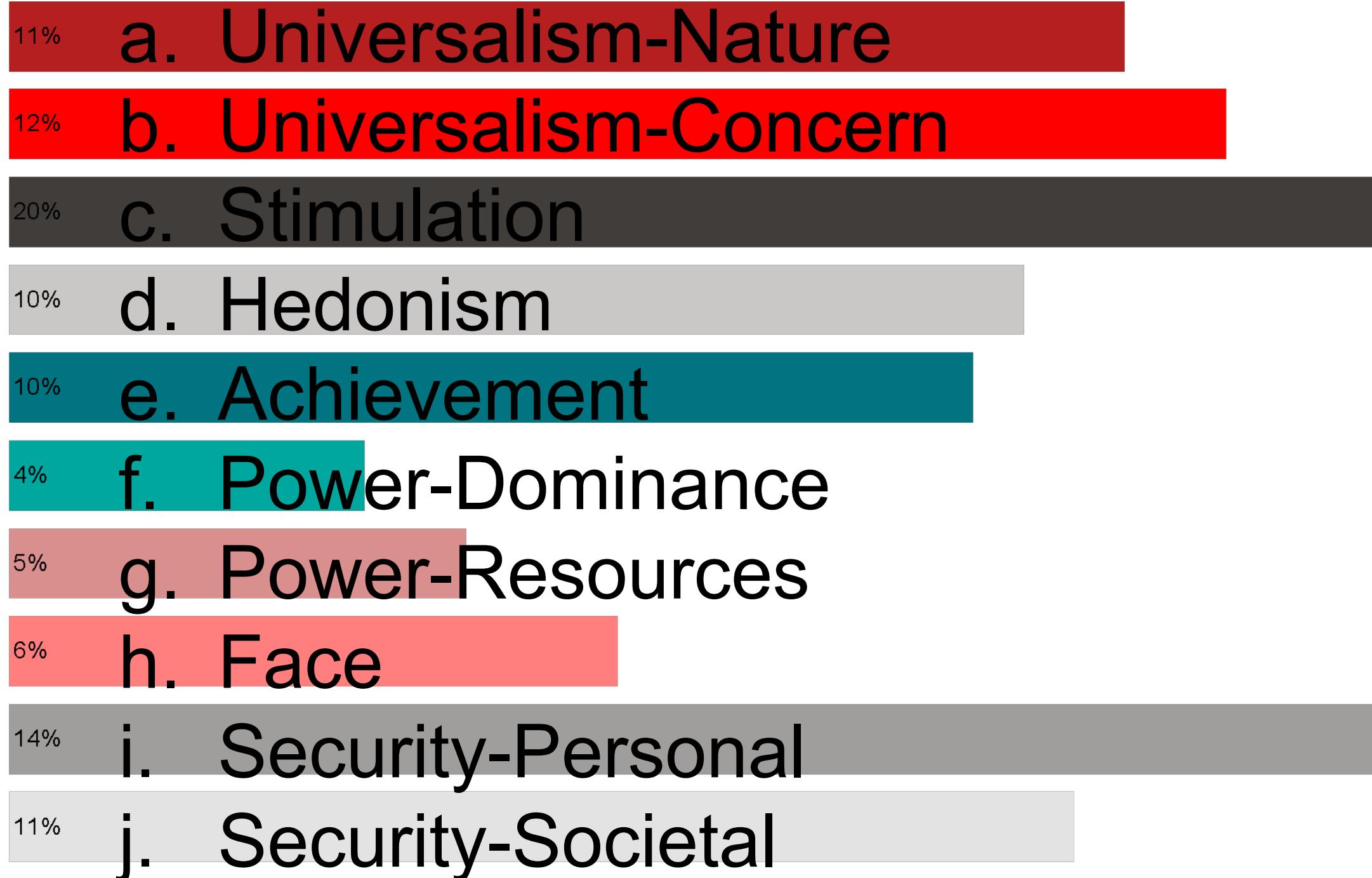
- Look at the advertisement website
<https://google.com/intl/en/maps/about>
- Explore the different features of Google Maps in a web browser (use a private window).

1. Indicate which values are visible
2. Indicate how they manifest

Values in Google Maps

URL: ttpoll.eu
Session ID: cs290

Select the values you have identified among these:



Part 3a – Stakeholder values

1. Read the stakeholder description
2. For each stakeholder, identify 2 values that you can relate to Google Maps
3. Specify whether they are benefited or harmed by Google Maps and how in the table:

Stakeholder	Key values	Manifested	Benefits	Harms

Part 3b – Value tensions

Draw the value tension map:

1. Place the values
2. Add the stakeholders concerned and indicate if it's a harm (red) or benefit (red)
3. 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

Overall debriefing of the strategy

- Products are not neutral: they embed and manifest values
- **Stakeholders have values, which may be harmed or supported by products**
 - 👉 Value tensions help identify how the design of a product may negatively affect some stakeholders
- It can happen that some value conflicts cannot be resolved
 - Identify core values and redlines
 - Analyze tradeoffs

Inclusive Design

We will go back to this case next week.

What's next?

Indicative evaluation

- The feedback system is open until midnight Sunday 13 October
 - 👉 log on ISA
 - Single question: “The running of the course enables my learning and an appropriate class climate”
 - Field for comments
- **Your feedback is important to improve the course!**
 - It is collected early so we have a chance to adjust what can be adjusted this year
 - Your comments are appreciated!

We start Fairness 2!

Tomorrow, Tuesday 8: notebook on recidivism risk prediction

By Monday 15:

- Watch **videos 4.1 to 4.5** + do the **quizzes**
- Finish the notebook
(and any other leftover from previous weeks)