

Künstliche Intelligenz und Ethik

Aktuelle Themen (253504a)
Wintersemester 21/22
Hochschule der Medien Stuttgart

Dozent: Prof. Dr. Andreas Koch

Bastian Fritz
bf033@hdm-stuttgart.de
Matr.-Nr.: 42784

Einführung

Das gesellschaftliche Interesse um Künstliche Intelligenz steigt und sinkt phasenweise. Mit der exponentiellen Entwicklung von Informationstechnologien steigt auch die Anzahl der Anwendungen von KI. Der Ausbau von Regelungen rund um Informationstechnologien und KI kann jedoch nicht mit der Entwicklung der einzelnen Technologien mithalten.

In der vorliegenden Arbeit wird die Beziehung zwischen Künstlicher Intelligenz und Ethik aufgezeigt, untersucht und bewertet. Trotz der rasanten Entwicklung neuer Technologien bleibt die Frage offen, ob neue Technologie auch automatisch besser ist. Nachdem grundlegende Begriffe erläutert wurden, wird die Problematik rund um Ethik und KI beschrieben und dargestellt, welche Folgen daraus entstehen. Dabei betrifft der Einsatz von Künstlicher Intelligenz nicht nur Unternehmen in der Technologiebranche, sondern kann Einflüsse auf den Alltag von Privatpersonen haben.

Grundlagen

Bevor die Beziehung zwischen Künstlicher Intelligenz und Ethik untersucht werden kann, müssen beide Begriffe ausführlich definiert werden. Das Oxford Languages Dictionary bezeichnet Ethik als »philosophische Disziplin oder einzelne Lehre, die das sittliche Verhalten des Menschen zum Gegenstand hat;«¹. Künstliche Intelligenz ist im Gabler Wirtschaftslexikon als »Erforschung „intelligenten“ Problemlösungsverhalten sowie die Erstellung „intelligenter“ Computersysteme [definiert]. Künstliche Intelligenz (KI) beschäftigt sich mit Methoden, die es einem Computer ermöglichen, solche Aufgaben zu lösen, die, wenn sie vom Menschen gelöst werden, Intelligenz erfordern.«². Wenn menschliches Verhalten Sitten und Moral unterliegt, ist zu bestimmen, welche Richtlinien für Künstliche Intelligenz gelten, wenn KI als „intelligentes“ Problemlösungsverhalten betrachtet wird. Aus diesem Grund wirft der Einsatz, sowie der Umgang mit Künstlicher Intelligenz ethische Fragen auf. Folglich ist zu analysieren, wie Ethik im Bereich der KI definiert wird und wie die Rolle dieses Themenbereiches zu gewichten ist. Die Stanford Enzyklopädie für Philosophie beschreibt den Umgang mit Ethik in KI wie folgt: »ethics of AI and robotics is often focused on “concerns” of various sorts, which is a typical response to new technologies.«³.

¹ Oxford Languages Dictionary, 2022

² Lackes/Siepermann, 2018

³ Bringsjord/Govindarajulu, 2018

Ethische Fragen

Aus der Beschreibung von Ethik der Stanford Enzyklopädie für Philosophie lässt sich die grundlegende Problematik ermitteln, wenn es zu Ethik in KI kommt. Im folgenden Kapitel wird diese Problematik anhand von drei Themenbereichen vorgestellt und analysiert. Diese Analyse bildet die Basis für die weiterführende Darstellung von Ethik in KI und möglichen Lösungsansätzen, die in der vorliegenden Arbeit vorgestellt werden. Die drei Bereiche sind unterteilt in: Privatsphäre, Transparenz und Voreingenommenheit.

Transparenz

Die erste Frage, die der Einsatz von Künstlicher Intelligenz aufwirft, ist die nach der Transparenz, meist durch das Black Box Modell repräsentiert.

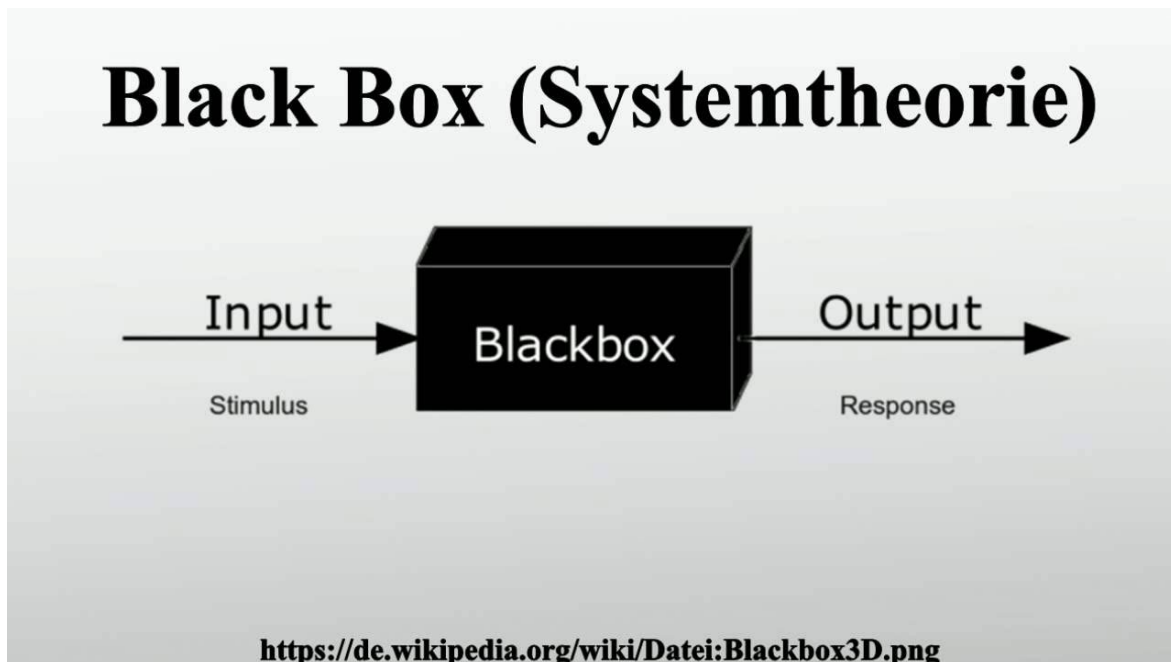


Abbildung 1: Darstellung des Black-Box-Modells

Dabei wird der Entscheidungsprozess der KI als Blackbox beschrieben. Es ist bekannt, welche Daten die KI erhält, diese schließen Trainings- und Arbeitsdaten ein. Außerdem ist das Ergebnis der Künstlichen Intelligenz ersichtlich. Problematisch ist jedoch die Verbindung zwischen Input und Output. Es ist nicht verfolgbar, welche Datensätze für das Endergebnis eine Rolle spielen, geschweige denn, wie die einzelnen Informationen gewichtet werden.

Das führt zum Aufkommen der nächsten ethischen Fragen. Wenn der Entscheidungsfindungsprozess einer KI nicht verfolgbar ist, wie kann eine faire und unvoreingenommene Entscheidung garantiert sein?

Voreingenommenheit

Wie bereits erwähnt, arbeiten Künstliche Intelligenzen sowohl mit Trainingsdaten als auch mit den zu verarbeitenden Daten, um eine gewünschte Aufgabe zu lösen. Hierbei ist jedoch zu beachten, dass die Entscheidung bzw. Lösung einer KI nicht unvoreingenommen ist. Grund dafür sind die bereits belasteten Daten, die die KI zur Entscheidungsfindung nutzt.

Obermeyer et al. analysierten die Ergebnisse einer KI, die Zuschüsse für die Krankenversicherung amerikanischer Patienten verteilte. Dabei konnte festgestellt werden, dass diese Verteilung auf Grund von voreingenommenen Daten nicht ausgeglichen war: »The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients.«⁴. Folglich erhalten weiße Menschen höhere Zuschüsse für ihre Krankenversicherung im Gegensatz zu Personen aus benachteiligten Minderheiten.

Privatsphäre

Künstliche Intelligenz bearbeiten eine große Menge Daten zur Lösung von Aufgaben. Wird die Verteilung von Krankenzuschüssen als Beispiel betrachtet, müssen spezielle Krankenakten, Kostenaufstellungen und private Ansprüche bearbeitet werden. Dabei kann es sich teilweise um sehr sensible Daten handeln. Es ist zu beachten, dass eine Künstliche Intelligenz keine Unterschiede zwischen öffentlich verwertbaren und privaten Daten macht.⁵

Unklarheit

Die zuvor beschriebenen Bereiche der ethischen Probleme, sind die Folgen eines kaum geregelten Einsatzes von KI. Ein Beispiel für die ungenauen Prinzipien rund um Künstliche Intelligenz ist Project Maven. Dieses Projekt des amerikanischen Verteidigungsministeriums, Pentagon, nutzt Künstliche Intelligenz zur Analyse von Drohnenaufnahmen, um Personen und Objekte zu erkennen. Diese Ergebnisse werden für militärische Operationen genutzt, um Informationen über »militärisch bedeutungsvolle Objekte«⁶ zu sammeln.

Das Projekt erlangte Aufmerksamkeit, als Mitarbeiter des Technologiekonzerns Google gegen die Teilnahme Googles am Projekt protestieren⁷. Infolge von Petitionen und Kündigungen

⁴ Obermeyer et al., 2019

⁵ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

⁶ Peitz, 2018

⁷ Peitz, 2018

beendete Google die Zusammenarbeit mit dem Pentagon für Projekt Maven und verfasste darauf die sogenannten Google AI Principles. Diese Prinzipien bestimmen einen, laut Google, korrekten Umgang mit Künstlicher Intelligenz: »We will not pursue AI applications or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.«⁸.

Dieser Auszug aus oben genannten Prinzipien ist jedoch kritisch zu betrachten. Im Zuge der Veröffentlichung der AI Principles bestätigte Google eine weiterführende Zusammenarbeit mit dem Pentagon an unterschiedlichen Projekten, so auch »search and rescue missions«⁹. Dabei ist anzumerken, dass die technischen Elemente von search and rescue Missionen, denen von target Missionen, wie Project Maven, sehr ähneln.¹⁰

Dieses Verhalten spiegelt die Unklarheit wider, wenn es zu Regelungen rund um den Einsatz von Künstlicher Intelligenz kommt. Das betrifft nicht nur die ethischen Aspekte von KI. In Googles AI Principles heißt es weiter: »We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control.«¹¹

Dabei ist nicht beschrieben, wie eine solche Kontrolle oder menschliche Überwachung gestaltet ist, geschweige denn in welchem Maße diese »appropriate«¹² Kontrolle Einsatz findet. Zusammenfassend lässt sich die momentane Regelung von Künstlicher Intelligenz als wenig effektiv bezeichnen, wobei es sich im schlimmsten Fall um reine Marketingmaßnahmen handelt.

Rollen der Ethik

Um den Einsatz von Künstlicher Intelligenz ethisch vertretbarer zu machen, gibt es unterschiedliche Ansätze. So beschreiben die »Triple Roles of Ethics for IT« den Einbezug ethischer Fragen in unterschiedlichen Schritten der Künstlichen Intelligenz.¹³

⁸ Google AI, 2021

⁹ vgl. TED, 2019

¹⁰ vgl. TED, 2019

¹¹ Google AI, 2021

¹² Google AI, 2021

¹³ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

Die drei Rollen sind auf die einzelnen Nutzungsaspekte von KI unterteilt: die Entwicklung, die Verwendung und der Algorithmus.¹⁴ Dabei ist festgelegt, wie Ethik in diesen Bereichen angewendet werden kann und welche Konsequenzen dabei entstehen.



Abbildung 2: Darstellung der drei Rollen der Ethik für KI

Ethics of Profession bezeichnet die Entwicklung von Künstlicher Intelligenz mit Hinsicht auf mögliche ethische Fragen und dem Versuch diese Probleme bereits vor der Anwendung zu lösen.¹⁵ »The ethics of profession ist the first angel of looking at the intersection of computer science broadly conceived and ethics.«¹⁶ Damit beschreibt Simon die erste Rolle von Ethik im Bereich der KI als Hilfsmittel für Gestalter und Entwickler von Algorithmen, um sich ethisch korrekt zu verhalten und dieses Verhalten auch im Hinblick auf ihre Arbeit anzuwenden.¹⁷ Dabei handelt es sich in den meisten Fällen um sogenannte "Codes of Conduct", in denen Richtlinien festgelegt sind, um Software zu entwickeln, die ethisch vertretbar sind.¹⁸ Diese sind vergleichbar mit Richtlinien, die in der Medizin für Ärzte und Pflegepersonal festgelegt sind.

Die zweite Möglichkeit, eine Verbindung zwischen Ethik und Künstlicher Intelligenz zu betrachten, ist Ethik in der Nutzung zu analysieren. Dabei wird untersucht, wie die Schnittstelle zwischen den Nutzern und der Informationstechnologie aufgebaut ist.¹⁹ Es werden Fragen zur Verwendung von KI und den Berührungspunkten mit den Nutzern gestellt: »For what purpose can we use customer data? (...) How to weigh different values and interests?«.²⁰ Kunden können in diesem Fall sowohl Privatpersonen als auch Unternehmen oder Regierungen sein.²¹

¹⁴ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

¹⁵ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

¹⁶ Alexander von Humboldt Institut für Internet und Gesellschaft, 2021 00:13:50 – 00:14:00

¹⁷ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

¹⁸ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

¹⁹ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

²⁰ Alexander von Humboldt Institut für Internet und Gesellschaft, 2021, 00:14:50 – 00:15:10

²¹ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

Der dritte Bereich beschreibt eine ethische Betrachtung der Künstlichen Intelligenz. Es wird hier zwischen der ethischen Analyse bestehender Technologie und der Zukünftiger unterschieden.²² Grundlegend für diese Analyse ist die Annahme, dass Künstliche Intelligenz nicht unvoreingenommen in ihrem Entscheidungsprozess ist. Brey beschreibt Technologie und Werte wie folgt: »computer systems and software are not morally neutral and (...) it is possible to identify tendencies in them to promote or demote particular moral values and norms.«²³

Aus dieser Aussage ist im weiteren Schritt zu untersuchen, wie Künstliche Intelligenz oder zukünftige Technologien entwickelt werden müssen, um gesellschaftliche und ethisch vertretbare Werte zu sichern. Im Bereich des Values in Design wird oben genannte Problematik kontinuierlich erforscht, um das ethische Dilemma in der Technologie zu lösen. Durch die Kooperation von Philosophen, Gestaltern und Entwicklern von Künstlicher Intelligenz können Werte, wie Transparenz, Privatsphäre und Fairness während der Entwicklung von Technologien beachtet werden, um so einen ethisch vertretbaren Einsatz von KI zu gewährleisten.²⁴ Die Diversität des Teams spielt laut Simon eine große Rolle, da die Entwicklung der KI durch unterschiedliche Blickwinkel profitiert.²⁵

Lösungsansätze

Nachdem ein Überblick über die Problematik von Ethik und KI dargestellt wurde, werden im Folgenden zwei Lösungsansätze vorgestellt, die den Versuch unternehmen, Künstliche Intelligenz ethisch vertretbarer zu machen.

Ethik Label

Die Vergabe von Ethik Label wurde von der AI Ethics Impact Group entwickelt. Die Ethik Label sehen eine kategorische Bewertung unterschiedlicher Algorithmen in Bezug auf eine ethische Korrektheit vor. Diese Bewertung basiert auf einem dreistufigen Modell: Einem philosophischen Modell zur Kategorisierung einzelner ethischer Werte, einer Risikoanalyse anhand einer Matrix, sowie einer visuellen Einordnung durch sogenannte Labels.

²² vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

²³ Floridi, 2010, S. 42

²⁴ vgl. Alexander von Humboldt Institut für Internet und Gesellschaft, 2021

²⁵ vgl. ebd.

Die AI Ethics Impact Group nutzt das sogenannte VCIO-Modell zur kategorischen Bewertung von Künstlicher Intelligenz. VCIO umfasst Werte, Kriterien, Indikatoren und Beobachtungen. Die Impact Group beschreibt die Zusammensetzung des Modells wie folgt: »As values are abstract, often in conflict with each other, and do not include means to evaluate their implementation, it is essential to have other components to fulfil these tasks.«²⁶

Abbildung 3 stellt die Bewertung von KI durch das VCIO-Modell dar und eine darauf basierende Einordnung der unterschiedlichen Labels.

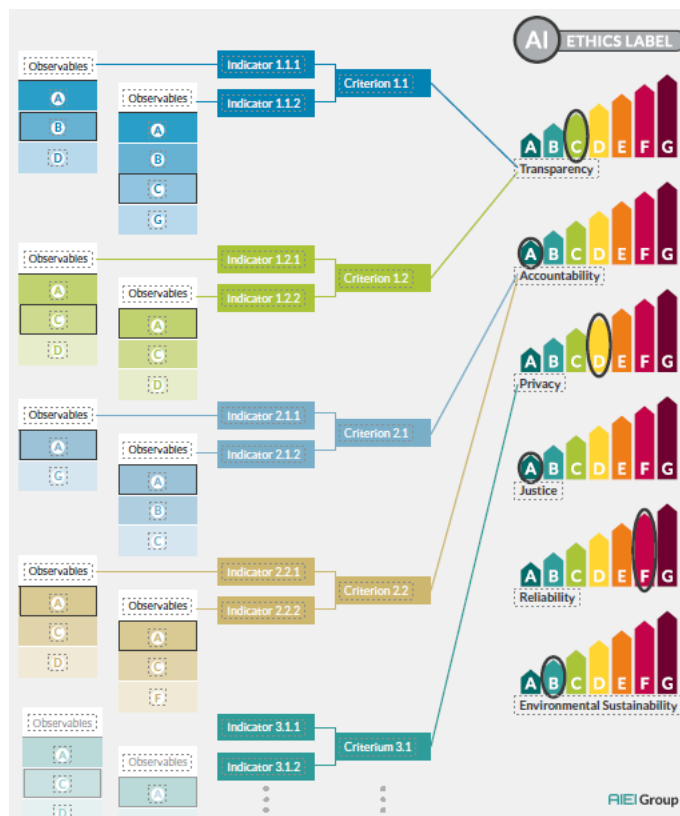


Abbildung 3: Verteilung des VCIO-Modells auf die visuellen Labels

Die Anwendung einer KI kann sich in manchen Fällen von dem vorhergesehenen Ergebnis unterscheiden, weshalb die Impact Group die Einteilung der KI in eine Risikomatrix vornimmt.²⁷ Dabei wird ein Algorithmus nach der Intensität des möglichen Schadens und der Abhängigkeit von betroffenen Personen an der Entscheidung der KI bewertet.²⁸

Explainable AI (XAI)

Um Künstliche Intelligenz und ihre Ergebnisse transparenter und verständlicher zu gestalten, wird die Möglichkeit einer erklärbaren KI erforscht. Abbildung 4 zeigt das Ergebnis einer KI zur Bestimmung der Fahrtdauer für eine Radtour anhand von Wetterdaten. Die Statistik stellt die

²⁶ AI Ethics Impact Group, 2020, S. 15

²⁷ vgl. AI Ethics Impact Group, 2020, S. 35

²⁸ vgl. ebd., S. 37-38

einzelnen Werte dar, die die KI zur Berechnung ihres Ergebnisses benötigt hat. Dabei erhält jeder eine Gewichtung, je nach Einbezug auf die schlussendliche Berechnung.²⁹

Predicted duration: 11.1651134 minutes
Actual duration: 10.0 minutes

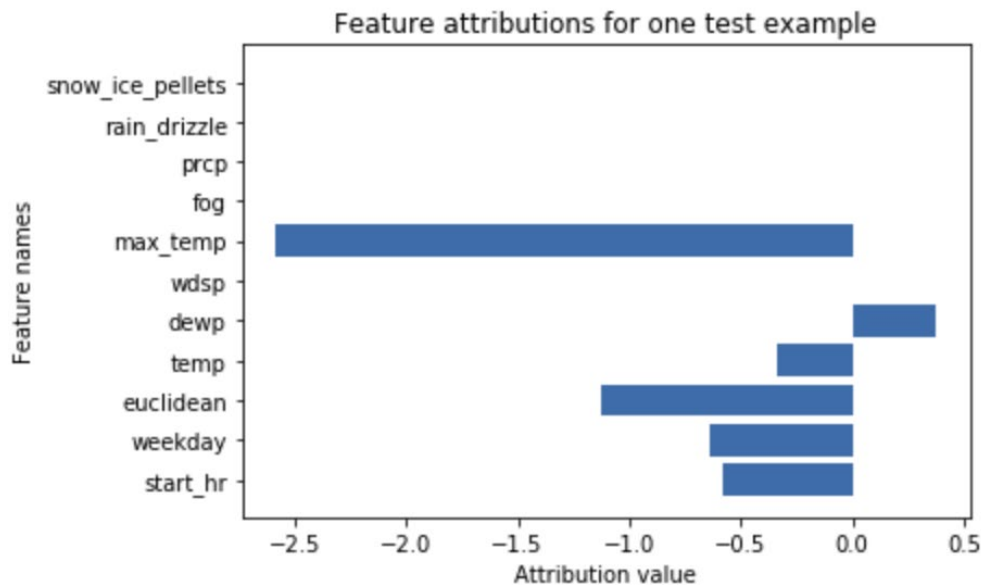


Abbildung 4: Gewichtung von Attributionswerten durch eine KI

Mit der KI von Google wird der Versuch unternommen, die Problematik des Black-Box-Modells zu eliminieren und dadurch den Entscheidungsfindungsprozess einer Künstlichen Intelligenz ersichtlicher zu machen.

Fazit

Der Einsatz von Künstlicher Intelligenz findet immer mehr Einzug in den Alltag der Gesellschaft und so auch die damit einhergehenden ungelösten Probleme. Die Entwicklung von Informationstechnologien wird in Zukunft weiter rasant voranschreiten. Dabei ist es wichtig, dass die Regelungen und Vorschriften nicht hinterherhinken.

Wie bereits die KI-Ethik-Forscherin Timnit Gebru im Interview mit Vice sagte: » (...) there needs to be certain oversights and certain measures, the corporations are not going to do it on their own.«³⁰. Der Einsatz von KI kann in vielen Fällen sehr nützlich sein. Es muss jedoch auch gefragt werden, ob dieser trotz möglicher Nachteile notwendig ist.

²⁹ vgl. Google Cloud, 2022

³⁰ VICE News, 2020

Literaturverzeichnis

AI Ethics Impact Group: » An interdisciplinary framework to operationalise AI ethics«. URL:

<https://www.ai-ethics-impact.org/resource/blob/1961130/>

c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf [Zugriff:

12.03.2022]

Alexander von Humboldt Institut für Internet und Gesellschaft: » Judith Simon: The ethics of AI and Big Data« [Video]. Youtube. URL: <https://www.youtube.com/watch?v=ZYBtyFnxIAE>

[Stand: 08.09.2021; Zugriff: 07.03.2022]

Bringsjord, Selmer/Govindarajulu, Naveen Sundar: »Artificial Intelligence«. In: Stanford Encyclopedia of Philosophy. URL: <https://plato.stanford.edu/entries/artificial-intelligence/>

[Stand: 12.07.2018; Zugriff: 12.02.2022]

Floridi, Luciano: » The Cambridge Handbook of Information and Computer Ethics«. URL:

<https://books.google.de/books?hl=de&lr=&id=iK3UjU9yfiEC&oi=fnd&pg=PA41&dq=Values+i>

n+Technology+and+Disclosive+Computer+Ethics.&ots=1da7Cy4FFm&sig=_9fWQc_xLln-

q4DIp9vxADT9c20#v=onepage&q=Values%20in%20Technology%20and%20Disclosive%20C

omputer%20Ethics.&f=false [Stand: 2010; Zugriff: 05.03.2022]

Google AI: » Artificial Intelligence at Google: Our Principles «. URL: <https://ai.google/principles/>

[Zugriff: 07.03.2022]

Google Cloud: » Einführung in Vertex Explainable AI«. URL: [https://cloud.google.com/vertex-](https://cloud.google.com/vertex-ai/docs/explainable-ai/overview)

ai/docs/explainable-ai/overview [Stand: 23.02.2022; Zugriff: 12.03.2022]

Lackes, Richard/Siepermann, Markus: »Künstliche Intelligenz (KI)«. In: Gabler

Wirtschaftslexikon. URL: [https://wirtschaftslexikon.gabler.de/definition/kuenstliche-](https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285/version-263673)

intelligenz-ki-40285/version-263673 [Stand: 19.02.2018; Zugriff: 07.03.2022]

Obermeyer, Ziad/Powers, Brian/Vogeli, Christine/Mullainathan, Sendhil: » Dissecting racial

bias in an algorithm used to manage the health of populations « URL: [https://www.science.org/](https://www.science.org/doi/full/10.1126/science.aax2342)

doi/full/10.1126/science.aax2342 [Stand: 25.10.2019; Zugriff: 07.03.2022]

Oxford Languages: »Ethik«. URL:

[https://www.google.com/search?q=ethik+definition&client=firefox-b-](https://www.google.com/search?q=ethik+definition&client=firefox-b-d&ei=oTAmYs_WC8TTkgWCpabQDQ&ved=oahUKEwiPq57cs7T2AhXEqaQKHYSKSCdoQ4dUD)

d&ei=oTAmYs_WC8TTkgWCpabQDQ&ved=oahUKEwiPq57cs7T2AhXEqaQKHYSKSCdoQ4dUD

CAO&uact=5&oq=ethik+definition&gs_lcp=Cgdnd3Mtd2l6EAMyCAgAEIAEELEDmgUIABCAB

DIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQ

gAEIAEOgclABBHELADogclABCwAxBDogolABDkAhCwAxgAOhUilHdHARDRAxDUAhDIAxCwAxBDGAE6EgguEMcBENEDEMgDELADEEMYAToPCC4Q1AIQyAMQsAMQQxgBOgQIABBDOgoIlHdHARDRAxBDOhEILhCABBCxAXCDARDHARDRAzoLCAAQgAQQsQMgE6DgguEIAEELEDEMcbENEDogQILhBDOhAILhCxAXDHARDRAxDUAhBDogclABCxAXBDogolHcXAXDUAhBDogolABCxAXCDARBDogUilHcABDoNCC4QgAQQxwEQrweEQckoECEEYAEoECEYYAVCBBVjCGWCKHGgBcAF4AIABuAGIAYIQkgEEMi4xN]gBAKABAcgBEsABAdoBBggAEAEYCdoBBggBEAEYCA&scient=gws-wiz [Zugriff: 10.03.2022]

Peitz, Dirk: »"Google wird einfach ersetzt" «. URL: <https://www.zeit.de/digital/internet/2018-06/maven-militaerprojekt-google-ausstieg-ruistungsexperte-paul-scharre> [Stand: 05.06.2018; Zugriff: 07.03.2022]

TED: » How To Solve AI's Ethical Puzzles «. [Video]. URL: https://www.ted.com/talks/cansu_canca_how_to_solve_ai_s_ethical_puzzles [Stand: 11.2019; Zugriff: 07.03.2022]

VICE News: » AI Ethics Researcher Timnit Gebru's Firing Doesn't Look Good For Google«. [Video]. Youtube. URL: <https://www.youtube.com/watch?v=fWqpPsMXVsl&t=570s> [Stand: 28.12.2020; Zugriff: 12.03.2022]