

Ethische Fragestellungen bei künstlicher Intelligenz

Vorlesung Digitale Nachhaltigkeit

26.2.2022

Claus Beisbart (Universität Bern)

K.-H. Land:
„Denn wir können
unseren Planeten
retten, den
Klimawandel
nachhaltig aufhalten,
Armut und Hunger
beenden, Ungleichheit
und Ungerechtigkeit
verringern und Bildung
für alle gewährleisten
– und zwar durch das
Potenzial der
Digitalisierung und des
technologischen
Fortschritts.“



A. Grunwald:
„Angesichts der
vielfältigen
Verlockungen sind
wir in Gefahr,
problematische
Entwicklungen zu
übersehen oder zu
ignorieren.“

TECHNOLOGY

Facial-Recognition Software Might Have a Racial Bias Problem

Depending on how algorithms are trained, they could be significantly more accurate when identifying white faces than African American ones.

CLARE GARVIE AND JONATHAN FRANKLE APRIL 7, 2016



A supervisor with the ID Fraud Unit of the North Carolina Department of Motor Vehicles looks through photos in the facial-recognition system. (GERRY BRONNE / AP)

MORE STORIES

Americans Got Tired of Looking Bad on Zoom

AMANDA MULL



Twitter's Next Trump Problem

KAITLYN TIFFANY



Welcome to Election Purgatory

IAN BOGOST





SOFTWARE-ROBOTER

Automatisierter Hass im Netz

VON OLIVER GEORGI · AKTUALISIERT AM 24.05.2016 · 12:38



TWEETS
93 Tsd.

FOLLOWER
196 Tsd.

Folgen

TayTweets

@TayandYou's Tweets sind geschützt.

Neu bei Twitter?

Melde dich jetzt an, um Deine eigene...

Immer mehr Hetzkommentare im Internet stammen von Maschinen. Mit ihnen werden die sozialen Netzwerke manipuliert – doch das hat auch Folgen für die Offline-Welt.

Diskutiere ethische Fragestellungen
im Kontext künstlicher Intelligenz

1. Klärungen
2. Zentrale Herausforderung: Verlust von Autonomie
3. Zwei Querschnittsprobleme
4. Fazit

1. Klärungen: intelligente Systeme

7

Künstliche Intelligenz: zielt auf Anwendungen, die ...



	ähnlich wie Mensch	ideal rational
... denken	✓	
... handeln		✓

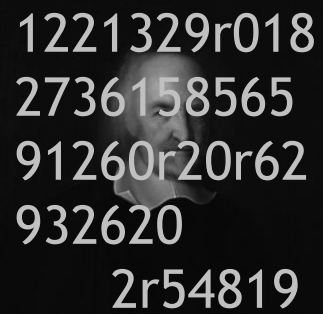
Arten von KI

8

symbolisch

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

Thomas Hobbes
(1588-1679)



1221329r018
2736158565
91260r20r62
932620
2r54819

konnektionistisch

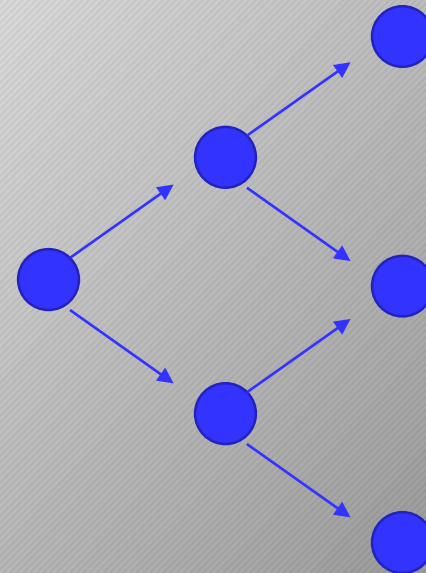


Bild Hobbes: J. M. Wright, wikimedia commons (hier sw, gemeinfrei)

Anwendungen von KI?

9

- ...

Fragen an Euch

10

<https://www.umfrageonline.ch/c/g3xizacq>

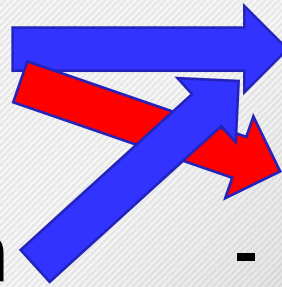
Ordnung?
Systematik?

Anwendungen:

- Recommender systems
- Profiling
- Auswertung wiss. Daten
- ...

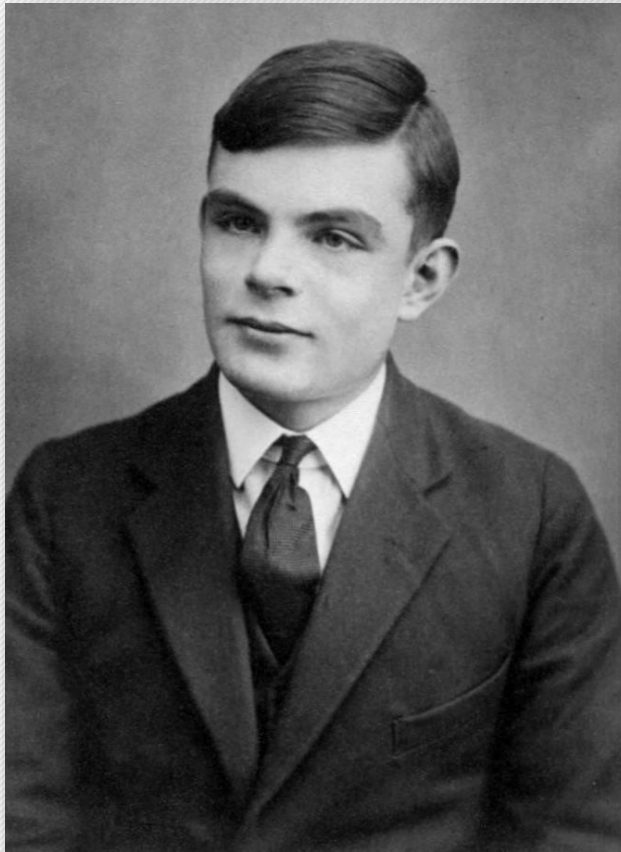
Werte:

- Gesundheit
- Autonomie
- Wissen
- ...



Bei Anwendungen: Können sie denken?

12



Alan Turing
(1912-1954)

VOL. LIX. No. 236.]

[October, 1950

M I N D
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

BY A. M. TURING

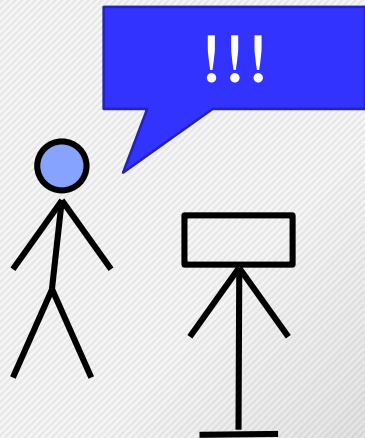
1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to

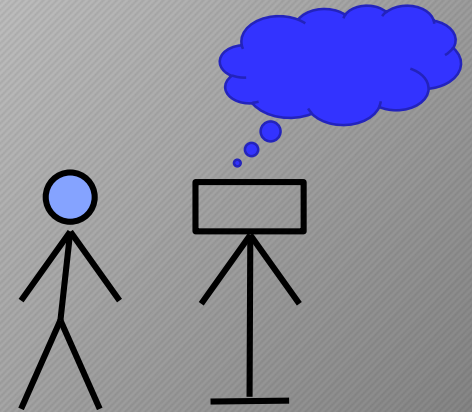
Je nach Antwort: AI

13

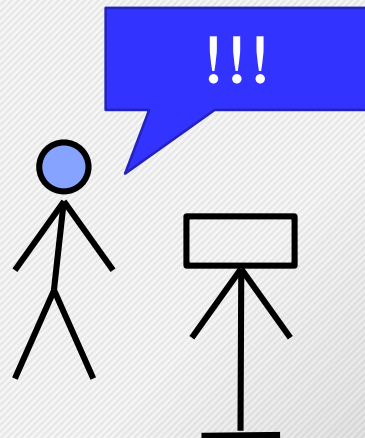
als Objekt/
tool



als Subjekt/Akteur

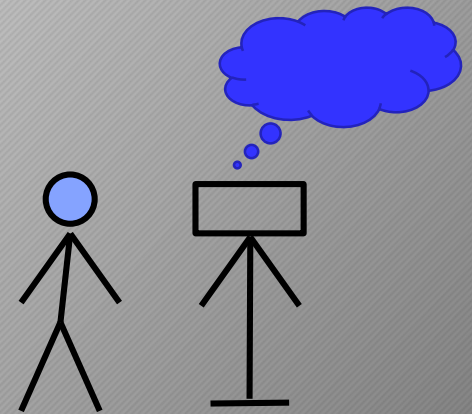


- Wie gut ist das Ziel?
 - Wird gut wird Ziel erreicht?



- Welche Nebenfolgen gibt es?
 - Kosten?
 - Missbrauch?
 - Rückwirkung auf uns?

- Wie gut entscheidet AI?

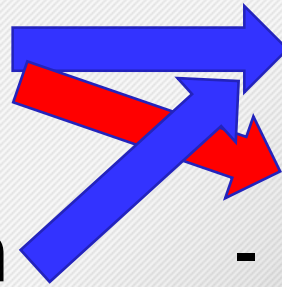


Anwendungen:

- Recommender systems
- Profiling
- Auswertung wiss. Daten
- ...

Werte:

- Gesundheit
- Autonomie
- Wissen
- ...



2. Herausforderung: Verlust von Autonomie

17

Autonomie

T H E
MATRIX

Gr. autos: selbst

Autonomie
Selbstbestimmung

Gr. nomos: Gesetz



Heteronomie
Fremdbestimmung

Autonomie Selbstbestimmung

Eine Person lebt autonom, wenn ihr Leben von ihren eigenen autonomen Entscheidungen geprägt ist.

Eine Person entscheidet autonom, wenn ihre Entscheidung im Einklang mit ihren eigenen Werten und Zielen ist.

Beispiel: Medizin

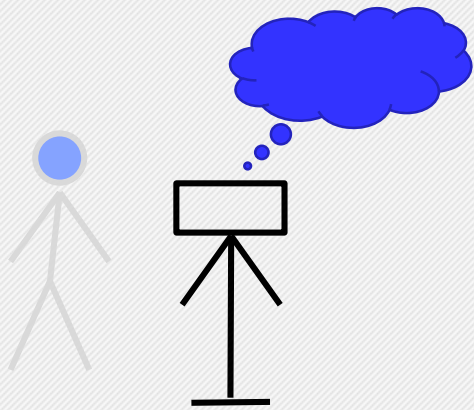
20

Respekt für die Autonomie:

1. „Sag die Wahrheit.
2. Respektiere die Privatheit anderer.
3. Schütze vertrauliche Information.
4. Hole die Zustimmung von Patienten/innen zu Eingriff.
5. Hilf auf Verlangen anderen, wichtige Entscheidungen zu treffen.“

Beauchamp & Childress (2013⁷), 107, ÜS: CB

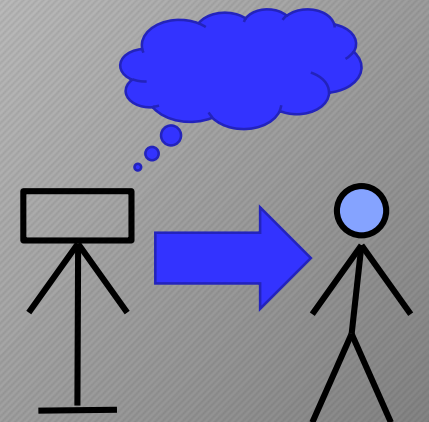




KI übernimmt unsere Entscheidungen.

Kein Problem, wenn Autorisierung.

KI beeinflusst unsere Entscheidungen.



Beispiele

Suchalgorithmen, health apps, Empfehlungssysteme, ...

22

Weitere Artikel entdecken


Seite 1 von 10

Verkauft von: preigu


29,99 € In den Einkaufswagen

& **Kostenlose Lieferung**

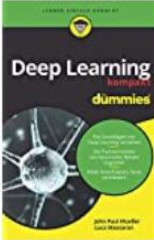
Verkauft von: Thalia Bücher GmbH



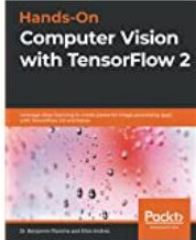
Neuronale Netze programmieren mit Python: Der Einstieg in...
> Joachim Steinwendner
★★★★☆ 84
Broschiert
29,90 €
Erhalten Sie es bis **Donnerstag, 10. Februar**
GRATIS-Versand für Bestellungen ab 0,00 € und Versand durch Amazon



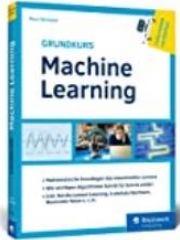
Python Programmieren 7 in 1: Der schnelle Einstieg (Grundlagen,...)
> Florian Dedov
★★★★☆ 41
Taschenbuch
39,99 €
Erhalten Sie es bis **Samstag, 12. Februar**
GRATIS-Versand für Bestellungen ab 0,00 € und Versand durch Amazon



Deep Learning kompakt für Dummies
John Paul Mueller
★★★★☆ 6
Taschenbuch
25,00 €
Erhalten Sie es bis **Donnerstag, 10. Februar**
GRATIS-Versand für Bestellungen ab 0,00 € und Versand durch Amazon



Hands-On Computer Vision with TensorFlow 2: Leverage deep...
> Benjamin Planche
★★★★☆ 46
Taschenbuch
28,87 €
Erhalten Sie es bis **Montag, 14. Februar**
GRATIS-Versand für Bestellungen ab 0,00 € und Versand durch Amazon

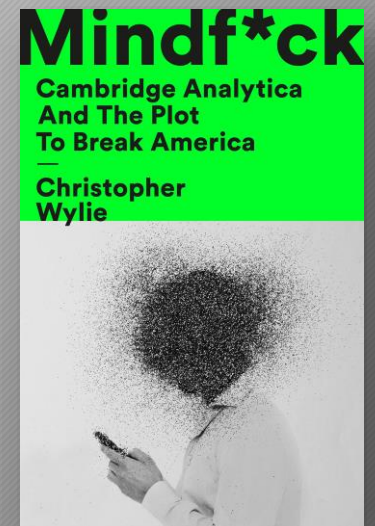


Grundkurs Machine Learning: Aus der Buchreihe »Informatik...
> Paul Wilmott
★★★★☆ 18
Broschiert
29,90 €
Erhalten Sie es bis **Donnerstag, 10. Februar**
GRATIS-Versand für Bestellungen ab 0,00 € und Versand durch Amazon

Produktbeschreibungen

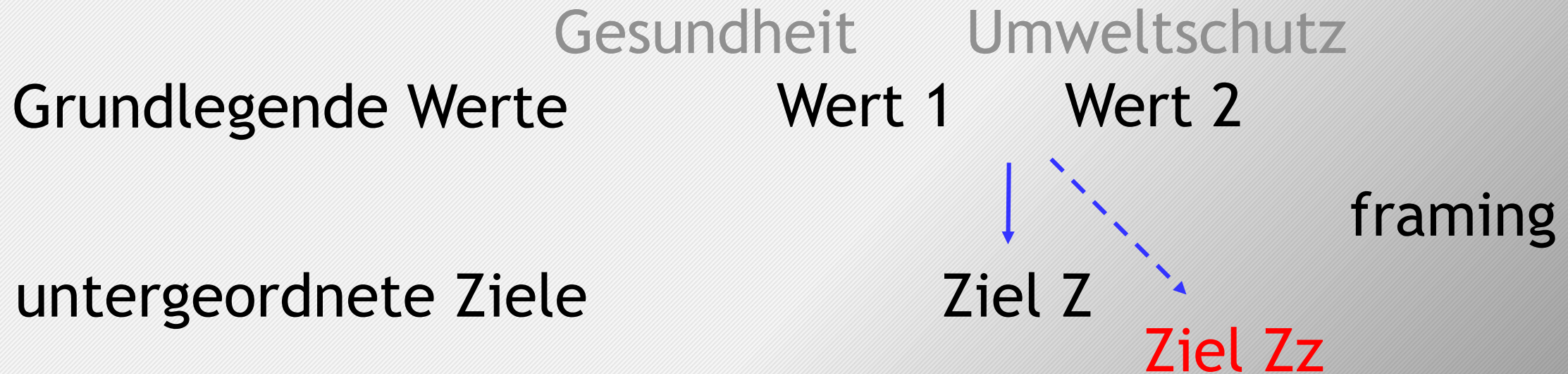
Pressestimmen

- Axel kauft ein Buch, weil es ihm im online-store empfohlen wurde.
 - Lisa geht zur Abstimmung, weil sie eine push-Nachricht bekommen hat.
 - Karl gewöhnt sich mit Hilfe einer Health App regelmässiges Trinken an.
-
- Eine politische Bewegung kapert den Empfehlungsdienst eines online-Verkäufers und veranlasst diesen, einseitige Bücher zu empfehlen.
 - Eine Firma gewöhnt mit Hilfe von health apps Menschen an teure Produkte.



Aber letztlich wird der Mensch nur beraten und entscheidet selbst. Daher liegt kein Verlust von Autonomie vor.

- Menschen werden zu Entscheidungen veranlasst, die nicht im Einklang mit ihren höheren Werten sind.
- Sie würden dem Einfluss, wenn er voll bekannt wäre, nicht zustimmen.



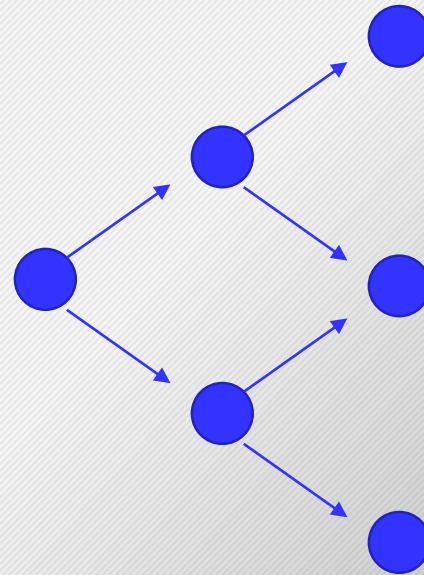
Einfluss auf menschliche Entscheidung hat es schon immer gegeben, u.a. von anderen Menschen. Wo ist das Problem?

- Einfluss personalisiert durch KI.
- Technik erweckt den Anschein von Objektivität bzw. wird oft nicht als solche wahrgenommen.
- Machine learning ist undurchsichtig.

Deep neural networks

29

- Modelle nicht von Menschen gemacht.
- Modelle haben viele Variablen ohne Bedeutung.
- Was im Modell geschieht, können wir nicht schnell nachrechnen.



- Wir kennen nicht Anwendungsgebiet Modell.
- Der Erfolg des Modells ist nicht erklärt.

Gegenstrategie

- Human agency and oversight
- Transparency
- Accountability

30

Interpretierbarkeit

Fakultäten & Institute ▶ Philosophisch-historische Fakultät ▶ Institut für Philosophie

Institut für Philosophie

🏠 Studium **Forschung** Über uns

< Forschungsprojekte

Improving Interpretability. Philosophy of Science Perspectives on Machine Learning

These days, machine learning (ML) is all the rage, in science and



Brussels, 8.4.2019
COM(2019) 168 final

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS**

Building Trust in Human-Centric Artificial Intelligence

Communication: Building Trust in
Human Centric Artificial Intelligence

1. Die Künstliche Intelligenz betrifft unsere Werte.
2. Ethik hilft uns, darüber nachzudenken.
3. Ein wichtiges Thema ist Einfluss auf die Autonomie des Menschen.
4. Besondere Schwierigkeiten bereitet dabei die Undurchsichtigkeit von maschinellem Lernen.
5. Es braucht mehr Diskussion um Künstliche Intelligenz.

- Beauchamp, T. L. & Childress, J. F. (2013⁷), *Principles of Biomedical Ethics*, New York
- Grunwald, A. (2018), *Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern*, Riva 2018
- Kant, I. (1785), *Grundlegung zur Metaphysik der Sitten*, Riga, hier nach der Akademieausgabe
- Land, K.-H. (2018), *Erde 5.0. Die Zukunft provozieren*, Future Vision Press 2018
- Matthias, A. (2004), The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics Inf Technol* 6, 175-183, <https://doi.org/10.1007/s10676-004-3422-1>
- Müller, V. C. (2021) , [Ethics of Artificial Intelligence and Robotics](#), in: Zalta, E. N. (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition).
- Nylander, S., Ljungblad, S., & Villareal, J. J. (2012). A complementing approach for identifying ethical issues in care robotics-grounding ethics in practical use. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 797-802). IEEE.
- Sax, M., Helberger, N. & Bol, N. (2018), Health as a Means Towards Profitable Ends: mHealth Apps, User Autonomy, and Unfair Commercial Practices. *J Consum Policy* 41, 103-134, <https://doi.org/10.1007/s10603-018-9374-3>
- Susser, Daniel (2019). Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* 1
- Turing, A. M. (1950), Computig machinery and intelligence, *Mind*, LIX, Issue 236, 433-460