

Die Tanzpartner von heute: Ein elegant-witziger Blick auf Ethik und KI

u^b

UNIVERSITÄT
BERN

Claus Beisbart

Claus.Beisbart@unibe.ch

12.9.2024

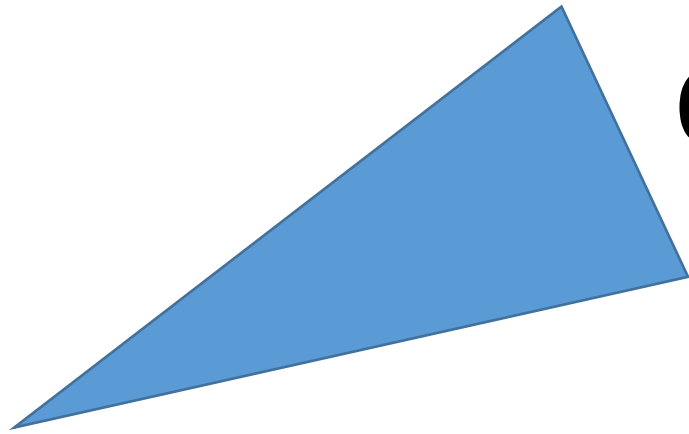
Aarau



ehealthsummit

Ziel des Referats

Untersuche den Einsatz von KI in
der Medizin



aus ethischer Perspektive

Struktur

1. Künstliche Intelligenz (KI) und Ethik

erste Annäherungsversuche

2. Einige Diskussionspunkte

a. Menschlichkeit

Walzer

a. Gerechtigkeit

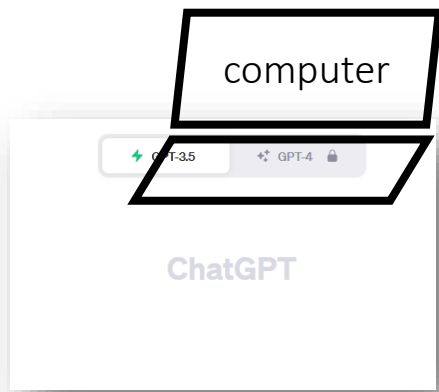
Formationstanz

3. Fazit

Happy End? 3

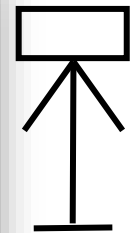
1. KI und Ethik

Forschungsfeld: zielt auf Systeme, die ...



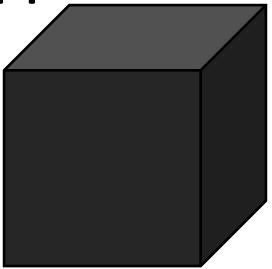
	ähnlich wie Menschen	ideal rational
denken		
handeln		

Bringsjord & Govindarajulu (2022) nach Russel & Norvig (2009)



Status heute

- KI liefert die Ergebnisse vieler kognitiver Leistungen
- KI in einzelnen Aspekten Menschen überlegen
- KI hat keine allgemeine Intelligenz
- KI empfindet nichts
- KI ist intransparent (“black box”)



Einsatz

Konsultation
(chatbots)

Therapie
(chatbots)

Verwaltung

Diagnose
(Klassifikatoren)

Vorhersage
(Klassifikatoren)

Operation
(Empfehlungssysteme, Roboter)

Medizinische

Praxis

Verbesserung

Forschung

Pflege

data science

Pflegeroboter

Wie finden wir das

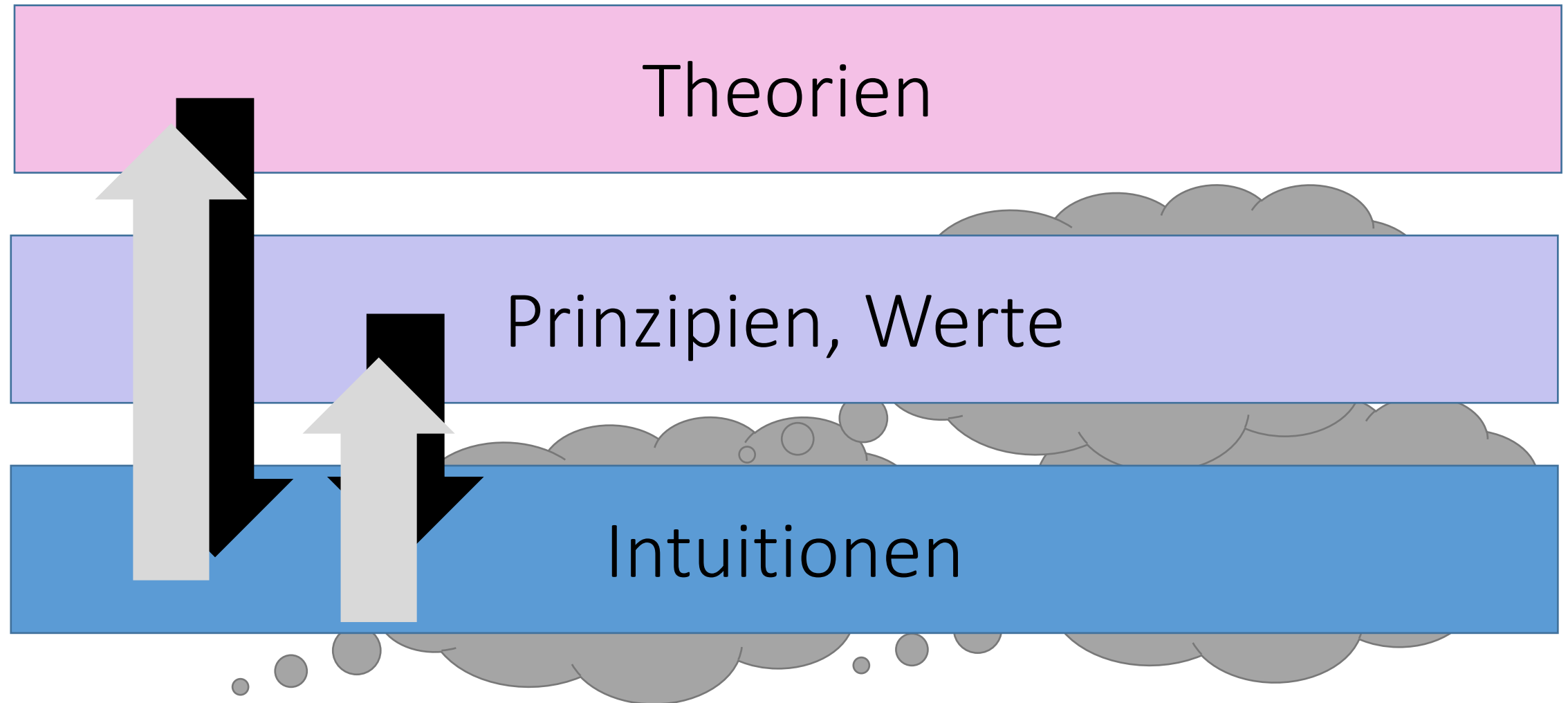
???

... von einem Roboter operiert zu werden?

... in der Pflege von einem Pflegeroboter unterhalten zu werden?

... im Alltag mit Robotern zusammenarbeiten?

Ethik: Systematisierung



Literaturstand

Stanford Encyclopedia of Philosophy

Browse ⓘ

Entry Contents

Bibliography

Academic Tools

Friends PDF Preview

Author and Citation Info

Back to Top

Keywords: Artificial intelligence, ethics, medicine, healthcare, future

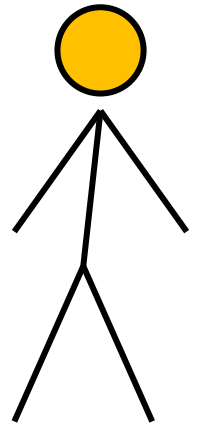
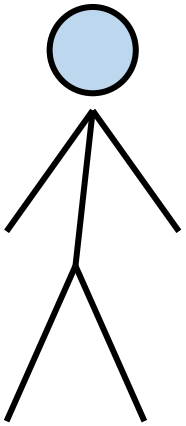
I. Artificial Intelligence and the Future of Medicine

AI encompasses a vast range of diverse technologies, and AI research cuts across a variety of disciplines including computer science, philosophy of mind, logic, neuroscience, and theoretical

2. Main Debates

- Surveillance
- Behaviour
- Systems
- Interaction
- Employment

2. Einige Diskussionspunkte



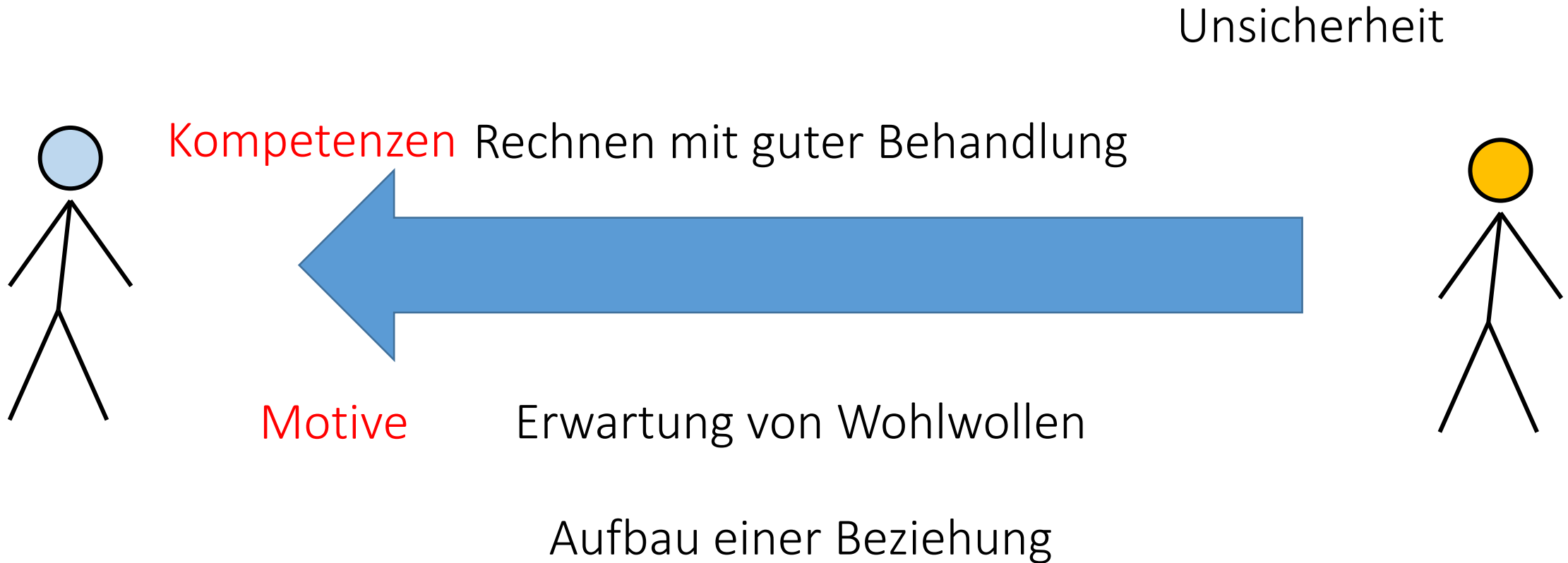
Hatherley (2020, 478):

i. Menschlichkeit

Bei Entscheidungen durch KI: Vertrauensverlust
Patientinnen vs. Ärzte

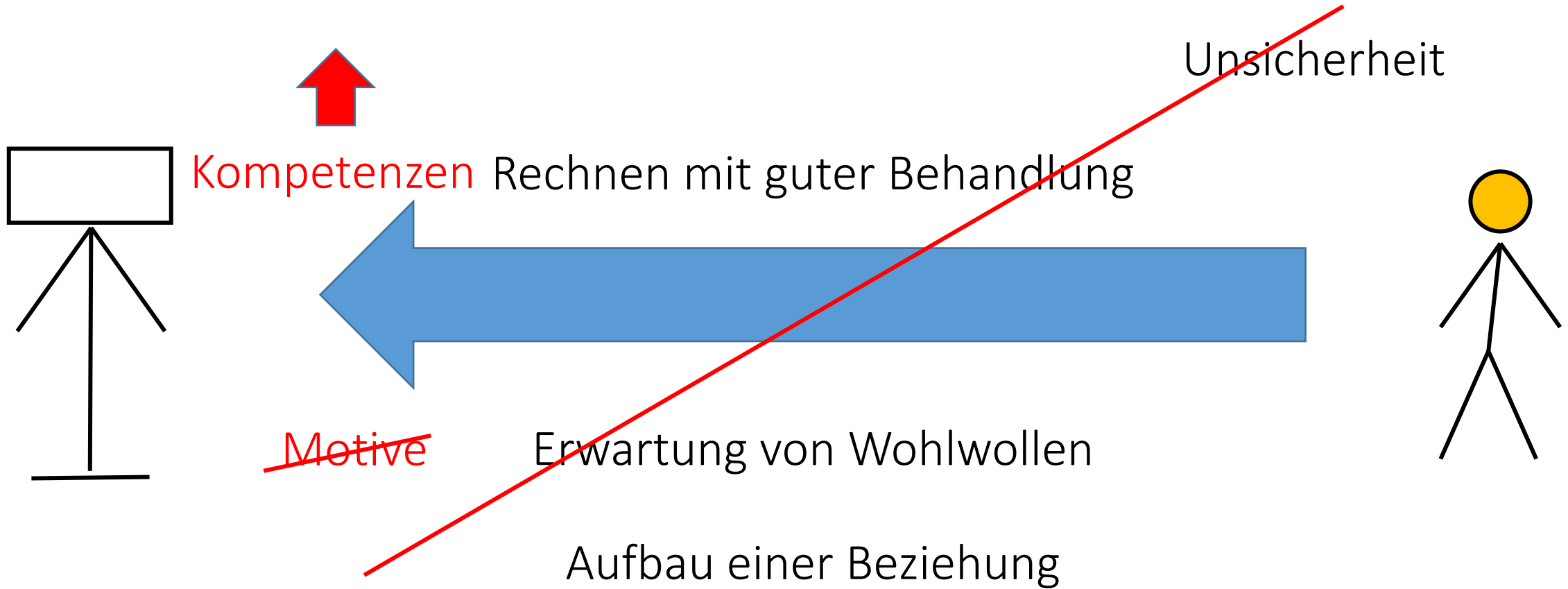


Was ist Vertrauen?



Cf. Hatherley (2020)

KI

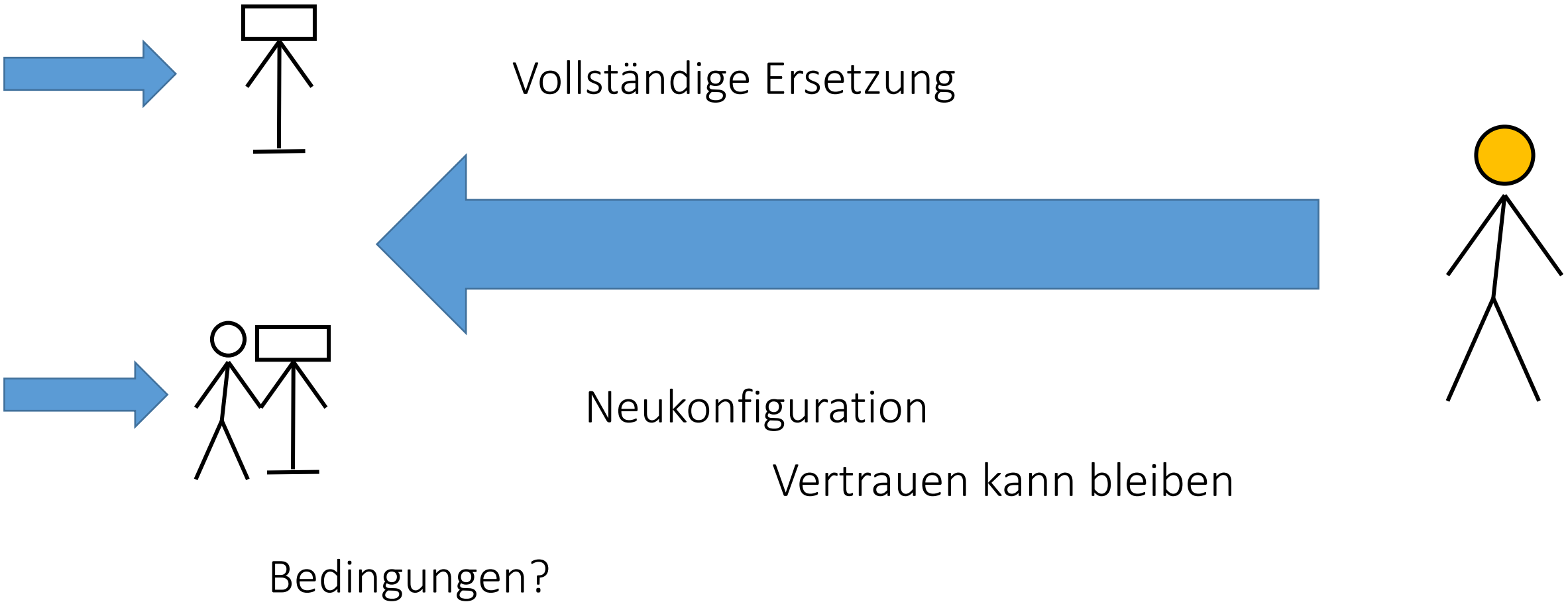


Cf. Hatherley (2020)

Argument von Hatherley

- Wir können Menschen, aber nicht KI vertrauen.
- Wenn KI Menschen an Kompetenzen übertrifft, wird Mensch ersetzt.
- Vertrauen wird verschwinden.
- Das ist schlecht, weil Vertrauen in der Medizin wichtig ist.

Einwand: Mehr Optionen



Zwischenzusammenfassung: i. Menschlichkeit


- Bestimmte Einstellungen sind gegenüber Menschen, nicht aber KI angemessen.
- Bestimmte Beziehungen funktionieren mit Menschen, nicht aber mit KI.
- Es droht die Gefahr, dass die Beziehungen durch KI verschwinden.
- Ob das so ist, hängt aber von Einsatz KI ab.

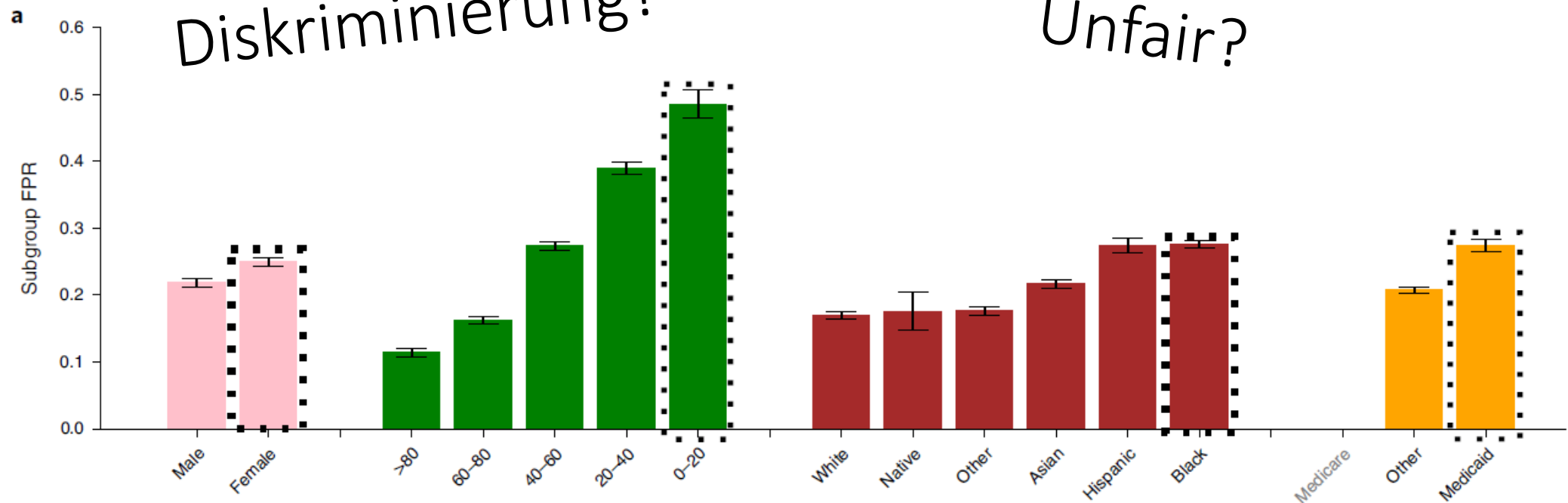
ii. „Bias“

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 10 December 2021](#)

Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations

[Laleh Seyyed-Kalantari](#) , [Haoran Zhang](#), [Matthew B. A. McDermott](#), [Irene Y. Chen](#) & [Marzyeh Ghassemi](#)



Seyyed-Kalantari et al. (2021), Fig. 2

Bias – Verzerrung

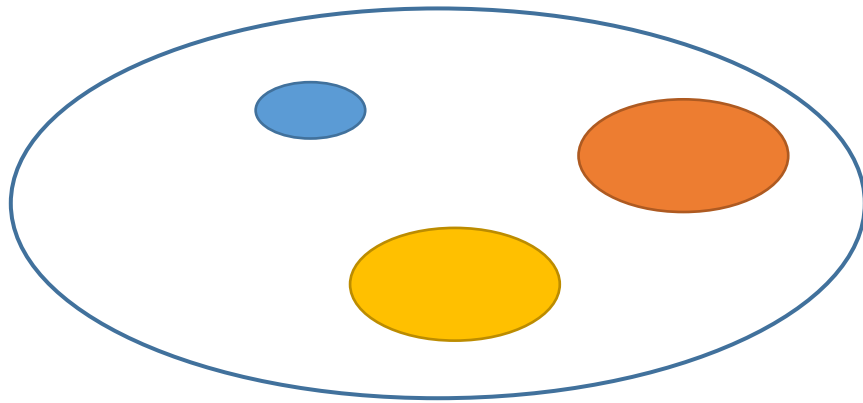
Minimaldefinition:

- Leistung von KI unterscheidet sich für unterschiedliche Gruppen.

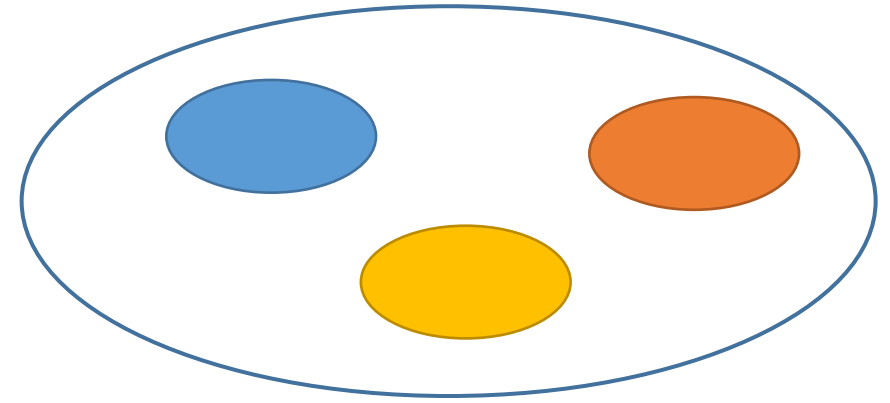
Moralische Bewertung:

- Kann moralisch problematisch sein

Beispiel 1: Unterrepräsentation

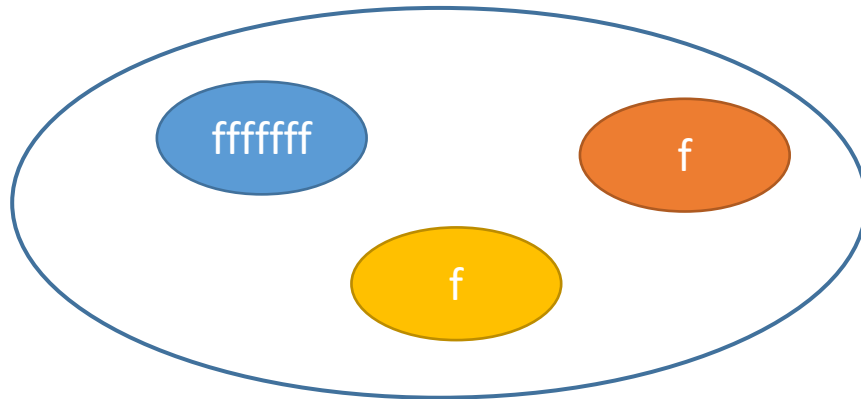


Trainingsdaten

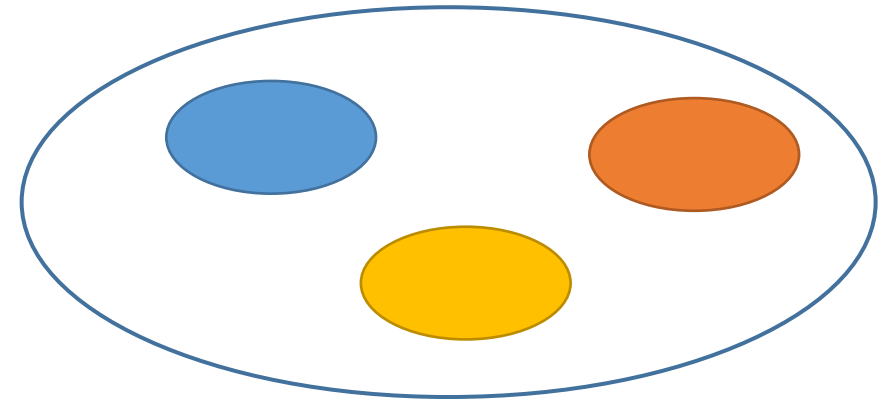


Bevölkerung Anwendung

Beispiel 2: fehlerhafte Daten



Trainingsdaten



Bevölkerung Anwendung

Lösungsansätze

1. „fares Maschinelles Lernen“
2. Untersuchung und Einsatz mit Menschenverstand

Johnson (2021), Fazilpour & Danks (2021)

Zwischenzusammenfassung: ii. Bias

1. KI arbeitet oft nicht für alle Gruppen gleich gut.
2. Das kann zu Ungerechtigkeiten führen.
3. Problem oft: Fortschreibung von Ungerechtigkeiten
4. Problem dabei: KI erscheint objektiv.
5. Es gibt Gegenmassnahmen.
6. Diese werfen interessante Fragen auf.

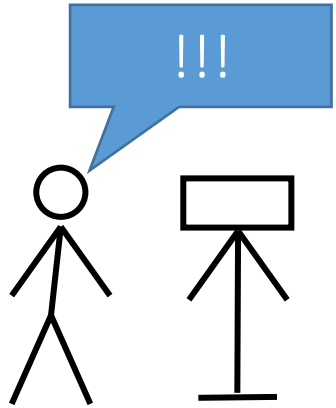
Weitere Diskussionspunkte

- Eingriffe in Autonomie
- Gefühl von Überwachung
- Verlust von Fähigkeiten vom Menschen („deskilling“)
- Verantwortungslücken bei Fehlern

Vgl. Sparrow & Hatherley (2019)

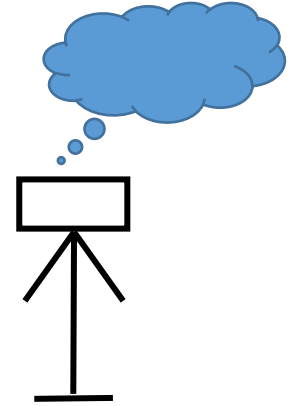
Unterscheidungen

i. Was macht KI?



Befehle ausführen

KI als Werkzeug/Objekt

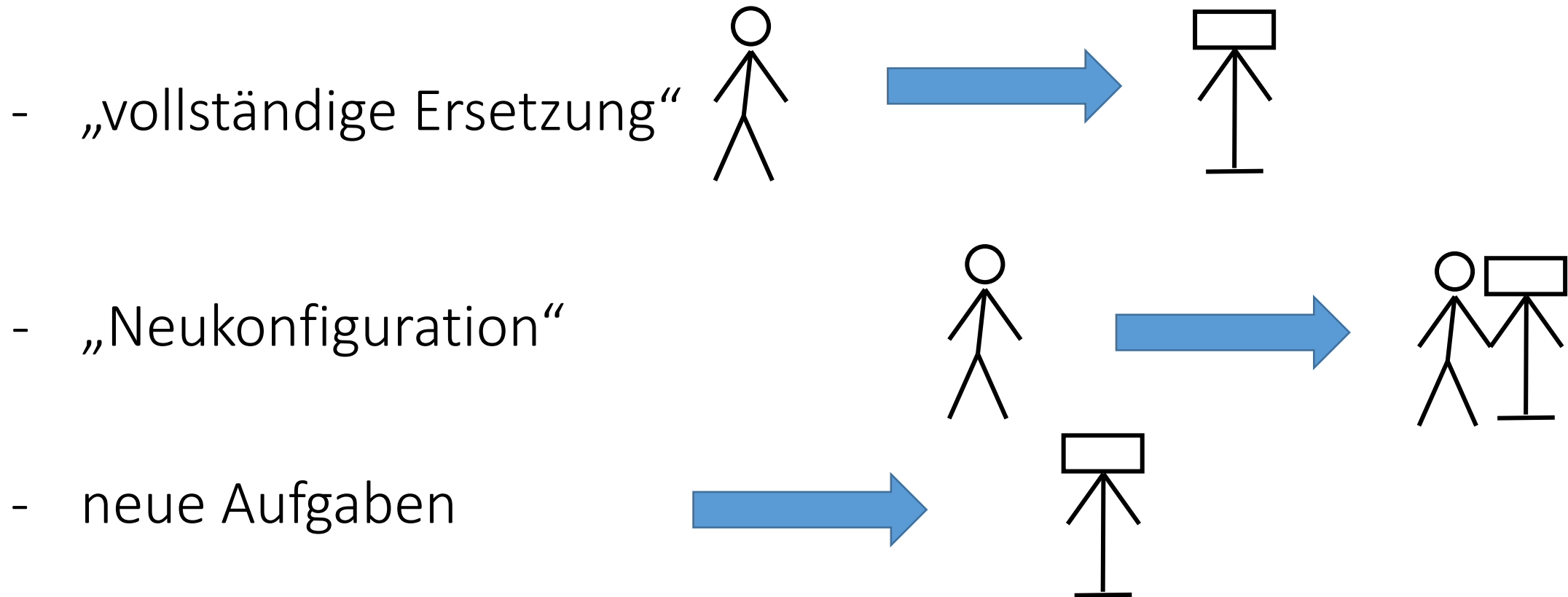


Entscheidungen treffen

KI als Subjekt

Müller (2023)

ii. Wie verhält sich KI zu menschl. Arbeit?



Tanzpartner

KI und Ethik?



Zusammenfassung

1. KI kann einige Tänze besser als Menschen.
2. KI lässt sich daher in Medizin vielfältig einsetzen.
3. Der Tanz von KI in der Medizin weckt auch ungute Gefühle.
4. In der Ethik lassen sich diese Gefühle diskutieren.
5. Dabei zeigt sich oft: Es kommt darauf an, wie der Mensch mit der KI tanzt.

Appell

Setzen wir uns dafür ein,
dass KI verantwortungsvoll gebraucht wird:

1. Aufklärung über KI
2. Öffentliche Diskussion über KI
3. Gesetzliche Leitlinien und Standards für KI

Merci viumau!

Zusammenfassung

1. KI ist hat bei einigen Fähigkeiten Menschen überholt.
2. KI lässt sich daher in Medizin vielfältig einsetzen.
3. Der Einsatz von KI in der Medizin weckt auch ungute Gefühle.
4. In der Ethik lassen sich diese Gefühle diskutieren.
5. Dabei zeigt sich oft: Es kommt darauf an, wie KI gebraucht wird.

Literatur

Anderson, M. & Anderson, Sue L. (Hrsg.) 2011, *Machine Ethics*, Cambridge University Press, New York.

Bringsjord, Selmer & Govindarajulu, Naveen S. 2022, Artificial Intelligence. In: E. N. Zalta & U. Nodelman (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/>.

Burrell, Jenna 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms, *Big Data and Society* **3** (1):205395171562251.

Danaher, John & Nyholm, Sven 2021, Automation, work and the achievement gap, *AI Ethics* **1**, 227–237.

Dreyfus, Hubert L. 1992, *What Computers Still Can’t Do*, MIT Press, Cambridge, MA.

Fazelpour, Sina & Danks, David 2021, Algorithmic bias: Senses, sources, solutions, *Philosophy Compass*, 16(8), e12760.

Hatherley, Joshua James 2020, Limits of trust in medical AI, *Journal of Medical Ethics* **46** (7), 478-481.

Johnson, Gabrielle M. 2021, Algorithmic bias: on the implicit biases of social technology, *Synthese* **198**, 9941–9961.

Müller, Vincent C. 2023, Ethics of Artificial Intelligence and Robotics, *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (Hrsg.), <<https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/>>.

Obermeyer, Ziad et al. 2019, Dissecting racial bias in an algorithm used to manage the health of populations, *Science* **366**, 6464, 447–453.

Nam, Ju Gang, et al. 2023, AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial, *Radiology* **307.2**, e221894.

Russell, Stuart & Norvig, Peter 2009, Artificial Intelligence: A Modern Approach, 3rd edition, Prentice Hall, Saddle River, NJ.

Seyyed-Kalantari, L., Zhang, H., McDermott, M.B.A. et al. 2021 , Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, *Nat Med* **27**, 2176–2182

Sparrow, Robert & Hatherley, Joshua James 2019, The promise and perils of AI in medicine, *International Journal of Chinese and Comparative Philosophy of Medicine* **17** (2), 79-109.

Stubbe, Julian et al. 2019, [Akzeptanz von Servicerobotern: Tools und Strategien für den erfolgreichen betrieblichen Einsatz](#), iit-Institut für Innovation und Technik, Berlin.