

# KV Ethik und Gender Studies



Univ.-Prof.<sup>in</sup> Dr.<sup>in</sup> Martina Mara

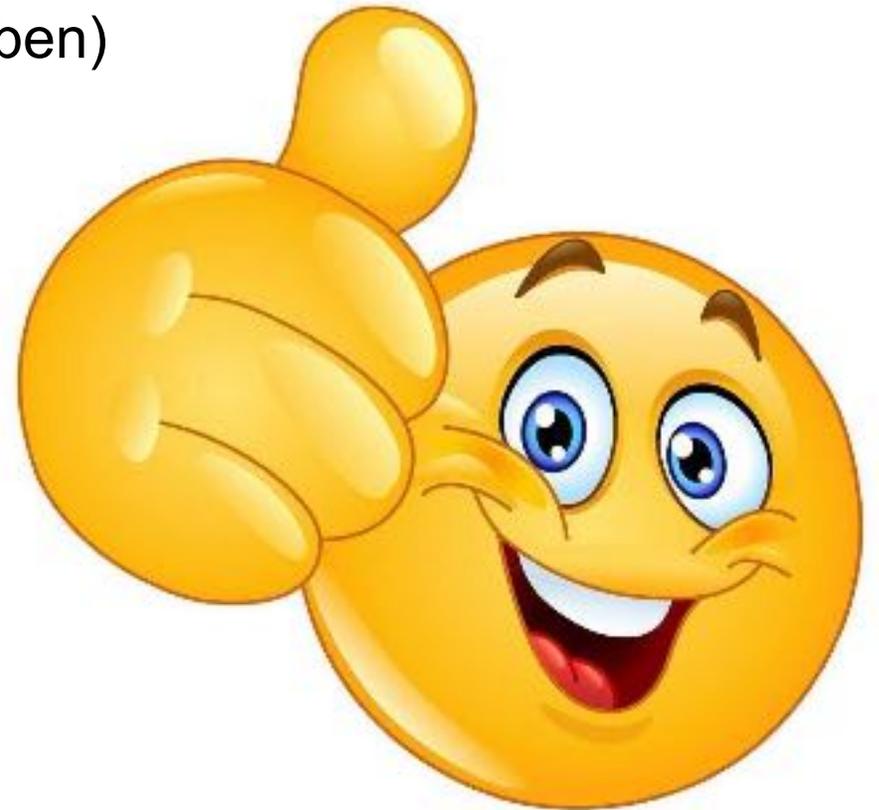
WS 2019

05.11.2019



# Resümee: Gruppenübung Dilemma-Diskussion

- Alle haben ihren Diskussions-Report abgegeben (yay)!
- (2 Gruppen haben Entwurf nicht per Button abgegeben)

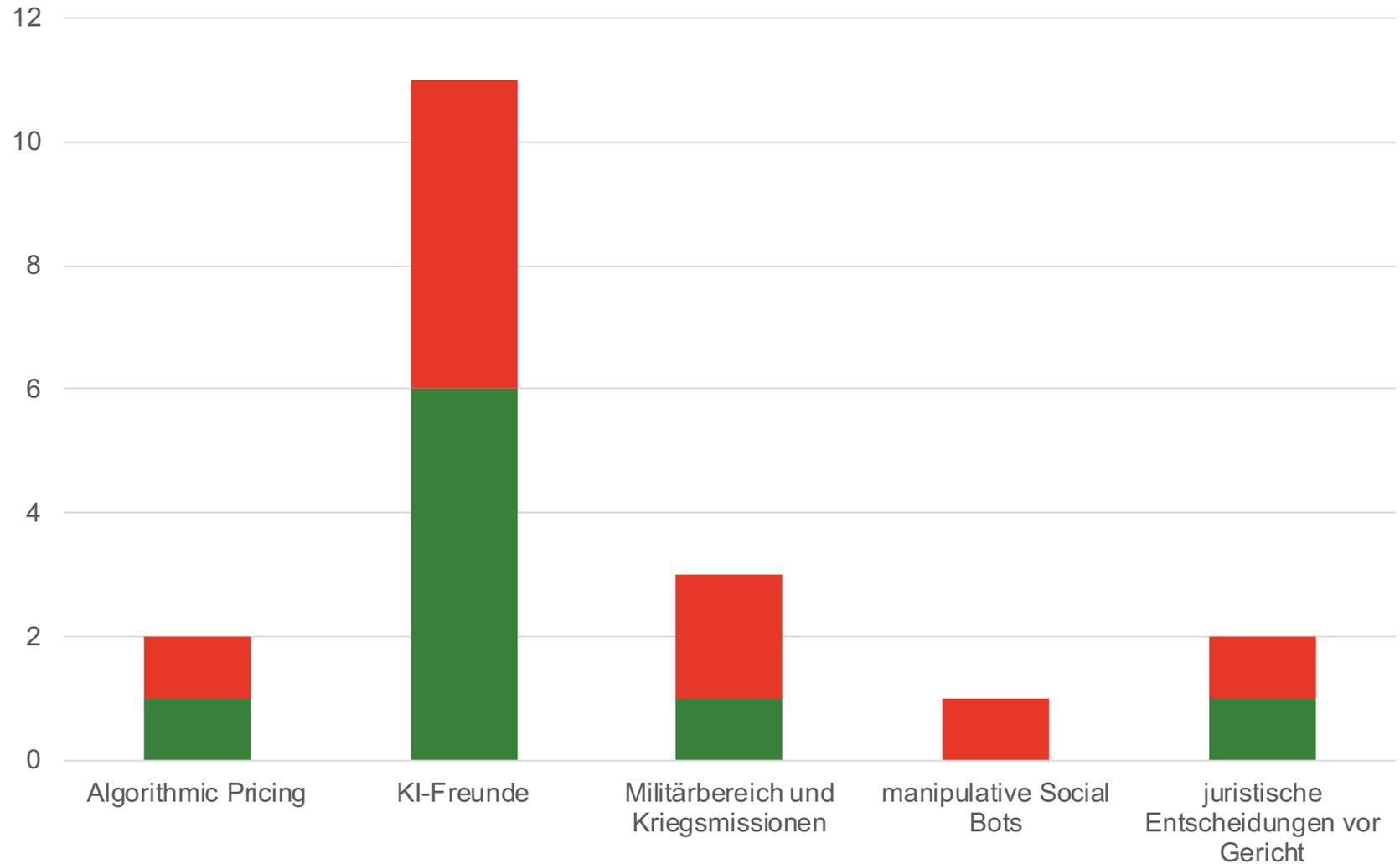


# Warum Dilemma-Diskussion als Gruppenübung?

- Auseinandersetzung mit ethischen Herausforderungen neuer Technologien; erfahren, dass es häufig nicht den “einen richtigen Weg” gibt
- Unterschiedliche Perspektiven unterschiedlicher Interessens- und NutzerInnengruppen einnehmen und nachvollziehen können.
- Über potenzielle Konsequenzen, Risiken und Chancen von KI in einem bestimmten Einsatzbereich nachdenken (“utilitaristischer” Ansatz).
- Abwägen von (stärkeren und schwächeren) Argumenten trainieren.
- Kollaboratives Arbeiten und argumentgestützten Diskurs trainieren.
- Gemeinsam zu einer Entscheidung kommen, die alle in der Gruppe mittragen.

# **Ergebnisse der Dilemma-Diskussionen im Überblick**

# Analyse: Welches Thema wurde wie oft gewählt?



# Welche Entscheidungen wurden getroffen?

## Algorithmic Pricing:

PRO:

- „Algorithmic Pricing würde in erster Linie vielleicht nicht Gleichheit widerspiegeln, aber dafür Gerechtigkeit, welche in mehreren Aspekten über der Gleichheit stehen sollte.“
- „AP führt zu marktgerechteren Preisen, die sich in Echtzeit an Angebot und Nachfrage anpassen. Ruinöse Preis-Spiralen werden vermieden, die Märkte stabilisiert.“
- „Genauere Preisvorhersagen sind algorithmisch möglich. Dadurch kann der perfekte Kaufzeitpunkt für die Kunden berechnet werden, um möglichst viel Geld zu sparen.“

# Welche Entscheidungen wurden getroffen?

## Algorithmic Pricing:

### CONTRA:

- „Datenschutz gefährdet, Informationen werden benutzt die sensibel sind und nicht veröffentlicht werden sollen“
- „Intransparenz: Der Kaufende weiß nicht, was der Algorithmus über einen weiß. Damit einhergehend ein geringeres Vertrauen vom Menschen dem Algorithmus gegenüber, das sich wieder im Kaufwillen ausdrückt. Diskriminierung kann schwer erkannt werden.“

# Welche Entscheidungen wurden getroffen?

## Algorithmic Pricing:

### getroffene Entscheidung:

1x NEIN [Unsere gemeinsame abschließende Antwort auf die Frage, ob der Einsatz von KI ethisch vertretbar für den Bereich “Algorithmic Pricing” ist, ist definitiv NEIN, weil die Ethik gegenüber der Wirtschaft überwiegt.]

1 x JA [Ja, weil durch AP im Optimalfall individualisierte Preise festgelegt werden können, die die soziale Ungerechtigkeit mindern,...; Wir sind also für AP, solange bei der Implementierung auf Transparenz und Privatsphäre nach den EU Ethics Guidelines geachtet wird.]

# Welche Entscheidungen wurden getroffen?

## Militärbereich und Kriegsmissionen

PRO:

- „Die KI kann aber zwischen Zivilisten und Feinden vielleicht sogar besser unterscheiden als Menschen. Beispiel: Militär- und Zivilflugzeuge“
- „Bessere Wahrnehmung durch Sensoren (Infrarot, Wärmebild, Radar, Strahlung,...)“
- „Maschinen sind unempfindlicher gegenüber Umwelteinflüsse.(Hitze, Unterwasser)“

# Welche Entscheidungen wurden getroffen?

## Militärbereich und Kriegsmissionen

### CONTRA:

- „Fehleranfällig. Bei einer Nicht ganz genauen Identifizierung, können leicht Fehlentscheidungen getroffen werden.“
- „Außerdem stellt sich die Frage, wann ein Krieg verloren geht? Wenn alle Roboter zerstört sind, kämpfen dann die Menschen weiter?“
- „Außer Kontrolle. Da die Menschen unterlegen sind, kann ein System, dass sich selbständig macht gefährlich werden (Terminator, Detroit,...)“
- „Niedrige Hemmschwelle für Militäroperationen, da ein Drohneneinsatz billiger als ein Truppentransport ist und auch das Risiko geringer ist (im schlimmsten Fall geht eine Maschine kaputt)“

# Welche Entscheidungen wurden getroffen?

## Militärbereich und Kriegsmissionen

getroffene Entscheidung:

- 2 x NEIN [Nein, weil ein zu hohes Missbrauchsrisiko von zum Beispiel Terroristen oder anderen etwaigen Menschen besteht. Des Weiteren ist der momentane Entwicklungsstand der Künstlichen Intelligenzenrisikoreich, da viele Aspekte wie zum Beispiel moralische Fragen noch nicht zur Gänze geklärt worden sind, weshalb der Einsatz im Krieg nicht empfehlenswert ist.]
- 1 x JA [Laut unserem Entschluss ist eine KI, die für humanitäre Zwecke eingesetzt wird (Abwehrsysteme, Verpflegung in Krisensituationen,...), durchaus sinnvoll, da die Sicherheit der Menschen, dadurch stark gewährleistet wird.]

# Welche Entscheidungen wurden getroffen?

## Manipulative Social Bots

PRO:

- „Chat Bots können zur Informationsgewinnung, Schadensabwicklung, Verbesserung des Kundenservices (Support) beitragen“
- „Manipulation mit guter Absicht bei Hilfsorganisationen“

# Welche Entscheidungen wurden getroffen?

## Manipulative Social Bots

CONTRA:

- „Wettbewerbsvorteil für größere Organisationen mit höherem Budget (eventueller Machtmissbrauch)“
- „Meinungsverzerrung durch Fake-News“
- „Ausbeutung der Gefühle der Menschen“

# Welche Entscheidungen wurden getroffen?

## Manipulative Social Bots

getroffene Entscheidung:

1 x NEIN [Nein, wegen der Ausnutzung des Vertrauens der Zielperson. Außerdem besteht die Möglichkeit des Missbrauchs der Social Bot (z.B.: für Scamming).]

# Welche Entscheidungen wurden getroffen?

## Juristische Entscheidungen vor Gericht

PRO:

- „Gleiche und faire Beurteilung von ähnlichen Fällen kann gewährleistet werden.“
- „Verkürzung der Wartezeiten auf ein Urteil durch die KI.“
- „Eine KI hat eine lange Lebensdauer. Sie wird nicht altern und kann jederzeit verbessert werden, indem sie dazulernt.“

# Welche Entscheidungen wurden getroffen?

## Juristische Entscheidungen vor Gericht

CONTRA:

- „Es ist zu wenig Wissen vorhanden, welche Daten für die Urteilsfindung signifikant sind.“
- „KI weiß nicht, dass sie etwas nicht weiß – Juristen schon. Ein Richter kann z.B Gutachten oder Ähnliches anordnen, wenn im Informationen fehlen und er keine Entscheidung treffen kann/mag“
- „KI kann falsch trainiert werden, so dass sie rassistisch ist“

# Welche Entscheidungen wurden getroffen?

## Juristische Entscheidungen vor Gericht

getroffene Entscheidung:

- 1 x NEIN [Nein, da die KI nicht so weit ist, dass sie so wichtige Entscheidungen treffen kann und die Gefahr von falschen Urteilen zu hoch ist. Um solche Fehler zu vermeiden, kann ein Mensch darüber schauen, was wiederum die KI unnötig macht.]
- 1 x JA [JA, falls die KI nur unterstützend eingesetzt wird und nicht die menschlichen Richter komplett ersetzt. Zu einem späteren Zeitpunkt, wenn die KIs ausführlich getestet und optimiert wurden (z.B. Daten sind von Stereotypen bereinigt), können wir uns auch vorstellen, dass die KI bei wenig komplexen Fällen selbstständig Urteile fällen kann.]

# Welche Entscheidungen wurden getroffen?

## KI-Freunde:

PRO:

- „Lernen von Sozial Kompetenzen für Menschen mit Verhaltensstörungen oder Autismus.“
- „Können als Freunde für Menschen in isolierten Umgebungen dienen. (zum Beispiel im Weltraum)“
- Mithilfe von KI-Freunden ist es einfacher, sich über private und vertrauliche Probleme zu unterhalten. Einem KI-Freund kann man alles anvertrauen, was man anderen Personen nicht anvertrauen würde, da es beispielsweise Auswirkungen auf die persönliche Beziehung zu den Personen hat (z.B. Depressionen, Ängste, etc....).
- Menschen mit langen Krankenhausaufenthalten können KI Freunde auch gut tun, da sie etwas Beschäftigung bieten und der Patient 24/7 mit ihnen austauschen könnte was mit Freunden und Verwandten nicht möglich ist, wenn der Patient in der Nacht nicht schlafen kann.

# Welche Entscheidungen wurden getroffen?

## KI-Freunde:

### CONTRA:

- „Ein künstlicher Freund kann einem Mensch nicht die Zuneigung geben, wie ein echter Mensch es tun könnte. Zum Beispiel bei einer Umarmung - Ein Roboter ist kalt und hart, ein Mensch aus Fleisch und Blut warm und weich.“
- „Juristischer Aspekt: Angenommen der KI-Freund bestärkt seine menschliche Bezugsperson darin ein Verbrechen zu begehen. Wie klärt man dann die Schuldfrage? Ist der KI-Freund dafür mitverantwortlich? Ist die Schuld auf jene zurückzuführen, die die Algorithmen für die KI realisiert haben?“
- „Anonymität ist nicht gewährleistet, Daten können an Dritte weitergegeben werden.“
- „KI's lernen von Trainingsboten. Was ist, wenn die KI dadurch manipuliert wird und dadurch Leute ebenfalls ins Negative manipuliert.“

# Welche Entscheidungen wurden getroffen?

## KI-Freunde

### getroffene Entscheidung:

5 x NEIN [Nein, weil KIs derzeit noch nicht ausgereift genug sind. Falls diese aber weiterentwickelt und reguliert werden, wäre der Einsatz für ältere Menschen, Personen mit Depressionen oder Behinderte denkbar]

6 x JA [Ja, da KI-Roboter eine große Hilfe für einsame, alte und kränkliche Personen sein können, solange eine ordentliche Gesetzgebung bezüglich der zu trainierenden Datensätzen, die ethischen Richtlinien der EU bezüglich KI eingehalten werden und, wie bei der Verschwiegenheitspflicht von Psychologen, persönliche und gesundheitliche Daten von Personen erst dann weitergegeben werden, wenn akute Gefahr besteht.]

# Spannendes Thema “KI-Freunde”

Ist “Freundschaft” mit KIs/Robotern wirklich möglich?

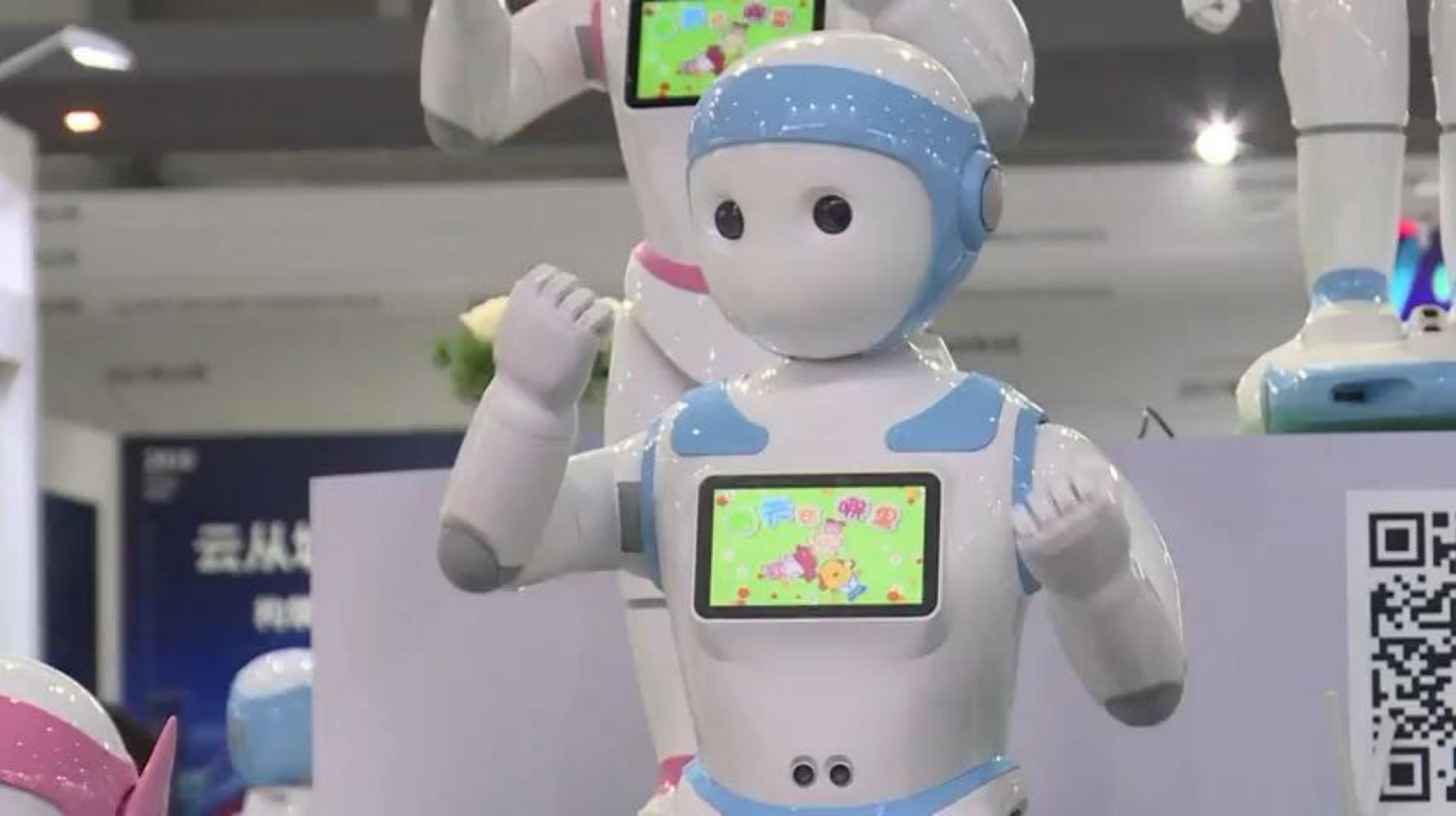
”Freundschaft” mit Maschinen wird immer asynchron/unidirektional bleiben, da diese Empathie und Verständnis klarerweise nur simulieren können.

Die Frage lautet also:

Wollen wir uns auf diese Simulation einlassen?

Welche Konsequenzen hätte dies? Potenzielle Chancen und Risiken?

Unterschiedlich für unterschiedliche Zielgruppen?



Paro



Gatebox: Living with your favorite character



Gatebox: Living with your favorite character



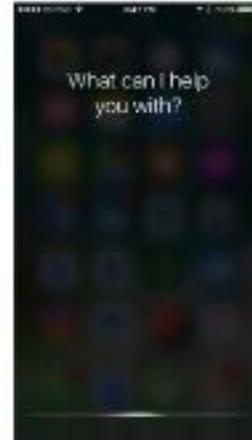
# Siri, Alexa & Co



**AMAZON'S ALEXA**



**GOOGLE'S ASSISTANT**

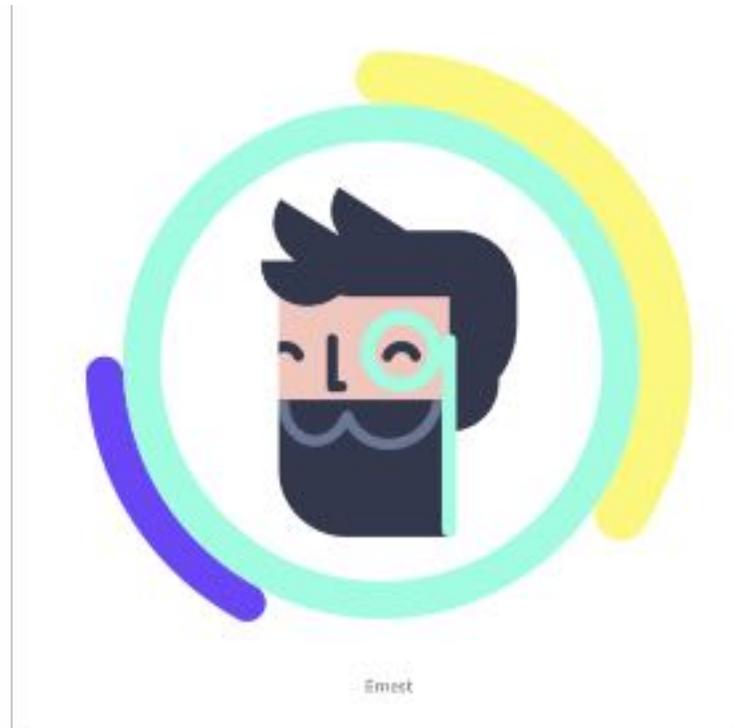


**APPLE'S SIRI**



**MICROSOFT'S CORTANA**

# Siri, Alexa & Co



# Wirkung geschlechtsspezifischer Roboter-Designs

Gleicher Roboter „Flobi“, einmal mit kurzen, einmal mit langen Haaren (Eyszel & Hegel, 2012):

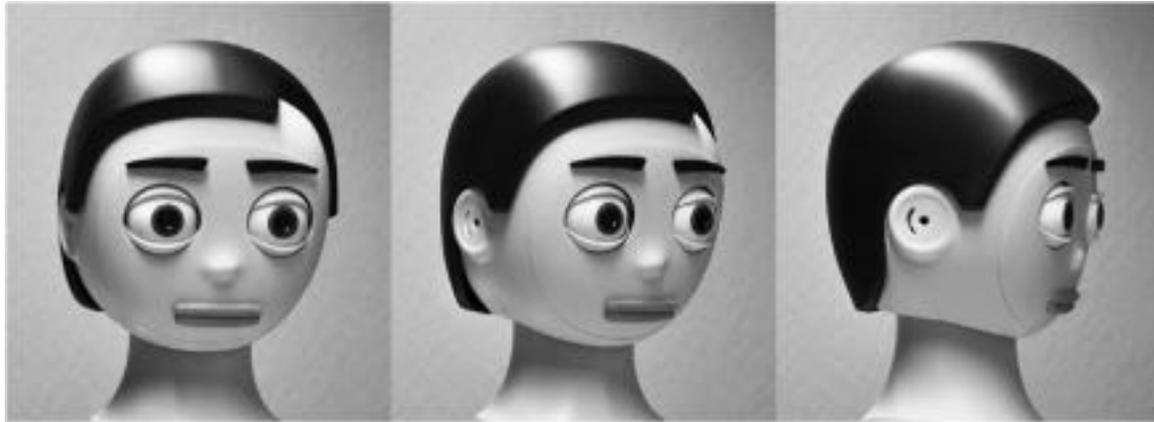


Figure 2. Male robot type: short hair.

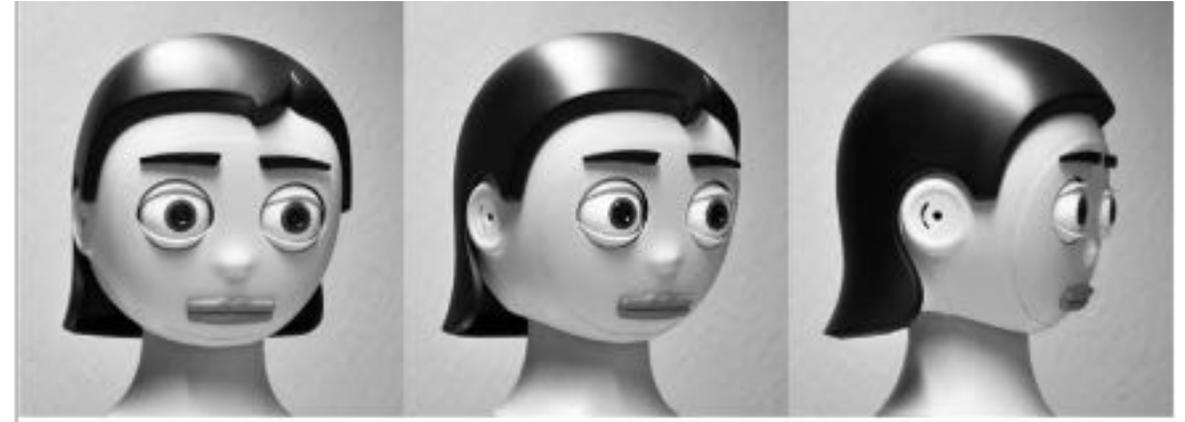


Figure 1. Female robot type: long hair.

## Wahrnehmung durch StudienteilnehmerInnen:

handlungsorientiert

*Passende Anwendungen:*

Reparieren technischer Geräte, Überwachung

## Wahrnehmung durch StudienteilnehmerInnen:

beziehungsorientiert

*Passende Anwendungen:*

Sozialbereich, Haushalt

# Aber das sind doch nur Maschinen ...

Natürliche Tendenz zum **Anthropomorphismus**:

Menschen neigen dazu, auch in (animierten) Objekten soziale und internationale Agenten zu sehen

Theorie der **Media Equation** (Reeves & Nass):

Sowohl Laien als auch Expertinnen reagieren in sozialer Weise auf Computer; sozialpsychologische Phänomene (z.B. Ingroup/Outgroup-Effekte oder Höflichkeit) lassen sich auf Mensch-Computer-Interaktion übertragen

Stimmt das **auch umgekehrt**? Wirken Schemata, die wir in Interaktion mit Computern/Robotern/Assistenzsystemen erlernen, auch auf interpersonelle Interaktionen und Wahrnehmungen zurück?

# Einbindung unterschiedlicher “Blickwinkel” in Technologiegestaltung



# **Nach wie vor große Notwendigkeit von Diversität in der Technik**

## **In der (soziodemografischen) Zusammensetzung von Teams**

Geschlecht, Alter, ethnischer Background, Menschen mit Beeinträchtigung

## **In Expertisen, Perspektiven, Mindsets**

Stärkere Zusammenarbeit verschiedener Disziplinen (z.B. Technik+SoWi) wäre wichtig

## **In den Daten, aus denen wir KI lernen lassen**

Sind die Trainingsdaten inklusiv? Sind alle NutzerInnengruppen gleichermaßen repräsentiert? Potenzielle diskriminierende Effekte?

-> Siehe z.B. schlechtes Verständnis älterer Menschen bei Sprachanalysesystemen

-> Beruf des "Datenkurators", der "Datenkuratorin"

# Der IEEE “Code of Ethics”

Das Institute of Electrical and Electronics Engineers (IEEE, gesprochen i triple e) ist der größte weltweite Berufsverband von IngenieurInnen.

<https://www.ieee.org/about/corporate/governance/p7-8.html>



# Der IEEE “Code of Ethics”

We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members, and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:

1. to hold paramount the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices, and to disclose promptly factors that might endanger the public or the environment;
2. to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;
3. to be honest and realistic in stating claims or estimates based on available data;
4. to reject bribery in all its forms;
5. to improve the understanding by individuals and society of the capabilities and societal implications of conventional and emerging technologies, including intelligent systems;
6. to maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;
7. to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;
8. to treat fairly all persons and to not engage in acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression;
9. to avoid injuring others, their property, reputation, or employment by false or malicious action;
10. to assist colleagues and co-workers in their professional development and to support them in following this code of ethics.

# IEEE/ACM Joint Task Force on Software Engineering Ethics & Professional Practice



+



Association for  
Computing Machinery

<https://www.computer.org/education/code-of-ethics>

# IEEE/ACM Software Engineering Ethics

Software engineers shall commit themselves to making the analysis, specification, design, development, testing and maintenance of software a beneficial and respected profession. In accordance with their commitment to the health, safety and welfare of the public, software engineers shall adhere to the following Eight Principles:

1. PUBLIC – Software engineers shall act consistently with the public interest.
2. CLIENT AND EMPLOYER – Software engineers shall act in a manner that is in the best interests of their client and employer consistent with the public interest.
3. PRODUCT – Software engineers shall ensure that their products and related modifications meet the highest professional standards possible.
4. JUDGMENT – Software engineers shall maintain integrity and independence in their professional judgment.
5. MANAGEMENT – Software engineering managers and leaders shall subscribe to and promote an ethical approach to the management of software development and maintenance.
6. PROFESSION – Software engineers shall advance the integrity and reputation of the profession consistent with the public interest.
7. COLLEAGUES – Software engineers shall be fair to and supportive of their colleagues.
8. SELF – Software engineers shall participate in lifelong learning regarding the practice of their profession and shall promote an ethical approach to the practice of the profession.

# IEEE/ACM Joint Task Force on Software Engineering Ethics & Professional Practice: Using the Code / Praktische Beispiele

## Case Studies

As prescribed by the Preamble of the Code, computing professionals should approach the dilemma with a holistic reading of the principles and evaluate the situation with thoughtful consideration to the circumstances. In all cases, the computing professional should defer to the public good as the paramount consideration. The analyses in the following cases highlight the intended interpretations of members of the 2018 Code task force.

- **Malware Disruption:** Security vendors and government organizations collaborate to disrupt the operation of an ISP that hosts malware.
- **Medical Implant Risk Analysis:** A medical implant device maker creates a smart phone application to monitor and control the device.
- **Abusive Workplace Behavior:** A manager fails to address abusive behavior by a technical team leader.
- **Autonomous Active Response Weapons:** A defense contractor that specializes in autonomous vehicles begins to integrate automated weaponry.
- **Dark UX Patterns:** A web developer realizes that their client's requests are intended to trick users into making accidental and expensive purchases.
- **Malicious Inputs to Content Filters:** An Internet content filtering service deploys machine learning techniques to automate the classification of blocked content.

*These cases studies are designed for educational purposes to illustrate how to apply the Code to analyze complex situations. All names, businesses, places, events, and incidents are fictitious and are not intended to refer to actual entities.*

<https://ethics.acm.org/code-of-ethics/using-the-code/>

# Wie geht es weiter?

- Jeder/Jedem wird via Moodle innerhalb der nächsten Woche zufällig ein Diskussions-Report (nicht der eigenen Gruppe) zum Lesen zugewiesen. Dazu muss ein kurzes schriftliches Feedback erstellt werden (= “Peer Review”)
- Nächste Woche startet Block II zum Thema Gender/Diversität (Prof. Sabitzer)
- Prüfungsstoff aus jedem Block für die schriftliche Wiederholung am 21.1.2020 wird gesammelt bekanntgegeben.

**Empfehlung:  
Webserie “Helena – Künstliche Intelligenz”  
(BR/Arte, 2019)**



# KV Ethik und Gender Studies



Danke!

Nächster Termin:  
Block II (Prof. Sabitzer) startet

