

# Künstliche Intelligenz

Der Deutsche Ethikrat fordert in seiner aktuellen Stellungnahme klare Leitlinien für den Einsatz von KI.

**Prau Simon, was ist der Unterschied zwischen künstlicher und natürlicher Intelligenz, wo liegt die Grenze?**

Zunächst einmal unterscheidet sich natürliche und künstliche Intelligenz in ihrer Materialität: Das Gehirn ist etwas anderes als Software. Aber künstliche Intelligenz funktioniert auch anders als natürliche: Auch wenn beide vielleicht Informationen speichern oder abrufen, sind die jeweiligen Prozesse andere. Was nun das Wesen menschlicher Intelligenz ist und ob es jemals eine sogenannte starke KI geben kann, daran scheiden sich die Geister.

**Was verstehen Sie unter „starker KI“?**

In der Stellungnahme unterscheiden wir zwischen enger, breiter und starker KI. Alle Technologien, die wir heute sehen, sind Beispiele enger KI. Das heißt: Sie simulieren menschliche Intelligenz in verschiedenen Bereichen, wie logisches Schließen oder Sprachproduktion, mit teils beeindruckender Qualität. Breite KI erweitert zunehmend das Spektrum oder die Domäne solcher Leistungen. Starke KI wiederum wird unterschiedlich definiert: Manche sagen, die perfekte Simulation aller Komponenten menschlicher Intelligenz sei bereits starke KI. Andere – so auch wir – verwenden den Begriff hingegen für die Vision einer KI, die menschliche Kognition nicht nur simuliert, sondern auch über Bewusstsein, Verständnis oder Emotionen verfügen würde. Manche vermuten, dass Letzteres jemals möglich sein kann, während andere sagen, dass es für die Zukunft nicht grundsätzlich ausgeschlossen werden kann. Nehmen wir als Beispiel ChatGPT, das mittlerweile ganz unterschiedliche Textarten in einer Qualität produziert, die es zunehmend schwer macht, zu erkennen, dass diese maschinell und nicht von einem Menschen erstellt wurde. Dennoch versteht ChatGPT nicht die Bedeutung des produzierten Texts, jener basiert rein auf statistischen Wortzusammenhängen und der Analyse und Produktion von Textmustern. ChatGPT hat also weder Verständnis noch Bewusstsein, auch wenn dies manchen Nutzer:innen so erscheinen mag.

**Wie definieren Sie Bewusstsein?**  
Bewusstsein beschreibt die Erfahrung der Welt und des Selbsts in der Welt. Mit der Frage, was Bewusstsein ist, beschäftigt sich die Menschheit seit Tausenden von Jahren und entsprechend gibt es keine einheitliche Definition davon.

**Die KI trifft heute schon selbstständig Entscheidungen. Sehen Sie das kritisch?**

Das ist eine ganz wichtige Fragestellung in unserer Stellungnahme. Wir delegieren zunehmend Entscheidungen an Maschinen. Kann man deren Tun als Handlung verstehen? Können Maschinen Verantwortung übernehmen? Wir sind zu dem Schluss gekom-



Nichts an diesem Bild ist Realität – noch nicht: Zwei künstliche Intelligenzen (vulgo: Roboter) philosophieren in klassischen Posen darüber, wohin sie die technische Entwicklung geführt hat.

## „Maschinen übernehmen keine Verantwortung“

Philosophin Judith Simon über die Revolution durch KI, ihre Gefahren – aber auch Chancen / Ein Interview von Lisa Berins

men: Im engen Sinne können Maschinen nicht handeln und dementsprechend auch keine Verantwortung übernehmen. Auf der anderen Seite können sie die Handlungsmöglichkeiten von Menschen massiv beeinflussen, sie erweitern oder sie vermindern. Wenn wir entscheiden wollen, wie KI-Technologien und Technologen im Allgemeinen zu bewerten sind, dann müssen wir ganz genau schauen, wozu dieses Delegieren von Entscheidungen führt. Erweitert sich die menschliche Autorschaft für das Leben?

Für wen ist die KI von Vorteil? Wenn ich beispielsweise im Personalbereich ein KI-Tool einsetze, dann kann das die Handlungsmöglichkeiten eines Personalers erweitern, aber die Person, die sich bewirbt, vielleicht vermindern, weil sie einem bestimmten Raster nicht entspricht. Was uns ganz wichtig ist zu betonen: Die Entscheidungen

darüber, was ein guter Einsatz, eine gute Nutzung von KI ist, kann immer nur kontextspezifisch und mit genauem Blick auf die Technik selbst, die beteiligten und betroffenen Personen, den Einsatzbereich, den institutionellen Rahmen und so weiter beantwortet werden. Allgemeine Empfehlungen können eine erste Orientierung geben, die detaillierte und genaue Analyse aber nicht ersetzen.

**Was sieht am Ende der Entwicklung? Schafft sich die menschliche Intelligenz durch die Weiterentwicklung von KI selbst ab?**

Das vollständige Ersetzen menschlicher Tätigkeiten durch Maschinen wäre der Endpunkt einer Delegation an Technik. Dies kann massive Auswirkungen haben, nicht zuletzt für die Personen, die ersetzt werden, aber auch für andere Betroffene. Aber: Von selbst übernimmt die Technik erst mal gar nichts. Es liegt in

den Händen von Menschen, die sie zu bestimmten Zwecken einsetzen. Ökonomische Motive können dazu führen, dass Aufgaben, die bisher Menschen erledigen, relativ zügig an KI abgegeben werden, weil sie einfach schneller und günstiger arbeitet. Andererseits beruht KI darauf, dass sie mit Daten trainiert wird – und diese Daten werden von Menschen unter oftmals sehr prekären Umständen produziert oder aufbereitet.

**Es werden viele Jobs wegfallen, oder?**

Es ist anzunehmen, dass zunehmend Tätigkeiten und damit auch Jobs überflüssig werden, von denen man immer gedacht hat, Menschen wären hier unersetzlich. Der Journalismus wird sicherlich unter Druck geraten, da KI gerade standardisierte Texte schneller und somit billiger als menschliche Redakteur:innen und

Redakteure verfassen kann. In allen Bereichen, in denen eine gewisse Standardisierung eine Rolle spielt, in denen nach bestimmten Schemata gearbeitet wird, ist KI ein Gamechanger: bei Vertragsprüfungen etwa – aber auch in der Kulturproduktion. Krimis, TV-Serien – die laufen nach bestimmten Mustern ab. Und KI ist nun mal ziemlich gut darin, Muster mit leichten Variationen zu re-produzieren.

**15 Expertinnen und Experten aus verschiedenen Bereichen haben an der Stellungnahme gearbeitet. Sie haben sicher auch heftig gestritten, oder?**

In der Tat, wir haben natürlich manchmal auch heftig diskutiert. Fast alles, was in der Stellungnahme steht, ist jedoch konsensuell. Aber es gibt wenige Punkte, die es eben nicht sind und an denen die Positionen unvereinbar bleiben. Diese Stellen sind als solche gekennzeichnet.

**Können Sie ein Beispiel geben?**

Ein Beispiel wäre „Classroom Analytics“ – genauer: die Frage, ob wir Technologen einsetzen wollen, die die Aufmerksamkeit oder Emotionen von Schüler:innen und Schülern im Unterricht über Videoaufzeichnung erfassen. Ist das etwas, das man kategorisch ablehnt, oder gibt es da möglicherweise zukünftig auch Chancen?

**Videoüberwachung im Klassenraum? Das klingt dystopisch. Ist das wirklich perspektivisch denkbar?**

Diese Technologien existieren ja schon. Es gibt Software zur Emotionserkennung, die in Videochatprogrammen eingebaut ist und vorgibt, anhand des Gesichtsausdrucks zu erkennen, wie man emotional verfasst ist, oder Sensoren, die die Aufmerksamkeit tracken sollen. Dies kann auch für das Live-Monitoring im Klassenraum eingesetzt werden, was teilweise in China bereits geschieht. Die Frage ist: Will man das für deutsche Klassenzimmer verhindern? Oder sagt man: Möglicherweise kann sich in Zukunft daraus auch ein Nutzen ergeben?

**Welcher könnte das sein?**

Manche argumentieren, dass es der Verbesserung des Lernprozesses dienen könne. Man könne auf sinkende Aufmerksamkeit reagieren, langweiligen Unterricht verändern und Verhalten anpassen. Auf der anderen Seite gibt es jedoch starke Bedenken, was die Privatsphäre und die Autonomie der aufgezeichneten Personen angeht. Auch Chilling-Effekte werden befürchtet, also negative Rückwirkungen auf die Schüler:innen und Schuler, die darum wissen, dass ihr Verhalten analysiert wird. Dadurch könnten Motivation, Autonomie und Handlungsfähigkeit beeinträchtigt werden, die Voraussetzung fürs Lernen sind. Eine weitere Kritik besteht darin, dass die theoretischen Grundlagen der Emotions-

## ZUR PERSON

Judith Simon ist Professorin für Ethik in der Informationstechnologie an der Uni Hamburg. Sie beschäftigt sich mit der Verschränkung ethischer, erkenntnistheoretischer und politischer Fragen im Kontext von Big Data, künstlicher Intelligenz und Digitalisierung.

**Sie ist Mitglied des Deutschen Ethikrates sowie anderer Gremien für wissenschaftliche Politikberatung und ist Sprecherin der Arbeitsgruppe „Mensch und Maschine“, die die Stellungnahme verfasste: bei**

wickelt, das einen Psychotherapeuten simulierte. Schon damals gab es das Problem, dass Patient:innen das Gefühl hatten, sie würden mit einem Menschen reden. Hier stellen sich natürlich zahlreiche Herausforderungen. Die Täuschung ist das eine, aber es stellen sich auch Fragen des Zugangs zu Psychotherapie und der Zugangsgerechtigkeit: Gerade in Anbetracht der massiven Unterversorgung im psychotherapeutischen Bereich gibt es einen großen Push, solche Technologien zu entwickeln, und dann kann man ketzerisch fragen, ob ein Chatbot besser ist als gar kein Angebot. Dass ich eher den Ausbau der psychotherapeutischen Versorgung anregen würde, ist da sicher kein Geheimnis.

**Wie kann man dieses Problem angehen?**

Es muss transparent gemacht werden, wie eine KI genau arbeitet, welche Datenbasis verwendet wurde, welche Schwachstellen sie hat. So sollte beispielsweise offen gelegt werden, wie eine medizinische Diagnose zustande kommt, wie hoch die jeweilige Unsicherheit ist. Ist sie zum Beispiel für Männer sicherer als für Frauen? Es müsste dann eine Art Warnhinweis geben: „Achtung, die Datenlage ist für Ihren Fall nicht so groß, deshalb ist das Ergebnis weniger sicher.“

**Im Bereich der Medizin stehen durch den stärkeren Einsatz von KI anscheinend viele Veränderungen an, unter anderem sollen sogar Psychotherapeuten bei einer KI möglich sein?**

Da gibt es tatsächlich viele Entwicklungen – und es ist auch keine neue Idee. Bereits 1966 hat Joseph Weizenbaum das interaktive Computerprogramm „Eliza“ ent-

Die Grundlagen der Technologie haben uns nicht überrascht, die waren schon länger bekannt. Wo wir aber nicht unbedingt gerechnet haben, war, dass dieses relativ unfertige Tool frei auf den Markt geworfen wurde und es auf einmal Millionen von Leuten nutzen. Als ChatGPT im November herauskam, waren wir mit unserer Stellungnahme fast fertig. Dennoch lassen sich die Fragen, die wir stellen, auch auf ChatGPT anwenden.

**Die KI entwickelt sich derzeit rasant. Jetzt wurde gerade GPT-4 vorgestellt, die Nachfolgegeneration des Programms, das für ChatGPT genutzt wird. GPT-4 verfügt über 100 Billionen künstlicher Synapsen und arbeitet nicht nur mit Text, sondern auch mit Bildern. Wie schätzen Sie das Potenzial von GPT-4 ein: Wird es der nächste „große Wurf“? Und was erwartet uns in der nächsten Zeit noch?**

Diese multimodale Verknüpfung wird eine Vielzahl neuer Anwendungen ermöglichen und wird derzeitige Prozesse der Text-, Bild- und Video-Generierung in vielen Bereichen noch einmal grundlegend und rasant herausfordern. Das große Disruptionspotenzial solcher Tools resultiert nicht nur aus ihrer Performance, sondern auch daraus, dass sie durch eine einfache Benutzbarkeit sehr schnell einen extrem hohen Verbreitungsgrad bekommen: Viele Personen können zunehmend ohne große Vorkenntnisse blitzschnell und nahezu ohne Aufwand Texte, Bilder und Videos produzieren, die kaum noch von Produkten zu unterscheiden sind, die zuvor menschliche Experten:innen erstellt haben. Das bringt sehr viele Potenziale mit sich – aber auch genauso viele Gefahren. Ein Trend in nächster Zeit dürfte der Einbau von GPT-4 und ähnlich multimodaler Tools in Roboter sein, um die Forschung und Entwicklung zu sogenannten verkörperter KI (embodied AI) voranzutreiben. Das heißt: Die Bewegung im und Wahrnehmung des Raumes können noch hinzu-

## Mensch, Maschine und das Dazwischen

Die KI wird viele Entscheidungen abnehmen – es sollten nicht allzu viele sein

VON LISA BERINS

Kameras und Mikrofone sind im Klassenzimmer montiert, tragen kleine Geräte am Körper, Wearables: die zeichnen ein genaues Bild jedes Individuums auf: Herzfrequenz, Körpertemperatur, emotionale und psychische Gesundheit, Mimik, Körperhaltung und Sprache, Sozialverhalten und natürlich die schulischen Leistungen – riesige Datenmengen. Mittels Künstlicher Intelligenz wird jede Unaufrichtigkeit erfasst, jeder Lernfortschritt dokumentiert.

So etwas war noch vor einigen Jahren Science Fiction gewesen, heute aber ist die komplette Durchleuchtung an einigen chinesischen Schulen Alltag. Denkbar wäre sie auch in Deutschland – eine Arbeitsgruppe des Deutschen Ethikrates hat jetzt über diesen Aspekt in ihrer Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ heiß diskutiert.

Die Menschheit befindet sich aktuell in einem „KI-Sommer“ heißt es darin: in einer Hochzeit der Entwicklungen dieser Technik. Die Verfeinerung des Chatbots ChatGPT vor vier Monaten verdeutlicht das extrem: Es generiert vermeintlich gut durchdachte, aber tatsächlich auch herbeihaltuzinierte Texte, die nach Wahrscheinlichkeiten von Sprache simuliert werden. Und es war noch relativ unausgereift, mittlerweile ist die nächste Generation auf dem Markt. Ein Vorgeschmack auf Kommendes: eine neue, immer noch mehr von Technik bestimmte Zeit.

Maschinell lernende KI und neuronale Netze, die Ummengen an Daten verarbeiten, sind kein kurzfristiger Trend, sondern ein wachsendes Business, das unseren Alltag durchdringt. Das wirft eine Menge Fragen auf: Wie geht man mit einer immer besser entwickelten künstlichen Intelligenz um, wo hilft sie, wo beschneidet sie Möglichkeiten? Welches Potenzial hat sie? Welche Aufgaben übernimmt sie, wen ersetzt sie, und ist das schlecht oder gut? Was muss getan werden, um KI nutzbar zu machen, ohne dass sie unkontrollierbar wird?

### Ein Startpunkt ist gesetzt

In seiner Analyse nimmt der Ethikrat – aufbauend auf die Stellungnahmen „Big Data und Gesundheit“ (2017) und „Robotik und Pflege“ (2019) – die Wechselwirkungen zwischen Mensch und Technik in den Blick. Seit 2020 hat der Ethikrat auf Bitte des Präsidiums des Deutschen Bundestags nun fast 300 Seiten Stellungnahme erarbeitet: die Grundlage für politische und gesellschaftliche Debatten, ein fundierender Beginn einer Diskussion, ein Startpunkt. Sie eruiert technische, philosophische und methodische Grundlagen und betrachtet ethische Fragestellungen in vier Bereichen, in denen sich KI stark ausbreitet: Medizin, Schule, öf-



Die Stellungnahme.

DFP

fentliche Kommunikation und Meinungsbildung sowie öffentliche Verwaltung.

Bei ihrer Analyse betrachtet die Arbeitsgruppe eingehend die komplexen Konzepte von Intelligenz und Kreativität, die nicht einfach auf Künstliche Intelligenz übertragbar sind. Es geht auch um Verständnis und Vernunft. Der Ethikrat postuliert, dass „die menschliche Urteilspraxis technisch nicht substituierbar“ sei – zumindest nicht durch heutige KI-Programme. Diese „verfügen nicht über die dafür relevanten Fähigkeiten des Sinnverstehens, der Intentionalität und der Referenz auf eine außersprachliche Wirklichkeit“.

Trotzdem vertrauen Menschen algorithmisch erzeugten Ergebnissen und automatisierten Entscheidungen oft mehr als menschlichen. Damit werde Verantwortung delegiert. Das kann dazu führen, dass „ein KI-System allmählich in die Rolle des eigentlichen ‚Entscheidungser‘ gerät und menschliche Autorschaft und Verantwortung ausgehöhlt werden“, warnt der Ethikrat.

In der öffentlichen Meinungsbildung sieht der Ethikrat die Gefahr, dass am wettbewerblichen Geschäftsmodell ausgerichtete Algorithmen durch eine stark personalisierte Informationsauswahl und aufmerksamkeitsbasierte Faktoren Fälschnachrichten, Hassrede, Hetze und strategische Desinformation begünstigen können. Durch die KI verändern sich „Informationsqualität und Diskursqualität“, so der Ethikrat; und damit eine wichtige Grundlage der öffentlichen Meinungsbildung – „mit potenziell weitreichenden Konsequenzen für Prozesse der politischen Willensbildung“.

Datenverzerrung und Diskriminierung durch Algorithmen könnten auch ein Problem beim Einsatz von KI in der Verwaltung werden. Der Rat steht aber ebenso mögliche positive Aspekte: Wird die Datenlage richtig genutzt, könnten Ungleichheiten aufgedeckt und korrigiert werden.

Die KI kann, dessen ist sich der Rat sicher, hilfreich sein. Aber: „Ziel und Richtschnur ethischer Bewertung muss immer die Stärkung menschlicher Autorschaft sein.“ Am Anfang steht der Mensch. Und auch am Ende.



Judith Simon (M.) und die Ethikratsmitglieder Julian Nida-Rümelin und Alena Buyx präsentieren ihre Stellungnahme am Montag in Berlin.

DFP