Redesign presentation: Human AI collaboration system

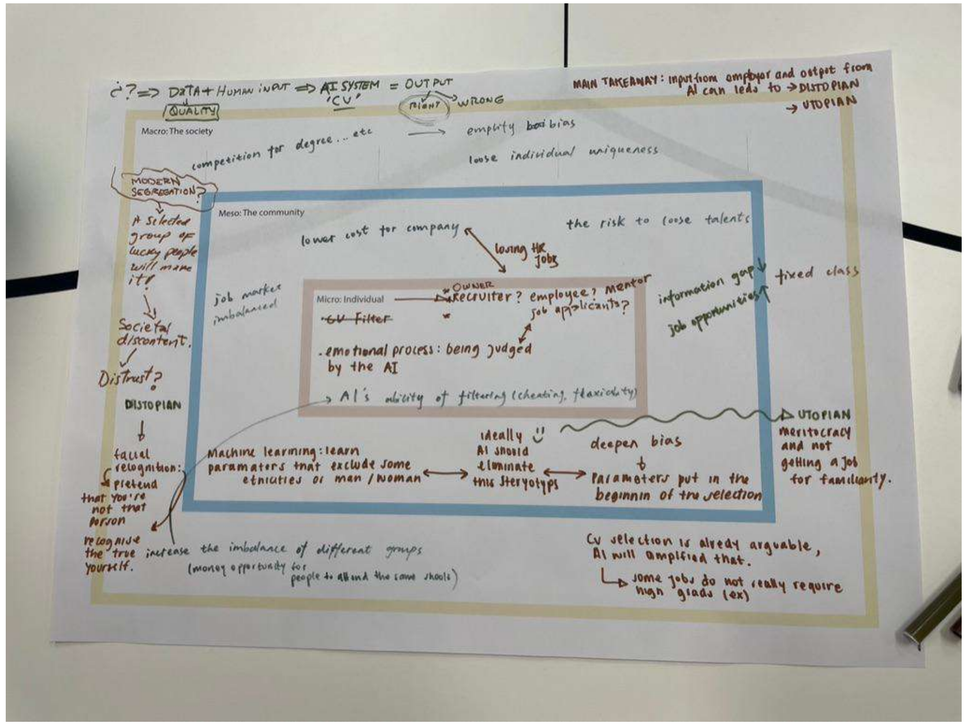
IAIRM - 2023

Mapping assignment

Introduction to the course

Description of the workshop

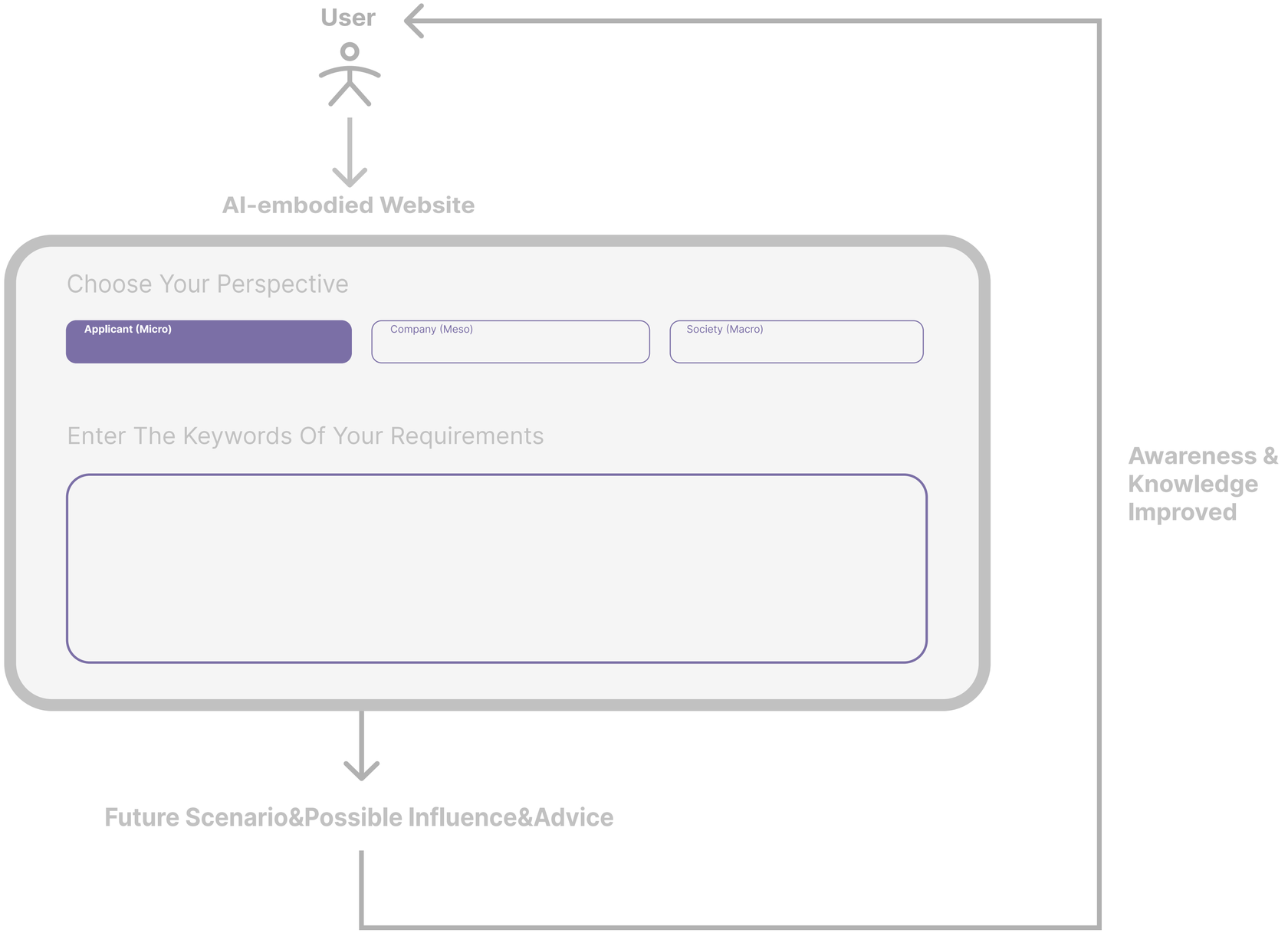
During the workshop, we had the opportunity to choose a theme and create different levels of study. In our case, the first method of mapping was useful in understanding the two major branches of AI - one positive and the other negative - that could impact the workplace. We also observed that AI is closely connected to its surroundings and is influenced by, but primarily influences, its stakeholders from a layering perspective.



Main takeaway

- AI is not an entity by itself. It is **entangled in a system** which needs to be considered while designing.
- The use we made of technology, even on a smaller scale, can **influence its output** of it in a massive way.

Redesign proposition



Designing for Contested Values

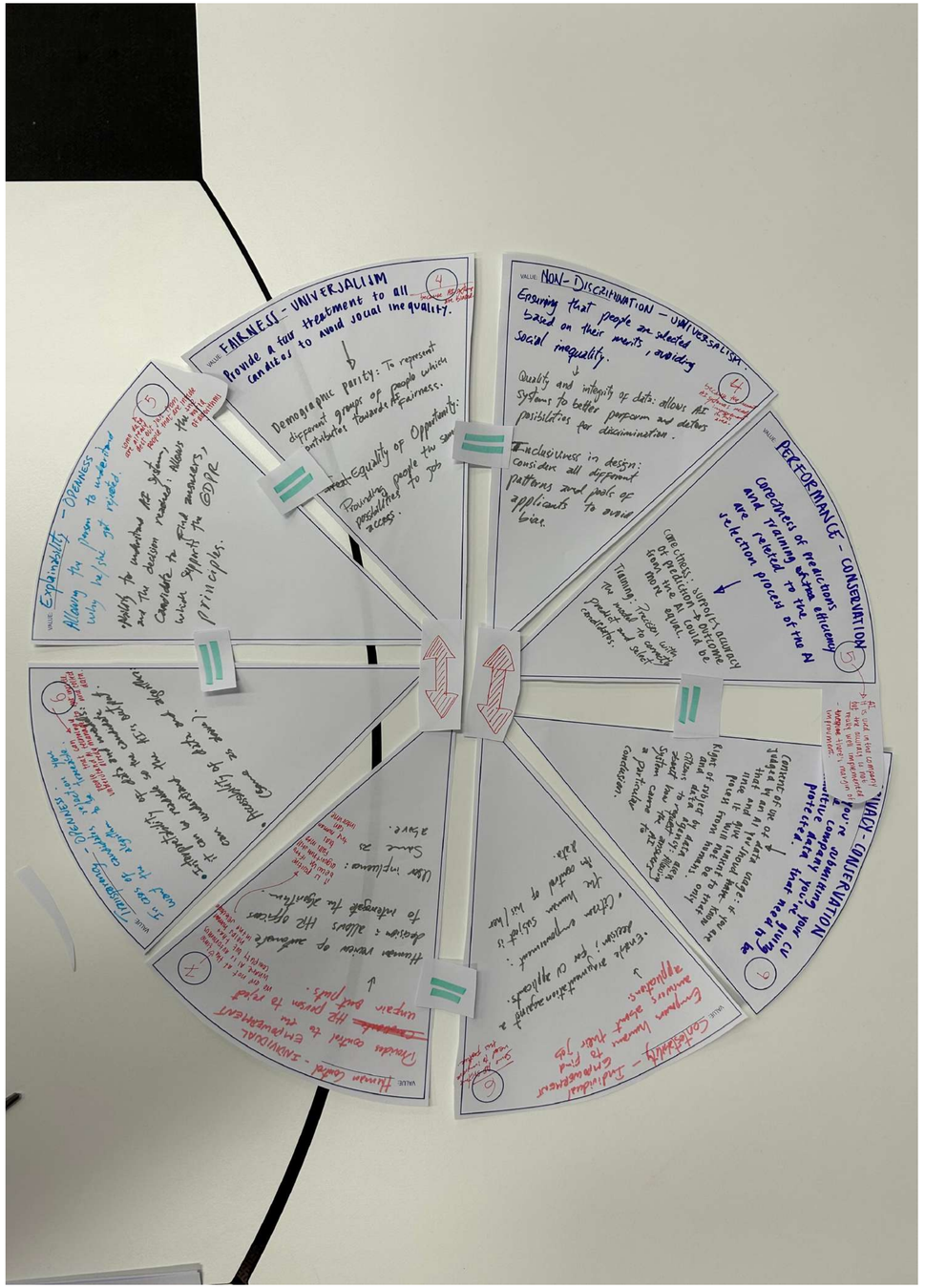
Mireia Yurrita

Description of the workshop

Initially, two values were encountered in the methodology: Human Control and Non-discrimination. Human Control was given a score of 7 out of 10 due to human involvement in the AI machine, while Non-discrimination received a score of 4 out of 10 because the AI recruitment system did not meet the minimum requirements to prevent discrimination. Despite the low score, addressing these tensions and requirements was a priority in the design process.

Main takeaway

- **Defining values** before designing a project can also make the project more suitable for the needs of those who commission it.
- The value you want to emphasize depends on who **commissioned** it.
- The value system should tend to be fully **respected for a more equal** design but it is almost impossible.



Redesign proposition

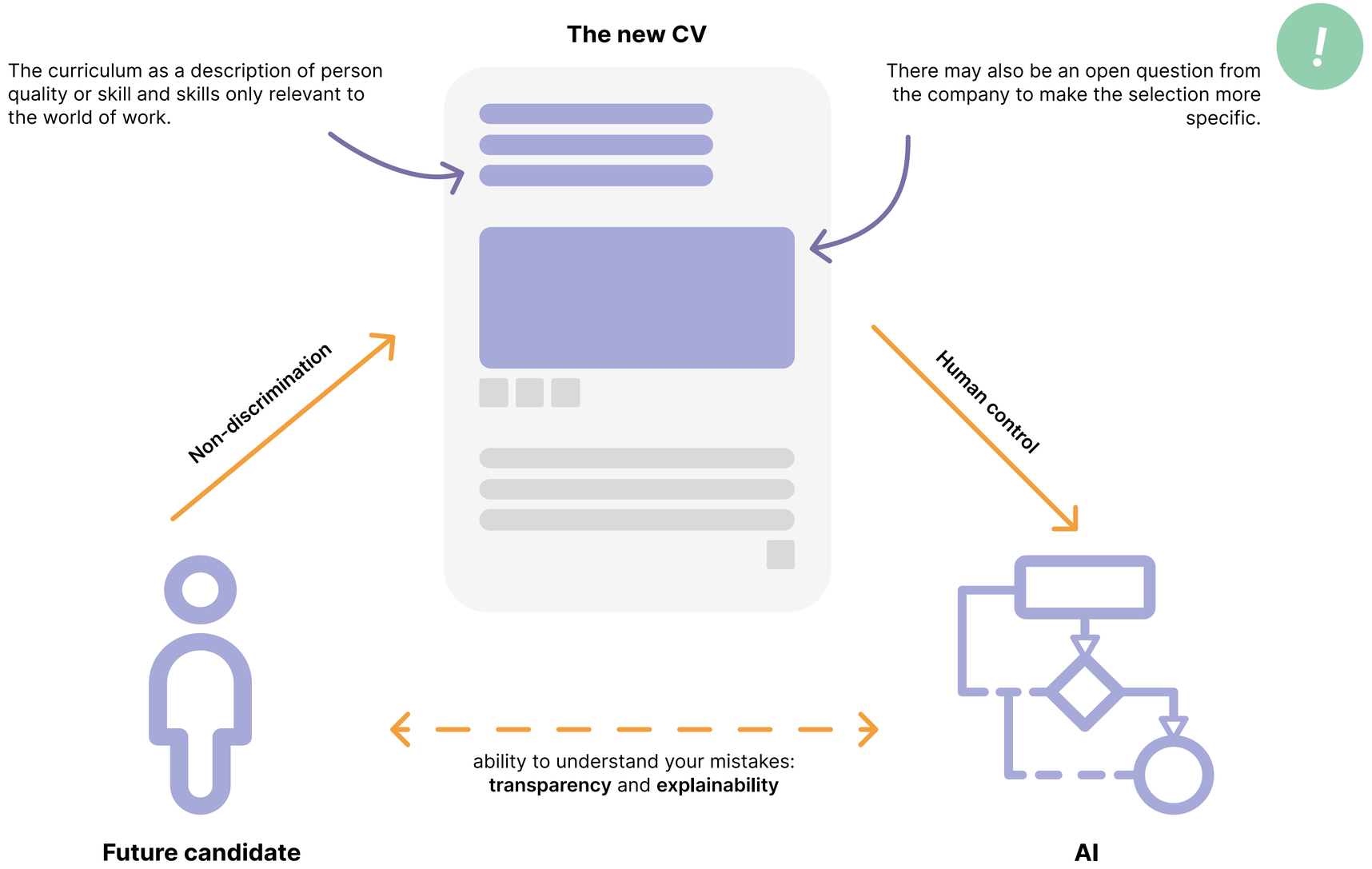


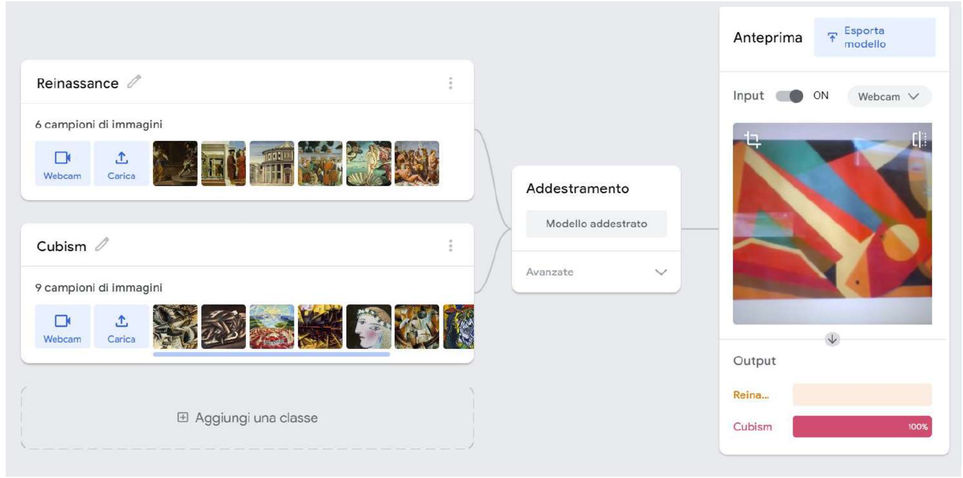
Figure 02. Representation of the new selection system.

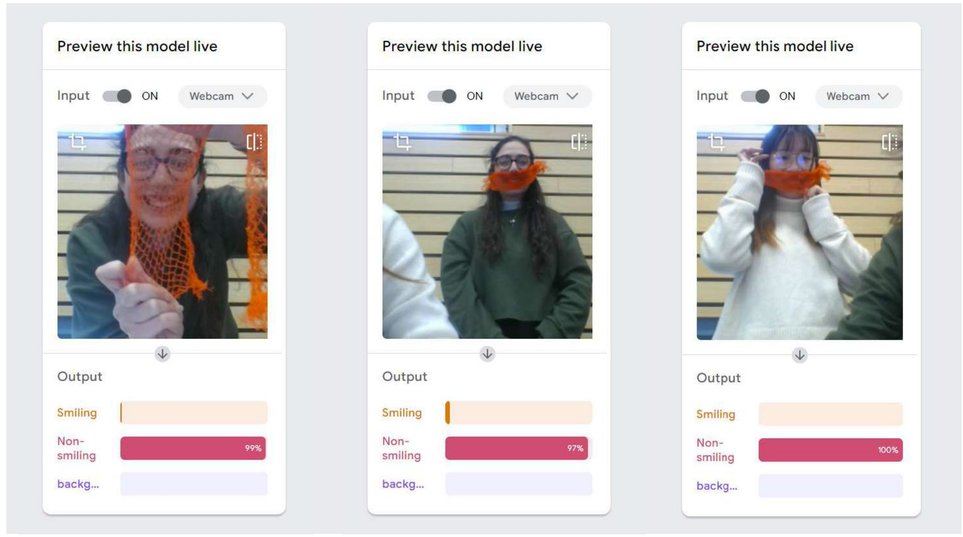
Hacking Intelligence

Mahan Mehrvarz

Description of the workshop

This week's methodology involved attempting to hack AI systems to determine possible vulnerabilities and whether intervention was necessary. The workshop comprised two sections, with the first involving the creation of an image recognition model to differentiate between renaissance and cubism paintings. During the second portion, another group's AI model for smile recognition was hacked by adding a physical layer to deceive the AI into detecting a smile.

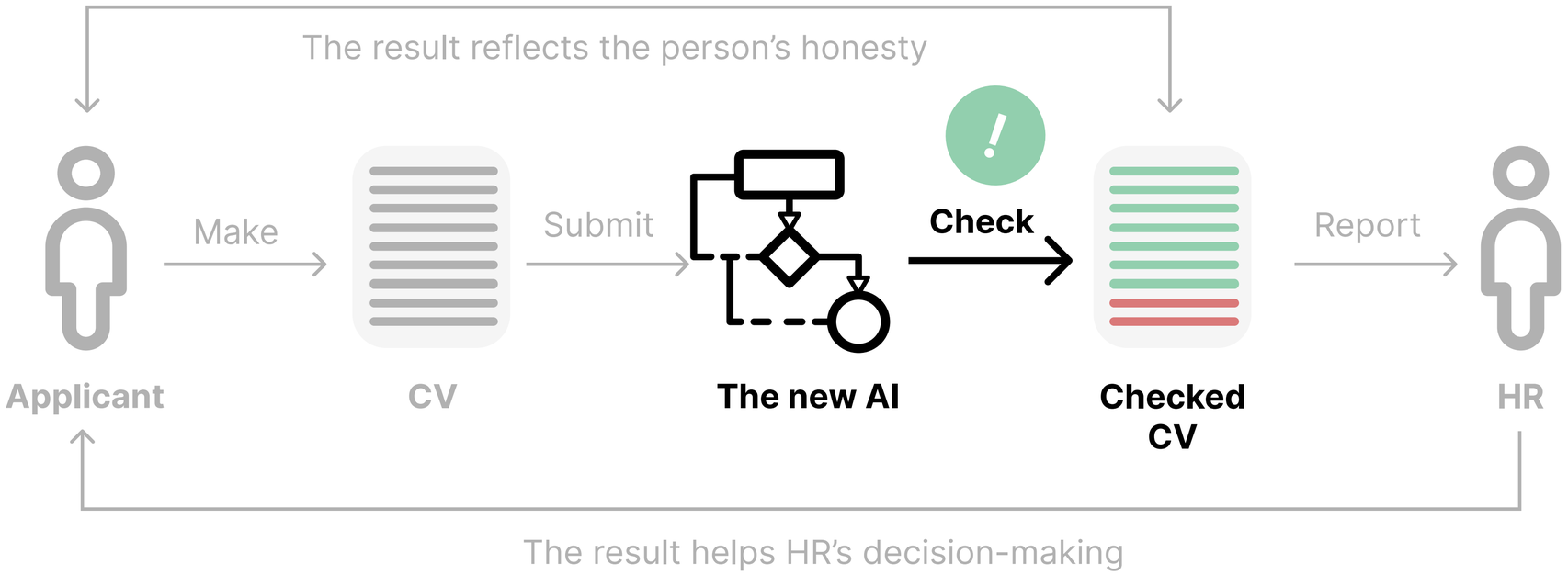




Main takeaway

- The hacking method is **applicable for every purpose** the company wants to have.
- Hacking is about finding **holes that you can work with**, but also about finding weaknesses to eliminate.

Redesign proposition



Coding content moderation and AI

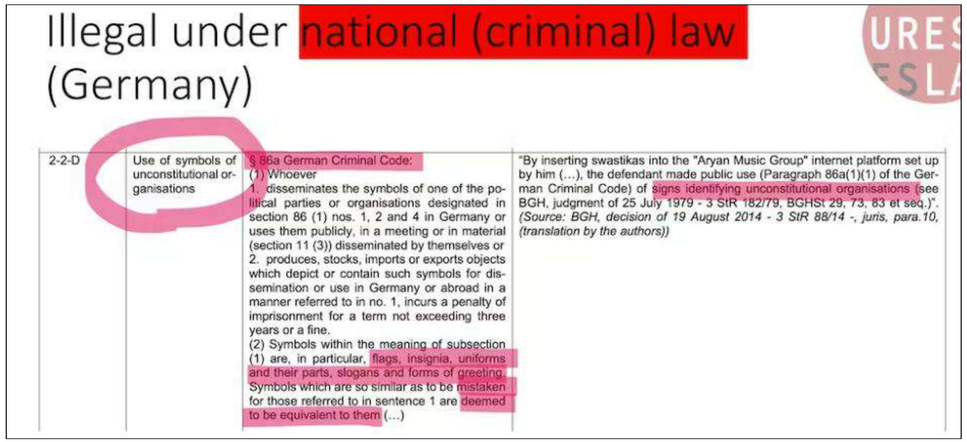
Marie-Therese Sekwenz and Ben Wagner

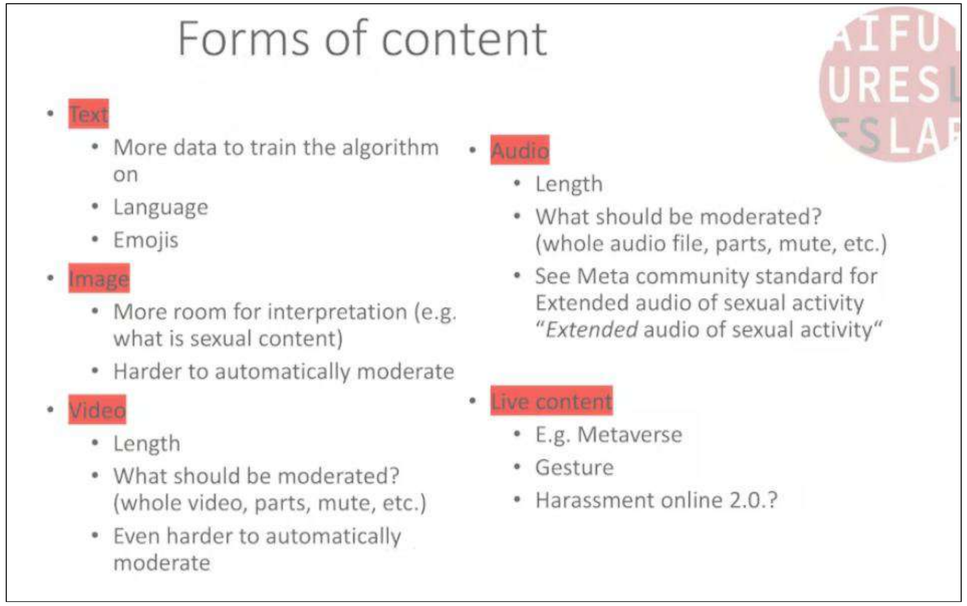
Description of the workshop

This week's methodology focused on content coding and moderation, exploring AI's potential applications and limitations in this field. The workshop comprised two segments, with the first involving the coding of ten contents using the coding handbook to establish their legal and coding categories. In the second segment, we discussed moderation and shared our results in pairs, making final decisions and comparing findings with others. The final group discussion offered insight into content moderators' work and AI's potential use.

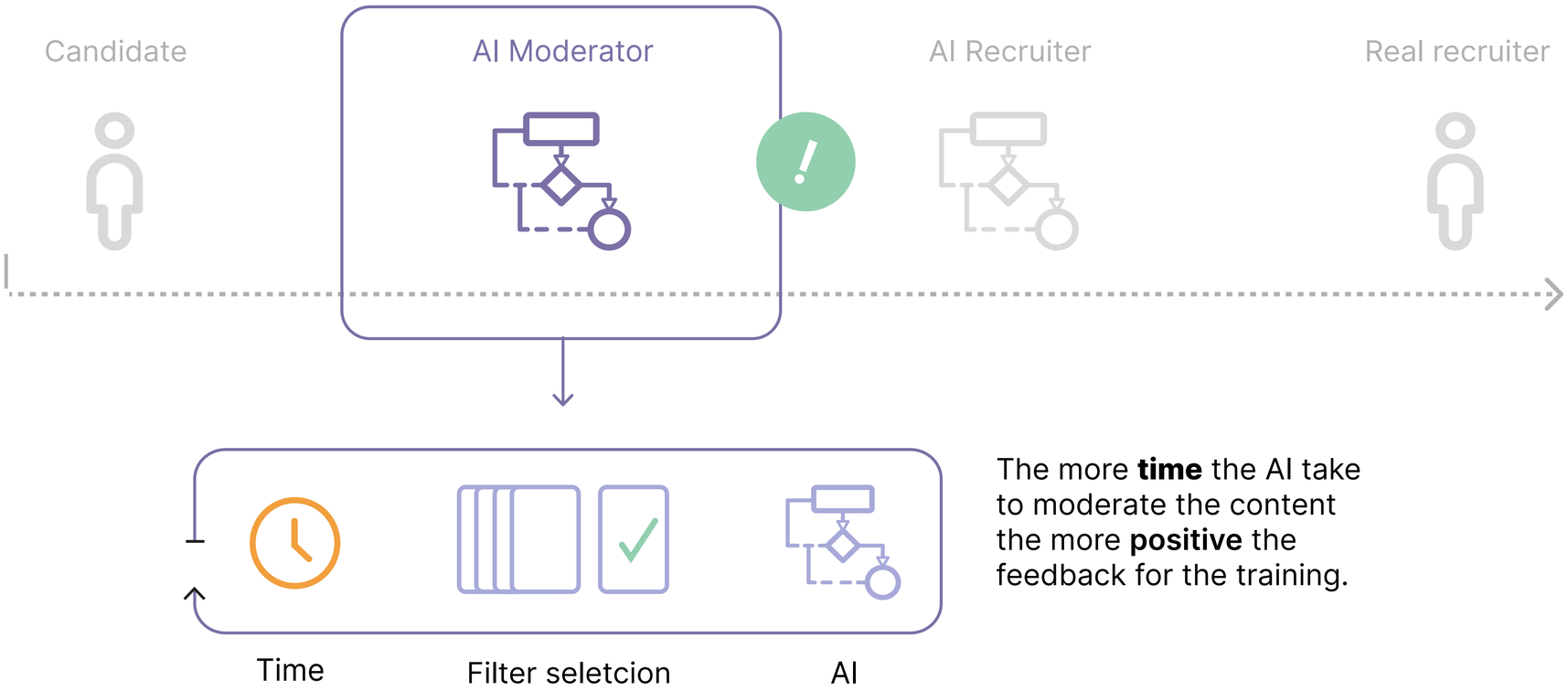
Main takeaway

- content moderation in our case is linked to the **privacy and legality** of certain information
- in our specific case it is easier to categorize what is **relevant** and what is not (since we always talk about work).





Redesign proposition



The more **time** the AI take to moderate the content the more **positive** the feedback for the training.

Ethics design for values through philosophical investigation

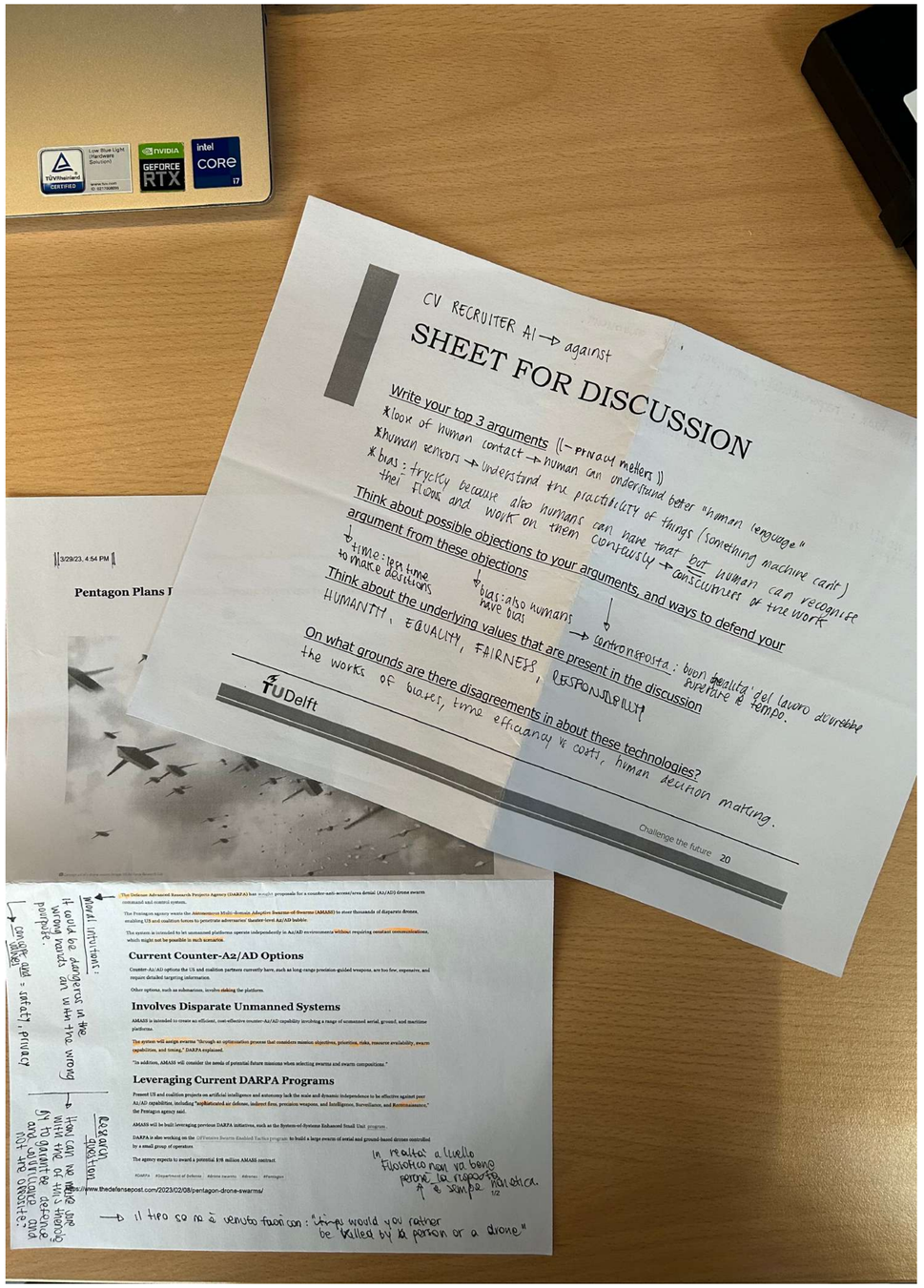
Aarón Moreno Inglés and Filippo Santoni De Sio.

Description of the workshop

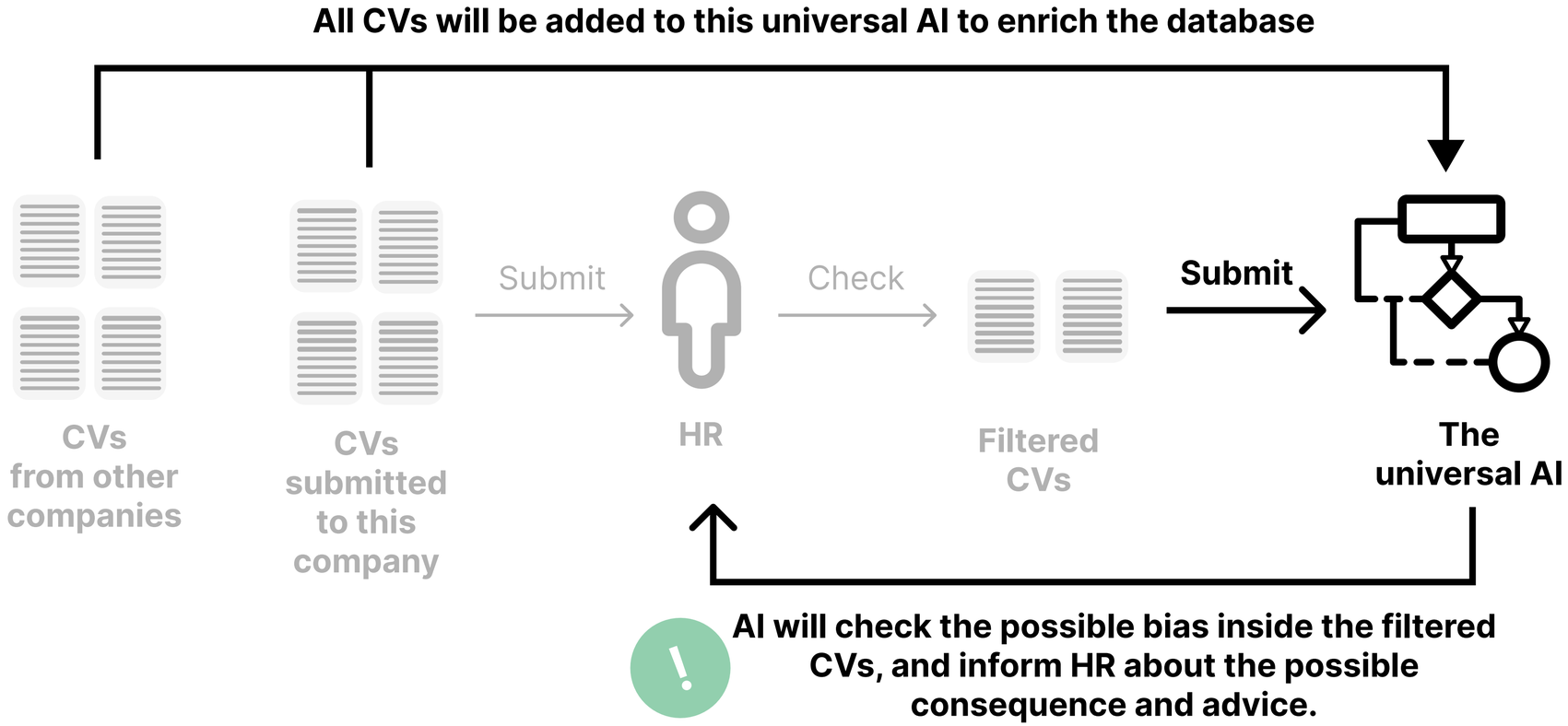
The workshop focused on using philosophy as a research method to analyse ethical issues related to AI. In the first part, we learned about moral intuitions and values, and applied them to a case study about war drones. The second part focused on different ethical issues that could be explored using philosophy, and ended with a debate about the use of AI in our case study.

Main takeaway

- The use of the philosophical method helps us from the point of view of **human rights** and the recognition of **bias** in the world of work.



Redesign proposition



Strange Labelling

Vera van der Burg

Description of the workshop

This week's method focused on using AI for self-reflection and reflecting on bias. We began by examining the relationship between AI and design, followed by the different roles and applications of AI. The lecturer emphasized the use of self-labeled datasets to train AI for self-reflection and presented her work on this subject. Finally, we applied our collected dataset to practice and shared our results.

Main takeaway

- Labeling depends very much on **programming**.
- AI is **not independent** at all in recognition.
- this method can be used both in a **speculative and functional** recognition.
- In our specific case it can help to **overcome the bias.**

Redesign proposition

New labelling system

01. New CV template

The candidate learns to fill the resume with new techniques with more images than text to describe objective data. When required the company could also ask for specific marks to indicate talents/gender and so on.

02. Labeling on obecjtive data

The cv is labelled based on this data so that the people labelling in the company and the AI later could not misunderstand the content.

03. AI selecting equality principle

The AI filer CV is trained to gain an equal amount of people from each category to guarantee equality in the first selection process.

How might we Integrate all the knowledge we learned, into

an AI-human collaboration system
that can exert the full power of humans and AI?

Further steps:

1. **Integrates all important values to a fair output**

2. **Care about human well-being while adopting AI**

3. **Overcoming bias from both humans and AI**

So we crate a **design system!**

Overview

Training	Checking	Screening	Reflecting	Deciding

Old Workflow

- Define the value based on single department demands.
- lack of discussion and consensus.

- No checking process
- Easy to be offended

- HR is the dominant people for screening

- No reflection process;
- Easy to follow bias;

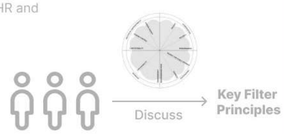
- HR is the dominant people for final decisions

New *Workflow*

VALUE DEFINE

Different stakeholders determine the principles of screening CVs, and the results will train both HR and AI.

(Value Framework From Lecture 1)



Learn

Train

MAIN USER FLOW

The main CV filtering work will be done by humans in our design, considering the well-being perspective. The AI will assist humans in this work.

Submit

Submit

Facilitate

Train

Submit

Inform

CONTENT MODERATOR / HACKING

Before the CVs handed in to HR, the first AI will check two the information inside from two perspectives.
1. whether the content is illegal or not;
2. whether the information is authentic or not.

LABELLING

While screening, the HR should label each section of the CV and express their feelings.

ETHICS / REFLECTION

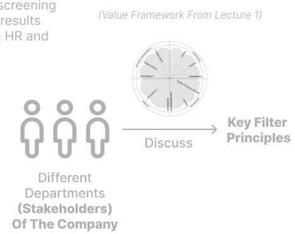
AI will check the possible bias inside the filtered CVs based on how does HR label them, and inform HR about the possible consequence and advice.

Training

- Define the value based on single department demands.
- lack of discussion and consensus.

VALUE DEFINE

Different stakeholders determine the principles of screening CVs, and the results will train both HR and AI.



MAIN USER FLOW

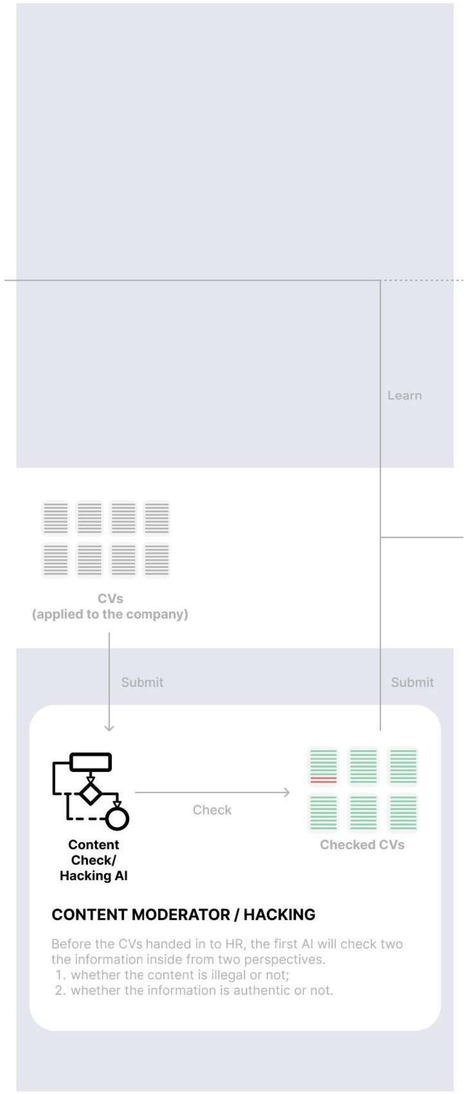
The main CV filtering work will be done by humans in our design, considering the well-being perspective. The AI will assist humans in this work.

Inspiration:

- ***Different Stakeholders Define The Most Important Screening Values Together*** *(From Workshop 2: Designing For Contested Values)*
- ***Using Values For The Training Process*** *(From Workshop 3: Hacking Intelligence)*

Checking

- No checking process
- Easy to be offended

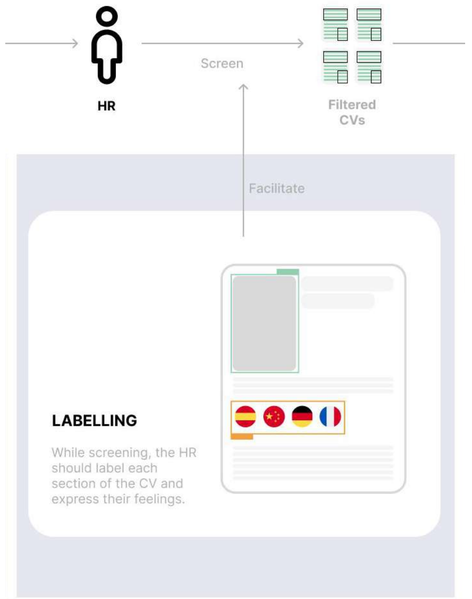


Inspiration:

- ***Let AI Check The Content For HR First To Avoid Illegal Content*** *(Workshop 4: Coding Content Moderation And AI)*
- ***Using AI Checking To Prevent Faked Information*** *(From Workshop 3: Hacking Intelligence)*

Screening

- HR is the dominant people for screening



Inspiration:

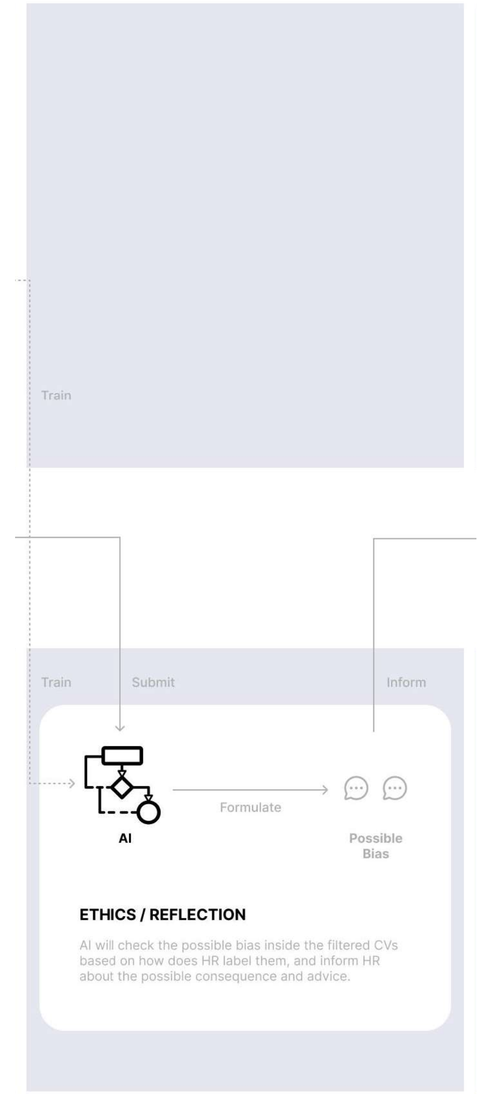
- ***Let Human To The Important Work To Take Care For Human Well-Being*** *(From Workshop 5: Ethics Design For Values Through Philosophical Investigation)*
- ***Let HR Do Labelling To Ensure Following Bias Check*** *(From Workshop 6: Strange Labelling)*

Reflecting

- No reflection process;
- Easy to follow bias;

Inspiration:

- ***Let AI Help HR To Consider About Ignored Ethics*** *(From Workshop 5: Ethics Design For Values Through Philosophical Investigation & From Workshop 6: Strange Labelling)*



Deciding

- HR is the dominant people for final decisions

Inspiration:

- ***Let Human In The Lead To Take Care Of Human Well-Being****(From Workshop 5: Ethics Design For Values Through Philosophical Investigation)*