

KV Ethik und Gender Studies



Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Martina Mara

WS 2019
22.10.2019



7 Key Requirements for Trustworthy AI

7 key requirements for ethical AI:

Human agency and oversight

Technically robustness & safe

Privacy and data governance

Transparency

Diversity, non-discrimination and fairness

Societal and environmental wellbeing

Accountable

**Will your algorithms pass the test?
Create AI humans can trust.**

#AI #ArtificialIntelligence





"Mm-hmm."



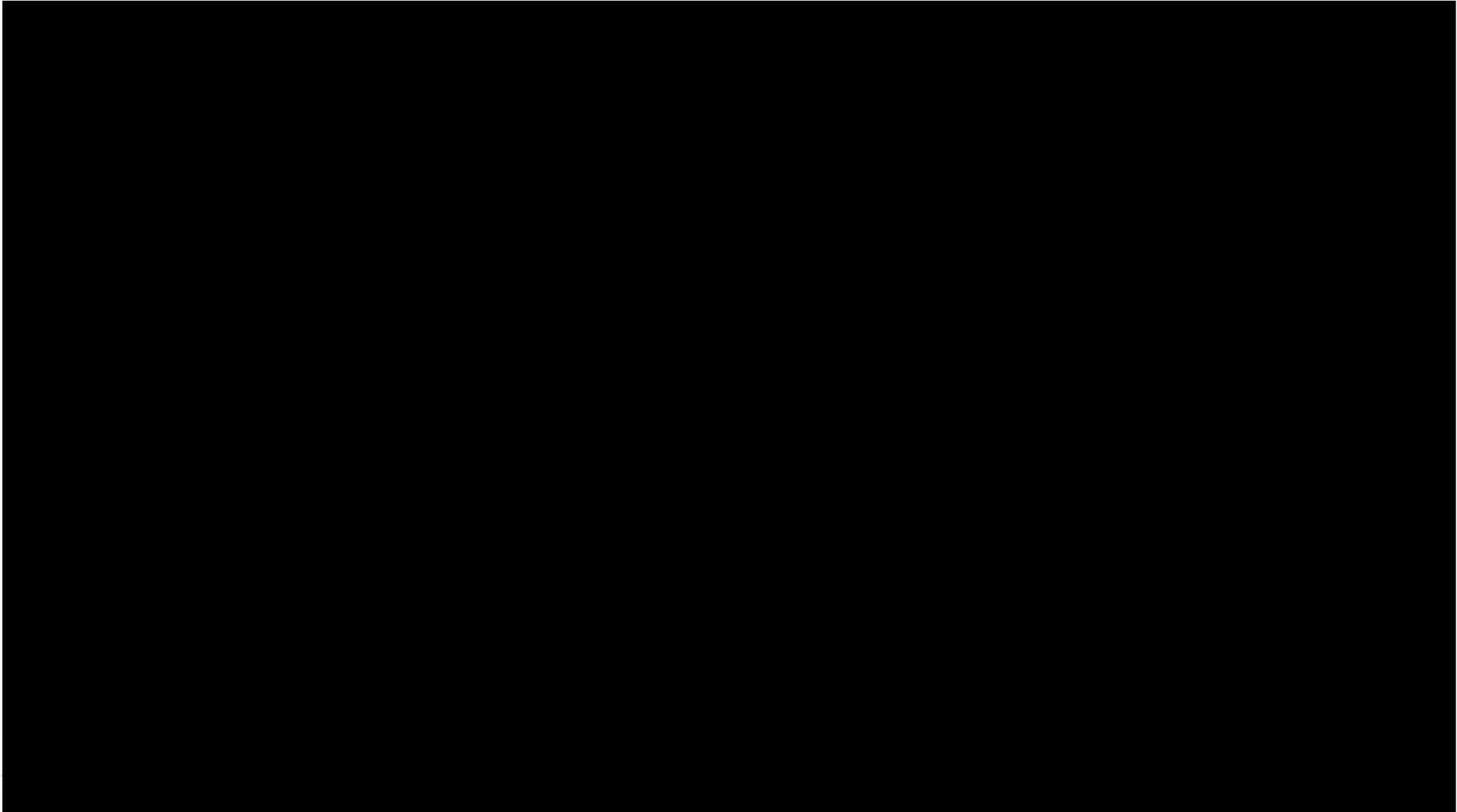
5) Diversity, non-discrimination and fairness

Schlüsselaspekte:

a) Vermeidung unfairer Verzerrungen (Bias)

Erkennbare diskriminierende Verzerrungen sollten nach Möglichkeit in der Phase der Datenerhebung beseitigt werden. Darüber hinaus sollte die Beschäftigung von Personen mit unterschiedlichen kulturellen Hintergründen, die aus verschiedenen Fachrichtungen stammen, gefördert werden, da dies zur Gewährleistung der Meinungsvielfalt beitragen könnte.

Gesichtserkennungs-Software weist eine Fehlerrate von 1% bei hellhäutigen Männern und 35% bei dunkelhäutigen Frauen auf (Buolamwini, 2018):



5) Diversity, non-discrimination and fairness

Schlüsselaspekte:

b) Barrierefreiheit und universeller Entwurf (Design)

Insbesondere in der Beziehung zwischen Unternehmen und Verbrauchern sollten die Systeme benutzerorientiert und so gestaltet sein, dass alle Menschen unabhängig von ihrem Alter, Geschlecht, ihren Fähigkeiten oder Merkmalen KI-Produkte oder -dienstleistungen nutzen können.

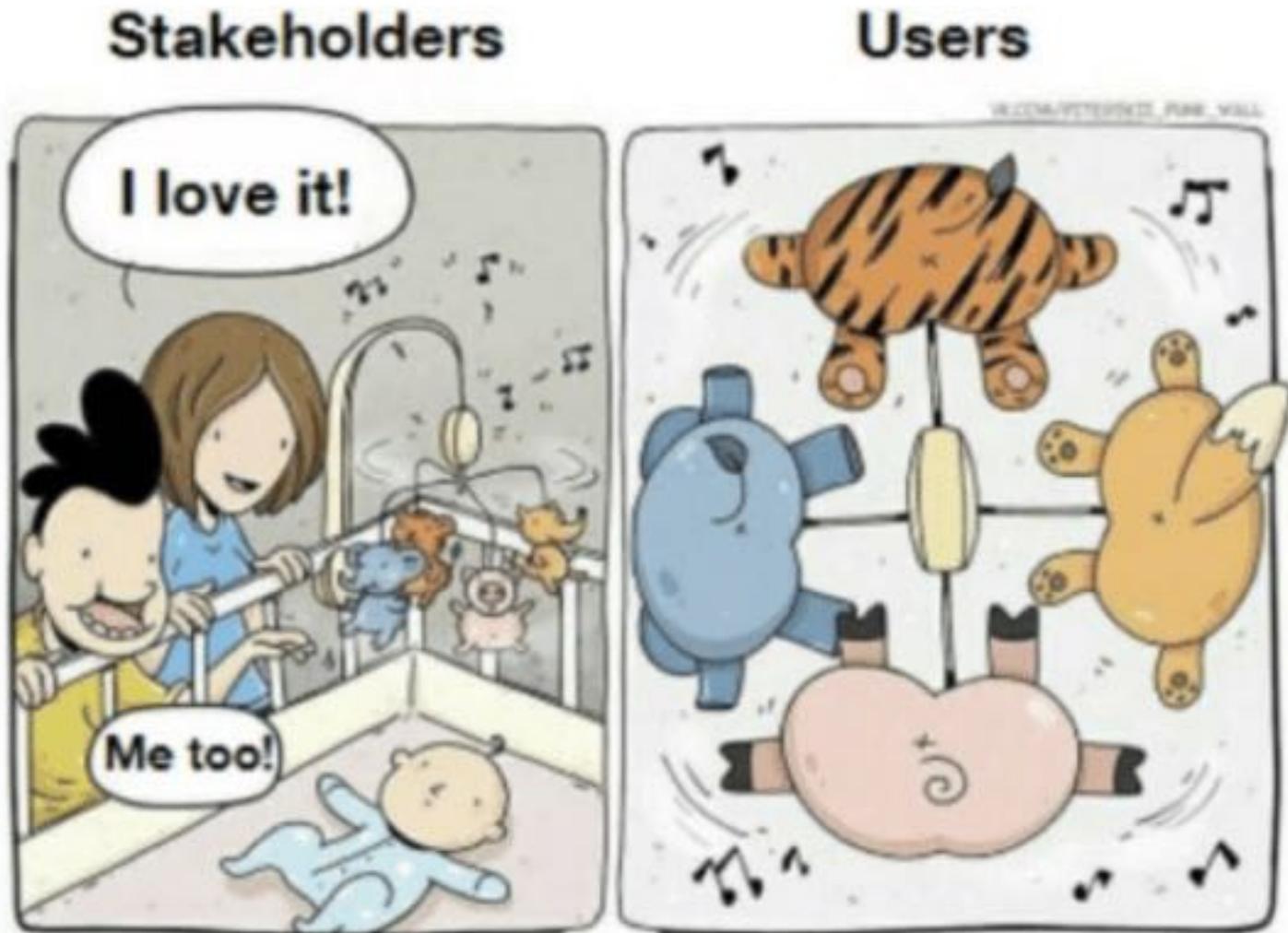
5) Diversity, non-discrimination and fairness

Schlüsselaspekte:

c) Beteiligung der Interessenträger

Zur Entwicklung vertrauenswürdiger KI-Systeme ist eine Konsultation der Interessenträger ratsam, die möglicherweise während des gesamten Lebenszyklus des Systems direkt oder indirekt von diesem betroffen sind.

Stichwort Stakeholder-Involvierung



Oranex



6) Societal and environmental wellbeing

Schlüsselaspekte:

a) Nachhaltige und umweltfreundliche KI

Die Erforschung von KI-Lösungen in Bezug auf globale Belange, wie z. B. die Ziele für eine nachhaltige Entwicklung, sollte ausgebaut werden. KI-Systeme versprechen, einen Beitrag zur Bewältigung einiger der drängendsten gesellschaftlichen Probleme zu leisten, doch muss sichergestellt sein, dass dies auf möglichst umweltfreundliche Art und Weise geschieht.



SUSTAINABLE DEVELOPMENT GOALS



Limitation: Wieviel Strom braucht eigentlich KI?

Creating an AI can be five times worse for the planet than a car



TECHNOLOGY 6 June 2019

By [Donna Lu](#)



6) Societal and environmental wellbeing

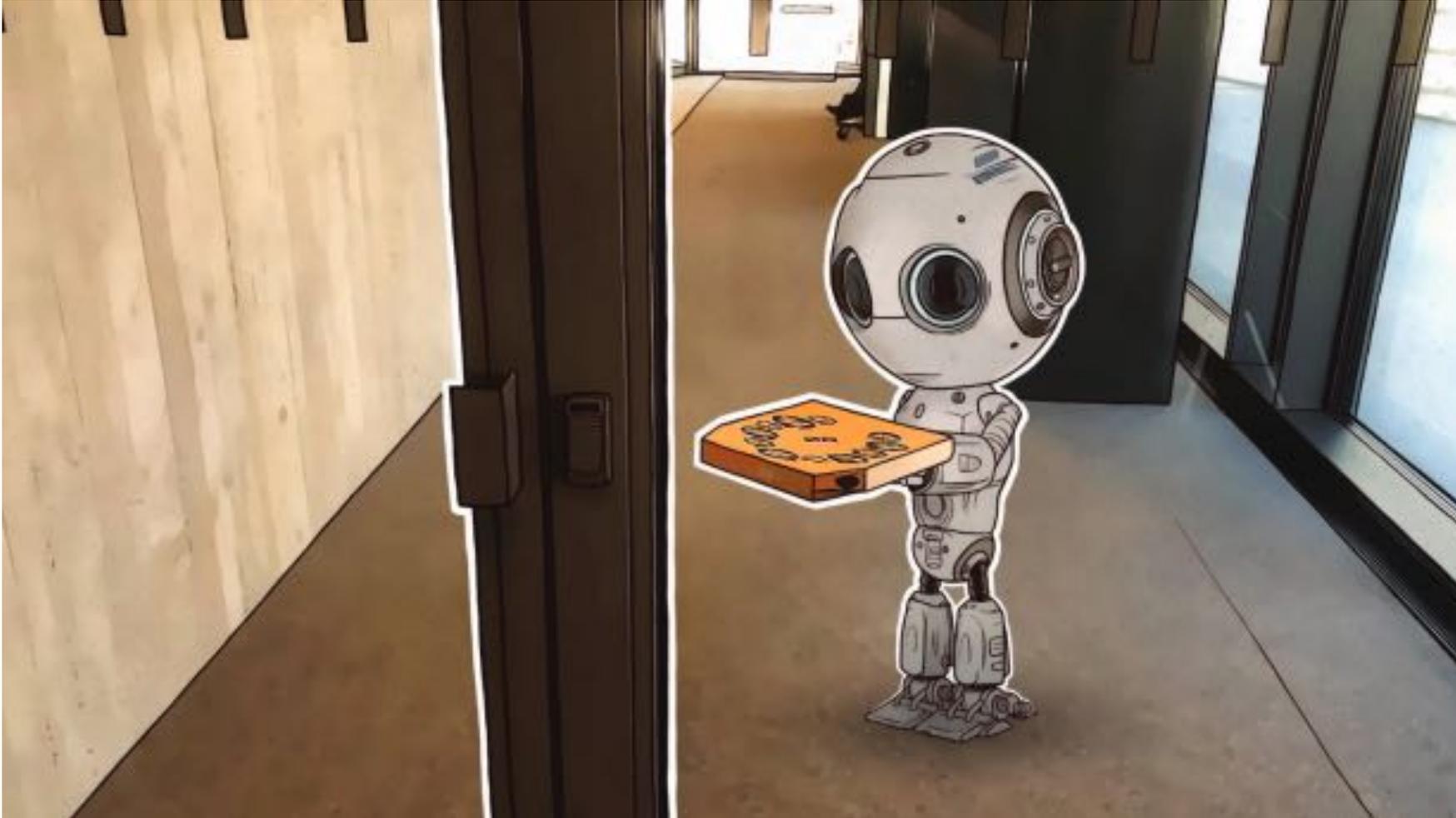
Schlüsselaspekte:

b) Soziale Auswirkungen

So wie KI-Systeme zur Verbesserung sozialer Kompetenzen eingesetzt werden können, können sie auch zu deren Verschlechterung beitragen. Dies könnte sich ebenso auf das körperliche und geistige Wohlergehen der Menschen auswirken.

**Meet your
new friend**

40% der MitarbeiterInnen öffneten einem niedlichen sozialen Roboter die Tür zum Sicherheitstrakt eines Bürogebäudes (Belpaeme et al, 2019):



6) Societal and environmental wellbeing

Schlüsselaspekte:

c) Gesellschaft und Demokratie

Der Einsatz von KI-Systemen sollte insbesondere im Zusammenhang mit dem demokratischen Prozess – einschließlich der politischen Entscheidungsfindung und in Bezug auf Wahlen – sorgfältig geprüft werden.

7) Accountability

Schlüsselaspekte:

a) Nachprüfbarkeit

Die Bewertung von KI-Systemen durch interne und externe Prüfer und das Vorliegen solcher Bewertungsberichte kann beträchtlich zur Vertrauenswürdigkeit der künstlichen Intelligenz beitragen.

Die externe Nachprüfbarkeit sollte insbesondere bei Anwendungen sichergestellt sein, die sich auf die Grundrechte auswirken, sowie bei sicherheitskritischen Anwendungen.

7) Accountability

Schlüsselaspekte:

b) Minimierung negativer Auswirkungen und Meldung

Der Einsatz von Folgenabschätzungen (z. B. „*Red Teaming*“ oder Formen der Algorithmen-Folgenabschätzungen) sowohl vor als auch während der Entwicklung, Einführung und Nutzung von KI-Systemen kann hilfreich sein, um negative Folgen möglichst gering zu halten. Diese Bewertungen müssen in einem angemessenen Verhältnis zu dem Risiko stehen, das die KI-Systeme darstellen. Ein angemessener Schutz für Informanten, Nichtregierungsorganisationen, Gewerkschaften und andere Stellen, die berechtigte Bedenken hinsichtlich eines KI-gestützten Systems äußern, muss gewährleistet sein.

7) Accountability

Schlüsselaspekte:

c) Rechtsmittel

Sollte es schließlich doch zu ungerechten und nachteiligen Auswirkungen kommen, sollten Vorkehrungen für einen angemessenen Rechtsschutz getroffen werden. Zu wissen, dass Rechtsbehelfe möglich sind, wenn etwas schief geht, ist der Schlüssel zur Vertrauensbildung.

Wer ist noch keiner Gruppe zugeteilt?

- Ca. 8-10 Personen je Gruppe - Einteilung bleibt für alle 3 Blöcke bestehen.
- Gruppenname festlegen
- Gruppenmitglieder in Formular eingeben: <http://bit.ly/33cSKWH>



Gruppenübung 1: Dilemma-Diskussion

USE OF AI?



Was ist ein (ethisches) Dilemma?

Ein ethisches Dilemma ist eine problematische Entscheidungssituation, in der zwei mögliche Alternativen zur Verfügung stehen und keine der anderen eindeutig vorzuziehen ist.

Ein Klassiker: Das Trolley-Problem

Situation:

You see a runaway trolley moving toward five tied-up people lying on the main track. You are standing next to a lever that controls a switch. If you pull the lever, the trolley will be redirected onto a side track, and the five people on the main track will be saved. However, there is a single person lying on the side track. You have two options:

- a) *Do nothing and allow the trolley to kill the five people on the main track.*
- b) *Pull the lever, diverting the trolley onto the side track where it will kill one person.*

Question:

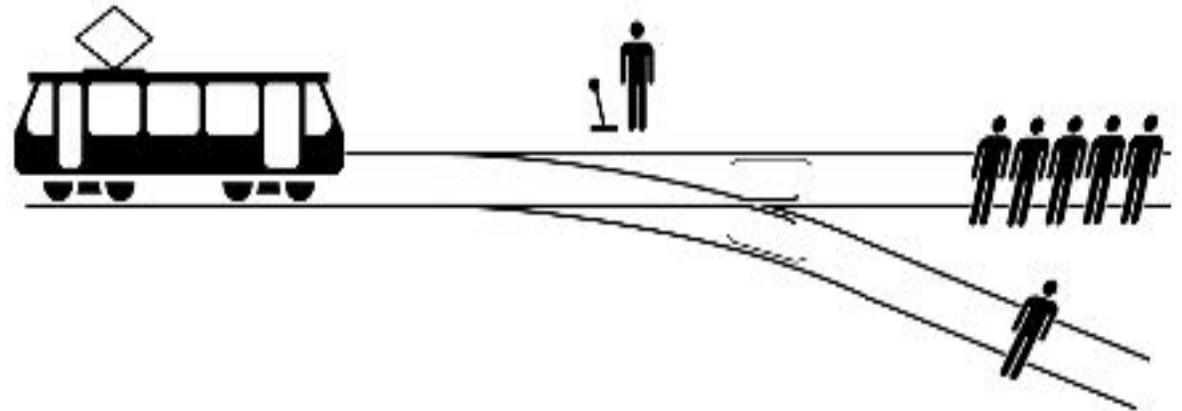
Pull the lever – yes or no?

Exemplary argument PRO:

It's better if instead of 5 people only 1 has to die.

Exemplary argument CON:

Moral distinction between “killing” and “letting die”.



Ein Klassiker: Das Trolley-Problem

BBC
RADIO

4

RIGHT & WRONG

The Trolley Problem

Narrated by



Harry Shearer

Was ist ein (ethisches) Dilemma?

Ein ethisches Dilemma ist eine problematische Entscheidungssituation, in der zwei mögliche Alternativen zur Verfügung stehen und keine der anderen eindeutig vorzuziehen ist.

In unserem Fall:

Ist der Einsatz von KI (definiert als lernendes System, z.B. ML/DL/NN) in einem bestimmten (kritischen) Bereich ethisch vertretbar?

Wieso ja?

Wieso nein?

Zur Methode der Dilemma-Diskussion

- Die Methode der moralischen Dilemmadiskussion (MDD) wurde vor mehr als 40 Jahren von Moshe Blatt & Lawrence Kohlberg (1975) eingeführt und zur Konstanzer Methode der Dilemmadiskussion (KMDD) weiterentwickelt (Lind, 2003).
- Grundannahmen: Moral ist lehrbar; moralisches Denken kann trainiert werden, indem moralische Dilemmata diskutiert und Argumente in Gruppen gegeneinander abgewogen werden; es werden aber geeignete Lernmöglichkeiten benötigt, in denen die TeilnehmerInnen ihre Argumente und Wertvorstellungen frei ausdrücken können.
- (K)MDD ist in der psychologischen und pädagogischen Forschung gut fundiert; positive Effekte auf die Förderung moralisch-demokratischer Kompetenzen wurden in vielen empirischen Studien nachgewiesen.
- Üblicherweise werden MDDs nach einem standardisierten Verfahren durchgeführt (Dilemmageschichte – individuelle Meinungsbildung und Argumentesammlung – Abstimmung – Bildung von Pro- und Contragruppen – Diskussion von Argumenten – finale Abstimmung)
- Unsere Gruppenübung folgt dem Standardprozess nicht ganz, ist aber von der KMDD inspiriert!

Dilemma-Fragen für die Gruppenübung 1:

Jede Gruppe sucht eines der folgenden Themen für die Dilemma-Diskussion aus:

Ist der Einsatz von KI ethisch vertretbar ...

- 1) ... für “Algorithmic Pricing” (z.B. individuelle Preisanpassung bei Flugtickets oder sogar Lebensmitteln)?
- 2) ... um personalisierte “KI-Freunde” zu entwickeln (z.B. für einsame Menschen oder ältere Personen)?
- 3) ... im Militärbereich und für Kriegsmissionen (z.B. intelligente Drohnen für zielgerichtete Angriffe)?
- 4) ... für die Entwicklung manipulativer Social Bots, die User überreden, Geld zu spenden (z.B. für wohltätige Zwecke)?
- 5) ... um juristische Entscheidungen vor Gericht zu fällen (z.B. um die Dauer einer Haftstrafe festzulegen)?

Nächste Schritte

- 1) Koordinieren Sie sich in der Gruppe und teilen Sie sich in eine PRO-Subgruppe und eine CONTRA-Subgruppe.
- 2) Jedes Gruppenmitglied liest sich in das Thema ein und sammelt individuell Argumente PRO oder CONTRA.
- 3) Sprechen Sie sich innerhalb Ihrer Subgruppe ab und stärken Sie Ihre Argumente.
- 4) Vereinbaren Sie einen Termin und führen Sie die Dilemma-Diskussion durch (persönlich oder online). Während der Diskussion wechseln sich Mitglieder der PRO- und CONTRA-Subgruppen mit ihren Argumenten ab und versuchen, Mitglieder der anderen Subgruppe zu überzeugen. Am Schluss findet eine Abstimmung über die finale Gruppenentscheidung statt.
- 5) Verfassen Sie den schriftlichen Diskussions-Bericht auf Basis der Vorlage.
- 6) Reichen Sie den Bericht via Moodle ein (Deadline 4.11., 18 Uhr) und erhalten Sie Ihre Note.
- 7) Peer Review: Später werden Sie eine Abgabe einer anderen Gruppe lesen und Feedback geben)



How to argue & discussion rules

Tip: You can base your arguments, among other things, on

- ... the assumed effects of the decision (negative or positive)
- ... relevant ethical principles (e.g. EU Guidelines)
- ... the viewpoint of a certain group of stakeholders/users
- ... law, regulations, what authorities would say
- ... examples from the past

Discussion rules:

- Mutual respect: arguments are fought only by counter-arguments – not by talking down or insulting others
- Equal say: everyone has the same right to make an argument, each person's argument has equal weight
- Argue in turns: pro and con arguments alternate one another (at least in the beginning of your discussion)

Abgabe: Schriftlicher Diskussions-Bericht

- Jede Gruppe erstellt einen schriftlichen Diskussions-Bericht und gibt diesen via Moodle ab (Abgabe durch ein Mitglied reicht)
- Benennung Bericht: "thema_gruppenname_uebung1"
- Deadline ist **MO, 4.11.2019, 18 Uhr**
- Eine Vorlage für die Erstellung des Berichts steht auf Moodle bereit

Kommunikation/Koordination in der Gruppe

Soll für jede Gruppe ein Moodle-Forum (nur sichtbar für Gruppenmitglieder) angelegt werden?

Diskussions-Bericht gemeinsam anfertigen

Bitte organisieren Sie selbst, wie Sie den Bericht innerhalb der Gruppe gemeinsam erstellen.

Beispielsweise können Sie aber eines der folgenden Tools für die simultane Bearbeitung des Textdokuments verwenden:

- Google Docs
- MS Office Online
- Dropbox Paper
- Cryptpad
- Nuclino

Evaluierung des schriftlichen Berichts

- Welche Entscheidung Sie am Ende als Gruppe treffen (Pro oder Contra) hat keinen Einfluss auf die Benotung! Eine eindeutig korrekte Antwort gibt es bei Dilemma-Diskussionen per Definition nicht.
- Die Note wird dadurch bestimmt, wie umfassend Pro- und Contra-Argumente gesammelt wurden, wie nachvollziehbar über die Nutzung von Quellen und den gesamten Diskussionsprozess berichtet und reflektiert wird, wie gut die schriftliche Qualität ist (klare, kurze Formulierungen) und wie nachvollziehbar es ist, dass alle Gruppenmitglieder mitgewirkt haben.



Termin am 29.10.20: Keine Präsenz gefordert

- Nutzen Sie die Zeit für die Gruppenübung.
- Angebot: Individuelle Konsultation für Gruppen
 - 10-Minuten-Slots bei Prof. Mara für Gruppen, die das in Anspruch nehmen möchten (persönlich/f2f, telefonisch oder via Skype)
 - Ein Link zu einem Doodle-Formular mit verfügbaren Slots wird bis MI, 23.10., in Moodle hochgeladen
 - Informationen zu Raumnummer, Telefonnummer und Skype-ID im Doodle-Formular
 - Wenn Sie Ihre Gruppe in Doodle für einen Termin eintragen, fügen Sie Ihrem Gruppennamen bitte als Kürzel „f2f“, „tel“ oder „skype“ hinzu.
 - Nicht alle Gruppenmitglieder müssen an der Konsultation teilnehmen.

Fragen?

KV Ethik und Gender Studies



Danke!

Nächster Termin:
Gruppenarbeit mit Online-Konsultation

