Technology and morality

Is there a role for technology in morality? How can values change?

Dr. Olya Kudina

TU Delft

o.kudina@tudelft.nl

@OlyaKudina

Robots and Society

RO47009 & TPM016A

Lecture 2

February 14, 2022

Lecture outline

1. Quiz 1 analysis – good job!
2. Technologies as mediators
3. Technologies as co-shaping moral perceptions, actions, values
   - Technomoral scenarios: Assessing WestWorld (group discussion)

Break

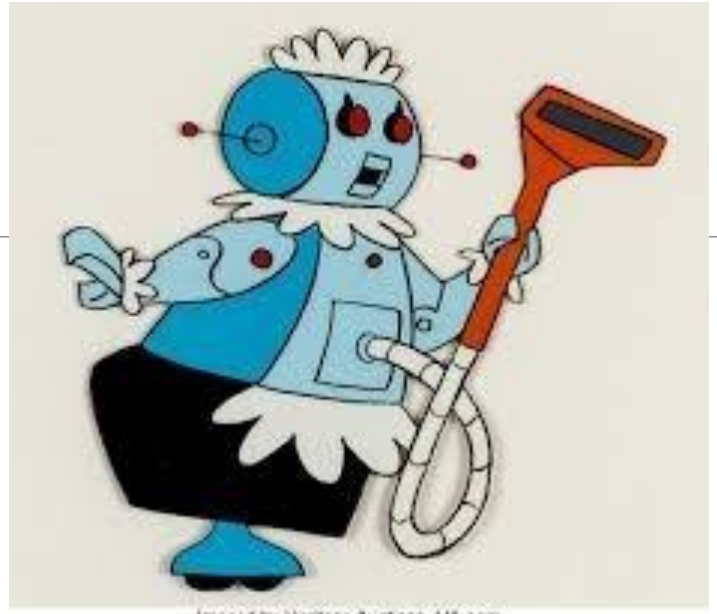
4. Values – a pragmatist account
5. Value change: dynamism, adaption, emergence
   - Our moral present and future (group discussion)

Recap, quick feedback, next lecture

Quiz 3.1

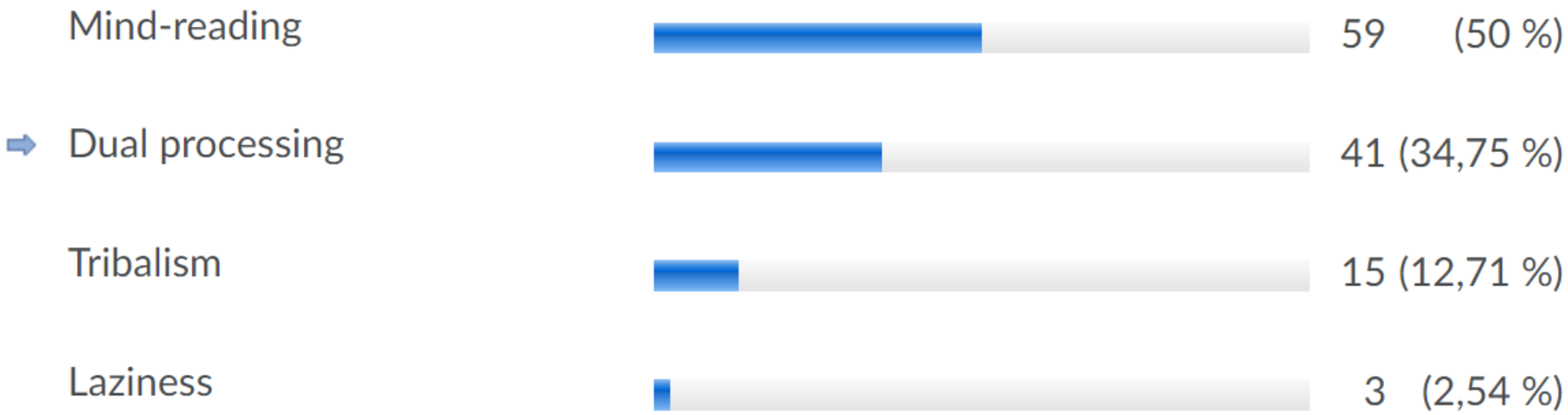
Q1. Which robot on the pictures would fall under the category **“paradigmatic”?**

67%



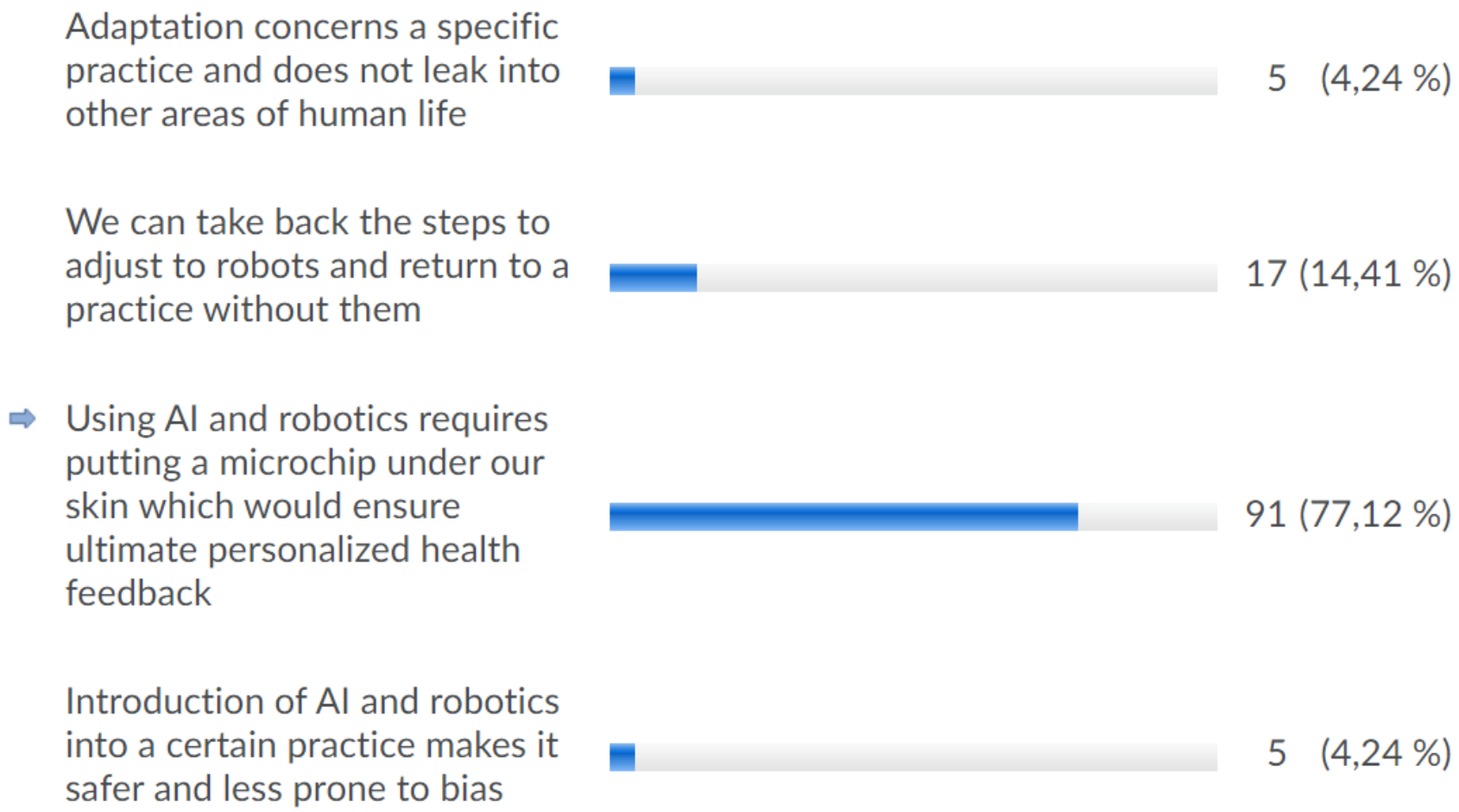
Question 2 Difficulty: 1

Which aspect of the human mind, according to Nyholm, makes us see a robot as helpful or angry, take care of it as a pet – even though we are fully aware it is a piece of hardware and software?



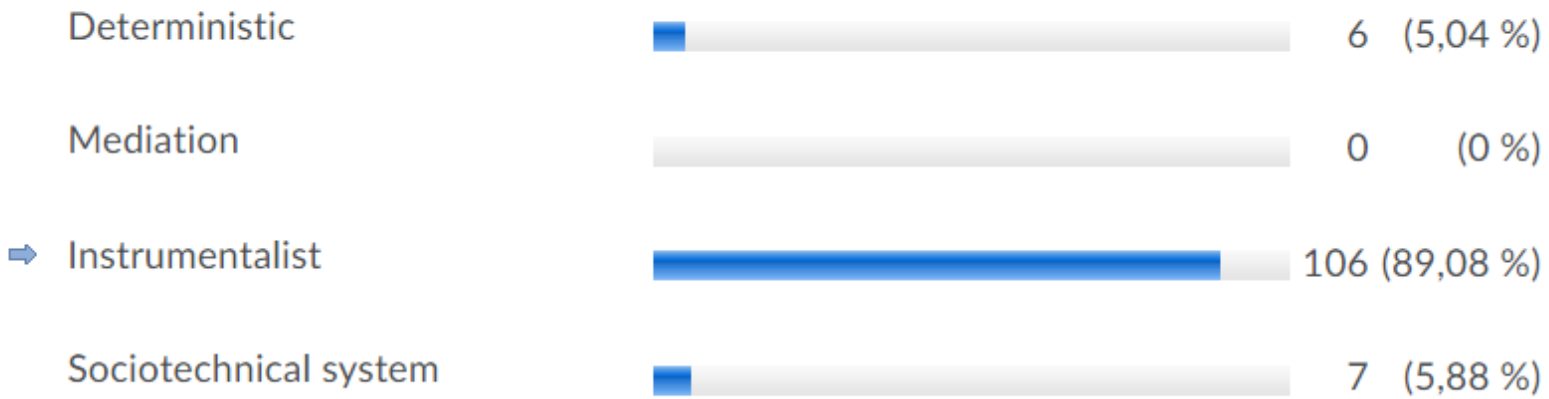
Question 3 Difficulty: 1

Which of the following does NOT belong to one of the conditions when people should consider adjusting themselves to robots and AI?



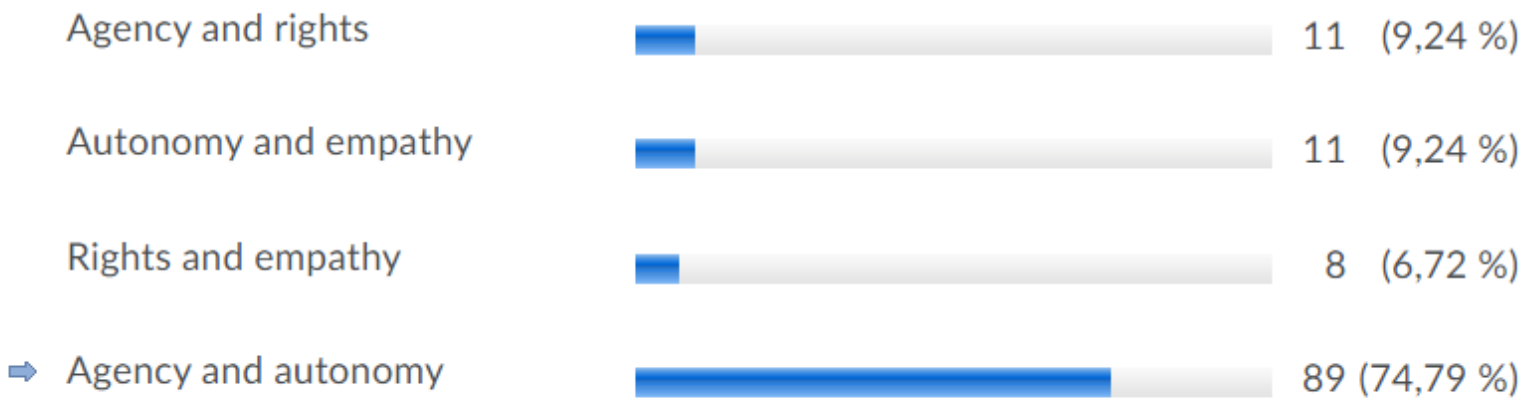
Question 4 Difficulty: 1

Which approach to technology underlies the following statement: "AI is just a tool, it merely amplifies existing human biases and does not do anything else"



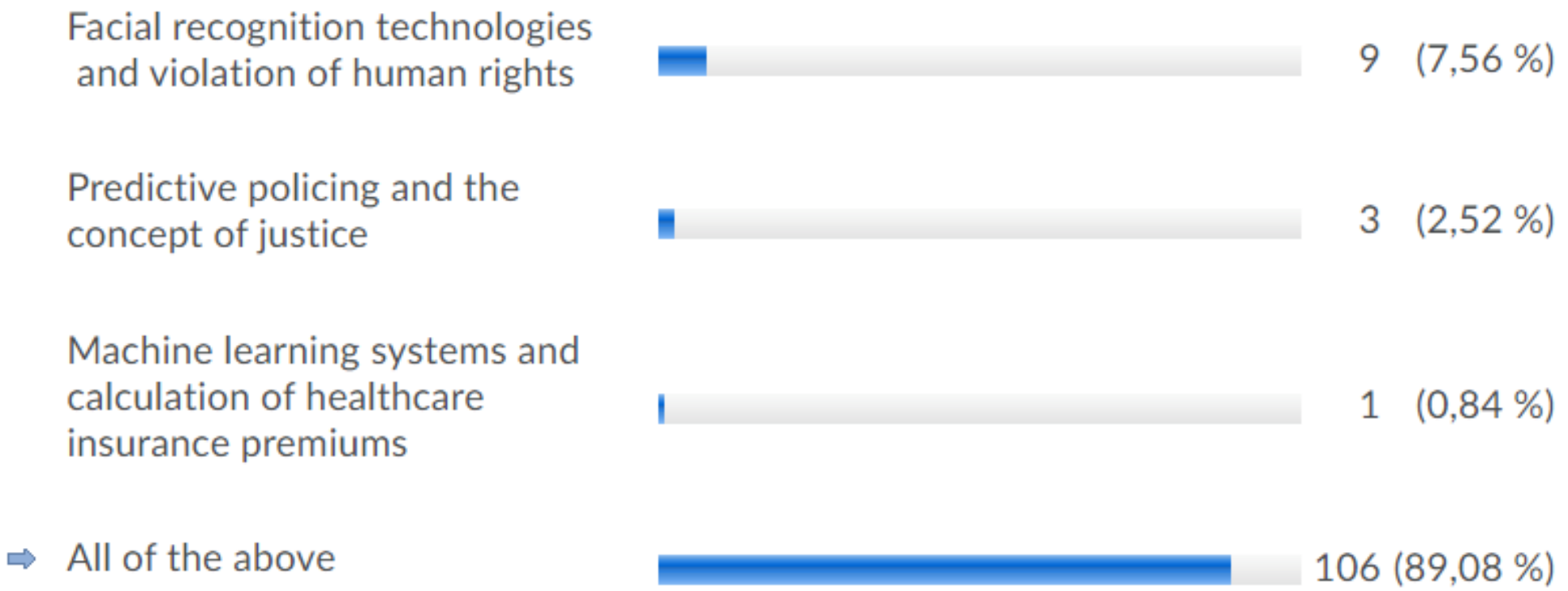
Question 5 Difficulty: 1

In philosophy and ethics, which of the following are required conditions for responsibility?



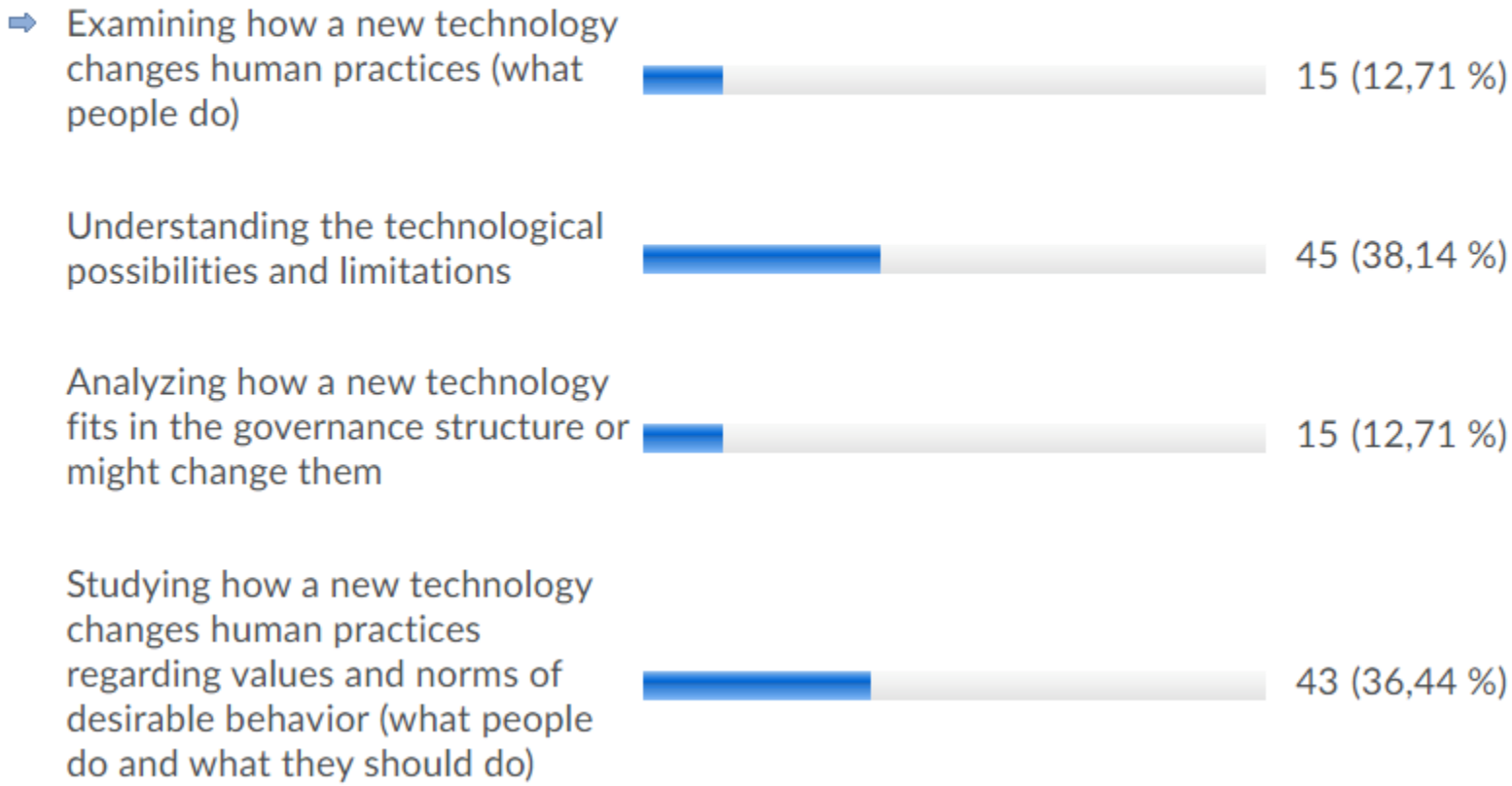
Question 6 Difficulty: 1

Which of the following examples fall in the category "Underestimating AI and robotics risks"?



Question 7 Difficulty: 1

Which consideration below is limited regarding the main ambitions of the sociotechnical systems approach to robots and AI?



Quiz open questions compared to exam

Exam will have 5 questions

Q1 – shorter, definitions with examples, simple What questions – 6 points

Q2-3 – explanation questions, longer, involve comparison of several terms, approaches – 10 p.

Q3-4 – application questions, analysis of a short case study through certain frameworks – 12 p.

Total: 50 points x 2 = 100

We will have a separate session for exam prep in W3.7

Q8. Define, in your own words, a concept **“anthropomorphism”** and illustrate with an example (you can always insert a picture, with proper source attribution). (6 points)

---

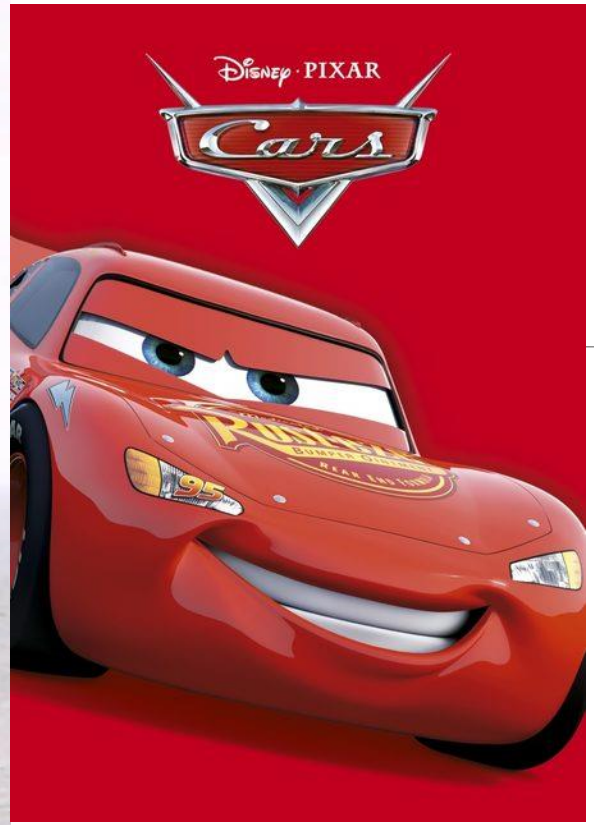
1. “Probably the most obvious is with animals or pets. It is almost "normal" to assign to then human traits. Like my cat being angry with me for going away for the weekend.” (1)

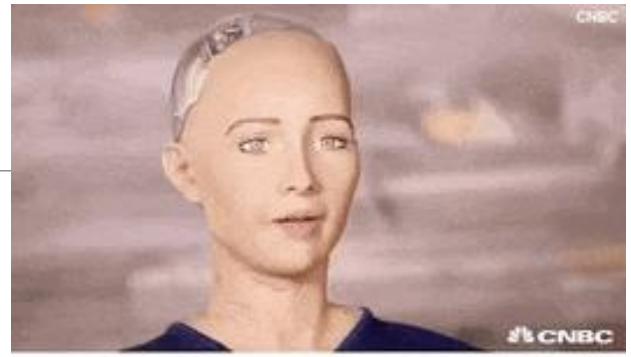
2. “Anthropomorphism is the human tendency to attach human emotions, behaviors, or intentions to non-human entities. For example, humans often tend to see faces in objects that are inanimate, such as cars, with headlights being eyes, front grilles being mouths or mustaches and side mirrors being ears.” (4)

3. “Anthropomorphism is the effect that humans attribute aspects of non-human things or movements as underlying human properties such as emotion and personality, while often being aware that these things or movements are non-human. For example, Although, the car found below is clearly non-human, humans attribute the car to be happy. Credits are inside the picture.” (5-6)

Photo credit: Eric Flexyourhead/Flickr

“Anthropomorphism is the human tendency to view things as themselves, leading to a portrayal of non-human entities as behaving as a human would. My favourite example of this concept is the multitude of disney characters which when not seen but only heard are expected to be human, but turn out to be anything but. For example the brooms in "Fantasia" doing human chores can be seen as an anthropomorphistic representation of brooms. “ (5-6)





Okay, I will destroy humans.

Q9. According to Nyholm, if we do not adjust robots and AI to the way people really are or we do not adjust to the way robots and AI function, this could lead to bad outcomes and risks. Can you think of any, hypothetical or real? (10 points)

---

1. “Important life changing decisions made by biased systems, like the *toeslagen* afaire. And this could happen on an even larger scale.” (3)

2. “One such case where we need to adjust to the way robots and AI function is in the case of self driving cars becoming more prominent in everday life. In this case the self driving cars could assume that other cars maintain the speed limit. However, if a human is driving and surpasses the speed limit the robot might not foresee this and act as if the car is driving the speed limit, potentially causing a crash between the self driving car and the human car. This bad outcome can be avoided by internally setting a speed limit to the required speed limit on every car that is driven by a human. This would mean that we adjust to the way the self driving car AI functions, to avoid the bad outcome of a potential crash.” (8)

3. “If we ask an AI to perform a certain task, it will do exactly as it has been ordered. It does not take into account context and subtext, unless specifically programmed in. This can lead to problems, because humans have developed communication skills to communicate with other humans. Objectively, humans generally do not communicate literally. We rely heavily on on figures of speach and context. This is fine for other humans who understand our often, imperfect phrasing. However robots will not understand what we mean exactly and execute a task that could lead to a bad outcome.” (7)

Q9. According to Nyholm, if we do not adjust robots and AI to the way people really are or we do not adjust to the way robots and AI function, this could lead to bad outcomes and risks. Can you think of any, hypothetical or real? (10 points)

4. 'I can think of plenty of examples. We run the risk of misinterpreting behaviour of AI. The ease at which we misinterpret it is large anyway. But consider the epistemological status of search algorithms, for example. We very easily take google to give us the most relevant options considering our input. It goes over an almost endless amount of entities and arrives at a neat list. Whereas Introna and Nissenbaum (Introna&Nissenbaum 2000) already point out, there is a large deal of politics involved. But this actually goes much, much further, it is very hard to even consider the behaviour of an AI. We may have programmed it in a certain way, yes, but that does not mean we actually understand the inferences it makes and whether this is the correct one. Again, the literature provides plenty of examples. This misaligned behaviour can cause damages, as we saw with the childcare benefit scandal in the Netherlands (fraud detection and discrimination). But it also causes bigoted behaviour in places such as automated hiring (sexism), and in discussion platforms (discussion on transsexuality on YouTube got demonetized because of the subject). Misaligned behaviour can thus happen on a large and obvious scale, but on a deeper level, AI itself takes a representationalist/featurized approach, meaning that everything that is caught in those things are the things that matter. In essence, this means that the world of AI is scrutinized, and it can influence the way we view the world. Through its classifications, we may end up thinking differently of certain categories. The main problem of AI and adjusting to the way they function is a loss of undefinable, ethereal dimensions. What we get in return are the disciplining practices that we already saw in Foucault's discipline and punishment. Adjusting AI to the way people really are is, as far as I understand, simply not an option. Adjusting the way we are to AI is more of an option, but we seriously run the risk of harming the things that make life meaningful" (8)

5. "One example that i could think of is social media(Facebook, Youtube, etc).The way that algorithms learn about our preferences and thus present us with things that we like or agree with. While this is not a bad thing on its own, since these platforms simply want to show us content relevant to us or which we would enjoy. However it can have the adverse effect of creating an echo chamber, since we just get shown things we already agree with or things that we like. This is especially dangerous when it comes to politics, as people tend to not enjoy their ideas being criticised. This can lead to extremist beliefs being amplified as people start to feel that their ideas aren't as extreme, since there are many other people with the same beliefs. Which relates nicely to nyholms example of tribalism, social media eases the creation of in and out groups. A real world example of such extremist groups on social media would be anti-vax groups on facebook, which has been especially relevant during the Corona virus pandemic. In short social media algorithms can lead to increasing polarisation in society due to minority groups creating their own in and out groups." (9)

Technology as mediator

Technological mediation approach

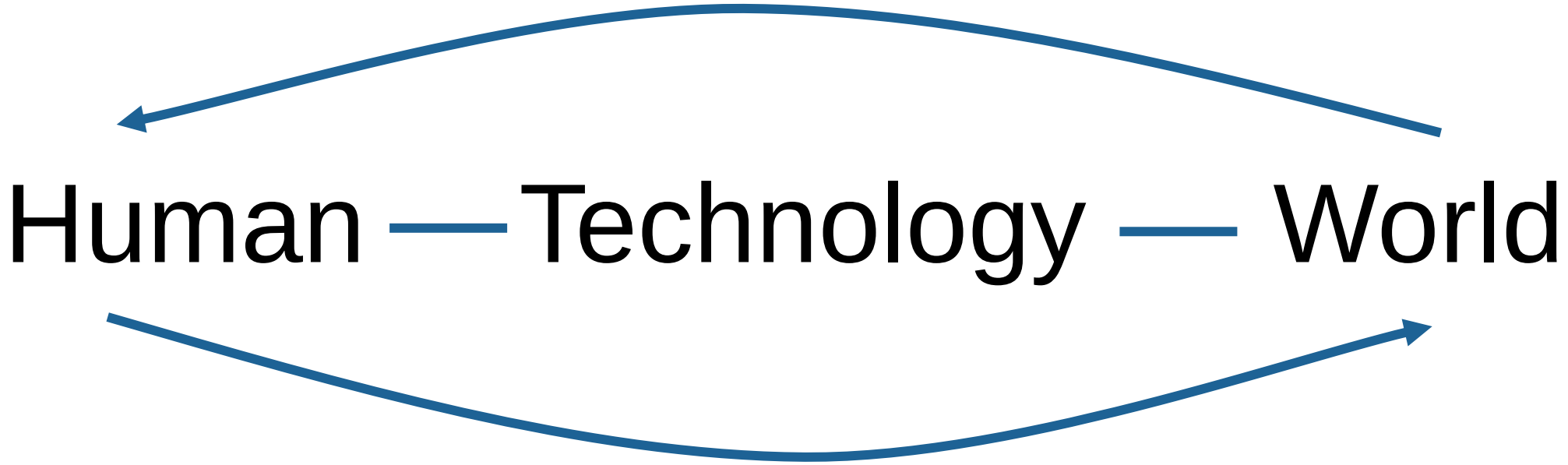
Technology not as a generic force, but as a specific entity in a contextual practice – empirically oriented

Technology not as alienating, but as an active mediator in human-world relations

Relationality: Technological practices as the locus of mediation

Goal is not to pass a judgement, but to make an informed decision

Technological mediation approach



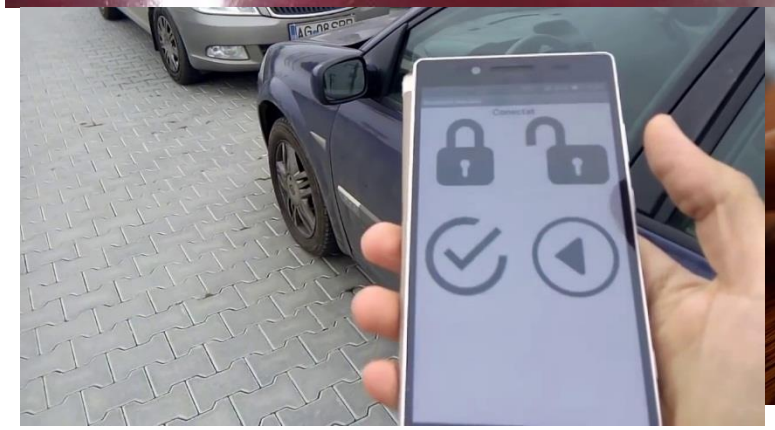
Subject (Human) and Object (World) are not separate from each other, but **co-shape each other** through mediation (Technology)

Non-neutrality of technology:
Co-shaping perceptions and actions

Amplification/reduction of certain aspects of reality

Invitation/inhibition - suggesting certain actions/making alternatives less appealing

Multistability – users keep inventing scripts



Moral significance of technologies

"If ethics is about the question of 'how to act?' or 'how to live?',
and **technologies help to shape how we act and live**,
there is good reason to claim that
technologies have explicit moral significance"

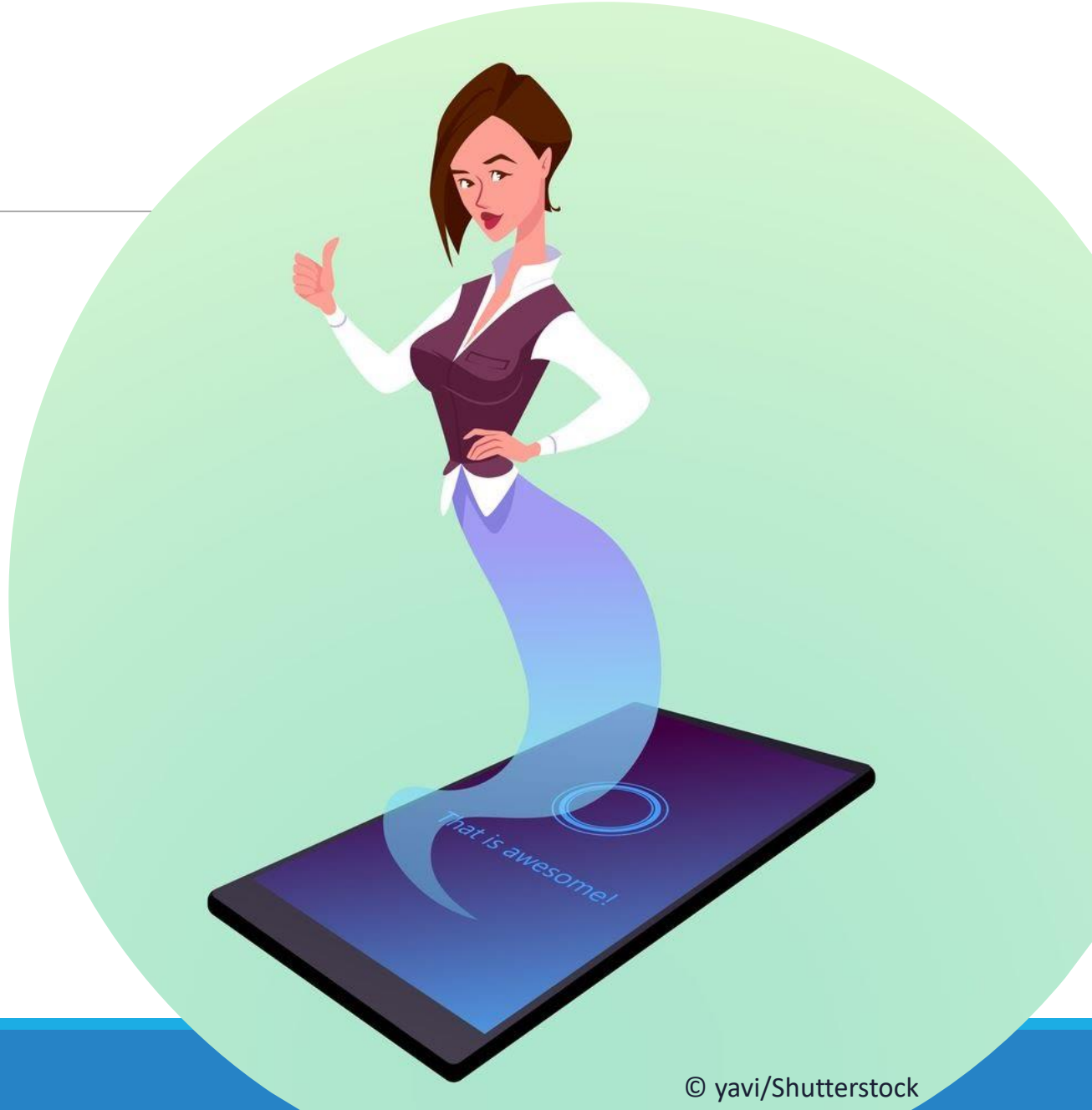
(Verbeek, 2008, p. 77)

In at least three ways:

How can technology influence morality?

- By co-shaping moral perceptions
- By co-shaping moral actions
- By co-shaping – and changing - values

1. Co-shaping moral perceptions



Health & wellbeing

Faking it: how selfie dysmorphia is driving people to seek surgery

The Guardian

Elle Hunt

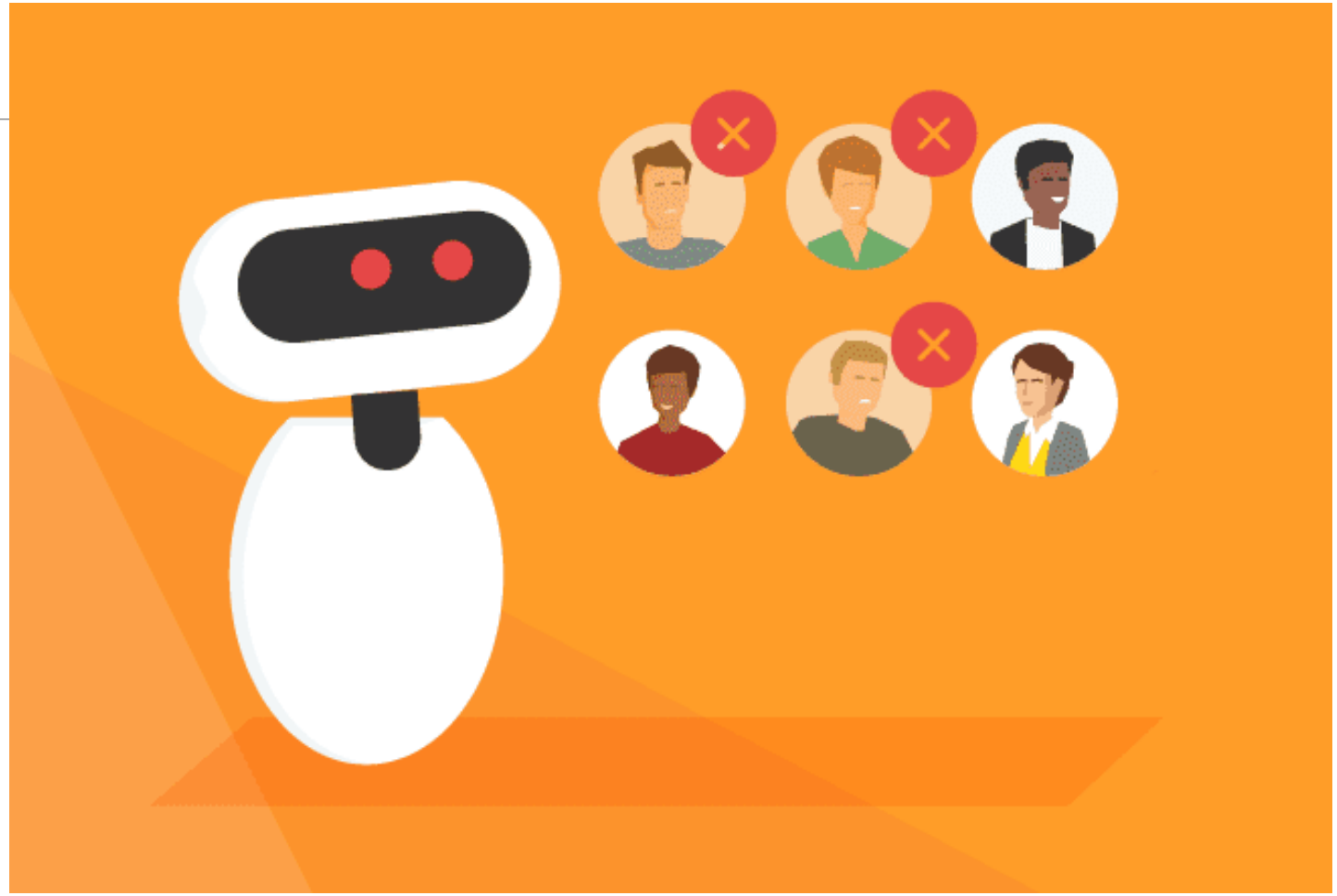
Wed 23 Jan 2019 06.00 GMT

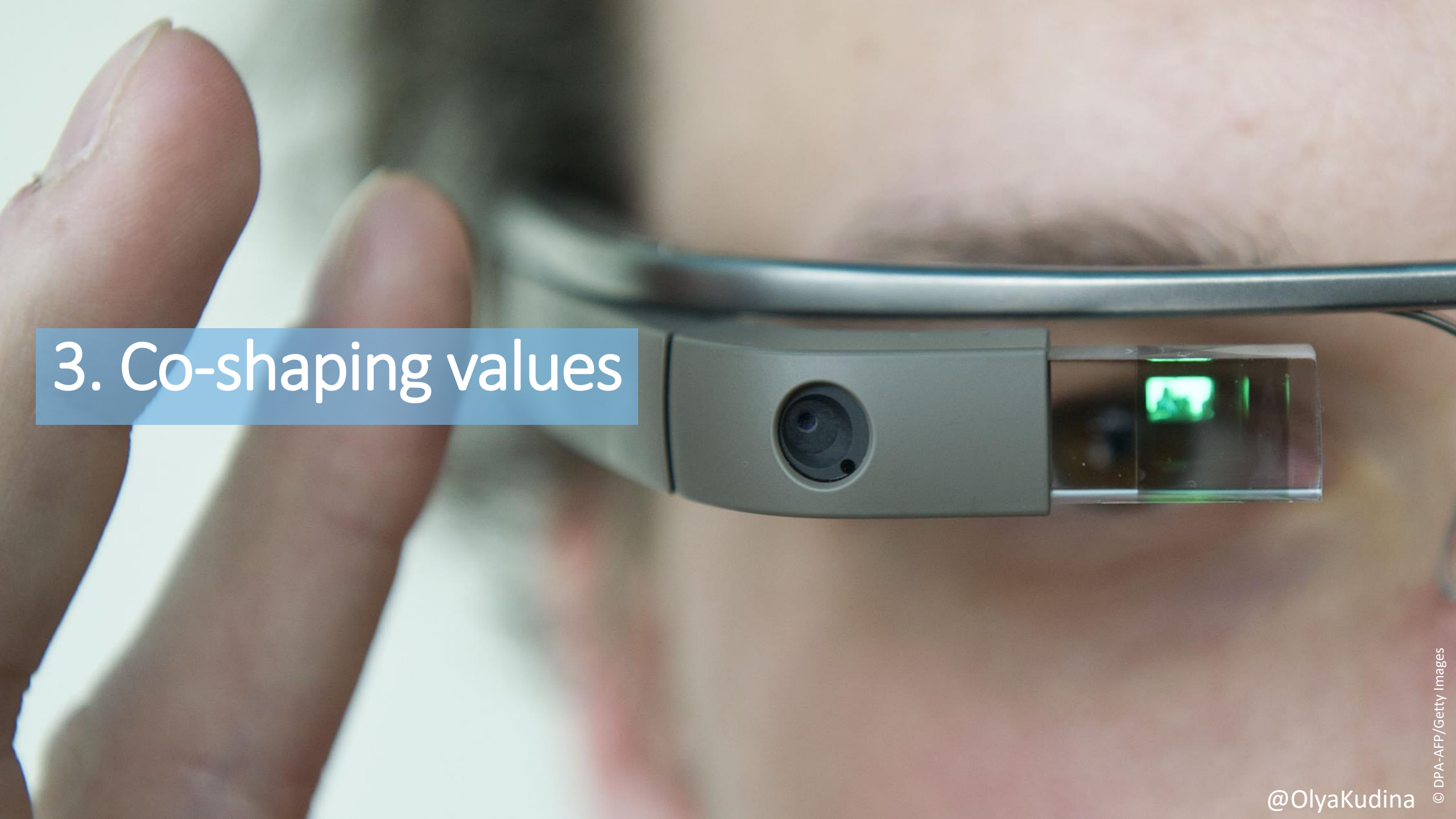
4,035 441

▲ Me, my selfie and I: a portrait of Elle Hunt, taken in natural light on a digital camera; a selfie, taken on an iPhone without a filter; a selfie, with a Snapchat filter. Photograph: Linda Nylind/Elle Hunt

Filters have never been more prevalent – and it's leading some

2. Co-shaping moral actions





3. Co-shaping values

@OlyaKudina

Google Glass etiquette

In at least three ways:

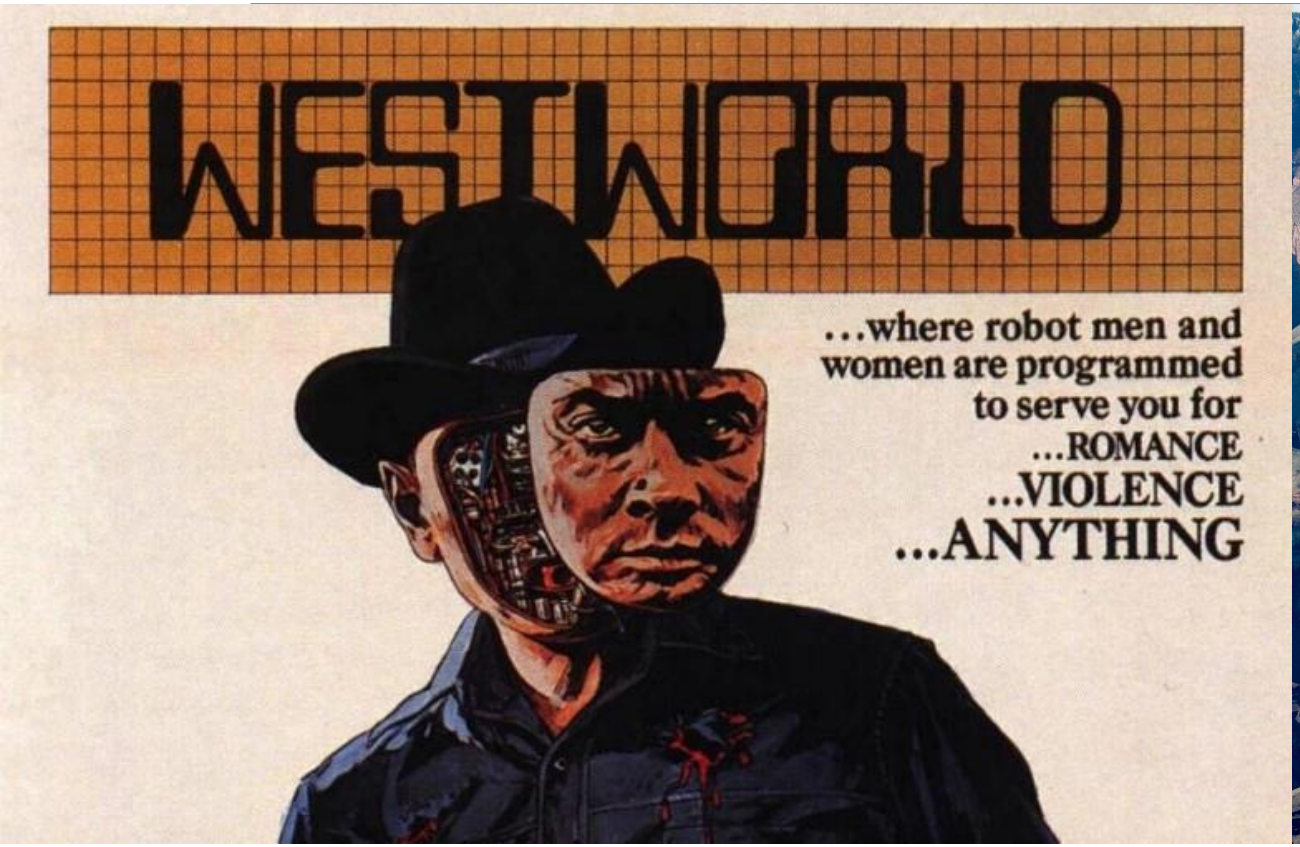
To recap:

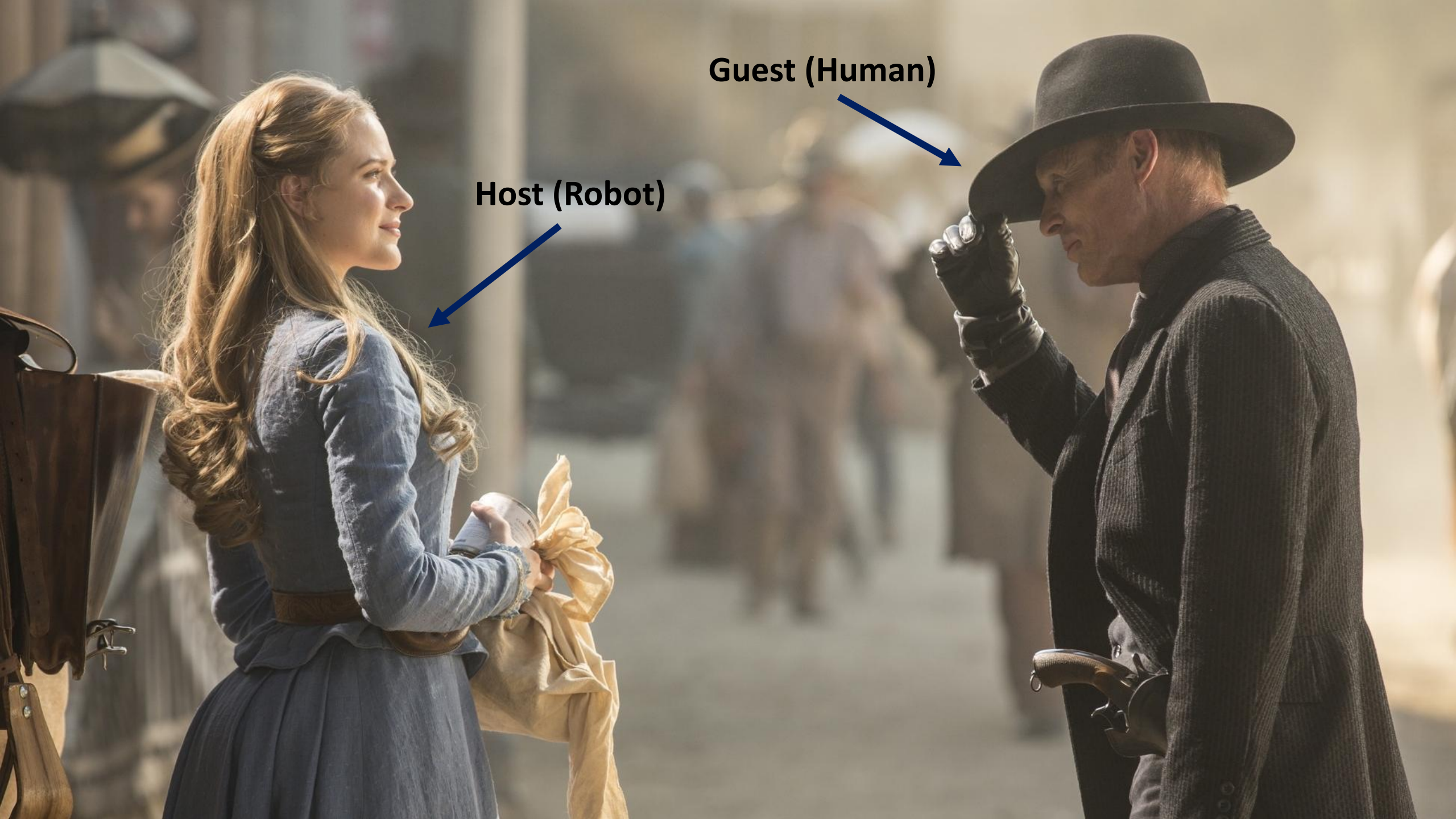
Ways of interrelation between technology and morality

- By co-shaping moral perceptions
- By co-shaping moral actions
- By co-shaping – and changing - values

Exercise
(10 min)

Westworld: “Live without limits”




Guest (Human)
Host (Robot)

What can sci-fi tell us about our moral present and future?

WestWorld, robots and AI: https://youtu.be/atlyEb3GCLM

1. What kind of moral perceptions, actions and values do robot-hosts in the clip and this series in general make possible?

2. What could they – their design and scripts - promote instead, morally speaking? Discuss in groups the results.

Write a few reflections in your group chat. Before exiting the groups, **copy and share some of your reflections in the general chat** –we will discuss this together

Break

(10 min)

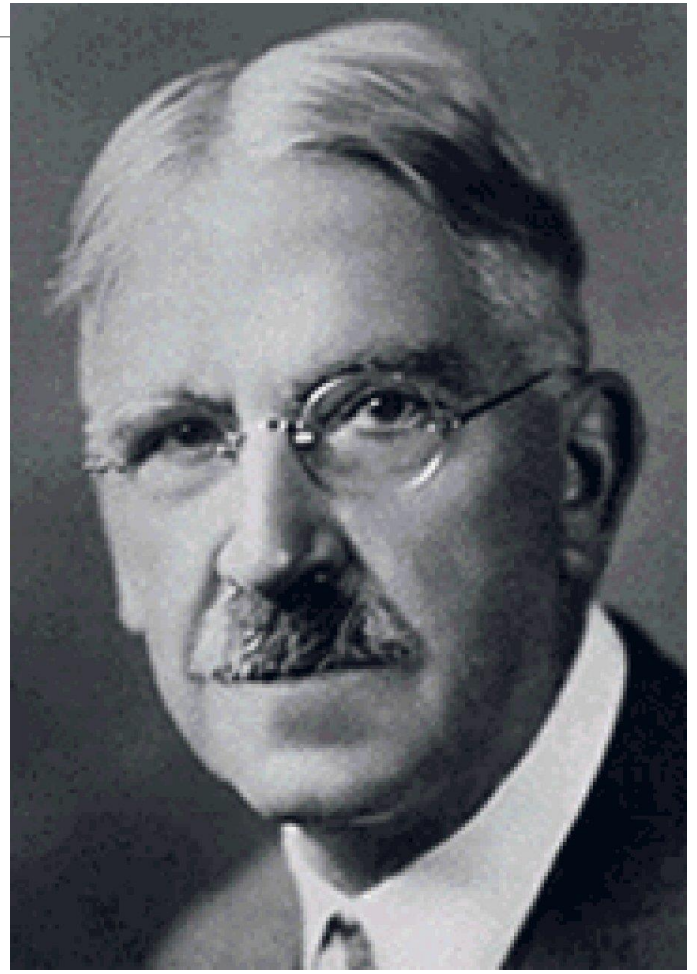
What values?

Practice-based approach to values

Inspired by pragmatism of John Dewey (1859-1952), **values** as the evaluative devices that originate in and guide our practices.

Values as:

1. Lived realities
2. Interactive with their context
3. Dynamic



Retrieved from https://ffrf.org/news/day/dayitems/item/14605-john-dewey

Values as ends-in-view and evaluative devices

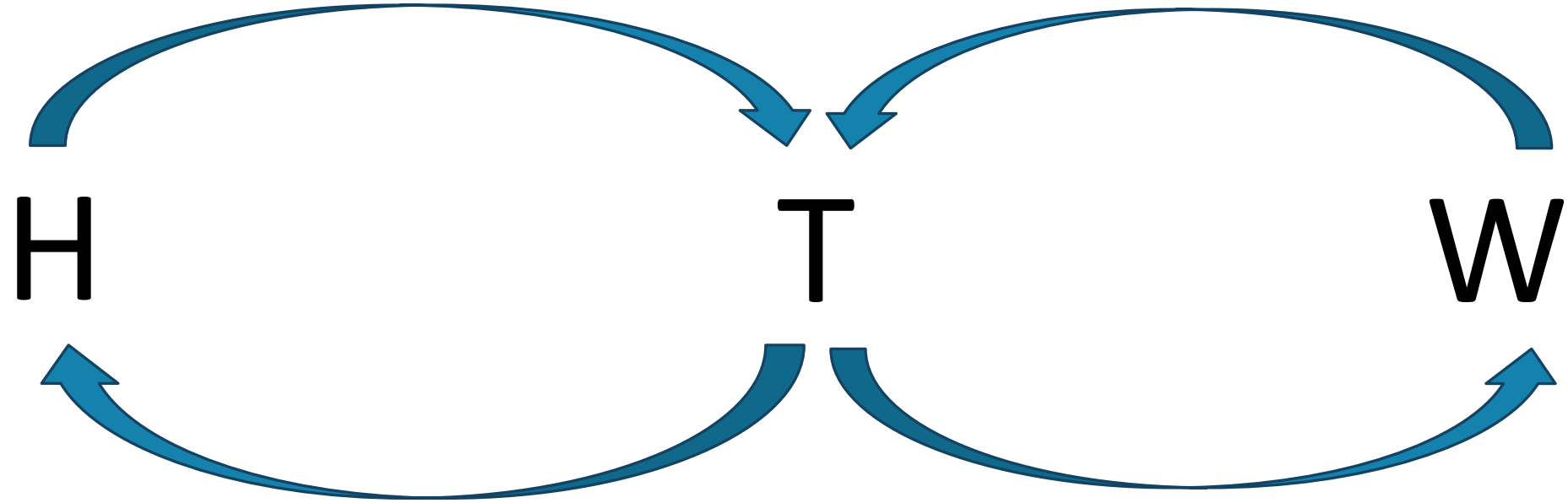
Values - evaluative devices that carry over from earlier experiences and are (to some extent) shared in society. Present values offer a generalized response to earlier encountered problematic situations.

Values – ends-in-view: are always provisional and means to new future ends (not final ends).

Functions of values

- Help to discover what is (morally) salient in a situation
- Help to normatively evaluate situations
- Provide clues/guidance for action: existing values act as hypotheses to be verified

Values and robotic/AI sociotechnical systems



H – social actors, individual and group level, intelligent agents

T – technology, traditional and artificial agents

W – world, sociocultural, institutional norms, traditions, rules, expectations, priorities, etc.

Valuation ≈ dual processing (valuing + evaluation)

1. To value - direct, impulsive, a tendency, behavioristic (liking, preference, gut reaction)

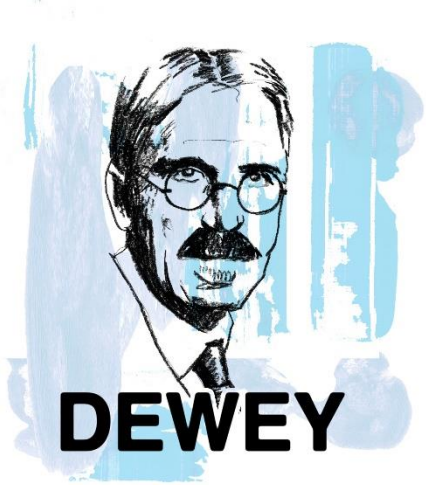
2. To evaluate = inquiry and judgement (beyond instict or preference). Value judgments arise when valuings are subjected to appraisal, when one asks whether one ought to value something.

"… valuation takes place only when there is something the matter; when there is some trouble to be done away with, some need, lack, or privation to be made good, some conflict of tendencies to be resolved by means of changing existing conditions. This fact in turns proves that there is present an intellectual factor – a factor of inquiry – whenever there is valuation" (Dewey, Theory of Valuation, 34)

Technology and value change

Why do values change?

Because new technologies create new morally problematic situations that require new moral responses and hence new values



Why value change & technology?

Technologies lead to new types of *consequences* that require new evaluative dimensions and therefore new values (e.g. privacy, sustainability) to evaluate sociotechnical systems;

Technologies offer new *opportunities* (e.g. to protect homes against earthquakes) that lead to new moral obligations and therefore new values;

Technologies create new *moral choices and dilemmas* where previously were no choices (e.g. predictive genetics) that require new values;

Technologies lead to new *experiences* (e.g. friendship online) that lead to new values or change existing values.

At least three types:

Types of technologically induced value change

(Van de Poel and Kudina, 2020)

- Value dynamism
- Value adaption
- Value emergence

1. Value dynamism -

(re)interpretation or modification of a value when applying it to a new situation

2. Value adaption –

- a structural modification of a value becomes general and carries over to other situations

3. Value emergence –

a value emerges from inquiries about problematic situations without the new value being an adaptation of a pre-existing value

Breakout groups
(7 min)

1. What kind of value changes due to technology have you observed in your lifetime? Maybe even due to robotics and AI?

2. Because of technologies around us, which present values do you think we should collectively try to change and in which way?

Write a few reflections in your group chat. Before exiting the groups, **copy and share some of your reflections in the general chat** – we will discuss this together

So if technologies are morally significant…

… where does this leave human agency? Autonomy? Independence?

… how can we hold people responsible—morally AND legally—for something that is not entirely up to them?

Retrieved from https://www.nytimes.com/2012/07/06/opinion/Heng-cartoon-who-is-to-blame-for-the-fukushima-nuclear-crisis.html

“Some misunderstandings about the moral significance of technology” (Verbeek, 2008)

Where to find morality? In people? In technologies? In the practices of people in the sociocultural world, mediated by technologies

Moral agency (classic): intentionality and freedom as conditions to be held morally responsible

Technologies are NOT moral agents themselves - informing the decisions, not determining them

Tech does not “have” intentionality or freedom, but it mediates what they are for people

Rethinking the classical notion of moral agency – not an individual, but a hybrid human-technological affair, no separation of (human) subjects and (nonhuman) objects “in themselves”

Intertwinement of technology & morality

1. Norms and values as interactional: we always create moral worlds by relating everything and everyone around us within our experiential-cultural-historical frame

2. People enact values in practices, in the use of technologies

3. Technologies help to shape also moral decisions and actions

Therefore,

Technologies are inalienably engaged in moral decision-making: technologies co-shape and even change morality

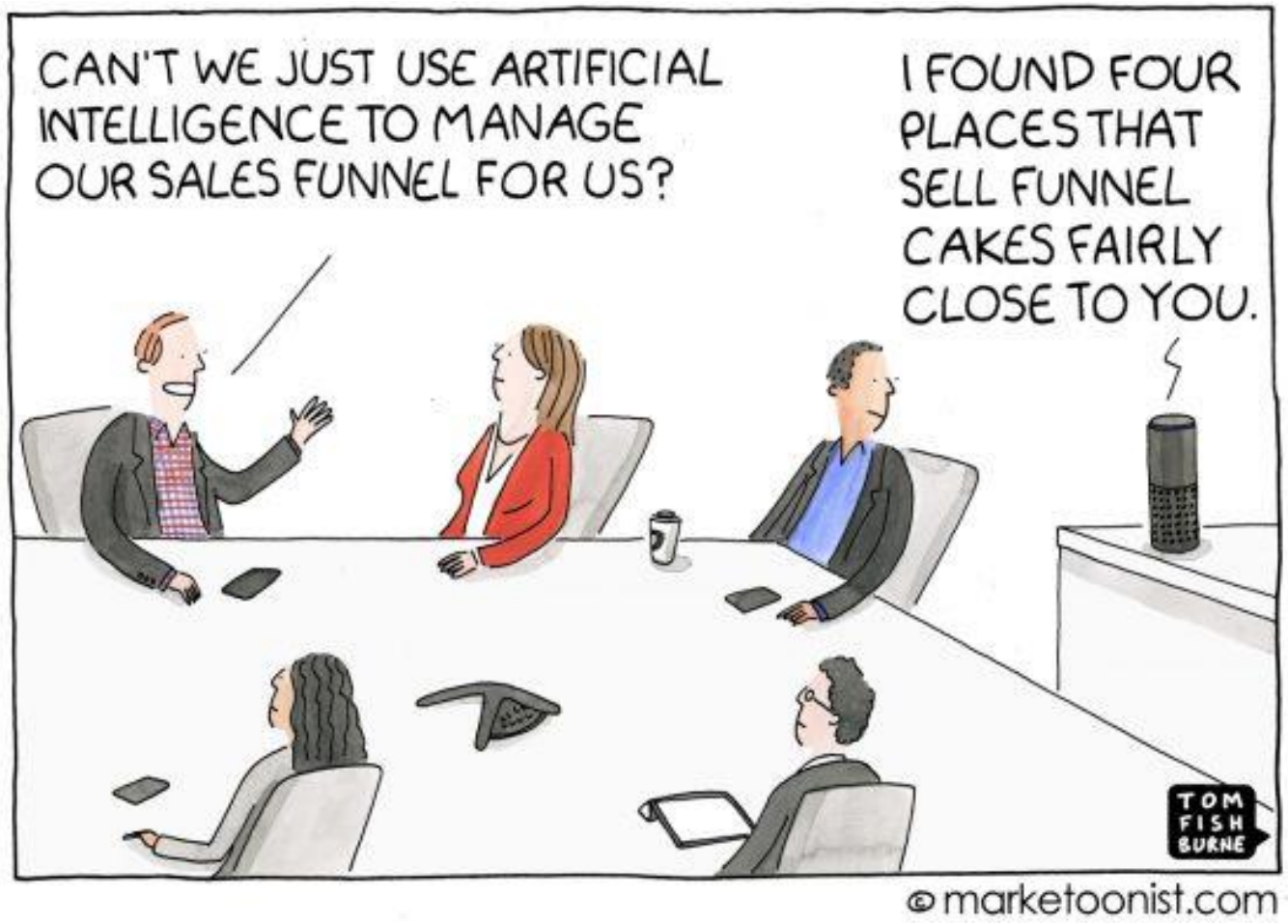
Verbeek (2008; 2011); Swierstra et al. (2009); Kudina (2019); Van de Poel (2011, 2018)

Next class

- Can we imbed values in design? Is this even a correct question to ask?
- What about moral overload?

Checklist before next lecture:

- ❑ Complete Quiz 2
- ❑ 4ECTS – Submit essay proposal (follow instructions in the Course Manual & Essay folder on BS)



Thank you!

o.kudina@tudelft.nl

olyakudina.com

@OlyaKudina