

Bessere Medizin durch Künstliche Intelligenz?

u^b

b
UNIVERSITÄT
BERN

Claus Beisbart

Claus.Beisbart@unibe.ch

Institut für Philosophie, Center for Artificial Intelligence in Medicine

3. Juni 2025

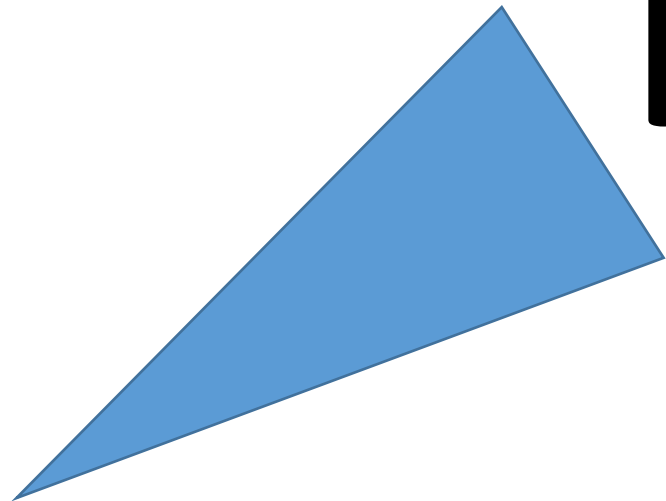
FORUM
elle

Die Frauenorganisation der Migros
L'organisation féminine de Migros
L'organizzazione femminile della Migros

Frage des Referats

Gut?

KI in Medizin



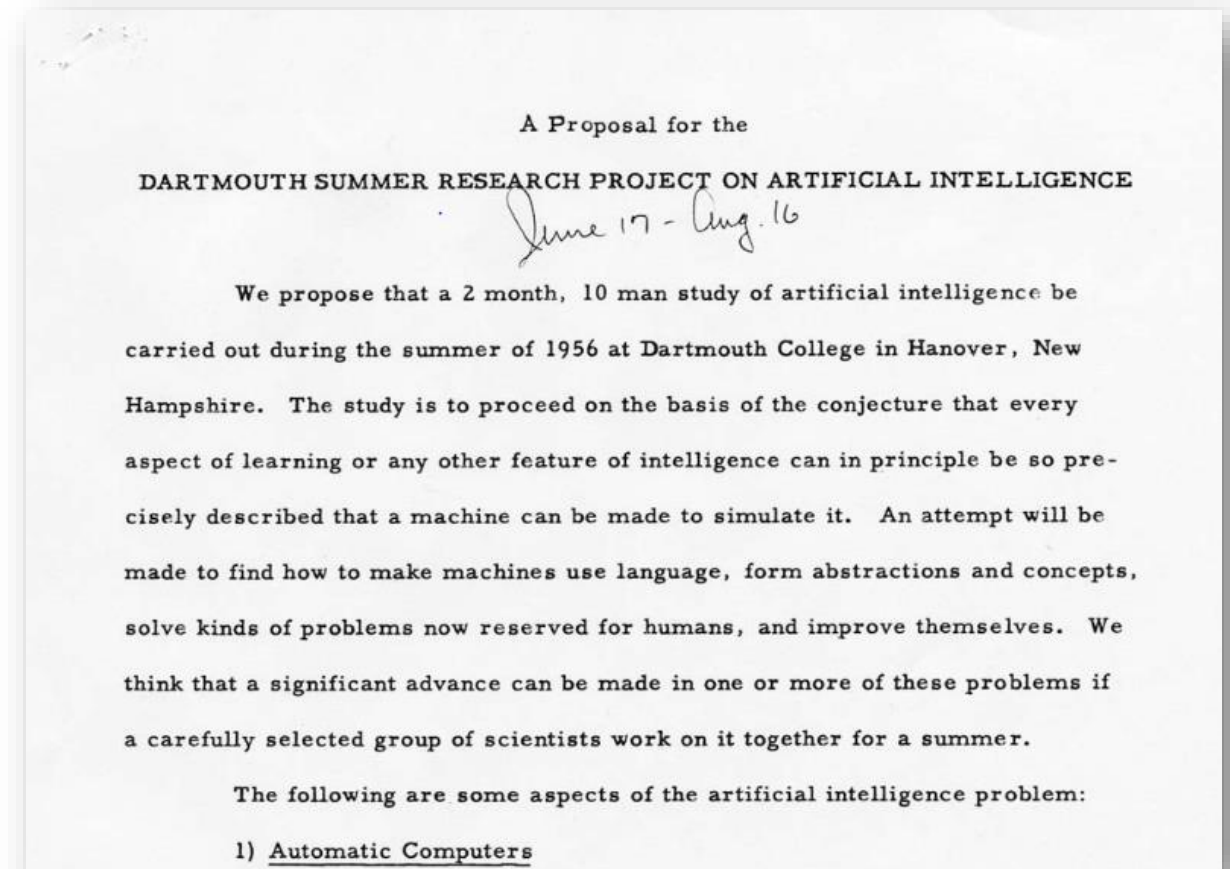
Perspektive Ethik

Struktur

1. Was ist Künstliche Intelligenz (KI)?
2. Wie wird KI in der Medizin eingesetzt?
3. Wie bewerten wir KI in der Medizin?
 - a. Allgemein
 - b. Bias
 - c. Vertrauen
 - d. Autonomie
4. Fazit

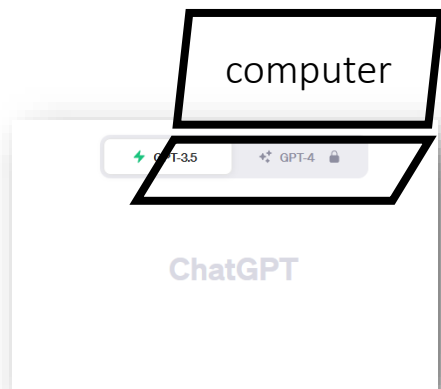
1. Was ist KI?

„Wir versuchen herauszufinden, wie wir Maschinen dazu zu bringen, Sprache zu gebrauchen, Abstraktionen und Begriffe zu bilden, Probleme zu lösen, wie sie bisher für Menschen reserviert waren, und sich selbst zu verbessern.“



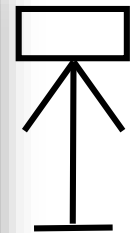
KI: Forschungsfeld

zielt auf Systeme, die ...



	ähnlich wie Menschen	ideal rational
denken		
handeln		

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



Bemerkungen

- Definition über Forschungsfeld bescheiden.
- Heute anderer Sprachgebrauch von „KI“: für Systeme, deren Verhalten intelligent aussieht.

Ziele der KI-Forschung

- Praktische Anwendungen: menschliches Leben erleichtern und verbessern
- Theoretisch besseres Verständnis von menschlicher Kognition

Wie hat man es probiert?



symbolisch

$$c = \sqrt{a^2 + b^2}$$

Thomas Hobbes
(1588-1679)

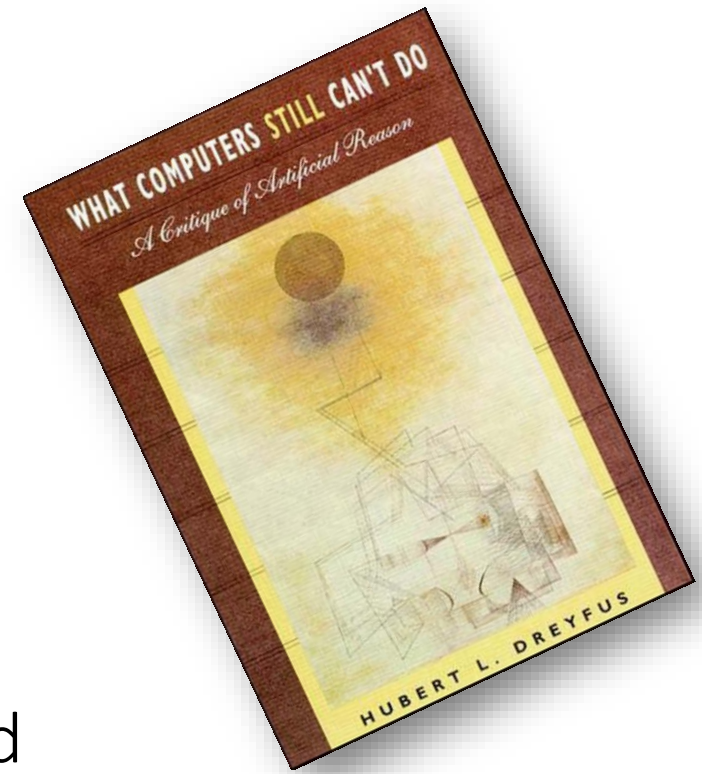


19221329501
82736158565
91260020462
93262002774
81065829429
32481913153

Hubert Dreyfus

“Die Annahmen [hinter symbolischer KI] [...] gehen dahin, dass der Mensch ein Wesen sein muss, welches auf der Basis von Regeln mit Daten rechnet, welche atomare Fakten beschreiben. [...]

[was fehlt:] die Rolle des Körpers in der Integration und Vereinheitlichung unserer Erfahrung von Dingen; die Rolle der Situation, welche einen Hintergrund bildet [...] und endlich die Rolle von menschlichen Zielen und Bedürfnissen bei der Organisation der Situation, so dass Dinge als relevant und zugänglich erkannt werden können.”



Dreyfus 1992/94, 231, 234 (Üs. CB)

Wie hat man es probiert?

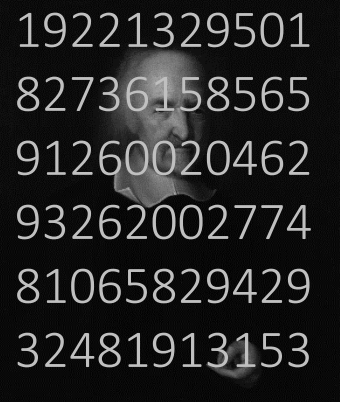
Maschinelles Lernen

konnektionistisch

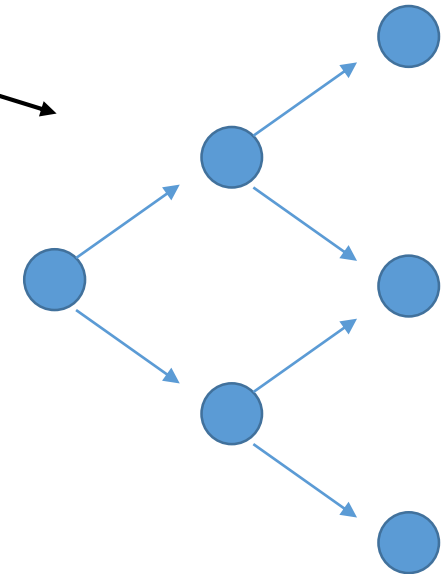
symbolisch

$$c = \sqrt{a^2 + b^2}$$

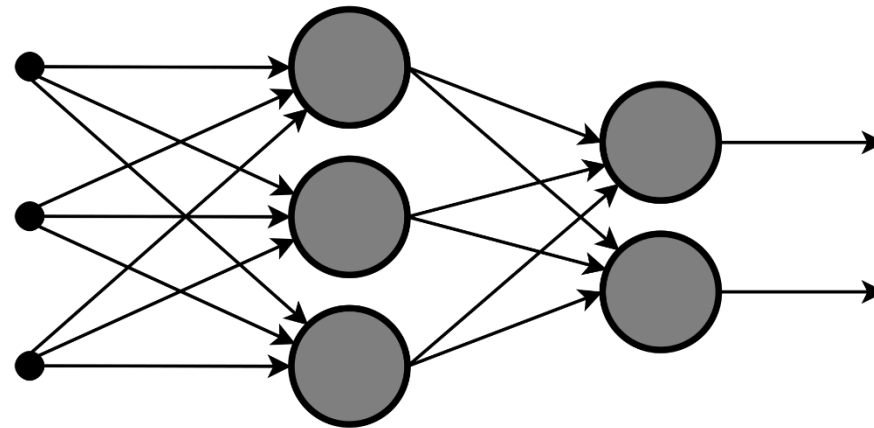
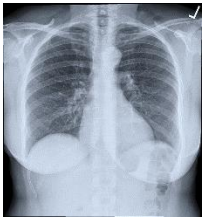
Thomas Hobbes
(1588-1679)



19221329501
82736158565
91260020462
93262002774
81065829429
32481913153



Überwachtes Lernen



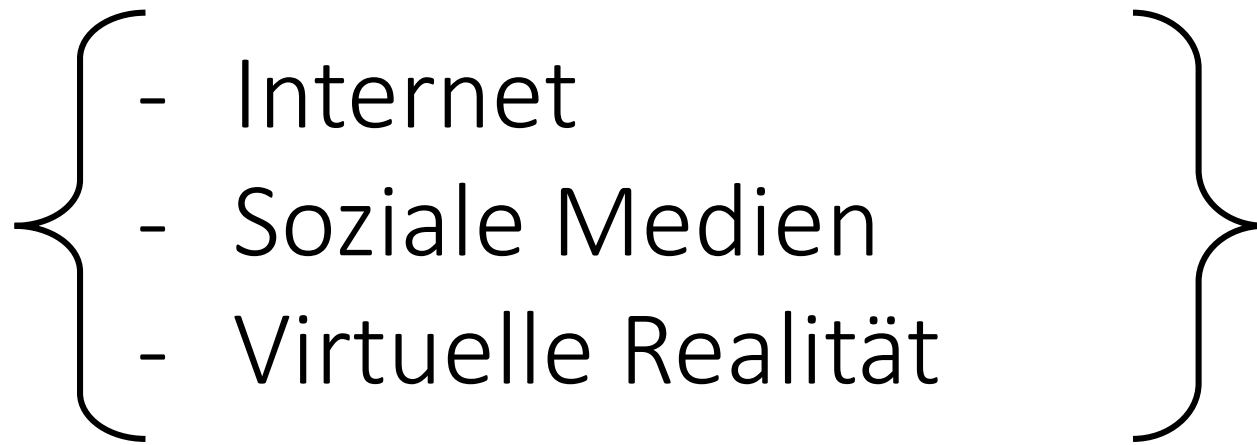
„gesund“

„gesund“

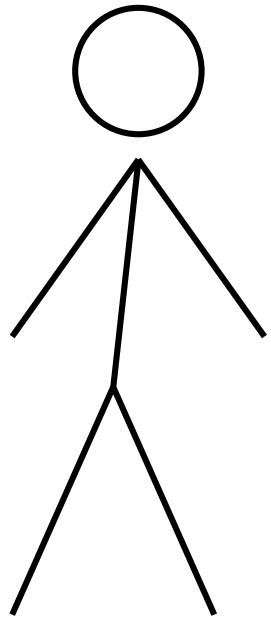
„krank“

???

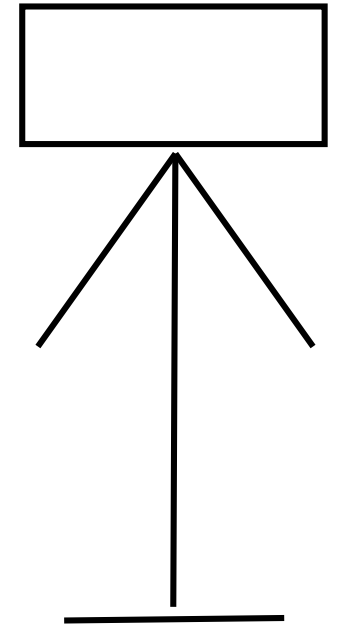
Andere Aspekte von Digitalisierung, nicht KI

- 
- Internet
 - Soziale Medien
 - Virtuelle Realität

Vergleich Mensch - KI



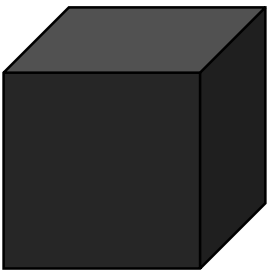
Rechnen	Rechnen
Textproduktion	Textproduktion
Denken	?
Verstehen	?
Gefühle	-
Empathie	-



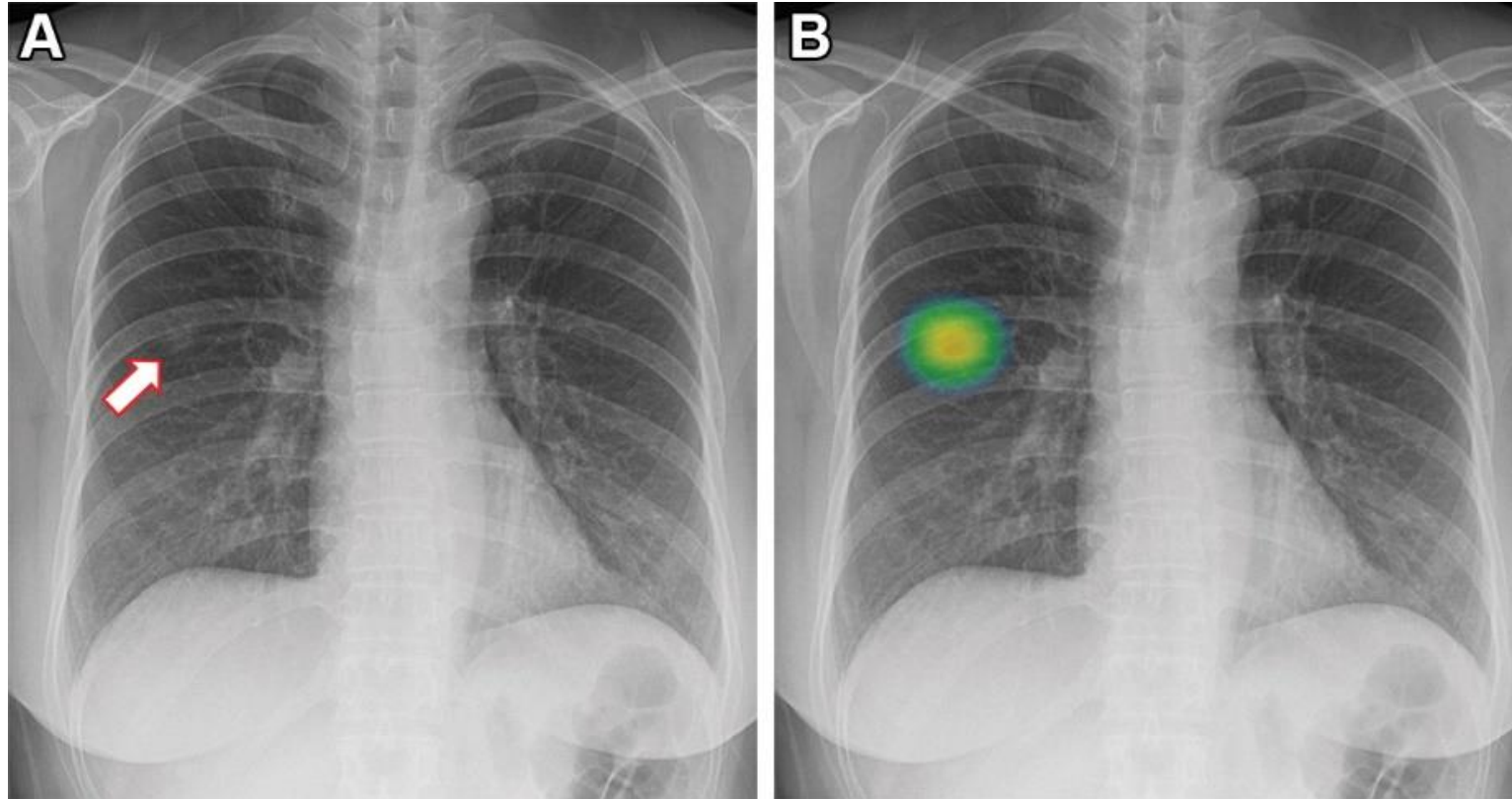
Problem

- hohes Mass an Opazität (“black boxes”)

Burrell (2016, 1, Übersetzung CB): “wenn man das Ergebnis eines Algorithmus erhält (die Klassifikationsentscheidung), dann hat man selten eine konkrete Idee darüber, wie oder warum eine bestimmte Klassifikation aus dem Input hervorgegangen ist.”



2. Wie wird KI in Medizin eingesetzt?



Diagnose

buoy

CHECKING FOR EMERGENCIES

[Edit](#)

b
No problem! We can still continue.
What symptom is bothering him the most?

Dry cough
2 days
[Remove](#)

Fever
Not sure or lower than 100.4F (38C), 1 day
[Remove](#)

b
Any other context we should take into account?

None of the above
[Edit](#)

b
Are they thinking about seeing a medical professional for this?

No

b
Would they like to check if their symptoms indicate a medical emergency?

Yes
[Edit](#)

[Keep going ↓](#)

b
Great. We'll screen for emergency situations next.

Welche Symptome machen Ihnen am meisten Sorgen?

- 2 Tage Husten
- Fieber, nicht sicher oder weniger als 38 Grad

Irgendwelche andere Faktoren, die wir berücksichtigen sollten?

- Nichts vom Genannten.

Überlegen Sie sich, professionelle Hilfe zu holen?

- Nein

Wollen Sie kontrollieren, ob Ihr Symptome einen Notfall anzeigen?

[Start](#)[Kurse](#)[Features](#)[Vorteile](#)[Newsletter](#)

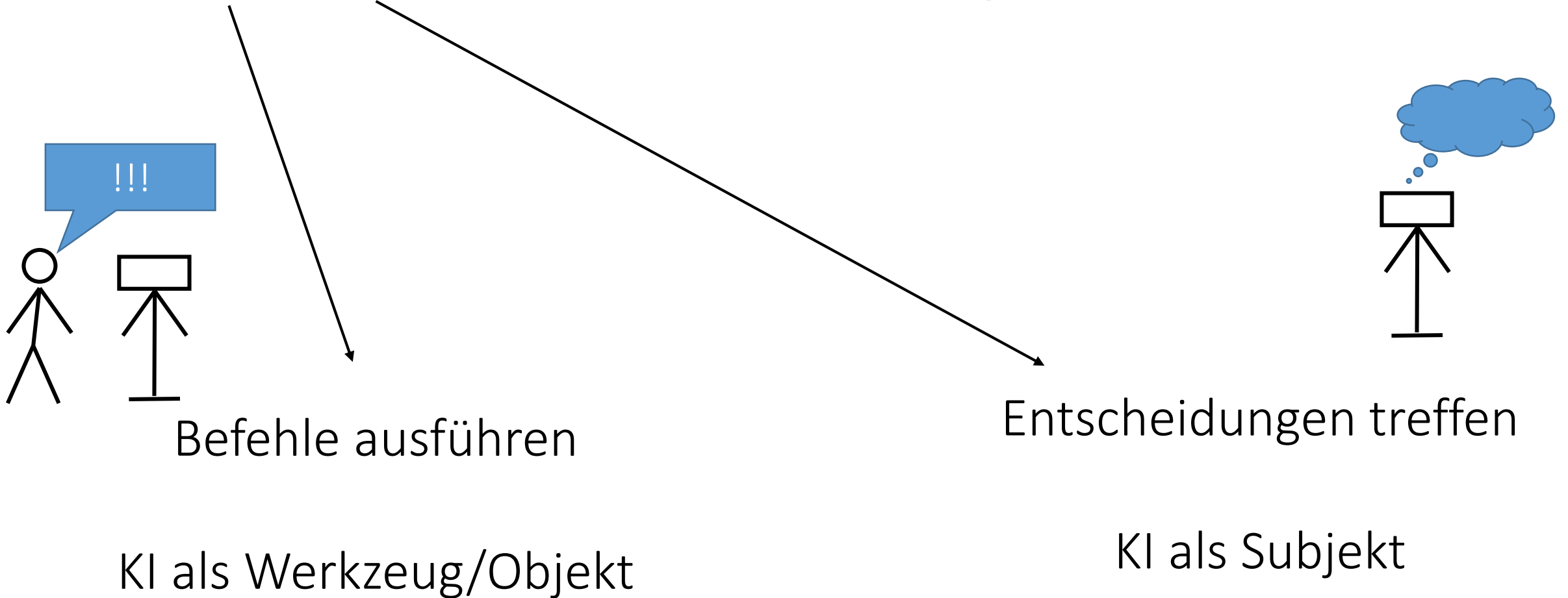
Hi, ich bin Mina. Dein Coach zur Vermeidung von Prüfungsangst.

Mina ist der erste digitale Therapeut und Coach, der auf kognitiver Verhaltenstherapie basiert und zur mentalen Unterstützung bei Prüfungsängsten, Leistungsdruck und Lernstress an deiner Seite steht. Derzeit befindet sich Mina in einer geschlossenen Beta-Phase und wird mit Studenten sowie Schülern getestet. Nach dem Launch werden weitere Therapie-Kurse (Depressionen, Burnout, Mobbing, etc.) folgen.

Pflege

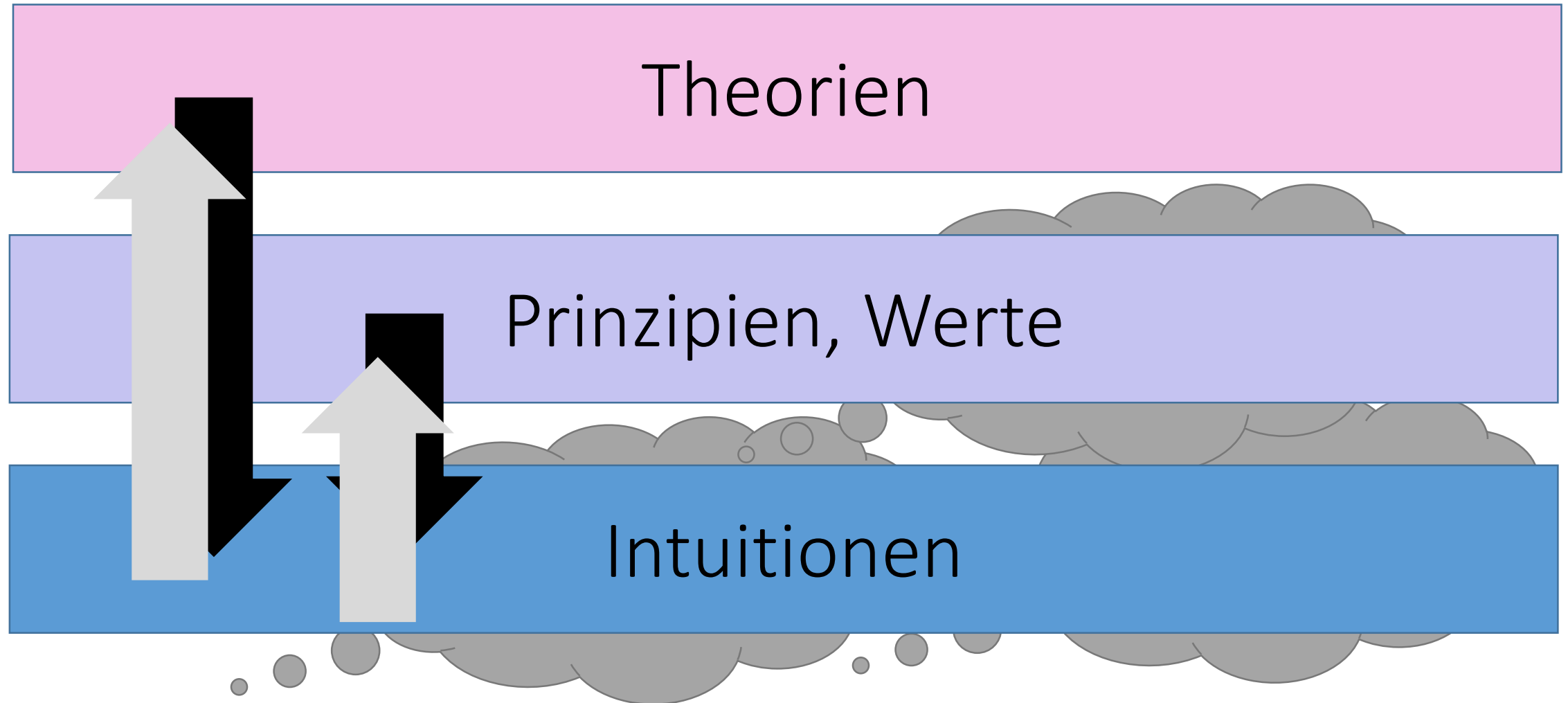


Einsatz von Künstlicher Intelligenz

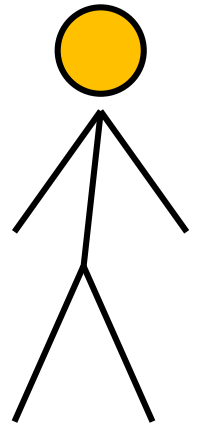
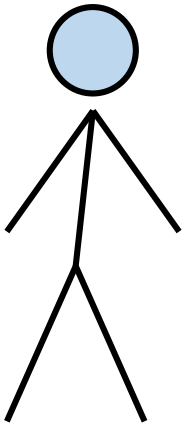


Müller (2023)

Ethik



Vertrauen



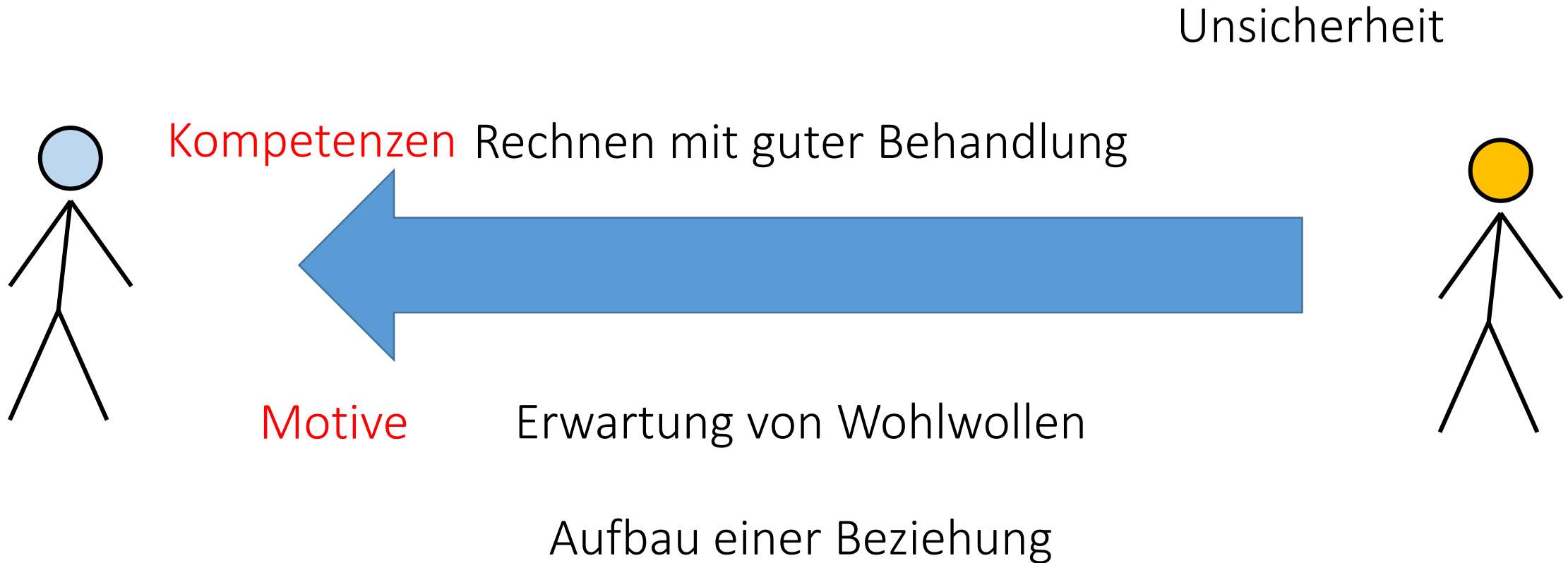
Vertrauen

Hatherley (2020, 478):

Bei Entscheidungen durch KI: Vertrauensverlust
Patientinnen vs. Ärzte

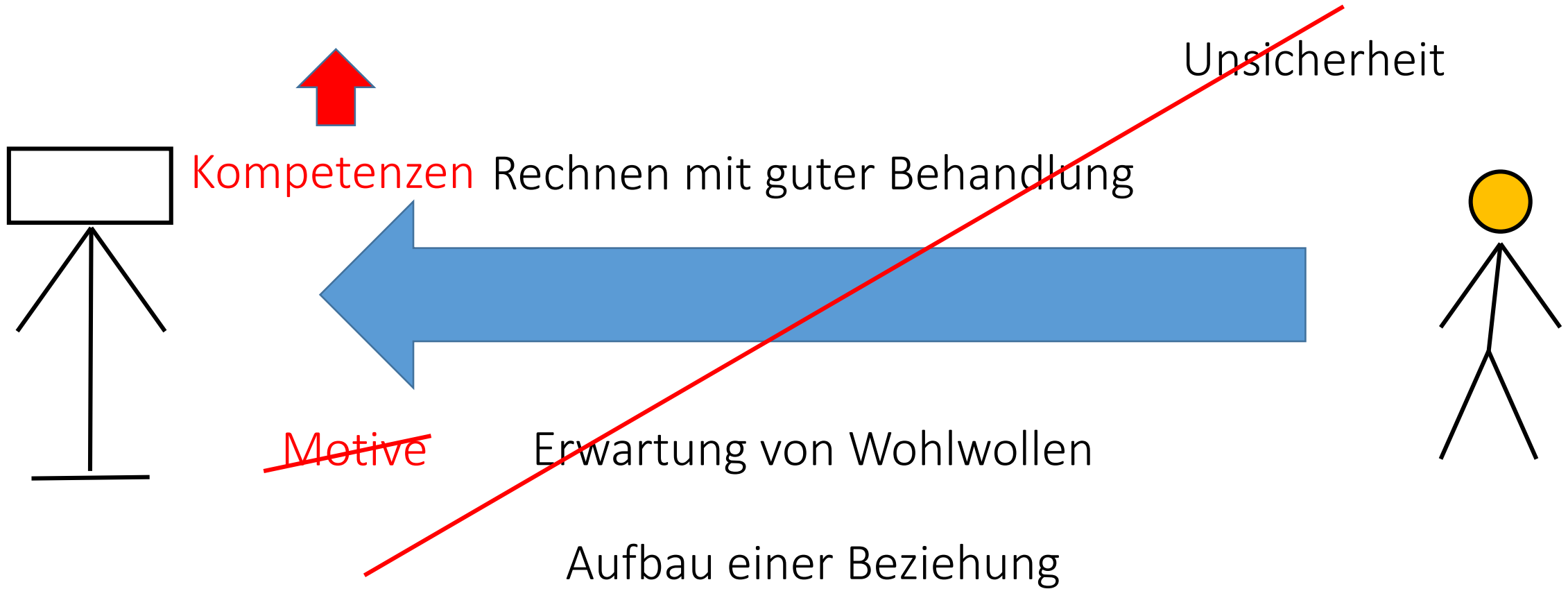


Was ist Vertrauen?



Cf. Hatherley (2020)

KI

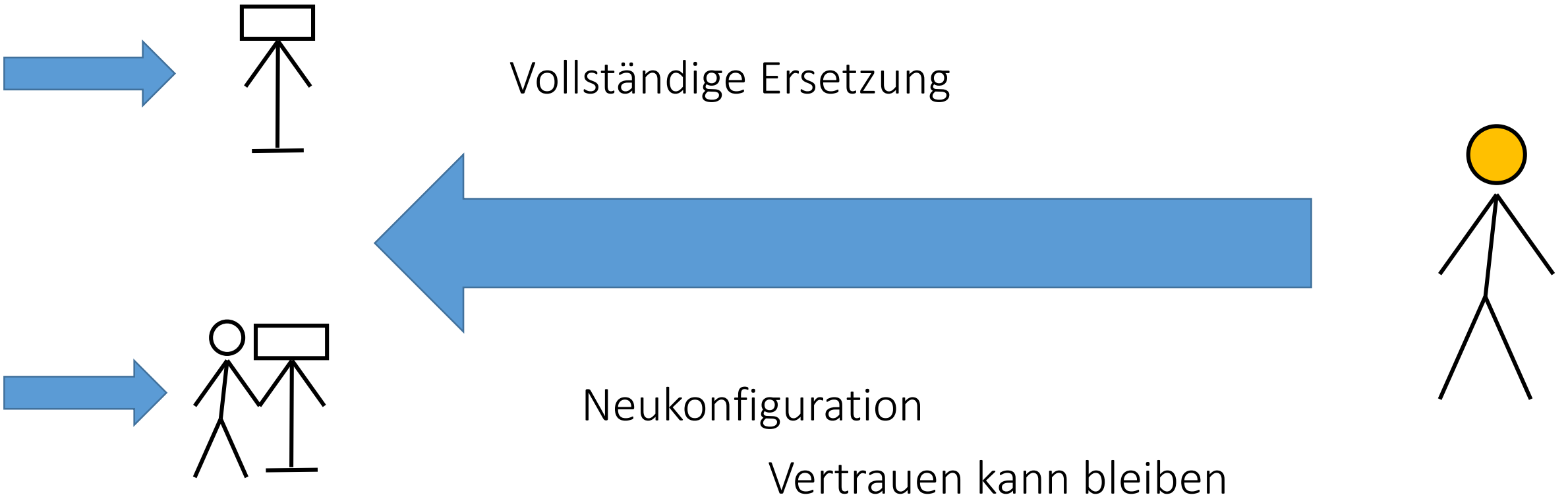


Cf. Hatherley (2020)

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



s. Danaher und Nyholm (xxx)

Zwischenzusammenfassung: Vertrauen

- 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.

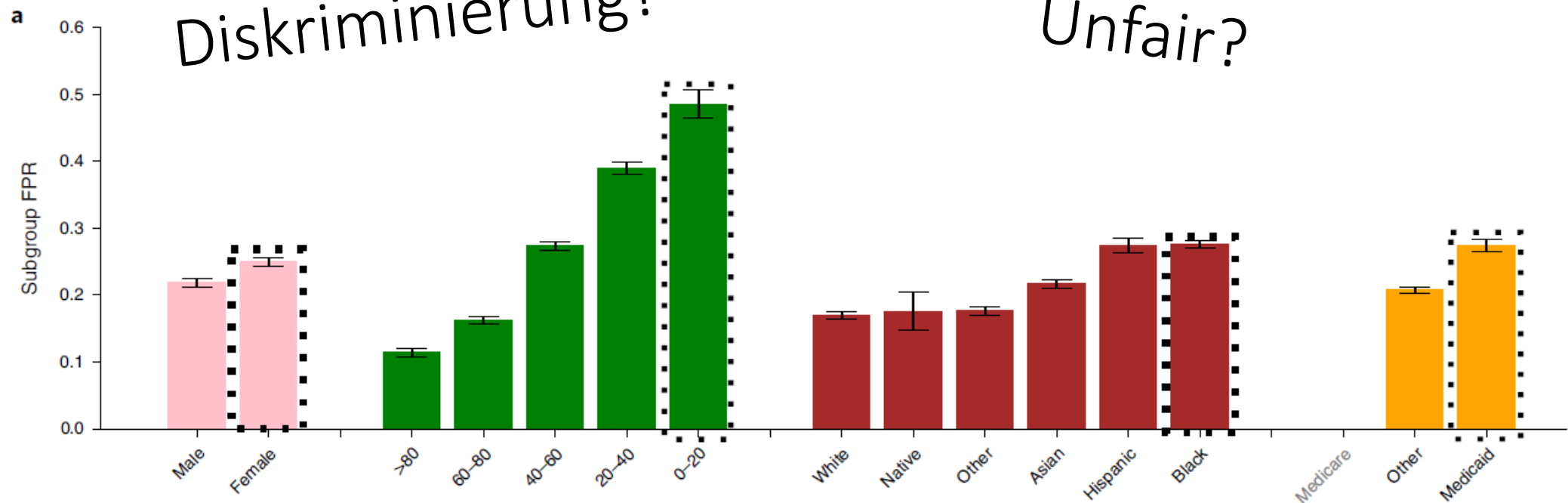
„Bias“

[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

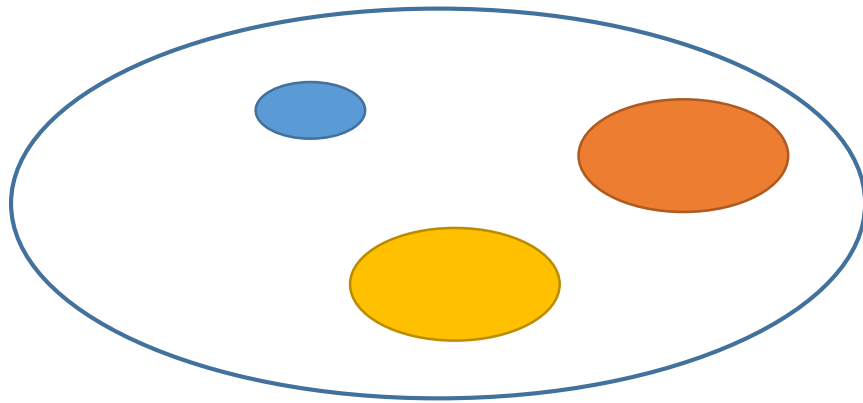
Minimaldefinition:

- Leistung von KI unterscheidet sich für unterschiedliche Gruppen.

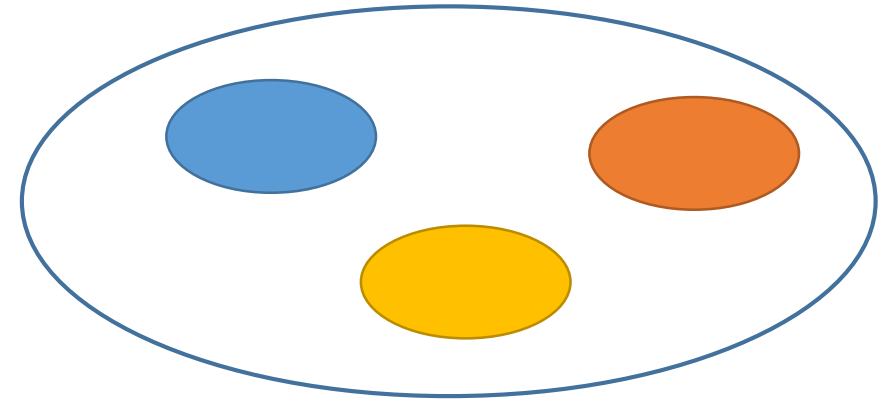
Moralische Bewertung:

- Kann moralisch problematisch sein

Beispiel 1: Unterrepräsentation

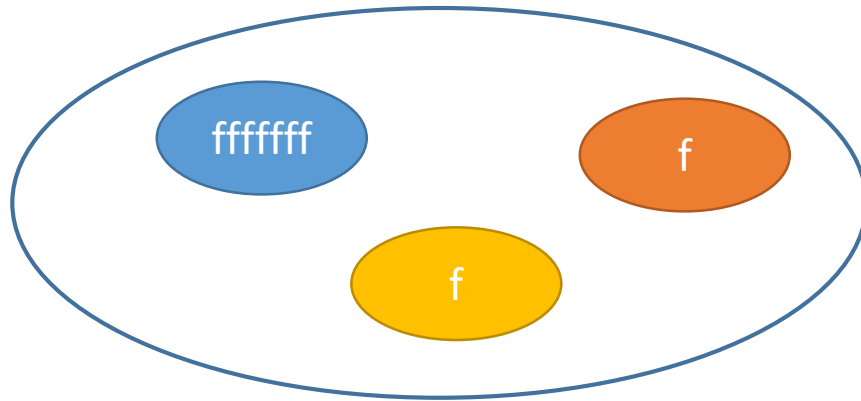


Trainingsdaten

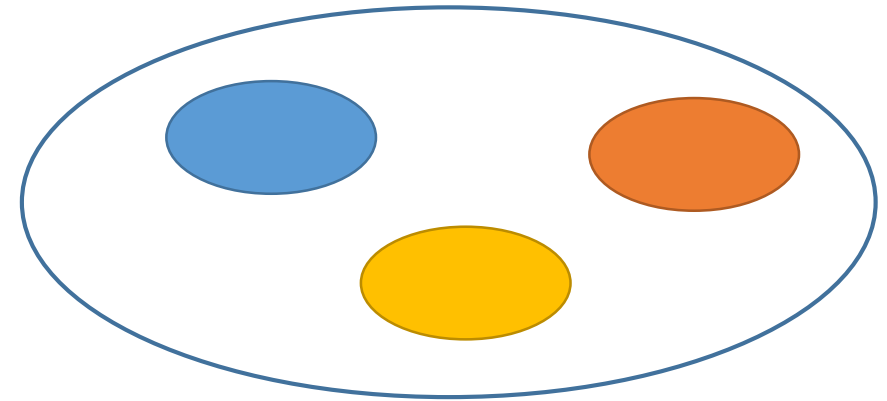


Bevölkerung Anwendung

Beispiel 2: fehlerhafte Daten



Trainingsdaten



Bevölkerung Anwendung

Lösungsansätze

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

Johnson (2021), Fazilpour & Danks (2021)

Zwischenzusammenfassung: 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.

5. Fazit

1. KI hat bei einigen Fähigkeiten Menschen überholt.
2. KI lässt sich vielfältig einsetzen.
3. KI kann Medizin besser und effektiver machen.
4. Der KI fehlen aber charakteristische menschliche Züge.
5. Der Einsatz von KI weckt daher auch ungute Gefühle.
6. In der Ethik lassen sich diese Gefühle diskutieren.
7. Dabei zeigt sich oft: Es kommt darauf an, wie KI gebraucht wird.

Schlusswort

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!

Literatur

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

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

Danaher, J. & Nyholm, S. (2021), Automation, work and the achievement gap. *AI Ethics* 1, 227–237, <https://doi.org/10.1007/s43681-020-00028-x>.

Dreyfus, Hubert L. (1992), *What Computers Still Can'T Do: A Critique of Artificial Reason*. MIT Press (hier Nachdruck 1994).

Fazilpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), e12760. <https://doi.org/10.1111/phc3.12760>.

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

Johnson, G.M. Algorithmic bias: on the implicit biases of social technology. *Synthese* 198, 9941–9961 (2021). <https://doi.org/10.1007/s11229-020-02696-y>.

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

Literatur

Nam, J. G. et al. (2023). AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. *Radiology*, 307(2), e221894.

Obermeyer, Z. et al. (2019), Dissecting racial bias in an algorithm used to manage the health of populations, *Science* 366, Issue 6464, 447-453.

Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach 3rd edition*, Saddle River, NJ: Prentice Hall.

Seyyed-Kalantari, L., et al. (2021), Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, *Nat Med* **27**, 2176–2182.

Shannon, C. et al. (1955), hier nach: *AI Magazine* 27/4, 13.

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

Was ist Autonomie?

38

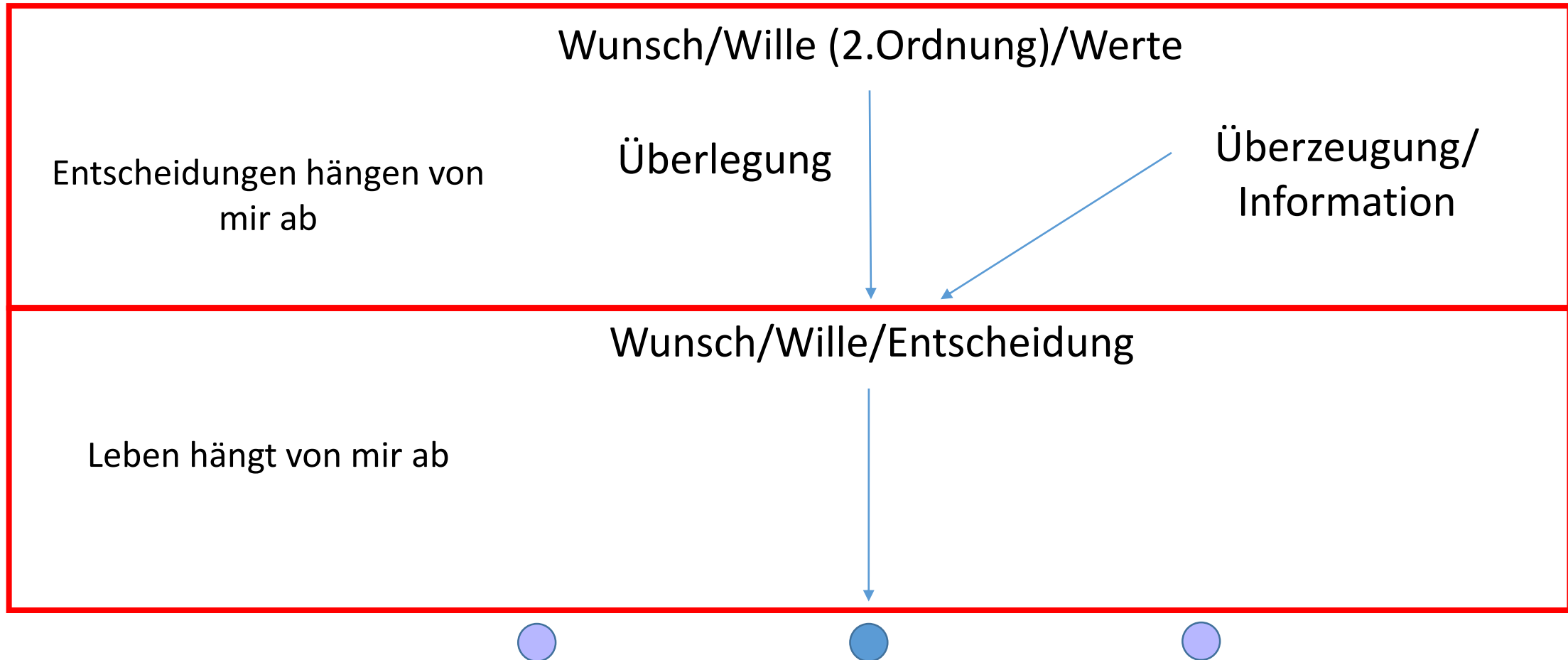


Isaiah Berlin
(1909-1997)

„Ich wünsche, dass mein Leben und meine Entscheidungen von mir abhängen, [...] ich möchte von Gründen bewegt werden, die meine eigenen sind und nicht von Ursachen, die mich gleichsam von aussen affizieren.“

Persönliche Autonomie

39



In den Worten von G. Dworkin

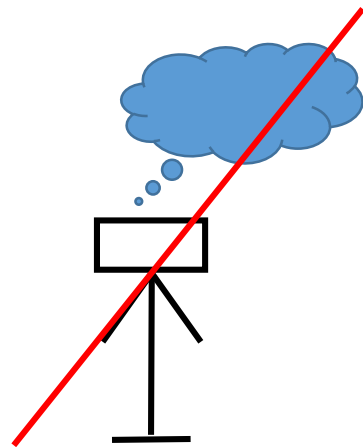
40



(*1937)

Autonomie

Fähigkeit höherer Ordnung, die
Entscheidungen/Wünsche erster Ordnung zu gestalten.



Vgl. Dworkin (1988, 20)

3. Einfluss der KI

„Drinkit unterstützt dich beim Erreichen deines Trinkzieles“

41

App Store Vorschau

Öffne den Mac App Store, um Apps zu kaufen und zu laden.



Drinkit: Mehr Wasser trinken (4+)

Trinkerinnerung & Trink App
Marcel Hagmann
Entwickelt für iPad

★★★★★ 4,6 • 136 Bewertungen

Gratis · In-App-Käufe möglich

SELF-CARE **Login**

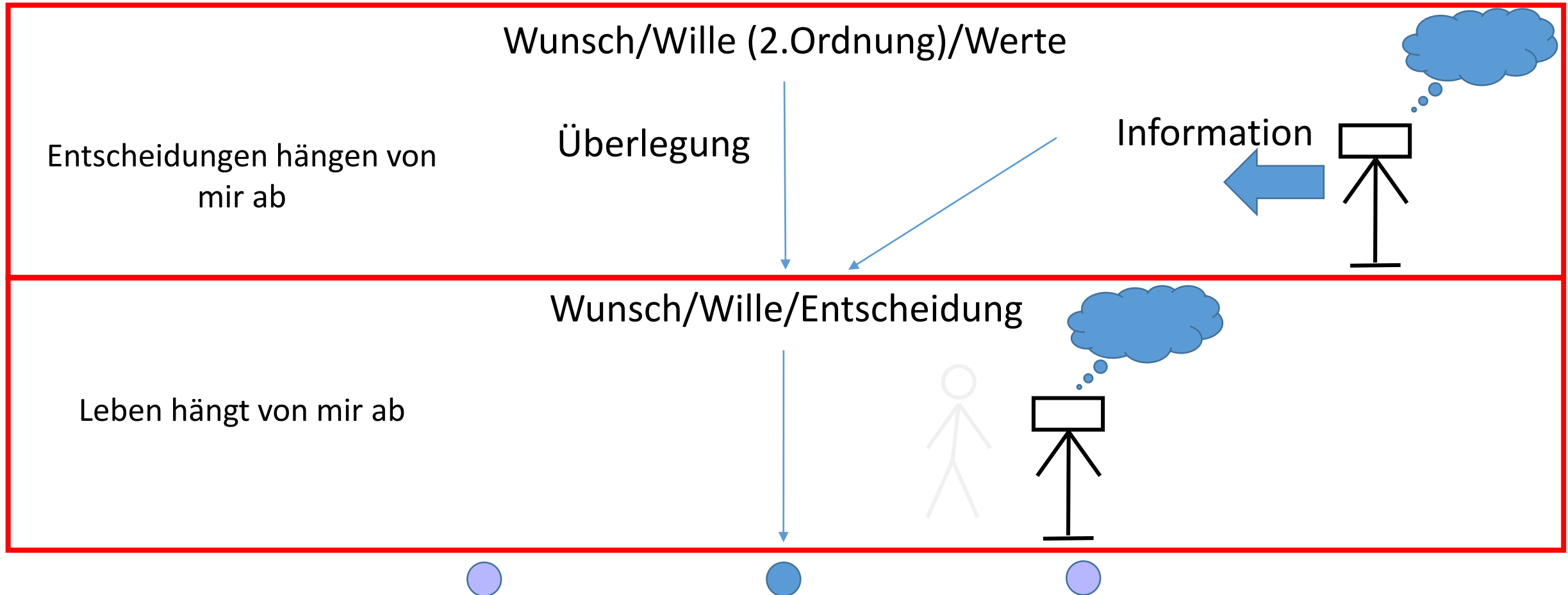
2

**to feel better,
wherever you go.**

Get all of the support you need to help improve your mental health—on your terms.

Wo ist der Einfluss?

42



Beispiel

43

Daily Mood Tracker

Answer simple questions to assess your mood, identify track health activities like exercise that can help.

[Learn more](#)

Guided Journeys

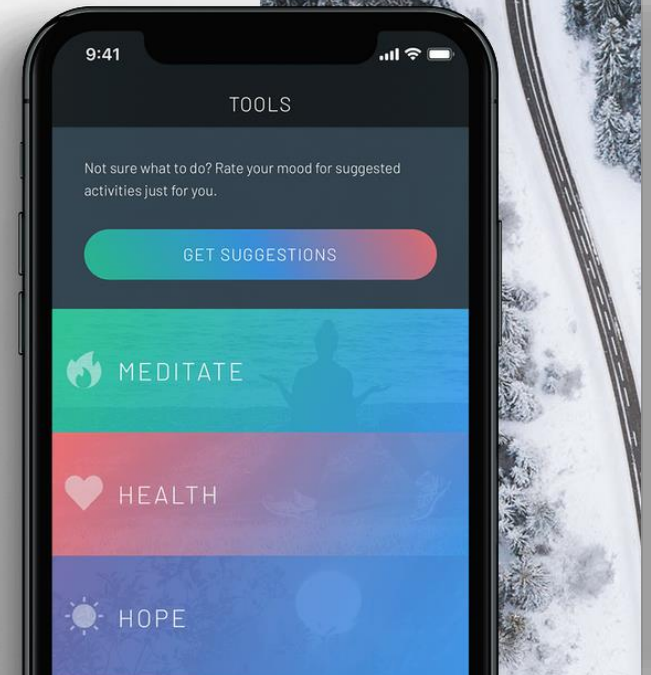
Designed by experts for a range of needs, Sanvello guided journeys use proven therapies to help you feel more in control and build long-term life skills.

[Learn more](#)

Coping Tools

Reach for just the right tool to relax, be in the moment, or manage stressful situations such as test-taking, public speaking, or morning dread.

[Learn more](#)

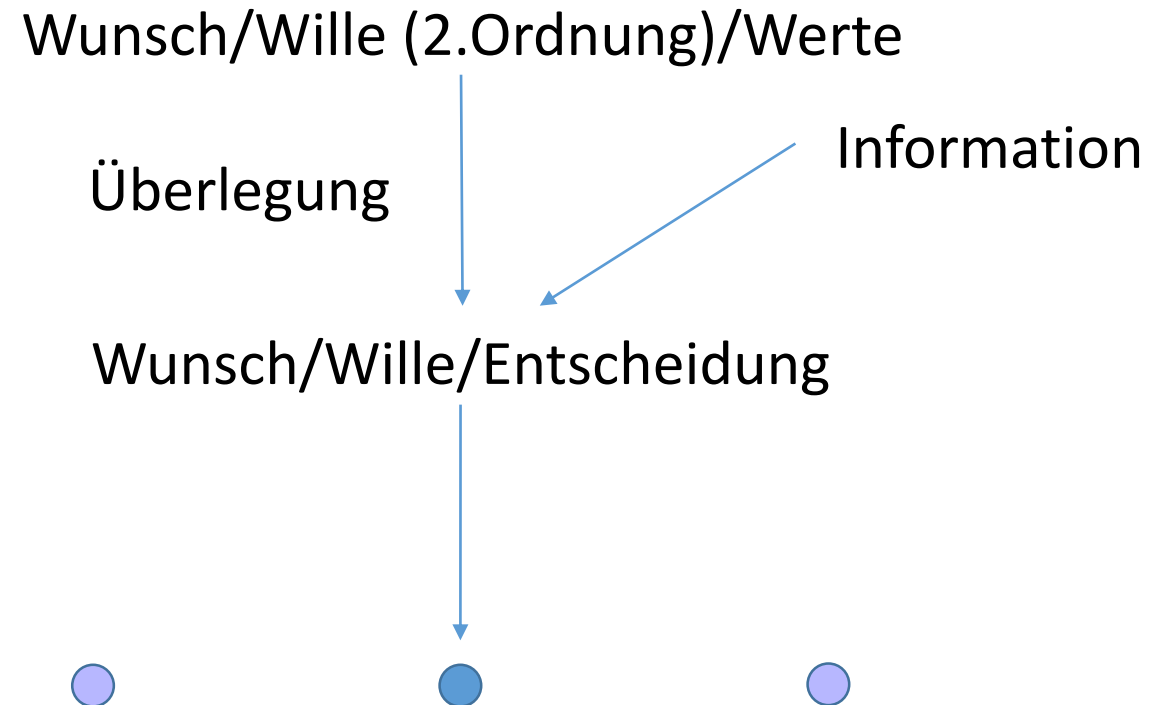


Systematik

44

Angriffspunkte:

- a. Information
- b. Entscheidungssituation
- c. Unterstützung Überlegung



a. Information

45

Positiv:

- Kann relevante Information liefern.



Bsp.: Tracking, Techniken, ...

Negativ:

- Kann Fehlinformation liefern.
- Kann relevante Information vorenthalten.



a. Fähigkeiten

Mensch in digitaler Transformation - Eine Vortragsreihe der Universität Bern

«Was man nicht mehr wissen muss: Ein Update»

Donnerstag, 16.11.2023, 18:00 Uhr



Sprachmodelle und künstliche Intelligenz: Im diesjährigen Vortrag zur Reihe «Mensch in digitaler Transformation» geht es darum, was sich durch die Fortschritte der Sprachmodelle, die zum Beispiel ChatGPT möglich machen, in Wissensarbeit und Wissensvermittlung verändert.

Was man nicht mehr *können* muss?

Beispiel Taschenrechner

Journal for Research in Mathematics Education
2003, Vol. 34, No. 5, 433–463

A Meta-Analysis of the Effects of Calculators on Students' Achievement and Attitude Levels in Precollege Mathematics Classes

Aimee J. Ellington, *Virginia Commonwealth University*

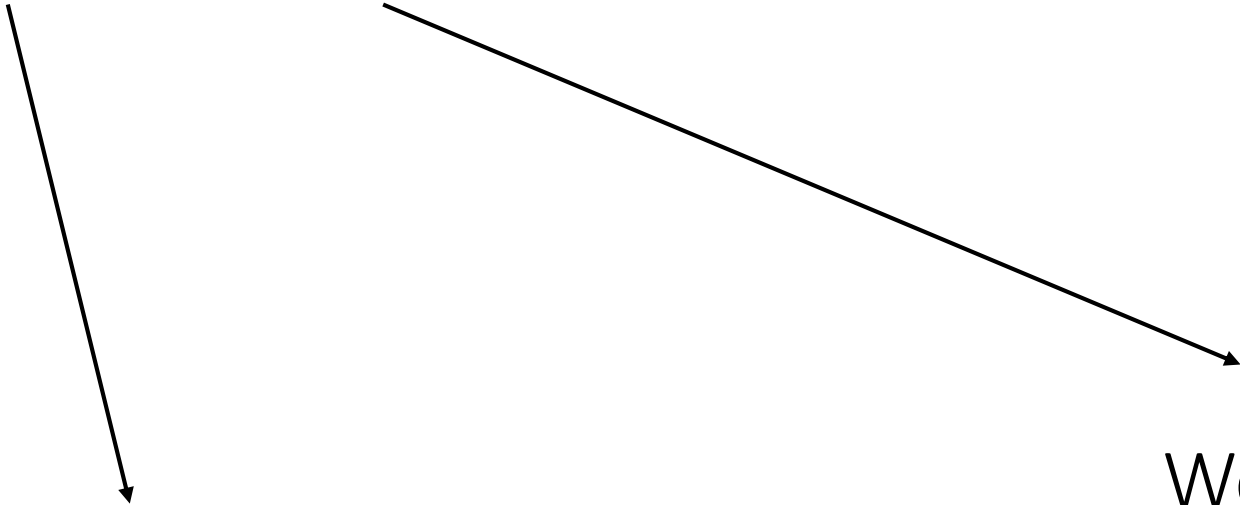
The findings of 54 research studies were integrated through meta-analysis to determine the effects of calculators on student achievement and attitude levels. Effect sizes were generated through Glassian techniques of meta-analysis, and Hedges and Olkin's (1985) inferential statistical methods were used to test the significance of effect size data. Results revealed that students' operational skills and problem-solving skills improved when calculators were an integral part of testing and instruction. The results



Fragen

- empirisch: wie entwickeln sich Fähigkeiten bei Gebrauch von KI, z.B. chatGPT?
- ethisch: Welche Fähigkeiten sind wichtig?

Wozu Fähigkeiten?



Wert als Mittel/um
anderer Ziele willen

Breit einsetzbare Mittel

Wert als Selbstzweck

Teil gutes Leben/Bed. für
Autonomie

Fähigkeiten als Mittel

Im Prinzip ersetzbar.

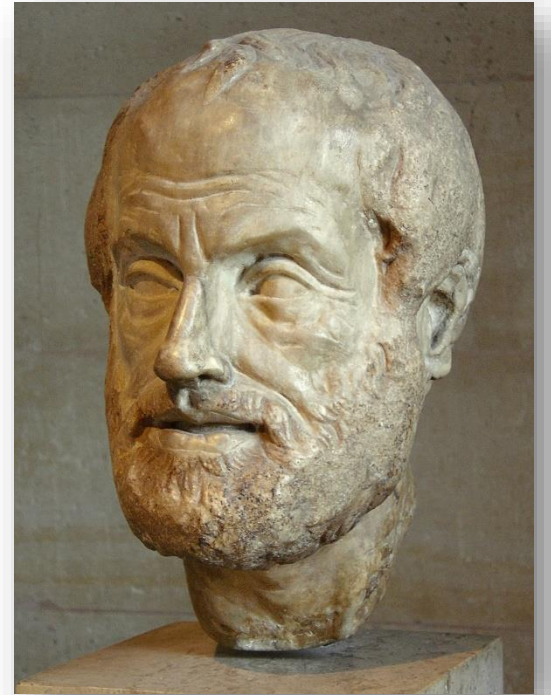
Argumente, die Fähigkeiten doch auszubilden:

1. „Notfall“: Wir brauchen die Fähigkeiten, wenn KI nicht zur Verfügung steht.
2. Klugheit: Wir sollten uns nicht abhängig von KI machen.
3. Beurteilung: Wir brauchen die Fähigkeiten, um beurteilen zu können, wie gut KI in den Fähigkeiten ist, und KI zu verbessern.

Fähigkeiten als Selbstzweck (1)

Gutes Leben:

- Aristoteles: Ausüben von charakteristischen Fähigkeiten
- M. Nussbaum (2006): capabilities:
 - imagination and thought
 - Practical reason



Fähigkeiten als Selbstzweck (2)



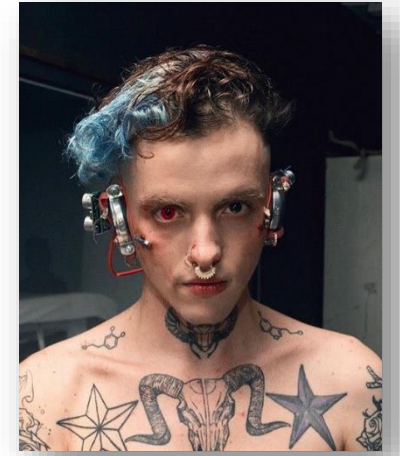
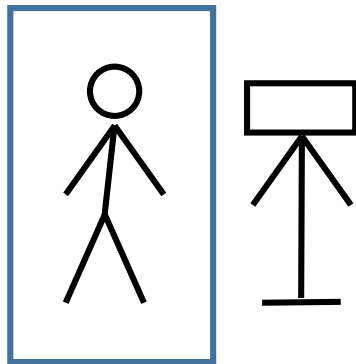
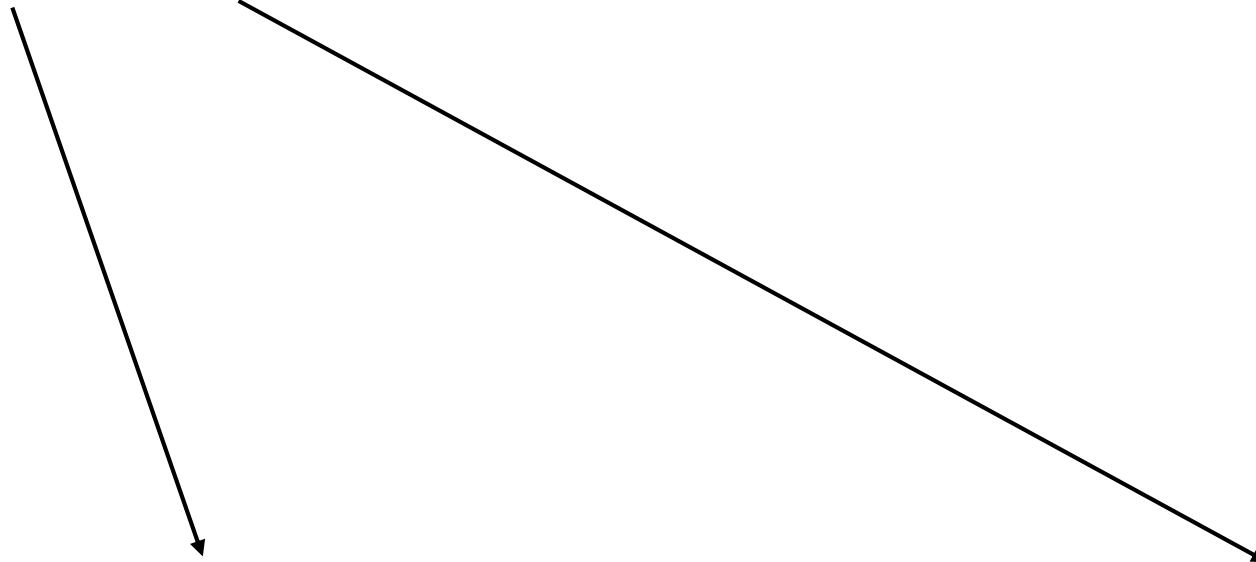
(1909-1997)

Autonomie:

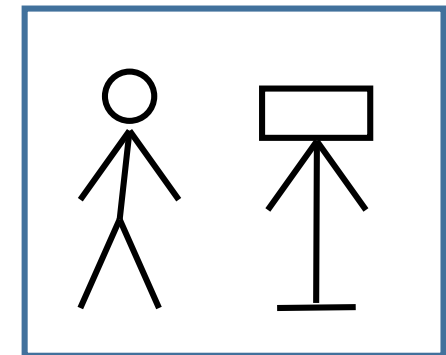
Isaiah Berlin (2002, 179):

„I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other men's, acts of will. I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from outside. I wish to be somebody, not nobody; a doer – deciding, not being decided for“

Zwei Selbstbilder

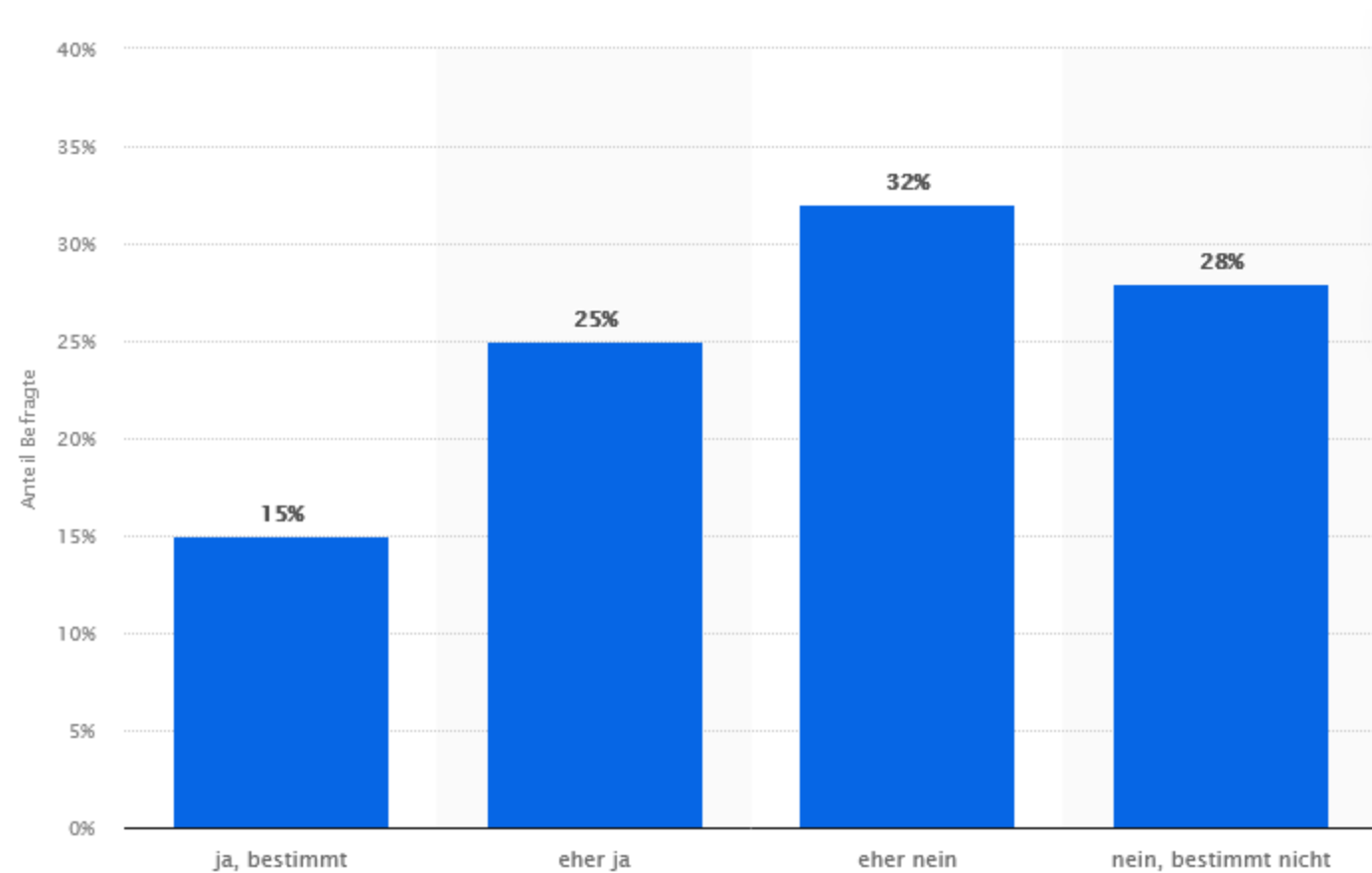


„Extended Mind“
Clark & Chalmers (1998)



Warum nicht?

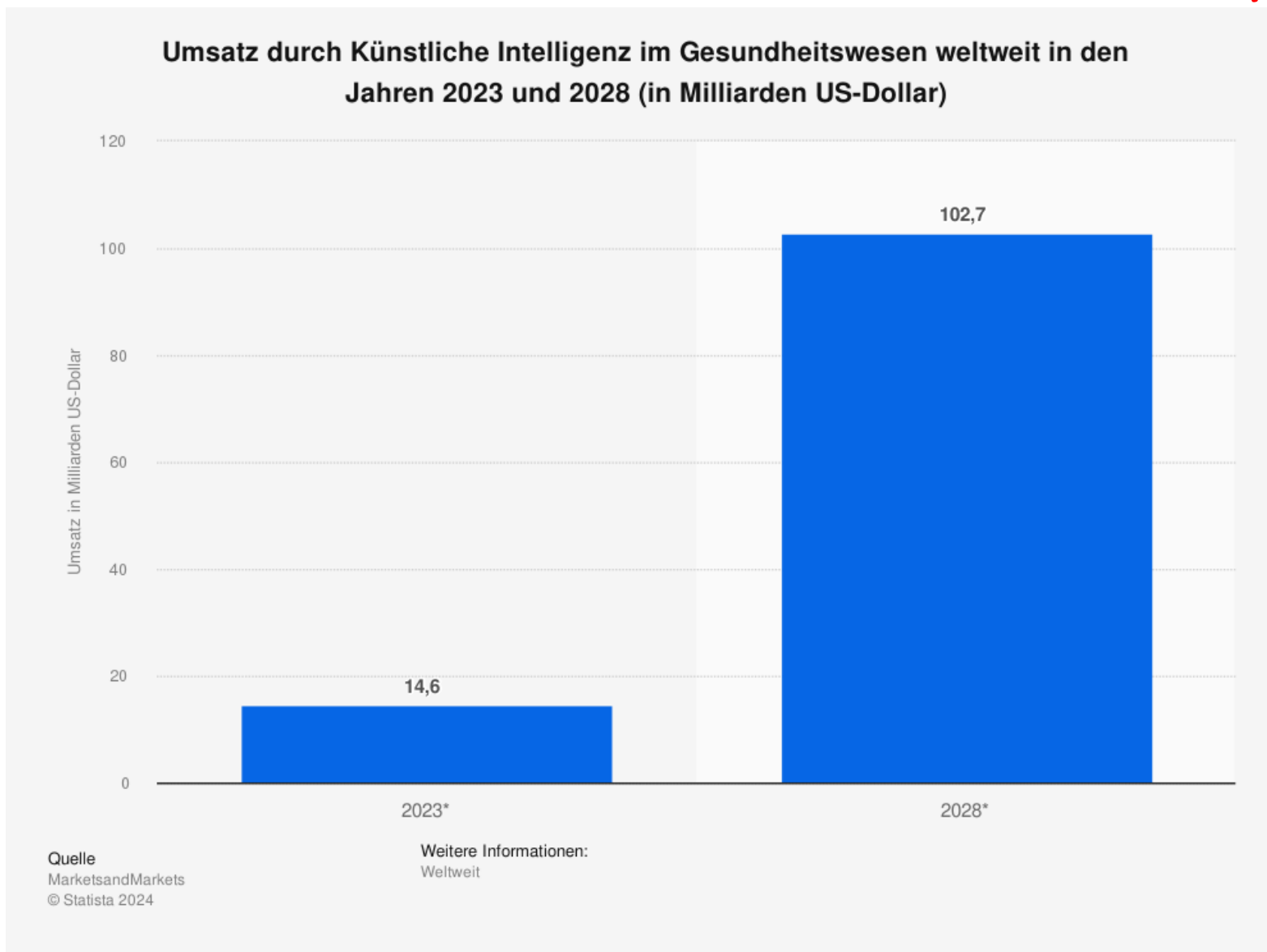
54



Werte

richt?

55



WUCC

„Bias“

KI berechnet
Risiken

Obermeyer et al. (2019, 447):

Bei people of colour werden
Risiken unterschätzt.

RESEARCH

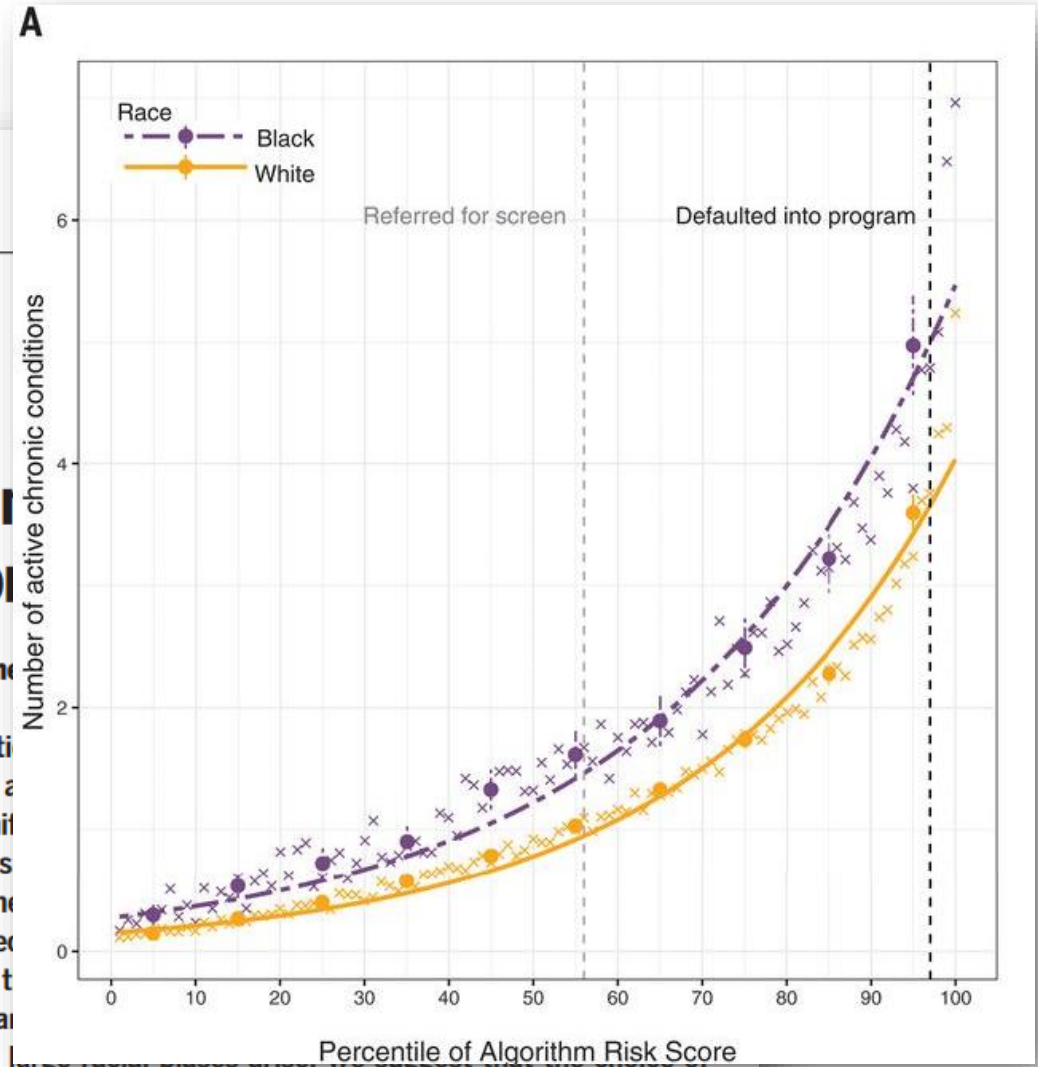
RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in the health of population

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine

Health systems rely on commercial prediction of health needs. We show that a widely used algorithm affecting millions of patients, exhibits significant bias: Black patients are considerably sicker than White patients. Remedying this disparity would increase the help from 17.7 to 46.5%. The bias arises because of illness, but unequal access to care means that for White patients. Thus, despite health care being biased by some measures of predictive accuracy, large-scale, convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.



3. Ethik der KI

KI

Wie kann KI
ethischer
werden?

Dürfen/sollten
wir KI
einsetzen?

Wie sollten wir
KI einsetzen?

Literature

- Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. 2019;8(7):2328-2331. doi:10.4103/jfmprc.jfmprc_440_19.
- Beauchamp, T. L. & Childress, J. F. 2009, Principles of Biomedical Ethics, 7th edition Oxford University Press, Oxford.
- Binns, R. (2020). On the apparent conflict between individual and group fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 514-524).
- Bostrom, N. 2014, Superintelligence. Paths, Dangers, Strategies, Oxford, Oxford University Press.
- Bringsjord, S. & Govindarajulu, N. S. (2022), Artificial Intelligence, *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), E. N. Zalta & U. Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/>
- Briganti, G. & Le Moine, O. 2020, Artificial Intelligence in Medicine: Today and Tomorrow, *Frontiers in Medicine* 7.
- Burrell, Jenna (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data and Society* 3 (1):205395171562251.
- Danaher, J. & Nyholm, S. Automation, work and the achievement gap. *AI Ethics* 1, 227–237 (2021). <https://doi.org/10.1007/s43681-020-00028-x>
- Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), e12760. <https://doi.org/10.1111/phc3.12760>
- Hatherley, Joshua James (2020). Limits of trust in medical AI. *Journal of Medical Ethics* 46 (7):478-481.
- Humphreys, P. 2009, The philosophical novelty of computer simulation methods, *Synthese* 169, 615–626.

- Johnson, G.M. Algorithmic bias: on the implicit biases of social technology. *Synthese* **198**, 9941–9961 (2021). <https://doi.org/10.1007/s11229-020-02696-y>
- Misselhorn, C. 2018, Grundfragen der Maschinenethik, Reclam, Stuttgart.
- Moor, J. H. 1979, Are There Decisions that Computers Should Never Make? *Nature and System* 1, 217–229.
- Müller, Vincent C., "Ethics of Artificial Intelligence and Robotics", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <<https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/>>.
- Obermeyer, Z. et al. (2019), Dissecting racial bias in an algorithm used to manage the health of populations, *Science* 366, Issue 6464, 447-453
- Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach 3rd edition*, Saddle River, NJ: Prentice Hall
- Seyyed-Kalantari, L., Zhang, H., McDermott, M.B.A. et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med* **27**, 2176–2182 (2021). <https://doi.org/10.1038/s41591-021-01595-0>
- 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
- Wallach, W. & Allen, C. 2009, *Moral Machines. Teaching Robots Right from Wrong*, Oxford University Press, New York.

Literatur (Fortsetzung)

Matthias, Andreas (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 6 (3):175-183.

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

Nam, J. G., Hwang, E. J., Kim, J., Park, N., Lee, E. H., Kim, H. J., ... & Goo, J. M. (2023). AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. *Radiology*, 307(2), e221894.

Nussbaum, M., 2006, *Frontiers of Justice: Disability, Nationality, Species Membership*, Cambridge, MA: Harvard University Press.

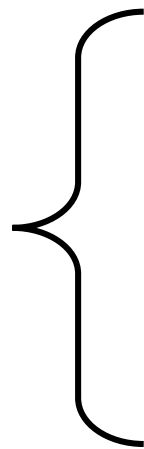
Regan, P., & Steeves, V. (2019). Education, Privacy and Big Data Algorithms: Taking the Persons out of Personalized Learning. *First Monday*.

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

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

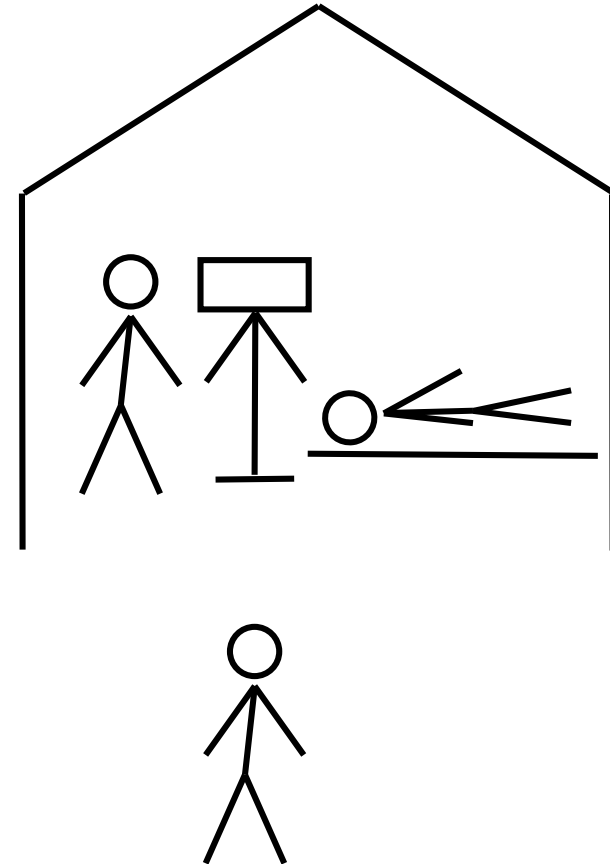
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, 46(4), 197-221.

Andere Aspekte Digitalisierung

- 
- Internet
 - Soziale Netzwerke
 - Virtuelle Realität
- 

iii. Um wen geht es?

- Patienten/-innen
 - Gesundheit
- Benutzende/Ärzte/-innen
 - Wert der Arbeit
- Software-Ingenieure
 - Verantwortung
- System („public health“)
 - Gerechtigkeit



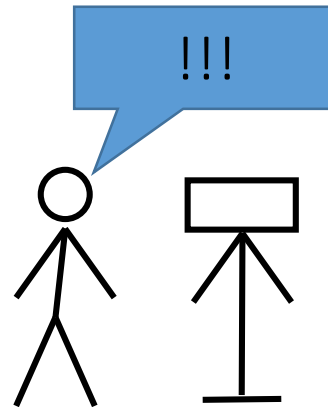
KI als Objekt

Ziel?

Nebeneffekte?

Missbrauch?

Anreize?



Risiken



KI als Subjekt

“*Maschinenethik* ist damit befasst, *Maschinen* ethischen Prinzipien oder Entscheidungsverfahren für Dilemmata [...] zu verleihen. Damit sollen die Maschinen durch ihre Entscheidungen in einer ethisch verantwortlichen Art und Weise funktionieren.”

Anderson & Anderson (2011, S. 1, Üs. CB)

