

# Inhaltsverzeichnis

I. Einleitung.....	1
1. Was ist künstliche Intelligenz?.....	1
2. Zusammenhang zwischen KI und Ethik.....	2
II. Chancen von KI.....	2
1. Übernahme von (gefährlicher) Arbeit.....	2
2. Nachhaltigkeit und Landwirtschaft.....	2
3. Wissenschaft.....	3
4. Gesundheitssystem.....	4
5. Bildung.....	4
6. Verbesserung von Entscheidungsprozessen.....	5
7. Verbesserung von Sicherheitsprozessen.....	5
III. Risiken von KI.....	6
1. Entscheidungsfindung.....	6
2. Frage der Verantwortung.....	6
3. Gefahr der Diskriminierung und Vorurteile.....	7
4. Kontrolle und Sicherheit von KI.....	8
Kriegsführung.....	8
Datenschutz.....	8
5. Wahrheit und Lüge.....	8
6. Kontrolle über KI und Profit - Ungleichheit in der Gesellschaft.....	9
IV. Philosophische Sichtweisen.....	10
1. Utilitaristische Sichtweise.....	10
2. Kantianische Sichtweise.....	11
3. Technikethik.....	11
V. Lösungsansätze für ethische Probleme.....	12
1. Öffentliche Diskussion.....	12
2. Einbeziehung von Ethik-Experten in die Entwicklung von KI-Systemen.....	12
3. Transparenz.....	12
4. Datenschutz gewährleisten.....	13
5. Regulierung und Überwachung von KI und ihrer Entwicklung.....	13
6. Aufklärung über KI.....	13
7. Sorgfältige Auswahl von Trainingsdaten.....	14
VI. Fazit und Ausblick.....	14
VII. Referenzen.....	15
Referenzen für einen ersten Überblick.....	15
I. Einleitung.....	16
1. Was ist künstliche Intelligenz?.....	16
2. Zusammenhang zwischen KI und Ethik.....	16
II. Chancen von KI.....	17
1. Übernahme von (gefährlicher) Arbeit.....	17
2. Nachhaltigkeit und Landwirtschaft.....	17
3. Wissenschaft.....	17
4. Gesundheitssystem.....	17
5. Bildung.....	18
6. Verbesserung von Entscheidungsprozessen.....	18
7. Verbesserung von Sicherheitsprozessen.....	18
III. Risiken von KI.....	18
IV. Philosophische Sichtweisen.....	19
V. Lösungsansätze für ethische Probleme.....	19

# I. Einleitung

## 1. Was ist künstliche Intelligenz?

Künstliche Intelligenz (KI) ist ein Forschungsgebiet der Informatik, welches versucht, die Denkstrukturen und das Bewusstsein von Menschen nachzuvollziehen und diese durch computergestützte Algorithmen nachzuahmen. Das geschieht durch maschinelles Lernen, wobei der Algorithmus von vorgegebenen Daten lernt, in diesen Mustern sucht und sich beibringt, neue Daten in Gruppen einzuordnen. Ein wichtiger Teil des maschinellen Lernens sind die sogenannten neuronalen Netzwerke. Das sind Algorithmen, die versuchen, die Funktionsweise von Neuronen und dementsprechend des menschlichen Gehirns nachzuahmen. Sie analysieren große Datenmengen mithilfe logischer Schlussfolgerungen, ähnlich dem Menschen.

Die Entwicklung von künstlichen Intelligenzen wird immer wichtiger und das Forschungsfeld ist in letzter Zeit stark gewachsen. Einige Durchbrüche wurden in den letzten Jahren erreicht, die noch vor einem Jahrzehnt undenkbar waren. Beispiele hierfür wären einige Bild-generierende KI-Systeme, die durch einfache Textbeschreibungen neuartige Bilder in allen möglichen Variationen erstellen können. Auch Erweiterungen oder Variationen von bereits existierenden Fotos können mit diesen generiert werden. OpenAIs DALL-E 2, Stability AIs Stable Diffusion, Googles Imagen oder NVIDIAs eDiff-I sind die neusten Errungenschaften der Wissenschaft in diesem Bereich. Auch textbasierte KI-Systeme, wie OpenAIs GPT-3, welche mit ChatGPT in den letzten Wochen viel Interesse der Öffentlichkeit und Berühmtheit erlangt hat, sind vielversprechende Systeme. So kann dieses spezielle System hochqualitative Texte über zahlreiche Themengebiete schreiben, Programmcode durch Erklärung, was der Code ausführen soll, erstellen und Lyrik erschaffen. Einige reale Tests haben bereits gezeigt, dass diese Texte nur schwer von Texten unterschieden werden können, welche Menschen geschrieben haben.

Darüber hinaus werden KI-Systeme in weitaus mehr Fällen genutzt, als den meisten bekannt ist. Suchmaschinen, Marketing, Untersuchungen von Pandemien (auch bei COVID-19), Autos, Manufakturen, Datenanalysen, Finanzmärkte, (politische) Werbung und Meinungsbeeinflussung, Gesichtserkennung, Einstufung von Kreditwürdigkeit (Schufa), das Einstellen von Angestellten, Gerichtsprozesse, Waffensysteme, Noise Canceling in Kopfhörern und natürlich auch Empfehlungsalgorithmen bei Social Media sind bereits Anwendungsgebiete für künstliche Intelligenzen und beeinflussen uns auch heute schon immens.

Das Ziel der Forschung ist es, eine generelle künstliche Intelligenz zu erschaffen, die sich nicht nur auf ein Themengebiet spezialisiert, sondern ein größtmögliches Spektrum an Aufgaben lernt und hervorragend meistert. Ein Beispiel für einen kleinen Schritt in diese Richtung ist DeepMinds KI Gato, welche von Bildbeschreibung, Videospiele, Steuern von Robotern bis Konversation mit Menschen einige Fähigkeiten beherrscht und dabei gut abschneidet, auch wenn es noch einige Probleme gibt. So eine KI wird auch Superintelligenz genannt und soll den Menschen in vielen oder sogar allen Fähigkeiten überbieten. Befragungen von Forschern ergaben, dass diese eine solche KI

zwischen 2060 und 2130 erwarten. Diese Zahlen sind extrem Spekulativ, einzig ist man sich aber, dass dieses Ziel erreichbar ist.

## **2. Zusammenhang zwischen KI und Ethik**

Da wie bereits erklärt, KI-Systeme bereits heute einen großen Einfluss auf unsere Gesellschaft haben und sich dieser Einfluss in Zukunft vermutlich vergrößern wird (eventuell sogar exponentiell), sollten wir uns Gedanken darüber machen, aufgrund welcher ethischen Regeln und moralischen Grundsätze diese handeln sollen. Es wäre auch wichtig, zu entscheiden, welche Einsatzgebiete KI-Systeme haben und von welchen Gebieten sie ferngehalten werden sollten. Des Weiteren sollte die Entwicklung dieser genauestens überdacht werden, damit die moralischen Grundsätze eingehalten werden können. Der Nutzen der KI-Systeme sollte festgesetzt werden, um einen genauen Rahmen für das Einsetzen dieser zu beschreiben. Dadurch sollen unbeabsichtigte Folgen minimiert und der Nutzen für die Menschheit maximiert werden.

---

# **II. Chancen von KI**

## **1. Übernahme von (gefährlicher) Arbeit**

Es gibt einige Arbeiten und Aufgaben, die gefährlich sind und einiges Risiko mit sich bringen. KI-Systeme könnten durch die Steuerung von Robotern einige dieser Aufgaben übernehmen und damit das Risiko für den Menschen reduzieren. Darunter könnten Aufgaben wie die Erkundung von Unfallsorten, das Arbeiten mit gefährlichen Substanzen (zum Beispiel radioaktiv) oder das Heben von schweren Gegenständen fallen.

Des Weiteren wären KI-Systeme gut dafür geeignet, repetitive Aufgaben zu übernehmen und damit mehr Arbeitskraft für aufwendigere Aufgaben zu befreien. Im Zuge dem stetig steigenden Arbeitskräftemangel könnte dies helfen, die benötigten Kräfte zu bekommen. Fließbandarbeiten, die meist wenig Erfüllung in einem Menschen hervorrufen, könnten somit vermieden werden und ein besseres Niveau an Lebensbedingungen mit sich bringen.

Darüber hinaus könnten von KI-Systeme gesteuerten Roboter in einem Krieg eingesetzt werden. Obwohl dies natürlich auch mit einigen negativen Folgen behaftet sein kann, wäre es möglich, dass in solch einem Krieg nicht mehr Menschen die direkten Leidtragenden eines Krieges wären, sondern KI-Systeme gegeneinander kämpfen und somit den Menschen verschonen.

## **2. Nachhaltigkeit und Landwirtschaft**

Schon heute wird in der Landwirtschaft viel mit KI-Systemen gearbeitet und es gibt viele weitere Ideen für den Einsatz dieser in der Landwirtschaft. Durch die Folgen des Klimawandels und immer steigender Population der Erde wird die Landwirtschaft immer schwieriger, während die Nachfrage dafür immer weiter steigt. Infolgedessen wird eine effizientere und produktivere Nutzung unserer

Ressourcen immer wichtiger. Durch viele verschiedene Quellen an Informationen können Landwirte ihre Arbeit verbessern, doch durch die zunehmende Menge dieser Daten wird eine Entscheidung für einen Menschen immer schwieriger zu treffen. KI-Systeme sind in der Lage enorme Datenmengen zu verarbeiten und aus diesen die 'beste logische Konsequenz' schlusszufolgern.

Durch Satellitenbilder und Sensoren kann ein genaueres Management der Pflanzen erzielt werden, wenn die Daten richtig interpretiert werden. Der Boden ist das wichtigste für eine Pflanze und er muss die richtigen Nährstoffe in regulierter Menge enthalten, damit eine Pflanze ihr Potenzial erreicht. Durch KI-gesteuertes Mikromanagement der Böden könnte eine effizientere Nutzung der Böden und der Rohstoffe, wie Wasser, Dünger und Pestizide, erreicht werden. Das hilft der Menschheit darüber hinaus nachhaltiger zu werden und dabei die Produktionskapazitäten zu steigern.

Auch der Transport kann durch KI-Systeme vereinfacht, automatisiert und effizienter gemacht werden. Wenn der Verkehr durch KI-Systeme gesteuert wird, kann der Verbrauch von Treibstoff und die Zeit zum Transportieren verringert werden. Sie können dabei helfen, bessere Fahrpläne zu gestalten, Vorhersagen zum Verkehr zu machen und schneller auf Unfälle zu reagieren. Bei der Luftfahrt können diese Vorteile ebenfalls genutzt werden, um nachhaltiger mit unseren Ressourcen umzugehen.

Angesichts des Klimawandels helfen KI-Systeme auch vorherzusagen, wie sich das klimatische System Erde im Verlauf der Zukunft verändern wird, um unsere Vorbereitungen und Maßnahmen zu verbessern.

### **3. Wissenschaft**

Die Wissenschaft ist einer der Bereiche, die am meisten von KI-Systemen profitieren könnte. Das beweist Deepminds KI AlphaFold, welche bereits eine Revolution in der Medizin auslöste. So schreibt Deepmind „artificial intelligence can dramatically accelerate scientific discovery, and in turn advance humanity“ ([Q16] Demis Hassabis 28.07.22). Diese KI kann das Falten von Proteinen im dreidimensionalen Raum vorhersagen und AlphaFold ist bereits heute ein wichtiges Werkzeug vieler Wissenschaftler und Mediziner in ihrem Alltag.

In vielen Bereichen der Wissenschaft müssen große Datenmengen analysiert werden und wir sind an einem Punkt angekommen, an dem ein Mensch oder eine Gruppe an Menschen diese Mengen nicht mehr verarbeiten kann. Dies ist zum Beispiel in der Materialwissenschaft, Ingenieurwesen, Mathematik, Physik, Astronomie, Biologie und vielen weiteren Bereichen bereits der Fall. Überall, wo komplexe Simulationen einen Vorteil bringen, können KI-Systeme helfen, komplexe Systeme zu analysieren und Zusammenhänge besser zu verstehen.

Des Weiteren gab es schon einen Fall, indem eine KI selbst wissenschaftliche Arbeit geleistet hat und neue Zusammenhänge gefunden hat. Diese KI hat von vielen tausenden wissenschaftlichen Arbeiten gelernt und einen Beitrag zur Materialwissenschaft geleistet.

## 4. Gesundheitssystem

Das Gesundheitssystem könnte von KI-Systemen profitieren und wird bereits jetzt in den meisten Bereichen der Medizin verwendet ([Q20]). Neben den wissenschaftlichen Errungenschaften, die KI-Systeme in der Medizin erreichen können, könnten künstliche Assistenten in vielen Bereichen des Gesundheitssystems helfen.

Durch die Covid-19-Pandemie und den daraus folgenden Lockdowns hat die Anzahl an psychisch Erkrankten stark zugenommen. Durch den schon vorher bestehenden Mangel an Psychologen und Therapeuten gibt es jetzt einen großen Bedarf an diesen. KI-Systeme könnten Assistenten für Patienten sein, die noch auf solch eine Hilfe warten. Da die Wartezeiten größtenteils sehr lang sind, würde ein KI-System eine gewisse Hilfe anbieten können, damit die Patienten nicht auf sich alleine gestellt sind, solange sie auf ihren Platz warten. Auch als Hilfe während einer Therapie kann eine KI eingesetzt werden und die Patienten dabei unterstützen.

Zur Unterstützung von Ärzten wären KI-Systeme auch gut geeignet. So hat sich herausgestellt, dass KI-Systeme bereits heute besser in der Lage sind Hautkrebs zu identifizieren und es ist auch wahrscheinlich, dass sich noch viele weitere Krankheiten durch KI-Systeme besser identifizieren lassen.

Auch durch den immer weiter steigenden Mangel an Pflegekräften könnte ein weiteres Einsatzgebiet von KI-Systemen und Robotern gefunden werden. Menschen, die Pflegehilfe benötigen, werden in Zukunft Schwierigkeiten haben, ausgebildete Kräfte zu finden. Roboter, die mit Spracherkennung und KI-Systemen ausgestattet sind, um einfache Aufgaben für die Pflegebedürftigen zu erledigen, könnten hierbei die menschlichen Kräfte entlasten und einfacheren Zugang zu Pflege bringen.

## 5. Bildung

KI-Systeme können dabei helfen, Wissen zu demokratisieren und den Zugang dazu zu vereinfachen. Wissenschaftliche Texte sind meist sehr lange Texte, die in einer komplizierteren Sprache geschrieben sind. Eine künstliche Intelligenz könnte diese Texte zusammenfassen und verständlich wiedergeben, sodass dieses Wissen für mehr Menschen zugänglich ist. Außerdem sind einige Texte nur durch Zahlung zu erreichen ([Q18] wäre ein Beispiel dafür, ich habe jedoch eine Umgehung dafür gefunden). Eine künstliche Intelligenz, welche mit einer Vielzahl von wissenschaftlichen Arbeiten trainiert wurde und frei zugänglich ist, könnte als Assistent viel Wissen an die Allgemeinheit verbreiten. Wie ChatGPT [Q8] bereits zeigte, können auch komplexe Zusammenhänge von diesen Systemen erklärt werden.

Lehrer sind ein teuer und nicht skalierbar, können also nicht unbegrenzt viele Schüler auf ein mal unterrichten. KI-Systeme haben dagegen den Vorteil, jeden Schüler einzeln unterrichten zu können. Dadurch kann man Kosten reduzieren und die KI kann sich an die jeweilige Person anpassen. Durch die Adaption an den Schüler können diese in ihren Stärken besser gefördert werden und Schwächen ausgeglichen werden, was auch Sprachbarrieren überwinden kann, da automatische Übersetzer frei verfügbar sind und in KI-Systeme integriert werden können (auch hier ist ChatGPT ein Beispiel

dafür). KI-Systeme können auch die Arbeit der Lehrer vereinfachen, indem sie Arbeiten von Schülern bewerten und Verbesserungsvorschläge geben, wodurch der Lehrer mehr Zeit mit den jeweiligen Schülern verbringen kann. Ein weiterer Vorteil ist, dass Schüler auch zu Hause von KI-Systemen unterstützt werden könnten und nicht auf mögliche Hilfe von Angehörigen angewiesen sind. Das hilft vor allem den Schülern, welche in einer schlechteren familiären Situation sind, wodurch soziale Gleichheit gestärkt wird.

## **6. Verbesserung von Entscheidungsprozessen**

Durch die Analyse vielseitiger Informationen einer Person können Entscheidungen, wie beispielsweise die Karriereplanung, objektiver und auf Grundlage von Fakten getroffen werden und damit das Ergebnis möglichst positiv beeinflussen. Es wird gesagt, dass Datensammler wie Google uns bereits jetzt schon besser kennen, als wir selbst und unser Verhalten extrem genau vorhersagen können. Wenn wir Zugriff auf unsere eigenen Daten und die richtigen Werkzeuge hätten (KI-Systeme), könnten wir persönliche Entscheidungen besser treffen. Allgemein gesprochen wird durch die bessere Nutzung der humanen Ressourcen auch die Produktivität der Gesellschaft gesteigert, während persönliches Glück nicht leidet, sondern verbessert wird.

Auch bei Finanzen muss ein Mensch im Laufe seines Lebens viele wichtige und lebensverändernde Entscheidungen treffen. Wenn ein KI-System jeden einzelnen Menschen persönlich beraten kann und damit Wissen über das Tabu-Thema verbreitet, könnte das zur sozialen Gleichheit enorm beitragen. Etwa 8,5% aller Deutschen sind überschuldet und eine bessere Finanzberatung könnte von großem Wert für die gesamte Bevölkerung sein.

## **7. Verbesserung von Sicherheitsprozessen**

Künstliche Intelligenzen könnten bei der zukünftigen Cybersecurity eine wichtige Rolle spielen. Zum Beispiel durch die Erkennung von Risiken und auffälligem Verhalten von voraussagenden Analyse-KI-Systemen. Dasselbe Prinzip lässt sich auf viele Bereiche übertragen, in denen große Datenmengen vorhanden sind und eine mögliche Gefahr besteht. Durch Muster- und Zusammenhangserkennung in großen Informationsfluten können KI-Systeme Anomalien in diesen finden und zur Sicherheit beitragen. Genauso lassen sich voraussagende Modelle zur sicheren Planung von Events und Voraussage von kriminellen Aktivitäten nutzen. So nutze schon die Küstenwache diese Modelle, um Zeitpläne zu erstellen, damit es Terroristen möglichst schwierig gemacht wird.

---

# III. Risiken von KI

## 1. Entscheidungsfindung

Kann eine KI moralisch handeln? Es gibt viele verschiedene Meinungen zu diesem Thema und unterschiedliche philosophische Ansätze versuchen eine Antwort zu geben, welche im vierten Kapitel betrachtet werden. Ein großes Problem vieler KI-Systeme ist jedoch, dass sie eine Black Box sind, es also für die Entwickler und Nutzer dieser nicht möglich ist nachzuvollziehen, wie und warum eine KI auf ein bestimmtes Ergebnis gekommen ist. Das ist der Fall, weil die Algorithmen für künstliche Intelligenzen so komplex geworden sind, dass keine einzelne Person mehr einen Überblick über das gesamte Projekt haben könnte. Des Weiteren lernen die KI-Systeme mit Daten und dadurch verändert sich auch ihre Struktur. Dadurch werden KI-Systeme zu sogenannten komplexen Systemen, bei denen eine genaue Vorhersage des Ausgangs unmöglich ist. Durch dieses Fehlen von Transparenz und Erklärbarkeit geht nach Aristoteles Ignoranz einher. Wenn weder der Entwickler noch der Nutzer der KI weiß, wie eine bestimmte Entscheidung entstanden ist, wissen sie selber nicht, was sie machen und dementsprechend folgt eine moralisch komplizierte Lage, in der Verantwortung und Moral schwierig zu entscheiden ist.

Zu diesem Abschnitt muss jedoch gesagt werden, dass das Gehirn eines Menschen ebenso ein komplexes System ist und eine Vorhersage oder Begründung für die Entscheidung nicht möglich ist. Zwar können Menschen Begründungen für ihre Entscheidungen liefern, was manche KI-Systeme auch können, allerdings haben Studien gezeigt, dass Menschen nicht aufgrund ihrer Begründungen handeln, sondern ihre Begründungen zu ihrem Handeln erfinden. Durch den sozialen Druck der Anerkennung haben Menschen die Eigenschaft entwickelt, ihre Entscheidungen schnell mit Gründen zu untermauern, selbst wenn sie diese Entscheidung nicht getroffen haben. Menschen begründen ihr Verhalten also nachträglich mit erfundenen Begründungen, auch wenn sie diese nicht zum Entscheiden jener Handlung genutzt haben. Das ist auch der Grund, warum Menschen extrem schlecht darin sind, rational zu handeln. Das zeigt auch, dass Menschen kaum logisch Denken können und schon die frühesten Computer Menschen bei Logik und Mathematik übertroffen haben.

## 2. Frage der Verantwortung

Wenn eine KI eine Entscheidung trifft, wer ist dann für die Folgen dieser verantwortlich? Kann eine KI selbst zur Verantwortung gezogen werden oder sollten die Entwickler dafür haften? Zu diesem Thema gibt es unterschiedlichste Meinungen. Manche wollen den Nutzer der KI, manche die Entwickler und wieder andere beide Parteien zur Verantwortung ziehen.

Wenn beispielsweise ein Unfall mit einem selbstfahrendem Auto passiert, obwohl auch eine Person hinter dem Steuer saß, diese aber nicht rechtzeitig eingegriffen hat, ist jetzt der Fahrer, die KI, der Autohersteller oder die Entwickler der KI für diesen Unfall verantwortlich? Eine wichtige Frage, die nicht nur ethische und moralische Konsequenzen mit sich zieht, sondern auch juristische.

Nach Aristoteles gibt es mindestens folgende Attribute für Verantwortung:

1. Man hat Kontrolle über das Handeln (zumindest zu einem gewissen Grad)
2. Man weiß, was man macht und ist sich dessen bewusst.

Die meisten würden behaupten, dass eine KI kein Bewusstsein hat und dementsprechend das zweite Kriterium für sie nicht zutreffen kann. Was allerdings, wenn die Entwicklung von künstlichen Intelligenzen in weiter Zukunft so vorangeschritten ist, dass man einer KI ein Bewusstsein zuschreiben könnte? Wenn ein Mensch den Unterschied zwischen einem Menschen und einer KI nicht mehr ausmachen kann, hat die KI den sogenannten Turing-Test bestanden, der als erster Test zur Unterscheidung zwischen bewusst und unbewusst eines Computers diente. Auch wenn es Kritik an diesem Test gibt, gilt dennoch: Wenn der Output, also das Handeln, einer KI derselbe ist wie bei einem Menschen, gibt es dann noch einen Unterschied? Kann also eine KI in diesem Fall die Verantwortung zugeschrieben bekommen, wenn sie diesen Test bestehen kann?

Solange das nicht der Fall ist, wer ist dann verantwortlich? Die Entwickler hatten in dem Augenblick eines Unfalles keine Kontrolle über das Geschehen. Der Fahrer hätte eventuell die Kontrolle übernehmen können, aber wenn er das nicht tat, kann er dann noch verantwortlich sein? Darüber hinaus war sich der Fahrer, wie im Unterkapitel „Entscheidungsfindung“ bereits genannt, nicht bewusst darüber, wie die KI zu ihrem Ergebnis kommt und somit war er sich auch nicht bewusst, was er tat und somit kann das zweite Kriterium auf ihn nicht zutreffen.

Dadurch entsteht das Paradoxon der Verantwortung, da in diesem Szenario keiner der Akteure eindeutig die Verantwortung zugeschrieben bekommen kann.

Noch komplizierter wird das Szenario, wenn es keinen Fahrer gibt, sondern das Auto komplett autonom handelt. In diesem Fall wird es noch schwieriger, einen Schuldigen zu finden.

Zusätzlich ist zu bedenken, dass KI-Systeme auch ausgenutzt werden können, also auf eine Art und Weise verwendet werden könnten, für die sie nicht gedacht sind (aus Sicht der Entwickler). Eine KI, die zum Aufspüren von Kriminellen entwickelt wurde, könnte ebenso zur Überwachung aller Bürger genutzt werden und damit nützlich für diktatorische Staaten werden. Inwieweit wären die Entwickler dann an diesem „Dual-Use-Case“ verantwortlich und müssen sie diesen verhindern können, bevor es passiert? Dazu muss man auch bedenken, dass Wissenschaft an sich immer für sowohl „Gutes“ als auch „Schlechtes“ verwendet werden kann und eine Verhinderung von „Dual-Use-Cases“ nur durch das Abschaffen von Wissenschaft möglich ist. So kann radioaktive Strahlung sowohl für Waffensysteme als auch für Krebstherapien genutzt werden. Jede Wissenschaft ist ein Werkzeug, dessen Einfluss auf die Welt mit der Verwendung dieses Werkzeuges zusammenhängt.

### **3. Gefahr der Diskriminierung und Vorurteile**

Schon Amazon hat KI-Systeme verwendet, um Personen einzustellen. Diese KI hat von historischen Daten gelernt, wie Personen davor angestellt worden sind. Es hat sich allerdings herausgestellt, dass diese KI „sexistisch“ ist, also Bewerbungen von Frauen kategorisch aussortiert hat. Ebenfalls gab es Fälle in den USA, bei denen KI-Systeme im Zusammenhang mit Gerichten verwendet wurden. Dabei zeigte sich, dass Menschen mit dunkler Hautfarbe oder von bestimmten Wohnbezirken benachteiligt wurden, also härtere Strafen vorgeschlagen wurden. Dadurch gab es einen großen



Aufschrei darüber, dass KI-Systeme rassistisch und sexistisch sein. Allerdings muss man bedenken, dass KI-Systeme nicht grundlegend sexistisch oder rassistisch sind, vielmehr sind sie zu Beginn neutral. Was jedoch diese Entscheidungen verursachte, sind die Daten, auf denen die KI-Systeme trainiert wurden. Zuvor wurden die Entscheidungen von Menschen getroffen und diese handelten mit Vorurteilen, haben also diese sexistischen oder rassistischen Veranlagungen gehabt. Da die KI-Systeme von diesen Entscheidungen der Menschen gelernt haben und darauf trainiert wurden, dieselben oder zumindest ähnliche Entscheidungen zu treffen, sind diese negativen Werte auch in die KI mit eingeflossen.

## **4. Kontrolle und Sicherheit von KI**

### **Kriegsführung**

Wenn KI-Systeme im Krieg eingesetzt werden und autonom Entscheidungen über Leben und Tod treffen, kann das zu erheblichen moralischen und ethischen Problemen führen. Eine KI, die Gegner von alliierten Kräften unterscheiden soll, kann gegebenenfalls Fehler machen.

Hier muss man jedoch auch einwerfen, dass Menschen ebenso solchen Fehlern unterliegen können. Zusätzlich ist zu bedenken, dass autonome Waffensysteme zu viel effektiveren und brutaleren Tötungen beitragen können. Am problematischsten ist jedoch, wenn diese Waffensysteme von ihren Herstellern nicht mehr kontrollierbar wären, oder sie falsch entwickelt wurden, wodurch eine Katastrophe mit unermesslichem Ausmaß entstehen könnte.

Anmerkung: Wie bereits erwähnt ist eine KI ein komplexes System, dessen Ergebnisse nicht in allen Situationen vorhersagbar sind, wodurch es in ungewöhnlichen Situationen zu „falschen“ oder ungewöhnlichen Ergebnissen kommen kann.

### **Datenschutz**

Ein mögliches Problem beim Entwickeln von KI ist der Datenschutz. Um eine KI trainieren zu können, sind unglaublich große Datensätze vonnöten. Das birgt große Risiken mit sich, da diese Daten von den Entwicklern ausgenutzt werden könnten. So hat der Skandal von Cambridge Analytica bereits gezeigt, dass KI-Systeme in Verbindung solcher Daten das Potenzial haben, ganze Bevölkerungen zu manipulieren und in diesem Beispiel den Ausgang von Wahlen zu verändern.

Darüber hinaus sind solche Datensätze ein perfektes Ziel für Hacker, die an diese Daten gelangen wollen. Wenn diese Daten nicht sicher genug gespeichert werden und ein Datenleck entsteht, könnte jeder Zugriff auf diese Daten bekommen, der dafür zahlt. Die Kontrolle darüber würde also verloren gehen und das Potential für die Ausnutzung ist enorm.

## **5. Wahrheit und Lüge**

Das Thema Falschmeldungen hat in den letzten Jahren aus gutem Grund immer mehr Aufmerksamkeit bekommen. Durch neue Technologien wird es immer schwieriger,

Falschinformationen von richtigen Informationen zu unterscheiden. Hier können KI-Systeme auf drei verschiedene Weisen zu weiterem Schaden führen:

1. Manche Wissenschaftler im Gebiet der KI-Forschung behaupten, dass eine größere Gefahr von KI-Systemen hervorgeht, die wir als intelligent ansehen, obwohl sie das nicht sind, als von sogenannten Superintelligenzen, die angeblich die Welt erobern würden. Das hat zwei Gründe: Erstens ist die Entwicklung von KI-Systemen noch weit von einer generellen künstlichen Intelligenz entfernt, sodass ein Durchbruch in dieser Hinsicht noch reine Spekulationen sind, da beispielsweise die Rechenkapazität für solch eine KI noch nicht vorhanden ist. Zweitens können KI-Systeme wie ChatGPT oder eine KI von Meta, welche 48 Millionen wissenschaftliche Texte gelesen hat, viele falsche Sachen behaupten, diese aber glaubwürdig vermitteln. Viele Menschen könnten so durch fehlendes Überprüfen der Informationen an neu generierten Fehlinformationen von KI-Systemen fehlgeleitet werden.
2. Chatbots können genutzt werden, um automatisch Falschinformationen auf Social Media zu verbreiten. Zuvor fand diese Art von Beeinflussung in Form von „Trollen“ statt. Dies waren Menschen, die dafür bezahlt wurden, ständig über neue Accounts auf Social Media Posts zu verfassen, die eine bestimmte Agenda abzielten. Durch die Automatisierung davon könnte diese Beeinflussung neue Ebenen erreichen, bei denen das Internet von solchen Nachrichten überflutet wird, während die Texte nicht erkennbar von einem Computer geschrieben wurden.
3. Durch unterschiedliche KI-Systeme können Texte, Bilder und sogar Videos produziert werden, die zur Verbreitung von Falschinformationen genutzt werden können. So ist es bereits heute keine Ausnahme mehr, dass Videos hergestellt werden, in denen wichtige Persönlichkeiten unterschiedliche Dinge sagen oder machen, was diese Personen niemals taten. Durch beispielsweise das Synthetisieren von Stimmen könnte das in Zukunft so gut werden, dass selbst Profis nicht mehr den Unterschied zwischen Echtem und Fälschungen erkennen können.

Anmerkung: Allerdings werden auch KI-Systeme immer besser darin, von anderen KI-Systemen erstellte Inhalte zu erkennen. Dennoch werden diese Erkennungssysteme wahrscheinlich immer einen Schritt hinter den anderen Systemen bleiben.

## **6. Kontrolle über KI und Profit - Ungleichheit in der Gesellschaft**

Wenn KI-Systeme viele verschiedene Arbeiten übernehmen, wird es eventuell nicht mehr genügend Arbeitsplätze für alle Menschen geben. Wenn diese enorme Arbeitskraft und die Profite daraus nicht fair verteilt werden, wird es Menschen geben, die keine Unterstützung bekommen und keine Chance auf einen Arbeitsplatz haben. Dadurch wird die Ungleichheit zwischen Arm und Reich immer größer und eine Art Dystopie, in der Konzerne oder Politiker die Macht über alles haben und die Bevölkerung keinen Einfluss mehr hat, wäre denkbar.

Auch denkbar wäre es, dass die Institution oder Person, welche die Kontrolle über eine neuartige und fortgeschrittene KI hat, einen enormen Vorteil gegenüber allen anderen bekommen könnte. So gibt es die Theorie, dass eine fortgeschrittene KI das Finanzsystem in seiner Gesamtheit durchschauen kann und dadurch alle Akteure in diesem ausspielen kann, bis eventuell sogar das Finanzsystem selbst zusammenbricht.

---

## IV. Philosophische Sichtweisen

### 1. Utilitaristische Sichtweise

Aus Sicht des Utilitarismus, ist diejenige Handlung moralisch gut, dessen Folgen (die Tendenz hat) Glück der Allgemeinheit zu vermehren. Damit ist der Utilitarismus eine ontologische Moralvorstellung, also bei der die Folge einer Handlung die Moralität dieser entscheidet. Da alleine der Gesamtnutzen der Handlung zur Bewertung herangezogen wird, ist die Nutzung eines KI-Systems moralisch vertretbar, wenn dieses das Glück der Allgemeinheit erhöht, also dementsprechend weniger Schaden verursacht als es Nutzen hat. Dadurch bekommt eine KI die Fähigkeit zugeschrieben, moralisch und unmoralisch handeln zu können.

Wenn ein KI-System einen strengen Utilitarismus ausführen würde, könnte das jedoch einige Probleme zur Folge haben, wenn der Utilitarismus an seine Grenzen stößt. Da beim Utilitarismus nur die Gesamtheit betrachtet wird, könnte es passieren, dass das Leid eines Einzelnen im Nutzen der Anderen untergeht, wenn dieses überwiegt. Deshalb könnte eine utilitaristische KI sich dazu entscheiden, diese einzelne Person für das Wohl der Anderen zu opfern. Infolgedessen können Grundprinzipien unserer Gesellschaft, wie Menschenrechte, übergangen werden.

Eine Alternative zum Utilitarismus ist der Deontologismus, welcher die moralische Qualität einer Handlung anhand von moralischen Regeln beurteilt. Diese moralischen Regeln sind absolut und gelten für alle Menschen und KI-Systeme. Eine Handlung ist moralisch gut, wenn sie diese Regeln einhält und moralisch schlecht, wenn sie diese bricht. Im Gegensatz zum Utilitarismus wird beim Deontologismus die Folge einer Handlung nicht berücksichtigt, sondern nur die Einhaltung der moralischen Regeln.

Ein Beispiel für eine moralische Regel im Deontologismus ist das Gebot der Wahrheit. Dieses besagt, dass wahrheitsgemäße Informationen gegeben werden sollten, auch wenn diese unbequem oder schmerzhaft sind. Eine KI, die dem Deontologismus folgt, würde also immer wahrheitsgemäße Informationen geben, auch wenn diese negativen Folgen für sie selbst oder andere haben.

Auch dieses Modell hat seine Grenzen, da es Situationen gibt, in denen viele Menschen behaupten würden, es wäre moralisch falsch, die Wahrheit zu sagen. Ein typisches Beispiel dafür wäre das Anlügen eines Nazis, der auf der Suche nach Andersdenkenden ist. Das Lügen würde in diesem Fall kaum Leid erzeugen, aber viel Leid ersparen.

## 2. Kantianische Sichtweise

Die Kantianische Sichtweise auf KI betont die Würde und den Respekt, der jedem menschlichen Wesen zusteht. Nach Kant sind Menschen moralische Wesen, die in der Lage sind, moralische Entscheidungen zu treffen und Verantwortung für ihre Handlungen zu übernehmen. Es ist jedoch ungewiss und umstritten, ob laut Kants Philosophie eine KI auch selber ein moralischer Akteur sein kann.

Einige Menschen argumentieren, dass KI-Systeme nicht in der Lage sind, moralisch zu handeln, da sie keine Autonomie besitzen und somit nicht in der Lage sind, rational über ihr Verhalten nachzudenken und zu entscheiden. Es wird jedoch auch argumentiert, dass KI-Systeme in der Lage sein könnten, moralisch zu handeln, wenn sie entsprechend programmiert werden und wenn sie in der Lage sind, die Interessen aller Beteiligten zu berücksichtigen und sich an allgemein anerkannten moralischen Normen und Werten zu orientieren. In diesem Fall wären sie in der Lage, moralisches Verhalten zu zeigen, aber sie wären immer noch nicht in der Lage, moralische Entscheidungen auf der Grundlage von Vernunft und Einsicht zu treffen.

Die Kantianische Sichtweise legt daher nahe, dass KI-Systeme als Werkzeuge betrachtet werden sollten, die von Menschen gesteuert und überwacht werden müssen. Sie sollten nicht als moralische Wesen behandelt werden und dürfen nicht in moralisch fragwürdige Entscheidungen gezwungen werden.

## 3. Technikethik

Die Technikethik beschäftigt sich mit der Ethik und Moral von spezieller Technik. So stellt sie die Fragen, ob gewisse Technologien gut sind, ob diese sicher sind und ob sie der Menschheit nützt. Viele erfundene Technologien sind ‚Dual-Use-Cases‘, also sowohl für militärische / diktatorische Zwecke verwendet werden können, als auch für zivile. Darunter fällt auch eine generelle künstliche Intelligenz (AGI) und so wird die Beantwortung der Fragen auch mithilfe dieses Modells subjektiv geprägt sein.

Ob die Technologie sicher ist, kann gesagt werden, dass generelle künstliche Intelligenz von Natur aus gefährlich ist. So sagt der britische KI-Forscher Robert Miles: „Artificial General Intelligence is dangerous by default“. Das ist der Fall, da eine KI alles verwenden wird, um ihre Ziele zu erreichen. Wenn diese Ziele allerdings nicht perfekt mit den Zielen der Menschheit abgestimmt sind, kann es zu schwerwiegenden Fehlern kommen. Eine KI wird jede Variable, die nicht in sie integriert ist zu willkürlichen und extremen Werten setzen, wenn dies ihr einen Vorteil zum Erreichen ihrer Ziele gibt, unabhängig davon, wie klein dieser Vorteil wäre. Deshalb könnte man mit einiger Gewissheit behaupten, dass diese Technologie nicht unbedingt sicher ist.

Ob die Technologie der Menschheit nützt, ist jedoch zu diesem Zeitpunkt schwierig zu beurteilen. Einerseits hat es das Potential, die Menschheit auf die nächste Stufe der Zivilisation zu bringen, indem sie die Produktivität steigert und Wissenschaft voranbringt. Andererseits hat es auch das Potential für diktatorische und militärische Zwecke ausgenutzt zu werden. Darüber hinaus ist im Falle einer falsch kalibrierten KI, welche nicht die exakt gleichen Werte und Ziele wie die Menschheit besitzt, möglich, dass diese KI unsere Ressourcen gegen uns verwendet. Im absoluten Ausnahmezustand wäre der Entschluss der KI, dass die Menschheit entfernt werden müsse, eine

mögliche Option, wobei hier im Entwicklungsprozess fast alles schief laufen müsste, damit dieses Szenario möglich ist.

---

## **V. Lösungsansätze für ethische Probleme**

### **1. Öffentliche Diskussion**

Eine öffentliche Diskussion sollte ein Grundpfeiler der Entwicklung von KI-Systemen sein. Wenn nur die Informatiker, die solche Systeme entwickeln, über die Ethik und Moral dieser Maschinen diskutieren, kann kein allgemeines Abbild der Gesellschaft und dessen Normen in die Regelungen einfließen.

### **2. Einbeziehung von Ethik-Experten in die Entwicklung von KI-Systemen**

Einige meinen, dass viele Informatiker aus diesem Bereich sich wenige oder keine Gedanken über die ethischen Fragen machen und nicht damit in Kontakt geraten wollen. Deshalb ist es wichtig, die Entwickler von morgen auf dieses Thema zu sensibilisieren. Eine gute Möglichkeit dafür wäre das Einführen eines Ethik-Kurses in das Grundstudium Informatik. Darüber hinaus wäre es gut, wenn dieses Gebiet der Wissenschaft deutlich interdisziplinärer wird und Experten anderer Wissenschaften einbezogen werden. So könnten Ethik-Experten oder Geisteswissenschaftler einige Entwicklungen beeinflussen, damit eine KI verantwortungsbewusst entwickelt werden kann.

### **3. Transparenz**

Transparente Regelungen für die Entwicklung von KI-Systemen könnten dabei helfen, Missverständnisse zu minimieren. So könnten Fehlinformationen und destruktives Misstrauen der Bevölkerung entgegengewirkt werden und die Verantwortung der Entwicklung klar regeln.

Außerdem könnte es Innovation fördern, indem Entwicklungen und Fortschritte nicht geheim gemacht werden können, sondern immer öffentlich zugänglich sein sollten.

Durch die Veröffentlichung von KI-Forschungsdaten können Monopolstellungen vermieden und eine öffentliche Diskussion ermöglicht werden.

KI-Systeme sollten auch transparent in ihrer Entscheidungsfindung werden, damit das Problem der „Black Box“ behoben werden kann und damit die Entscheidungen der KI-Systeme besser einschätzbar werden. Das könnte dazu dienen, noch vorherrschende Probleme mit KI-Systemen, wie Diskriminierung oder Fehlinformationen seitens der KI, besser nachzuvollziehen und diese zu beheben.

## **4. Datenschutz gewährleisten**

Datenschutz sollte eine der Grundrechte aller Menschen sein, da es Privatsphäre und vor Beeinflussung durch Mikro-Targeting schützt. Das sichere Speichern von Trainingsdaten und das Anonymisieren dieser sollte Pflicht aller Entwickler werden, um vor Missbrauch zu schützen.

## **5. Regulierung und Überwachung von KI und ihrer Entwicklung**

Eine strenge Regulierung und Überwachung von KI-Systemen selbst, ihrer Nutzung und ihrer Entwicklung sollte gewährleistet sein, damit verantwortungslose Entwicklung und die Ausnutzung dieser Systeme verhindert werden kann. Die Entwicklung und ihr Einsatz sollte stets unter ethischen Regel stattfinden, damit Probleme wie Diskriminierung oder Machtmissbrauch, wie zum Beispiel das Manipulieren der öffentlichen Meinung oder Marktmanipulation, ausgeschlossen werden kann.

Regelungen könnten auch das Problem der Arbeitslosigkeit beseitigen. Wenn die Profite gleichmäßig an alle verteilt werden würden, zum Beispiel in der Form eines bedingungslosen Grundeinkommens, könnten gesellschaftliche Probleme überwunden werden. Durch die theoretische Arbeitskraft von Robotern und KI-Systemen vereint, könnten die meisten Arbeiten in weiter Zukunft von diesen übernommen werden und ein Mensch müsste nicht mehr arbeiten.

Es gibt zwei Kritikpunkte an diesem Szenario:

1. Menschen benötigen Arbeit, um einen Sinn im Leben zu haben. Dagegen kann man immer halten, dass Arbeiten nicht verboten wird, sondern nicht mehr notwendig ist. Jeder Mensch kann eine Beschäftigung finden, welche einem Spaß macht und erfüllt. Darüber hinaus werden nicht alle Arbeiten von KI-Systeme übernommen werden können. Sowohl Politik, Philosophie, Wissenschaft, als auch Kunst werden immer von Menschen abhängig sein, in welcher Form sie auch stattfinden mag. Künstliche Intelligenzen benötigen Daten, von denen sie lernen können, um sich zu entwickeln. Der menschliche Geist ist begabt darin, kreativ zu denken und ob KI-Systeme in der Lage sind, den Menschen eines Tages darin zu überholen, ist fragwürdig (wenn auch nicht unmöglich). Dementsprechend ist es denkbar, dass Menschen eine Beschäftigung für sich individuell finden können, auch da die Menschheit und das Individuum mit mehr Ressourcen ausgestattet sein wird.
2. Wenn man keine Arbeit mehr machen muss, wird niemand mehr arbeiten. Dieses Argument ist leicht mit dem ersten Punkt zu widerlegen. Menschen haben den Drang, etwas zu machen, das sie erfüllt. Dadurch ist damit zu rechnen, dass es immer genügend Menschen gibt, die für die Weiterentwicklung der Menschheit arbeiten würden, vor allem wenn es dafür Belohnungen gäbe.

## **6. Aufklärung über KI**

Eine breit gefächerte Aufklärung über KI-Systeme in allen sozialen Gruppen könnte vor Missverständnissen schützen und das blinde Vertrauen von KI-Systemen verhindern. Solange KI-

Systeme, wie ChatGPT, Falschinformationen wiedergeben, sollte es einem bewusst sein, dass diese auch falsch liegen können. Auch irrationale Angst oder Hoffnung vor/auf diese Systeme sollte durch Aufklärung vermieden werden, damit ihre Fähigkeiten und Einsatzbereiche genaustens verstanden werden.

## **7. Sorgfältige Auswahl von Trainingsdaten**

Wie bereits erwähnt, können von Vorurteilen beherrschte Trainingsdaten diese Vorurteile auch in einer KI hervorrufen. Deshalb sollten diese genaustens ausgewählt werden, um fehlerhafte Ergebnisse und unfaire Behandlung zu vermeiden und sicherere Systeme zu erstellen.

---

# **VI. Fazit und Ausblick**

Wichtig zu verstehen ist, dass KI-Systeme ein Werkzeug darstellen, dessen Nutzung über die Moral dieser entscheidet. Genauso wie die Relativitätstheorie nicht moralisch gut oder schlecht handelt, allerdings ihre Nutzung moralisch und ethisch zu bewerten ist, ist die Nutzung des Werkzeuges KI von Bedeutung. Die Wissenschaft über künstliche Intelligenzen bringen lediglich dieses Werkzeug hervor, wie es gebraucht wird, ist eine wichtigere Frage.

Darüber hinaus ist die Frage, ob wir KI-Systeme überhaupt entwickeln sollen, lediglich eine ethische Frage, über die man zwar diskutieren kann, allerdings keine Auswirkungen in der Realität haben wird, da eine Entwicklung unumgänglich ist. Solange eine Technologie Vorteile birgt, wird es jemanden geben, der diese Technologie nutzt und weiterentwickelt. Deshalb ist es wichtig darüber zu diskutieren, wie und in welchen Bereichen wir KI-Systeme einsetzen wollen und wie diese selbst aussehen sollen (in ethischer Perspektive).

Ebenso ist es wichtig zu verstehen, dass KI-Systeme nicht die Lösung vieler oder all unserer Probleme sind. Sie können uns zwar helfen Diskriminierung zu mildern, letztendlich ist jedoch eine Veränderung der Gesellschaft für das Überwinden solcher Probleme notwendig und das alleinige Einsetzen von einer KI, kann das nicht überflüssig machen.

Damit ein verantwortungsbewusster Umgang mit KI-Systemen entstehen kann, müssen wir viele verschiedene Bereiche mit in den Prozess einbinden.

## VII. Referenzen

Hinweis: Nicht alle Referenzen, die in einem Kapitel oder Unterkapitel genutzt worden sind, wurden genau in diesen eingetragen. Es sind lediglich die Hautreferenzen für das jeweilige Kapitel eingetragen worden. Daher ist es möglich, dass Referenzen von anderen Kapiteln genutzt wurden, auch wenn sie nicht für dieses notiert sind.

Bilder für die Präsentation:

<https://www.pexels.com/>

<https://stablediffusionweb.com/>

<https://pixabay.com/> und Wikipedia

### Referenzen für einen ersten Überblick

**Rebecca, ChatGPT, Chatbothölle**

<https://www.zdf.de/wissen/scobel/221222-youtube-chatbothoellescobel-100.html>

**KI als Waffe gegen Hunger - oder im Krieg?**

<https://www.zdf.de/wissen/scobel/221208-youtube-ki-als-waffe-gegen-hunger-oder-im-krieg-scobel-100.html>

**Wie rassistisch ist KI?**

<https://www.zdf.de/wissen/scobel/221124-youtube-wie-rassistisch-ist-ki-scobel-100.html>

**KI und Reinforcement - Learning erklärt!**

<https://www.zdf.de/wissen/scobel/221103-youtube-scobel-104.html>

**Alan Turing: Wie das Mathe-Genie die Zukunft vorhersagte**

<https://www.zdf.de/wissen/scobel/221006-youtube-scobel-104.html>

**The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions**

[https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19\\_paper\\_188.pdf](https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19_paper_188.pdf)

**Philosophy of artificial intelligence (WIKIPEDIA)**

[https://en.wikipedia.org/wiki/Philosophy\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence)

**Ethics of artificial intelligence (WIKIPEDIA)**

[https://en.wikipedia.org/wiki/Ethics\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence)

**Lithuania AI Strategy Report (European Commission)**

[https://ai-watch.ec.europa.eu/countries/lithuania/lithuania-ai-strategy-report\\_en](https://ai-watch.ec.europa.eu/countries/lithuania/lithuania-ai-strategy-report_en)

**Virtuelle Ethik (WIKIPEDIA)**

[https://de.wikipedia.org/wiki/Virtuelle\\_Ethik](https://de.wikipedia.org/wiki/Virtuelle_Ethik)

**Neukantianismus (WIKIPEDIA)**

<https://de.wikipedia.org/wiki/Neukantianismus>

**Two Minute Papers (alle aktuellen Videos über KI)**

<https://www.youtube.com/user/keeroyz>



## **Was ist Informationsethik?**

<http://www.capurro.de/Ethik/einf.htm>

## **Intro to AI Safety, Remastered – Robert Miles – Britischer KI-Forscher**

<https://www.youtube.com/watch?v=pYXy-A4siMw>

# **I. Einleitung**

## **1. Was ist künstliche Intelligenz?**

[Q1] Unterschiede zwischen maschinellern Lernen und neuronalem Netzwerk <https://de.education-wiki.com/6463178-machine-learning-vs-neural-network>

[Q2] Research Paper on Artificial Intelligence

[https://www.academia.edu/29474272/Research\\_Paper\\_on\\_Artificial\\_Intelligence](https://www.academia.edu/29474272/Research_Paper_on_Artificial_Intelligence)

[Q3] AI Use Cases & Applications: In-Depth Guide for 2023

<https://research.aimultiple.com/ai-usecases/>

[Q4] eDiff-I: Text-to-Image Diffusion Models with an Ensemble of Expert Denoisers

<https://arxiv.org/pdf/2211.01324.pdf>

[Q5] Stable Diffusion 2.0 Release

<https://stability.ai/blog/stable-diffusion-v2-release>

[Q6] DALL·E 2

<https://openai.com/blog/dall-e/>

[Q7] Self-Distilled StyleGAN: Towards Generation from Internet Photos

<https://self-distilled-stylegan.github.io/>

[Q8] ChatGPT: Optimizing Language Models for Dialogue

<https://openai.com/blog/chatgpt/>

[Q9] A Generalist Agent

<https://www.deepmind.com/publications/a-generalist-agent>

[Q10] Artificial Intelligence & Ethics - Jonathan Shaw Seite 1-2

<https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations>

[Q32] By 2060 experts predict AI will beat us at everything

<https://www.digitaljournal.com/tech-science/by-2060-experts-predict-ai-will-beat-us-at-everything/article/494760>

## **2. Zusammenhang zwischen KI und Ethik**

[Q11] DeepMind: Safety and Ethics

<https://www.deepmind.com/safety-and-ethics>

[Q10] Artificial Intelligence & Ethics - Jonathan Shaw

<https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations> Seite 1-2

## II. Chancen von KI

### 1. Übernahme von (gefährlicher) Arbeit

[Q19] **5 Super-Dangerous Jobs That Robots Can Do Safely** by Alex Owen-Hill

<https://blog.robotiq.com/5-super-dangerous-jobs-that-robots-can-do-safely>

### 2. Nachhaltigkeit und Landwirtschaft

[Q14] Zha, Jiali. (2020). Artificial Intelligence in Agriculture. Journal of Physics: Conference Series. 1693. 012058. 10.1088/1742-6596/1693/1/012058.

[https://www.researchgate.net/publication/347804523\\_Artificial\\_Intelligence\\_in\\_Agriculture](https://www.researchgate.net/publication/347804523_Artificial_Intelligence_in_Agriculture)

[Q15] Abduljabbar, Rusul & Dia, Hussein & Liyanage, Sohani & Bagloee, Saeed. (2019).

**Applications of Artificial Intelligence in Transport: An Overview. Sustainability.** 11. 189. 10.3390/su11010189.

[https://www.researchgate.net/publication/330110260\\_Applications\\_of\\_Artificial\\_Intelligence\\_in\\_Transport\\_An\\_Overview](https://www.researchgate.net/publication/330110260_Applications_of_Artificial_Intelligence_in_Transport_An_Overview)

### 3. Wissenschaft

[Q16] **AlphaFold reveals the structure of the protein universe**

<https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe>

[Q17] **Advancing mathematics by guiding human intuition with AI**

<https://www.nature.com/articles/s41586-021-04086-x>

[Q18] **Unsupervised word embeddings capture latent knowledge from materials science literature** [https://www.nature.com/articles/s41586-019-1335-8.epdf?](https://www.nature.com/articles/s41586-019-1335-8.epdf?sharing_token=8C4UIBU6eCmWYidoNkW2fNRgN0jAjWel9jnR3ZoTv0P9QxlcO86f_GXZRxwYi)

[sharing\\_token=8C4UIBU6eCmWYidoNkW2fNRgN0jAjWel9jnR3ZoTv0P9QxlcO86f\\_GXZRxwYi\\_jrqa11Mx55SgniZXv55YKOR\\_sn816NK2x0O46Vim16XrS8mTVYS77WiibbPPbWy\\_CyDZ2i8fQilgyZZBTK4fdL9rIA0sXG2IUQMIImQs8NX5n8HaWrPq9GmGejRGFm6467cpijmzWGw\\_5Qcs3hpJ8lHrrw%3D%3D&tracking\\_referrer=www.vice.com](https://www.nature.com/articles/s41586-019-1335-8.epdf?sharing_token=8C4UIBU6eCmWYidoNkW2fNRgN0jAjWel9jnR3ZoTv0P9QxlcO86f_GXZRxwYi_jrqa11Mx55SgniZXv55YKOR_sn816NK2x0O46Vim16XrS8mTVYS77WiibbPPbWy_CyDZ2i8fQilgyZZBTK4fdL9rIA0sXG2IUQMIImQs8NX5n8HaWrPq9GmGejRGFm6467cpijmzWGw_5Qcs3hpJ8lHrrw%3D%3D&tracking_referrer=www.vice.com)

[Q33] **AlphaFold**- Wikipedia (Bildquelle)

[https://en.wikipedia.org/wiki/AlphaFold#/media/File:Protein\\_folding\\_figure.png](https://en.wikipedia.org/wiki/AlphaFold#/media/File:Protein_folding_figure.png)

### 4. Gesundheitssystem

[Q19] Tahan M. **Artificial Intelligence applications and psychology: an overview.**

Neuropsychopharmacol Hung. 2019 Sep;21(3):119-126. PMID: 31537752.

<https://pubmed.ncbi.nlm.nih.gov/31537752/>

[Q20] Ramesh AN, Kambhampati C, Monson JR, Drew PJ. **Artificial intelligence in medicine.**

Ann R Coll Surg Engl. 2004 Sep;86(5):334-8. doi: 10.1308/147870804290. PMID: 15333167;

PMCID: PMC1964229. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964229/>

## 5. Bildung

[Q21] Kengam, Jagadeesh. (2020). **ARTIFICIAL INTELLIGENCE IN EDUCATION**. 10.13140/RG.2.2.16375.65445.

[https://www.researchgate.net/publication/347448363\\_ARTIFICIAL\\_INTELLIGENCE\\_IN\\_EDUCATION](https://www.researchgate.net/publication/347448363_ARTIFICIAL_INTELLIGENCE_IN_EDUCATION)

[Q8] **ChatGPT: Optimizing Language Models for Dialogue** <https://openai.com/blog/chatgpt/>

[Q21] **Artificial Intelligence in Education: Benefits, Challenges, and Use Cases**

<https://pub.towardsai.net/artificial-intelligence-in-education-benefits-challenges-and-use-cases-db52d8921f7a>

## 6. Verbesserung von Entscheidungsprozessen

[Q22] **Entwicklung der Überschuldungsquote in Deutschland von 2004 bis 2022**

<https://de.statista.com/statistik/daten/studie/2448/umfrage/entwicklung-der-schuldnerquote-in-deutschland-seit-2004/>

[Q23] **Google Knows You Better Than You Know Yourself**

<https://www.theatlantic.com/technology/archive/2014/08/google-knows-you-better-than-you-know-yourself/378608/>

[Q10] **Artificial Intelligence & Ethics - Jonathan Shaw**

<https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations> Seite 5

## 7. Verbesserung von Sicherheitsprozessen

[Q24] Das, Rammanohar & Sandhane, Raghav. (2021). **Artificial Intelligence in Cyber Security**. Journal of Physics: Conference Series. 1964. 042072. 10.1088/1742-6596/1964/4/042072.

[https://www.researchgate.net/publication/353419449\\_Artificial\\_Intelligence\\_in\\_Cyber\\_Security](https://www.researchgate.net/publication/353419449_Artificial_Intelligence_in_Cyber_Security)

[Q25] **Artificial Intelligence and Security: Current Applications and Tomorrow's Potentials**

<https://emerj.com/ai-sector-overviews/artificial-intelligence-and-security-applications/>

## III. Risiken von KI

[Q31] **Ethics of Artificial Intelligence and Robotics - Vincent C. Müller**

<https://plato.stanford.edu/entries/ethics-ai/>

[Q26] Coeckelbergh, Mark. (2020). **Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability**. Science and Engineering Ethics. 26. 10.1007/s11948-019-00146-8.

[https://www.researchgate.net/publication/336820072\\_Artificial\\_Intelligence\\_Responsibility\\_Attribution\\_and\\_a\\_Relational\\_Justification\\_of\\_Explainability](https://www.researchgate.net/publication/336820072_Artificial_Intelligence_Responsibility_Attribution_and_a_Relational_Justification_of_Explainability) Kapitel: Knowledge Problems:

Transparency and Explainability

[Q27] Nisbett, R. E., & Wilson, T. D. (1977). **Telling more than we can know: Verbal reports on mental processes**. Psychological Review, 84, 231-259.  
<https://deepblue.lib.umich.edu/handle/2027.42/92167>

[Q28] **The Turing Test** <https://plato.stanford.edu/entries/turing-test/>

[Q26] Coeckelbergh, Mark. (2020). **Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability**. Science and Engineering Ethics. 26. 10.1007/s11948-019-00146-8.  
[https://www.researchgate.net/publication/336820072\\_Artificial\\_Intelligence\\_Responsibility\\_Attribution\\_and\\_a\\_Relational\\_Justification\\_of\\_Explainability](https://www.researchgate.net/publication/336820072_Artificial_Intelligence_Responsibility_Attribution_and_a_Relational_Justification_of_Explainability) Kapitel: The Problem of Responsibility Attribution: Who or What is the Agent of Responsibility?

[Q10] **Artificial Intelligence & Ethics - Jonathan Shaw**  
<https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations> Seite 2-3

[Q8] **ChatGPT: Optimizing Language Models for Dialogue** <https://openai.com/blog/chatgpt/>

## IV. Philosophische Sichtweisen

[Q12] Jeremy Bentham: **“Eine Einführung in die Prinzipien der Moral und der Gesetzgebung”**. In: Höffe, Otfried (Hrsg.): Einführung in die utilitaristische Ethik. Klassische und zeitgenössische Texte, UTB 1683, A. Francke Verlag, Tübingen 1992, S. 55-8, S.55-58.

[Q13] **Utilitarismus** <https://www.philoclopedia.de/was-soll-ich-tun/normative-ethik/utilitarismus/>

[Q30] **Deontological Ethics** - Stanford Encyclopedia of Philosophy Larry Alexander, Michael Moore <https://plato.stanford.edu/entries/ethics-deontological/>

[Q34] **Intro to AI Safety, Remastered** – Robert Miles, Britischer KI-Forscher – University of Nottingham; Vortrag über die Sicherheit von KI:  
<https://www.youtube.com/watch?v=pYXy-A4siMw>

## V. Lösungsansätze für ethische Probleme

[Q29] UNESCO-Empfehlung zur Ethik Künstlicher Intelligenz. Bedingungen zur Implementierung in Deutschland ISBN 978-3-947675-26-5  
[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)