

Hausarbeit im Rahmen des Studiums  
Data Science - Master of Science (M.Sc.)

# **Ethik und Recht in Data Science**

Künstliche Intelligenz (KI) braucht Ethik – Untersuchung der Rolle  
von Standards



*Alwine Schultze*

2. Mai 2023

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>I</b>
<b>Abbildungsverzeichnis</b>	<b>II</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Kontext und Motivation . . . . .	1
1.2 Zielsetzung und Fragestellung . . . . .	1
<b>2 Grundlagen der Künstlichen Intelligenz</b>	<b>2</b>
2.1 Definition und Geschichte . . . . .	2
2.2 Algorithmen und maschinelles Lernen . . . . .	3
2.3 Einsatzfelder und Anwendungsbeispiele . . . . .	4
<b>3 Diskriminierung und Verzerrungen in KI-Systemen</b>	<b>6</b>
3.1 Ursachen und Beispiele für Diskriminierung durch KI . . . . .	6
3.2 Methoden zur Verringerung von Verzerrungen . . . . .	7
<b>4 Ethikstandards für KI-Systeme</b>	<b>8</b>
4.1 Aktueller Stand der Ethikstandards . . . . .	8
4.2 Wichtigkeit und Funktion von Standards . . . . .	10
4.3 Herausforderungen bei der Umsetzung von Ethikstandards . . . . .	10
<b>5 Reflexion und Ausblick: Herausforderungen und Chancen</b>	<b>12</b>
5.1 Kritische Auseinandersetzung und Fortschritte . . . . .	12
5.2 Zukunftsperspektiven und die Rolle von Regulierungen . . . . .	12
<b>Literaturverzeichnis</b>	<b>IV</b>
<b>Online-Quellen</b>	<b>IV</b>
<b>Anhang</b>	<b>VI</b>

# Abbildungsverzeichnis

1	Wichtige Meilensteine der KI-Forschung (eigene Darstellung) . . . . .	2
2	Sieben Kernanforderungen an KI-Systeme (vgl. HEG-KI 2018, S. 18) . . . . .	8

# **1 Einleitung**

## **1.1 Kontext und Motivation**

In den letzten Jahren hat sich die künstliche Intelligenz (KI) rasant entwickelt und findet Anwendung in vielen Bereichen unseres täglichen Lebens, wie zum Beispiel im Autofahren, in intelligenten Funktionen und in der Kommunikation (vgl. Lichtenthaler 2021, S. 2). Die Integration von KI-Systemen in den Alltag hat das Potenzial, unsere Lebensqualität zu erhöhen, Arbeitsprozesse zu optimieren und neue wirtschaftliche Chancen zu eröffnen. Die im Jahr 2020 veröffentlichte Studie „Opportunities of Artificial Intelligence“ des Europäischen Parlaments zeigt, dass das Potenzial für eine Steigerung der Arbeitsproduktivität bis zum Jahr 2035 durch den Einsatz von künstlicher Intelligenz zwischen geschätzten 11 und 37 Prozent liegt. Darüber hinaus könnte KI dazu beitragen, die weltweiten Treibhausgasemissionen bis 2030 um geschätzte 1,5 bis 4 Prozent zu reduzieren (vgl. EAGER u. a. 2020).

Gleichzeitig ergeben sich jedoch auch ethische Fragestellungen und Herausforderungen, die sich durch den Einsatz dieser Technologie ergeben. Stephen Hawking betonte einst, wie wichtig es ist, die Auswirkungen der künstlichen Intelligenz, sowohl positiv als auch negativ, im Auge zu behalten und dabei besondere Aufmerksamkeit auf die ethischen Implikationen zu legen, indem er sagte: „KI ist wahrscheinlich das Beste oder das Schlimmste, was der Menschheit passieren kann“. Um die negativen Auswirkungen zu minimieren und einen verantwortungsbewussten Einsatz von KI zu gewährleisten, sind ethische Standards und Richtlinien von entscheidender Bedeutung. In diesem Zusammenhang spielt auch die geplante EU-Regulierung für Künstliche Intelligenz, der sogenannte „AI Act“, eine wichtige Rolle. Dieser soll einen rechtlichen Rahmen schaffen, der sowohl den Schutz der Grundrechte und die Sicherheit der Bürgerinnen und Bürger gewährleistet als auch Innovation und wirtschaftliches Wachstum im Bereich der KI fördert (siehe Algorithm Watch 2022).

## **1.2 Zielsetzung und Fragestellung**

Diese Arbeit bietet einen umfassenden Überblick über die ethischen Fragestellungen im Bereich der künstlichen Intelligenz und die Bedeutung von Standards in diesem Kontext. Sie behandelt die Grundlagen der Künstlichen Intelligenz sowie Diskriminierung und Verzerrungen im Zusammenhang mit KI-Anwendungen und präsentiert Ansätze zur Verringerung dieser. Darüber hinaus wird der aktuelle Stand der Ethikstandards betrachtet und deren Bedeutung und Rolle im Kontext von KI-Systemen erörtert. Es werden auch Herausforderungen bei der Umsetzung von ethischen Standards betrachtet. Abschließend erfolgt eine kritische Analyse der gegenwärtigen Situation im Bereich Künstliche Intelligenz und Ethik sowie eine Diskussion möglicher Entwicklungen und zukünftiger Perspektiven. Ein Fazit fasst die wichtigsten Erkenntnisse der Arbeit zusammen.

## 2 Grundlagen der Künstlichen Intelligenz

### 2.1 Definition und Geschichte

Der Begriff künstliche Intelligenz (KI) ist in der Literatur vielfältig interpretiert, da allein die Begrifflichkeit „Intelligenz“ schwer zu definieren ist. Laut Buxmann und Schmidt (2021) besteht jedoch Einigkeit darüber, dass KI ein bedeutendes Teilgebiet der Informatik darstellt, das sich auf die Erforschung und Entwicklung sogenannter „intelligenter Agenten“ konzentriert. Diese Agenten zeichnen sich dadurch aus, dass sie in der Lage sind, selbstständig Probleme zu lösen. In der KI-Forschung wird grundsätzlich zwischen starker und schwacher KI unterschieden (Buxmann und Schmidt 2021, S. 6).

Laut Buxmann und Schmidt (2021) konzentriert sich die schwache KI auf die Entwicklung gezielter Algorithmen, die speziell für bestimmte und abgegrenzte Problemstellungen konzipiert sind. Lernfähigkeit stellt dabei eine zentrale Anforderung an die schwache KI dar (vgl. Buxmann und Schmidt 2021, S. 7). Die starke Künstliche Intelligenz bezieht sich auf eine Art von KI, die menschenähnliche kognitive Fähigkeiten besitzt oder sie sogar übertrifft (vgl. VDI 2018, S. 8). Menschliche Eigenschaften wie Bewusstsein und Empathie werden oft als wesentliche Merkmale einer starken KI angesehen, bis heute gibt es keine vollständig entwickelte KI, die alle Aspekte der menschlichen Intelligenz und Kognition nachbildet oder übertrifft (vgl. Buxmann und Schmidt 2021, S. 7). Die meisten der heutigen KI-Systeme sind Beispiele für schwache KI, starke KI bleibt weiterhin ein langfristiges Ziel in der KI-Forschung. Trotz einiger Fortschritte ist man noch weit davon entfernt, eine KI zu entwickeln, die umfassend menschenähnliche Intelligenz und Selbstbewusstsein aufweist.

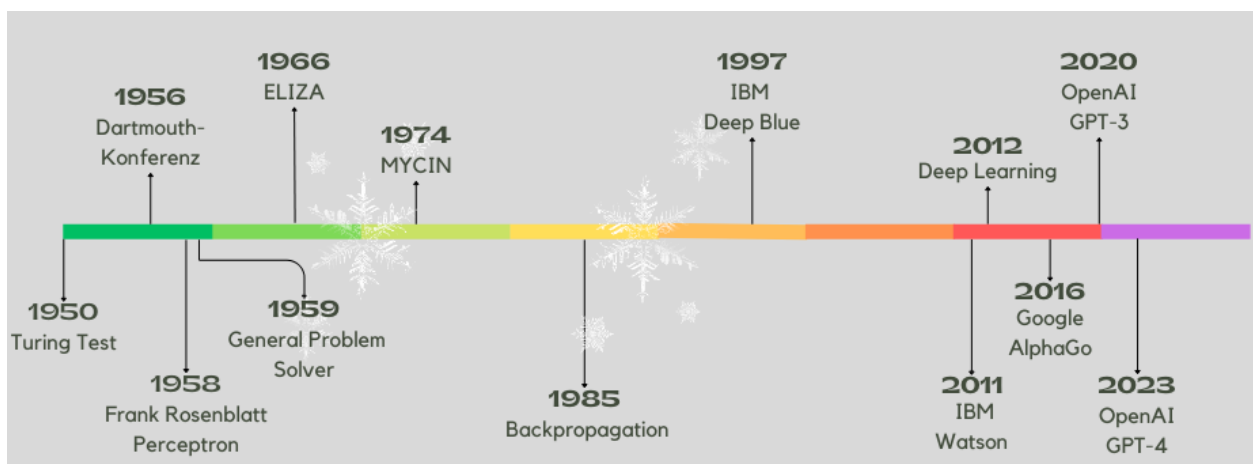


Abbildung 1: Wichtige Meilensteine der KI-Forschung (eigene Darstellung)

Um die Entwicklung der KI-Forschung besser zu verstehen, kann man die historischen Meilensteine betrachten: Abbildung 1 zeigt die wichtigsten Meilensteine der KI-Forschung von 1950 bis 2023. Als Geburtsstunde der KI gilt heute eine sechswöchige Konferenz, die 1956 am Darmouth

College stattfand. Von der Entstehung des Begriffs KI bis zum heutigen Tag sind zwei sogenannte KI-Winter verzeichnet worden, diese sind ebenfalls in der Abbildung erkenntlich. Der Ausdruck „KI-Winter“ bezieht sich auf Phasen, in denen die Investitionen in die Forschung erheblich zurückgingen. Gründe dafür könnten beispielsweise eine unzureichende Rechenleistung oder nicht erfüllte Erwartungen gewesen sein (siehe Wikipedia 2023 und Buxmann und Schmidt 2021, S. 4 ff.). Der Anhang 5.2 enthält weitere nennenswerte Errungenschaften der KI-Forschung. Im nächsten Abschnitt wird erläutert, welche zentralen Bestandteile in diesen Technologien enthalten sind, um ein tieferes Verständnis dafür zu vermitteln, wie Lernprozesse und Entscheidungsfindung in KI-Systemen funktionieren.

## 2.2 Algorithmen und maschinelles Lernen

Es ist von großer Bedeutung, die fundamentalen Konzepte der KI zu verstehen, um die damit verbundenen ethischen Fragen zu erfassen. Algorithmen und maschinelles Lernen bilden die Grundlagen der KI und sind daher von entscheidender Relevanz. Algorithmen sind eine Reihe von Anweisungen, die in einer bestimmten Reihenfolge ausgeführt werden, um ein bestimmtes Problem zu lösen. Sie sind das Herzstück von Computerprogrammen und ermöglichen es Maschinen, komplexe Berechnungen durchzuführen und Entscheidungen zu treffen. Ein einfaches Beispiel ist ein Algorithmus, der Zahlen der Größe nach sortiert. Maschinelles Lernen (ML), ein Teilbereich der künstlichen Intelligenz, verwendet Algorithmen wiederum, um Muster in Daten zu erkennen. Diese Muster helfen Vorhersagen oder gar automatisierte Entscheidungen zu treffen, ohne explizit dafür programmiert worden zu sein. ML-Algorithmen ermöglichen es Computern, aus Daten zu lernen und sich im Laufe der Zeit zu verbessern. Beim maschinellen Lernen wird zwischen drei wesentlichen Lernmethoden unterschieden:

**Überwachtes Lernen** Beim überwachten Lernen werden Algorithmen mit gelabelten Trainingsdaten gefüttert, bei denen die Eingabe- und Ausgabewerte bekannt sind. Der Algorithmus lernt, eine Beziehung zwischen den Eingabe- und Ausgabedaten herzustellen, um Vorhersagen für neue, ungesehene Daten zu treffen (vgl. Sheikh, Prins und Schrijvers 2023, S. 37).

**Unüberwachtes Lernen** Das unüberwachte Lernen bezieht sich auf Situationen, in denen Algorithmen nur Eingabedaten ohne zugehörige Ausgabewerte erhalten. Der Algorithmus muss eigenständig Muster und Strukturen in den Daten erkennen, ohne auf vorgegebene Labels zurückzugreifen (vgl. Sheikh, Prins und Schrijvers 2023, S. 37).

**Bestärkendes Lernen** Beim bestärkenden Lernen erlernt ein System, wie etwa ein Roboter, eigenständig Aufgaben, indem es durch fortlaufende Rückmeldungen seine Strategie optimiert, wie beispielsweise DeepMind mit AlphaGo Zero beim Go-Spiel demonstrierte (vgl. VDI 2018, S. 18).

Nach Sheikh, Prins und Schrijvers (2023) ist Deep Learning ein spezieller Bereich des maschinellen Lernens. Dabei werden künstliche neuronale Netze verwendet, die aus mehreren Schichten bestehen und Informationen sowohl senden als auch empfangen können. Diese Schichten verarbeiten Informationen hierarchisch und ermöglichen es der Maschine, komplexe und abstrakte Muster zu erkennen. Neuronale Netze sind von der Funktionsweise des menschlichen Gehirns inspiriert und bestehen aus Knoten, die Neuronen repräsentieren, und Verbindungen, die Synapsen darstellen (vgl. Sheikh, Prins und Schrijvers 2023, S. 31). Alle der oben genannten Lernmethoden können mit dem Deep Learning-Ansatz kombiniert werden (vgl. VDI 2018, S. 19).

## 2.3 Einsatzfelder und Anwendungsbeispiele

Die Künstliche Intelligenz (KI) hat in den letzten Jahren in zahlreichen unserer Lebensbereiche Einzug gehalten. Im Folgenden sind einige dieser Anwendungen exemplarisch dargestellt.

**Bild- und Spracherkennung** Ein bedeutendes Anwendungsgebiet der künstlichen Intelligenz ist die Bild- und Spracherkennung. Beispielsweise kann Google Lens Objekte, Text und Szenen in Bildern identifizieren und dem Nutzer relevante Informationen liefern (vgl. Google 2023). Ein weiteres Beispiel ist die Spracherkennungstechnologie, wie sie in Sprachassistenten wie Amazon Alexa oder Apple Siri verwendet wird. Diese Systeme verwenden KI, um gesprochene Sprache in Text umzuwandeln und darauf basierend Aktionen auszuführen oder Fragen zu beantworten.

**Gesundheitswesen** KI spielt auch eine entscheidende Rolle im Gesundheitswesen. Sie kann zur Verbesserung der Diagnose und Behandlung von Krankheiten eingesetzt werden, indem sie zum Beispiel Muster in medizinischen Bildern erkennt, wie es bei der Früherkennung von Krebs durch Algorithmen wie Google DeepMind's LYNA der Fall ist (vgl. Weekly 2023).

**Logistik und Verkehr** Im Bereich der Logistik und Verkehrssysteme ermöglicht KI die Optimierung von Transport- und Lieferketten. Beispielsweise verwendet das Unternehmen Tesla KI-Systeme in seinen autonomen Fahrzeugen, um sicherzustellen, dass sie effizient und sicher fahren können (vgl. Tesla 2023).

**Finanzen** Im Finanzsektor wird KI zur Erkennung von Betrug und zur Risikobewertung eingesetzt. Durch die Analyse von Transaktionsmustern kann KI potenzielle Betrugsfälle erkennen und rechtzeitig Alarm schlagen. Darüber hinaus nutzen Unternehmen wie Moody's Analytics KI-Modelle, um das Kreditrisiko von Kreditnehmern präziser zu bewerten und den Finanzinstituten bei der Entscheidungsfindung zu helfen (vgl. Analytics 2023).

**Kreativität** Die Entwicklung von KI hat auch den Bereich der kreativen Künste beeinflusst. So können KI-gestützte Systeme wie OpenAI's GPT-Modelle Texte verfassen, die menschenähnliche Schreibstile und Ideen aufweisen. In der Musikindustrie gibt es ebenfalls zahlreiche KI-Tools wie AIVA oder Amper Music, die eigene Musikstücke komponieren und so Künstler bei der Entstehung neuer Werke unterstützen können (vgl. TopTen.AI 2023).

Insgesamt hat die künstliche Intelligenz zahlreiche positive Anwendungsmöglichkeiten in verschiedenen Lebensbereichen und Branchen ermöglicht, die unser Leben und unsere Arbeitsweise erheblich verändert und verbessert haben. Allerdings hat die zunehmende Integration von KI-Systemen in unsere Gesellschaft auch Herausforderungen und Bedenken hervorgerufen, insbesondere in Bezug auf Diskriminierung und Verzerrungen. Im nächsten Kapitel setzt sich die vorliegende Arbeit mit den potenziellen Problemen auseinander, die durch Diskriminierung und Verzerrungen in KI-Systemen entstehen können, und mögliche Lösungsansätze diskutieren, um diese Herausforderungen zu bewältigen.



## 3 Diskriminierung und Verzerrungen in KI-Systemen

### 3.1 Ursachen und Beispiele für Diskriminierung durch KI

Die Ursachen von Diskriminierung und Verzerrungen in KI-Systemen können vielfältig sein. In Büchern wie „Ein Algorithmus hat kein Taktgefühl“ und „Hello World: Was Algorithmen können und wie sie unser Leben verändern“ werden zahlreiche Beispiele erwähnt, wie Algorithmen unseren Alltag bestimmen und warum sie manchmal nicht die gewünschten Ergebnisse liefern. Drei Hauptfaktoren für Diskriminierungen durch KI-Systeme werden sowohl hier als auch in einschlägiger Fachliteratur genannt: Voreingenommenheit in den Trainingsdaten, algorithmische Verzerrungen und mangelnde Diversität.

KI-Systeme lernen aus den Daten, die ihnen während des Trainings zur Verfügung gestellt werden. Wenn vorhandene Daten bereits Verzerrungen aufweisen, besteht häufig das Risiko, dass diese fest verankerte gesellschaftliche Voreingenommenheit auf die Software übergeht (vgl. Plattform Lernende Systeme 2019, S. 8). Beispielsweise könnte ein KI-System, das für die Personalbeschaffung entwickelt wurde, unbeabsichtigt diskriminierend sein, wenn es auf historischen Daten trainiert wurde, die eine Tendenz zur Bevorzugung von Bewerbern einer bestimmten Geschlechts- oder Ethniegruppe aufweisen. In solchen Fällen wird die KI dazu neigen, Entscheidungen zu treffen, die diese Voreingenommenheiten weiter verstärken. Das hätte eine systematische Diskriminierung von Bewerbern mit bestimmten Eigenschaften zur Folge (vgl. Fall 1 aus dem Whitepaper Plattform Lernende Systeme 2019, S. 3).

Die grundlegenden Algorithmen von Künstlicher Intelligenz können ebenfalls Diskriminierung und Verzerrungen hervorbringen, diese werden als Algorithmische oder Technische Verzerrungen bezeichnet. Ein Beispiel hierfür ist der sogenannte „Black-Box-Effekt“, bei dem die Entscheidungsprozesse eines KI-Systems undurchschaubar sind und nicht einfach verstanden werden können (vgl. Praevention-Aktuell.de 2023). Dies kann dazu führen, dass solche Systeme Entscheidungen basierend auf verborgenen Verzerrungen treffen, ohne dass Entwickler oder Nutzer dies bemerken. Ein bekanntes Beispiel dafür ist die Unterscheidung zwischen Hunden und Wölfen. In den Trainingsdaten waren Wölfe stets in einer schneebedeckten Umgebung abgebildet. Der Algorithmus erlernte daraufhin, Hunde und Wölfe anhand der Anwesenheit von Schnee zu unterscheiden. Dies führte dazu, dass ein Hund auf einem Bild mit Schnee fälschlicherweise als Wolf klassifiziert wurde (vgl. Praevention-Aktuell.de 2023).

Schließlich kann die mangelnde Diversität unter den KI-Entwicklern zu einer unzureichenden Berücksichtigung verschiedener Perspektiven und Bedürfnisse führen, was wiederum Diskriminierung und Verzerrungen in KI-Systemen begünstigt. Wenn Entwickler, die KI-Systeme entwerfen, eine homogene Gruppe bilden, können sie unbeabsichtigt ihre eigenen kulturellen, sozialen oder persönlichen Voreingenommenheiten in die KI-Systeme einbringen. Auch hier fordert UNESCO (2023) mehr Diversität in KI-Unternehmen. Ein Beispiel hierfür ist ein automatischer Seifenspen-

der, der lediglich mit weißen Testpersonen getestet wurde. Nach der Produktion stellte sich heraus, dass dieser keine Seife für nicht-weiße Personen ausgab (vgl. Plattform Lernende Systeme 2019, S. 6).

Um diese Herausforderungen zu bewältigen, ist es wichtig, Mechanismen zur Identifizierung und Behebung von Diskriminierung und Verzerrungen in KI-Systemen zu entwickeln.

### **3.2 Methoden zur Verringerung von Verzerrungen**

Angesichts der verschiedenen Ursachen und Beispiele für Diskriminierung, die durch den Einsatz von KI-Verfahren entstehen, lassen sich mehrere Ansätze ableiten, um diese Problematik zu reduzieren. Einerseits kann durch eine gründliche Untersuchung und Aufbereitung der Trainingsdaten das Auftreten von Verzerrungen verringert werden. Dies lässt sich beispielsweise durch das Erheben repräsentativer Daten, das Entfernen diskriminierender Attribute oder das Anwenden von Techniken zur Datenausgewogenheit erreichen. Eine ausgewogene Datenerfassung stellt sicher, dass alle relevanten Gruppen und Situationen angemessen berücksichtigt werden, wodurch das KI-System in der Lage ist, gerechte Entscheidungen zu treffen.

Darüber hinaus kann die Gestaltung von KI-Systemen, die nachvollziehbar und transparent sind, ebenfalls dazu beitragen, Verzerrungen zu erkennen und zu beheben. Wenn die Entscheidungsprozesse eines KI-Systems verständlich sind, können Entwickler und Nutzer mögliche Diskriminierungen und Verzerrungen besser identifizieren und entsprechende Maßnahmen ergreifen. Ansätze wie erklärbare KI und die Offenlegung von Bewertungsmetriken für Fairness können dazu beitragen, das Verständnis und die Kontrolle über die Entscheidungsfindung der KI zu verbessern.

Des Weiteren ist die Förderung von Vielfalt und Inklusion in den Entwicklerteams von Softwarelösungen, die Methoden der künstlichen Intelligenz einsetzen, von entscheidender Bedeutung. Dies trägt dazu bei, dass unterschiedliche Perspektiven und Bedürfnisse berücksichtigt werden. Diversität hilft dabei, potenzielle Voreingenommenheiten und Diskriminierungen frühzeitig zu identifizieren und geeignete Lösungen zu entwickeln, um Fairness und Gerechtigkeit in KI-Systemen sicherzustellen.

Ebenso kann die Einführung von Regulierungen und Leitlinien auf nationaler und internationaler Ebene dazu beitragen, Diskriminierung und Verzerrungen in KI-Anwendungen zu verhindern. Solche Leitlinien könnten Standards für Fairness, Transparenz und Verantwortung definieren, um sicherzustellen, dass KI-Systeme im Einklang mit ethischen und gesellschaftlichen Werten stehen.

Zusammenfassend erfordert die Erkennung und Behebung von Diskriminierung und Verzerrungen in KI-Systemen eine komplexe und facettenreiche Herangehensweise, die sowohl technische als auch soziale Aspekte einbezieht. Dabei können ethische Standards für KI-Systeme eine besondere Rolle spielen.

## 4 Ethikstandards für KI-Systeme

### 4.1 Aktueller Stand der Ethikstandards

Verschiedene politische und akademische Institutionen sowie Organisationen und Unternehmen aus der Privatwirtschaft arbeiten bereits an Ethikstandards für KI-Systeme. So hat die EU beispielsweise die „Ethik-Leitlinien für eine vertrauenswürdige KI“ von der Hochrangigen Expertengruppe für Künstliche Intelligenz (HEG-KI) erarbeiten lassen. In den Leitlinien werden ethisches Handeln, Robustheit und Rechtmäßigkeit als die drei essenziellen Komponenten für vertrauenswürdige KI aufgeführt (vgl. Krüger 2021, S. 421). Die HEG-KI hat darin sieben zentrale Anforderungen für KI-Systeme formuliert und miteinander verknüpft, die in Abbildung 2 dargestellt sind. Die HEG-KI fordert zudem eine kontinuierliche Bewertung und Berücksichtigung dieser Anforderungen während des gesamten Lebenszyklus eines KI-Systems (vgl. HEG-KI 2018, S. 18).

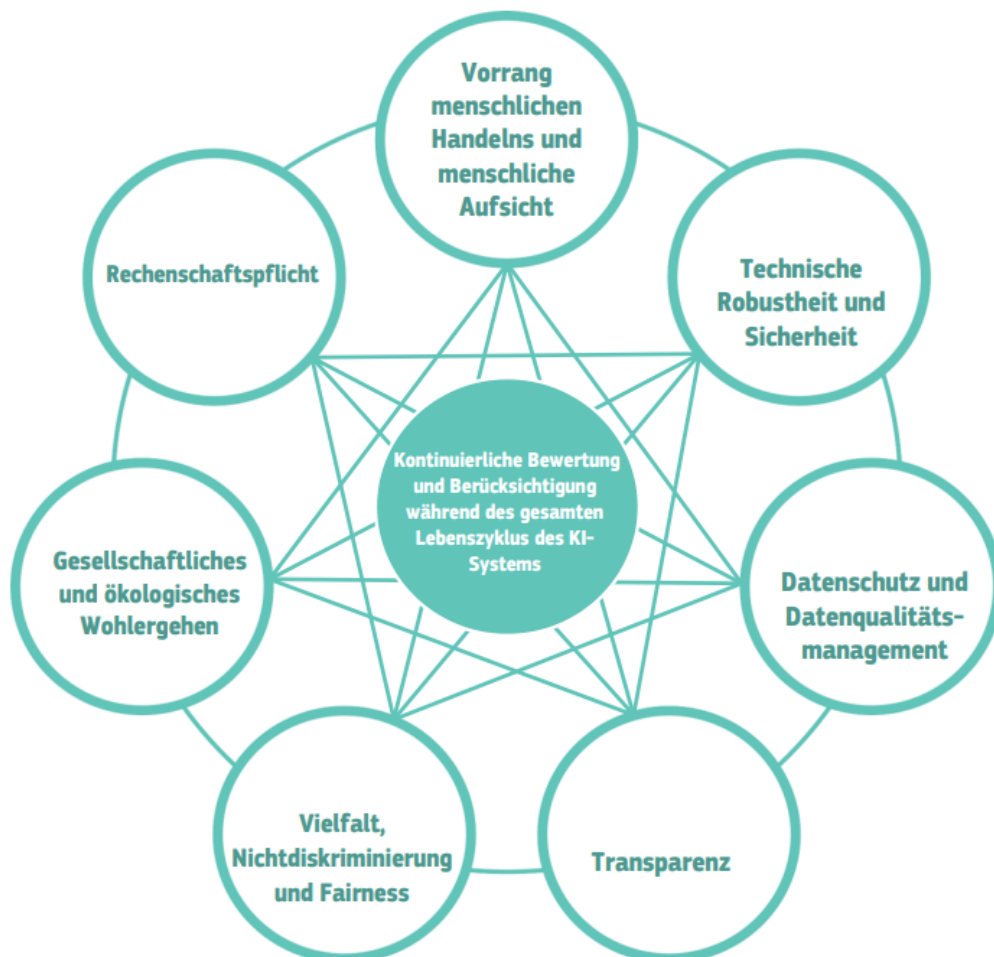


Abbildung 2: Sieben Kernanforderungen an KI-Systeme (vgl. HEG-KI 2018, S. 18)

Um diese Anforderungen auf europäischer Ebene durchzusetzen und zu harmonisieren, wurde ein wichtiger Schritt unternommen: Im April 2021 legte die Europäische Kommission einen abschlie-

ßenden Gesetzesvorschlag vor, der darauf abzielt, einen einheitlichen Rechtsrahmen für Künstliche Intelligenz in der EU zu etablieren. Dies soll Innovationen unterstützen, die Sicherheit von KI-Systemen gewährleisten und die Grundrechte der Bürger schützen. Dieser wird seitdem in den einzelnen Mitgliedstaaten sowie den Europäischen Institutionen kontrovers diskutiert. Der AI Act stellt einen wichtigen Schritt in Richtung einer umfassenden Regulierung von KI-Systemen in der Europäischen Union dar. Im Hinblick auf die Ethik der KI berücksichtigt der Gesetzesentwurf verschiedene Aspekte (vgl. Europäische Kommission 2021).

**KI-Risikoklassifizierung** Der Entwurf klassifiziert KI-Systeme entsprechend ihres Risikogrades für die Gesellschaft und die Grundrechte, um besonderen Fokus auf Anwendungen zu legen, die höhere ethische Bedenken aufwerfen.

**Regelungen für Hochrisiko-KI-Systeme** Der Entwurf stellt hohe Anforderungen an KI-Systeme mit hohem Risiko, wie Transparenz, Dokumentation, Qualitätssicherung, Sicherheit und menschliche Aufsicht, um ethische Bedenken und potenzielle negative Auswirkungen zu minimieren.

**Transparenz** Der Entwurf betont die Bedeutung der Transparenz von KI-Systemen, damit Nutzer wissen, dass sie mit KI interagieren und ihre Rechte sowie Möglichkeiten zur Anfechtung von KI-Entscheidungen kennen.

**Verbot bestimmter KI-Praktiken** Der Vorschlag verbietet bestimmte KI-Anwendungen, die eine klare Bedrohung für die Sicherheit, Grundrechte oder Menschenwürde darstellen, wie beispielsweise Echtzeit-Gesichtserkennung in öffentlichen Räumen.

**Regulierungsbehörden** Der Entwurf schlägt die Schaffung nationaler Aufsichtsbehörden und eines europäischen KI-Ausschusses vor, um die Einhaltung des Rechtsrahmens zu überwachen und durchzusetzen.

Nicht nur öffentliche Organisationen haben Ethikstandards definiert, sondern auch Technologieunternehmen wie Google und Forschungsinstitutionen wie OpenAI haben eigene Prinzipien und Leitlinien entwickelt, um ethisches Handeln, Transparenz und Verantwortung bei der Entwicklung und Implementierung von KI-Systemen sicherzustellen. Angesichts der wachsenden Bedeutung von KI-Systemen in unserer Gesellschaft ist es unerlässlich, die Relevanz und Funktion von Standards im Bereich der künstlichen Intelligenz genauer zu betrachten, um deren nachhaltige und ethische Integration in unterschiedlichen Anwendungsfeldern zu gewährleisten.

## 4.2 Wichtigkeit und Funktion von Standards

Ethikstandards für KI-Systeme sind von entscheidender Bedeutung, weil sie gewährleisten, dass die Entwicklung und Anwendung von KI-Technologien den menschlichen Werten, Grundrechten und gesellschaftlichen Normen gerecht wird. Sie fungieren als Leitfaden für verantwortungsvolles Handeln und bieten Orientierung für Entwickler, Unternehmen, politische Entscheidungsträger und weitere Interessengruppen.

Darüber hinaus stellen Ethikstandards Chancengleichheit für Unternehmen sicher, die KI-Systeme entwickeln oder einsetzen. Die Einhaltung gemeinsamer ethischer Leitlinien erlaubt es Unternehmen, auf einem fairen und ausgeglichenen Markt zu agieren, der Innovation und wirtschaftlichen Fortschritt gleichermaßen fördert. Indem ein einheitlicher Rahmen etabliert wird, dem alle Unternehmen folgen müssen, wird verhindert, dass einige Unternehmen unfaire Wettbewerbsvorteile erzielen, indem sie Datenschutz oder gerechte Entscheidungsfindung vernachlässigen.

Die Beachtung von Ethikstandards trägt auch wesentlich dazu bei, das Vertrauen von Verbrauchern, Geschäftspartnern und Aufsichtsbehörden in KI-Systeme zu stärken. Unternehmen, die sich an ethische Leitlinien halten, können sich als verantwortungsbewusste Akteure positionieren und somit das Vertrauen ihrer Kunden und Stakeholder gewinnen. Dies ist vergleichbar mit der Anwendung von DIN-Normen, die ebenfalls Vertrauen in Unternehmen und Systeme schaffen.

Des Weiteren helfen Ethikstandards Unternehmen, mögliche Risiken im Zusammenhang mit KI-Systemen frühzeitig zu erkennen und angemessen zu bewältigen. Dies trägt zur Minimierung von Reputations- und Haftungsrisiken bei und stellt somit Chancengleichheit für alle beteiligten Unternehmen sicher.

Ethikstandards sind entscheidend für die Verhinderung von gefährlichen KI-Systemen, da sie Datenschutz, Fairness, Transparenz, Verantwortlichkeit und Sicherheit betonen. Durch die Einhaltung dieser Grundsätze können Entwickler und Betreiber potenzielle Risiken minimieren, Vertrauen aufbauen und sicherstellen, dass KI-Systeme zum Wohl der Menschen eingesetzt werden.

## 4.3 Herausforderungen bei der Umsetzung von Ethikstandards

Die UNESCO hat Empfehlungen zur Ethik von Künstlicher Intelligenz entwickelt und Bedingungen für deren Umsetzung erarbeitet (vgl. Kettemann 2021). Daraus ergeben sich mehrere Herausforderungen für die Implementierung von Ethikstandards in KI-Systemen:

- Entwicklung geeigneter Testverfahren, die den Zielen des KI-Systems gerecht werden und sowohl in Laborumgebungen als auch in realen Situationen anwendbar sind.
- Ständige Überwachung aller Phasen des KI-Lebenszyklus, einschließlich der Funktionsweise von Algorithmen, verwendeter Daten und beteiligter KI-Akteure, um den Schutz der Grundrechte sicherzustellen.

- Errichtung einer angemessenen Kontrollstruktur und Überprüfbarkeit von KI-Anwendungen, um die Auswirkungen von KI-Systemen auf Menschenrechte, Rechtsstaatlichkeit und inklusive Gesellschaften zu identifizieren sowie mögliche negative Folgen zu verhindern oder abzumildern.
- Etablierung adäquater Überwachungsmechanismen, die eine Bewertung von Algorithmen, Daten und Entwicklungsprozessen ermöglichen und eine externe Kontrolle von KI-Systemen zulassen.
- Schaffung eines geeigneten rechtlichen Rahmens für ethische Folgenabschätzungen von KI-Systemen, der transparent und multidisziplinär ist und von verschiedenen Interessengruppen sowie multikulturellen, pluralistischen und integrativen Ansätzen unterstützt wird.
- Gewährleistung des öffentlichen Zugangs zu ethischen Folgenabschätzungen. Behörden sollten verpflichtet werden, die von ihnen eingesetzten und/oder genutzten KI-Systeme durch die Implementierung angemessener Mechanismen und Instrumente zu überwachen.

Unternehmen stehen vor vielfältigen Herausforderungen, wenn sie Ethikstandards für KI-Systeme umsetzen möchten. Die oben genannten Aspekte sind nur einige Beispiele dafür.

## **5 Reflexion und Ausblick: Herausforderungen und Chancen**

### **5.1 Kritische Auseinandersetzung und Fortschritte**

In den letzten Jahren ist die Nutzung von KI-Systemen in Bereichen wie Gesundheit, Verkehr, Bildung, Umwelt und Wirtschaft gestiegen. Dies unterstreicht die Dringlichkeit, ethische Leitlinien und Standards zu etablieren, um Technologien im Einklang mit Grundrechten und demokratischen Werten zum Wohl der Gesellschaft einzusetzen.

Datenschutz, Diskriminierung, Transparenz, Verantwortung und Arbeitsmarktauswirkungen sind zentrale Themen in der ethischen Diskussion um KI. Trotz Fortschritten bei der Umsetzung ethischer Leitlinien bleiben Herausforderungen bestehen, insbesondere in Bezug auf die Intransparenz und Kontrollierbarkeit vieler KI-Systeme.

Ein gemeinsames Verständnis und die Übereinkunft über ethische Grundlagen sind erforderlich, um grenzübergreifende Zusammenarbeit und die Entwicklung von Technologien, die den globalen Interessen gerecht werden, zu ermöglichen. Dabei sollten Datenschutz, Diskriminierungsfreiheit, Transparenz und Verantwortung gewährleistet und die positiven Auswirkungen von KI auf das gesellschaftliche Wohlergehen maximiert werden.

### **5.2 Zukunftsperspektiven und die Rolle von Regulierungen**

Eine fortlaufende Diskussion und Kooperation zwischen verschiedenen Beteiligten, wie Regierungen, Forschungseinrichtungen, Unternehmen und der Zivilgesellschaft, ist entscheidend, um gemeinsame ethische KI-Standards zu entwickeln und umzusetzen. Der AI Act zeigt, dass das Bewusstsein für die Bedeutung einer verantwortungsvollen und fairen Nutzung von KI-Technologien zunimmt, aber weitere Anstrengungen sind notwendig.

Die Etablierung von Ethikstandards und rechtlichen Rahmenbedingungen, wie dem AI Act, kann das Vertrauen der Bevölkerung in KI-Technologien erhöhen und gleichzeitig Diskriminierung, Datenschutzverletzungen und mangelnde Transparenz verhindern. Zudem sind globale Zusammenarbeit sowie Überwachungs- und Zertifizierungssysteme erforderlich, um Anwendungen, die KI-Verfahren nutzen, besser zu überwachen und zu regulieren.

In der Zukunft könnten weltweite Ethikstandards und rechtliche Rahmenbedingungen, wie ein globaler AI Act, Chancengleichheit fördern und zum Wohle der gesamten Menschheit beitragen. Diese Entwicklung könnte helfen, drängende globale Herausforderungen wie Klimawandel, Armut und Gesundheitsversorgung zu bewältigen und die negativen Auswirkungen von KI-Technologien zu reduzieren.

Die kontinuierliche Auseinandersetzung mit ethischen Fragestellungen der Künstlichen Intelligenz ist entscheidend, um eine nachhaltige und ethische Integration von KI in verschiedenen Anwendungsbereichen zu gewährleisten. Die Zukunft der Künstlichen Intelligenz liegt in unseren

Händen – es liegt an uns, die richtigen Entscheidungen zu treffen und die erforderlichen Maßnahmen zu ergreifen, um eine bessere und gerechtere Welt für alle zu schaffen.

Um die ethische Entwicklung und Anwendung von KI-Technologien zu fördern, müssen Bildungsinitiativen und öffentliche Debatten gestärkt werden, um das Bewusstsein für die Chancen und Risiken der Künstlichen Intelligenz zu erhöhen. Dies beinhaltet auch die Einbindung verschiedener gesellschaftlicher Gruppen in die Entscheidungsfindung, um eine vielfältige und umfassende Perspektive auf die ethischen Aspekte von KI zu gewährleisten.

Innovationen in KI sollten darauf abzielen, menschliche Fähigkeiten zu ergänzen und zu erweitern, statt sie zu ersetzen. Dies könnte dazu beitragen, eine ausgewogene Koexistenz zwischen Mensch und Maschine zu erreichen, die das menschliche Potenzial entfaltet und eine nachhaltige und gerechte Zukunft fördert.

Letztendlich ist es wichtig, dass wir uns weiterhin den ethischen Herausforderungen der Künstlichen Intelligenz stellen und gemeinsam eine Vision für eine technologische Zukunft entwickeln, die den Menschen in den Mittelpunkt stellt und die Lebensqualität für alle verbessert.



## Literaturverzeichnis

- Buxmann, Peter und Holger Schmidt (2021). *Künstliche Intelligenz - Mit Algorithmen zum wirtschaftlichen Erfolg*. 2., aktualisierte und erweiterte Auflage. Springer Gabler.
- Krüger, Sven (2021). *Die KI-Entscheidung - Künstliche Intelligenz und was wir daraus machen*. Springer.
- Lichtenthaler, Ulrich (2021). *Künstliche Intelligenz erfolgreich umsetzen - Praxisbeispiele für integrierte Intelligenz*. Springer Gabler.
- Sheikh, Haroon, Corien Prins und Erik Schrijvers (2023). *Mission AI - The New System Technology*. Springer.

## Online-Quellen

- Algorithm Watch (2022). *AI Act erklärt*. de. Zugriff am 16. März 2023. URL: <https://algorithmwatch.org/de/ai-act-erklart/>.
- Analytics, Moody's (2023). *Credit Risk Solutions*. Zugriff am 10. April 2023. URL: <https://www.moodyanalytics.com/solutions-overview/credit-risk>.
- EAGER, James u. a. (2020). *Opportunities of Artificial Intelligence*. en. Institution: European Parliament. Zugriff am 17. März 2023. URL: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL\\_STU\(2020\)652713\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf).
- Europäische Kommission (2021). *Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über einen Rechtsrahmen für Künstliche Intelligenz (Künstliche-Intelligenz-Gesetz)*. de. Zugriff am 27. April 2023. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206>.
- Google (2023). *Google Lens*. Zugriff am 10. April 2023. URL: <https://lens.google/>.
- HEG-KI (2018). *Ethik-Leitlinien für eine vertrauenswürdige KI*. de. Zugriff am 27. April 2023. URL: <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-de>.
- Kettemann, Prof. Dr. Matthias C. (2021). *UNESCO-Empfehlung zur Ethik Künstlicher Intelligenz. Bedingungen zur Implementierung in Deutschland*. de. Zugriff am 27. April 2023. URL: [https://www.unesco.de/sites/default/files/2022-03/DUK\\_Broschuere\\_KI-Empfehlung\\_DS\\_web\\_final.pdf](https://www.unesco.de/sites/default/files/2022-03/DUK_Broschuere_KI-Empfehlung_DS_web_final.pdf).
- Plattform Lernende Systeme, Arbeitsgruppe 3 der (2019). *Weiterbildung im Zeitalter Künstlicher Intelligenz: Herausforderungen, Strategien und Handlungsempfehlungen*. URL: [https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3\\_Whitepaper\\_250619.pdf](https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_250619.pdf).

- Praevention-Aktuell.de (2023). *Frage an KI: Hund oder Wolf?* Zugriff am 10. April 2023. URL: <https://praevention-aktuell.de/frage-an-ki-hund-oder-wolf/>.
- Tesla, Inc. (2023). *Tesla AI*. Zugriff am 10. April 2023. URL: <https://www.tesla.com/AI>.
- TopTen.AI (2023). *Top 10 AI Music Generators Review*. Zugriff am 10. April 2023. URL: <https://topten.ai/music-generators-review/>.
- UNESCO (2023). *Künstliche Intelligenz für mehr Geschlechtergerechtigkeit und gegen Diskriminierung - Handlungsansätze für die Politik*. Zugriff am 10. April 2023. URL: [https://www.unesco.de/sites/default/files/2023-01/Handlungsans%C3%A4tze\\_%20KI%20und%20Geschlechtergerechtigkeit\\_DUK.pdf](https://www.unesco.de/sites/default/files/2023-01/Handlungsans%C3%A4tze_%20KI%20und%20Geschlechtergerechtigkeit_DUK.pdf).
- VDI (2018). *VDI-Statusreport Künstliche Intelligenz*. Institution: Verein Deutscher Ingenieure (VDI). Zugriff am 18. März 2023. URL: <https://www.vdi.de/ueber-uns/presse/publikationen/details/vdi-statusreport-kuenstliche-intelligenz>.
- Weekly, Healthcare (2023). *Google AI's LYNA: Better Than Humans in Detecting Advanced Breast Cancer*. Zugriff am 10. April 2023. URL: <https://healthcareweekly.com/google-ais-lyna-better-than-humans-in-detecting-advanced-breast-cancer/>.
- Wikipedia (2023). *KI-Winter*. Zugriff am 18. März 2023]. URL: <https://de.wikipedia.org/wiki/KI-Winter>.

## Anhang

- **1950 - Turing Test**, ein Frage-Antwort-Spiel, das von einem Computerprogramm als Bestanden gilt, sobald dieser Test keinen Unterschied zu menschlicher feststellen kann. (vgl. Sheikh, Prins und Schrijvers 2023, S. 26)
- **1958 - Perceptron** entwickelt von Frank Rosenblatt ist der Klassifikationsalgorithmus, der selbständig lernte Buchstaben zu erkennen, ohne dass diese vorprogrammiert waren (vgl. Sheikh, Prins und Schrijvers 2023, S. 32).
- **1959 - General Problem Solver** war eines der ersten Programme, das in der Lage war, Probleme durch symbolisches Reasoning zu lösen (vgl. Sheikh, Prins und Schrijvers 2023, S. 29).
- **1966 - ELIZA** ein von Joseph Weizenbaum am MIT entwickeltes Programm, das als Vorläufer des heutigen Chatbots interpretiert werden kann. Hier wurde die natürliche Sprache für die Kommunikation zwischen Mensch und Maschine das erste mal erfolgreich angewendet (vgl. Buxmann und Schmidt 2021, S. 5).
- **1974 - MYCIN** war ein Expertensystem zur Diagnose und Behandlung von Infektionskrankheiten und gilt als Meilenstein für die medizinische KI (vgl. Buxmann und Schmidt 2021, S. 5).
- **1985 - Backpropagation-Algorithmus**, ein Algorithmus, der das effektive Training von neuronalen Netzen ermöglichte, indem mehrere Layer eines Neuronales Netzes trainiert wurden (vgl. Sheikh, Prins und Schrijvers 2023, S. 35).
- **1997 - Deep Blue**, ein von IBM entwickelter Schachcomputer, besiegte den damaligen Schachweltmeister Garry Kasparov (vgl. Buxmann und Schmidt 2021, S. 6).
- **2011 - Watson**, ein weiteres von IBM entwickeltes kognitives Computersystem, dem es gelang gegen zwei ehemalige Champions in der US-Spielshow Jeopardy! zu gewinnen (vgl. Sheikh, Prins und Schrijvers 2023, S. 139).
- **2012 - Deep Learning**, ein Zweig des maschinellen Lernens, revolutionierte die KI-Forschung und ermöglichte bahnbrechende Fortschritte (vgl. Sheikh, Prins und Schrijvers 2023, S. 38).
- **2016 - AlphaGo**, ein von Google entwickeltes Programm, besiegte den Go-Weltmeister Lee Sedol und gilt als bedeutender Fortschritt in der KI-Forschung (vgl. Sheikh, Prins und Schrijvers 2023, S. 36).

- **2020 - GPT-3 (Generative Pre-trained Transformer)** ist ein leistungsstarkes neuronales Netzwerk für natürliche Sprachverarbeitung, das auf großen Textdatenmengen trainiert wurde und in der Lage ist, automatisch Texte zu generieren, die von menschlichen Texten kaum zu unterscheiden sind (vgl. Sheikh, Prins und Schrijvers 2023, S. 140).
- **2023 - GPT-4 (Generative Pre-trained Transformer)** ist eine Weiterentwicklung von GPT-3, mit mehr Daten und einer größeren Anzahl an Parametern trainiert. Es soll neben Text auch Input im Bildformat akzeptieren und verarbeiten können.

Neben GPT-3 und GPT-4 gibt es auch weitere Large-Language-Models (LLM) von Unternehmen wie Google und Meta entwickelt, die hier nicht explizit als Meilensteine aufgelistet wurden, hier hat sich bislang der Erfolg noch nicht eingestellt.