ZEIT ONLINE

Künstliche Intelligenz

Der Fehler liegt im System

Künstliche Intelligenz leidet unter strukturellen Problemen. Gut meinende Menschen wollen die mit Werten und Gemeinwohl lösen. Das kann nicht funktionieren.

Ein Gastbeitrag von **Jürgen Geuter**

30. September 2021, 20:38 Uhr / 153 Kommentare /

Komm, ich helf dir: Bei künstlicher Intelligenz denken viele zuerst an Roboter. Tatsächlich stecken dahinter Softwaresysteme, die mittels statistischer Methoden zu Ergebnissen kommen – nicht immer zu unserem Vorteil. © Possessed Photography/unsplash.com [https://unsplash.com/@possessedphotography]

Wenn wir heute über moderne Technologie reden, dann dreht sich das Gespräch oft um künstliche Intelligenz [https://www.zeit.de/2021/35/kuenstliche-intelligenz-sprache-rechtsrahmen-textproduktion-gpt-3-ki-sprachmodelle] oder kürzer: KI. Der Begriff der künstlichen Intelligenz ist dabei ein reiner Marketingbegriff und sollte deshalb eigentlich grundsätzlich nur in Anführungszeichen geführt werden, da die von ihm beschriebenen Statistiksysteme und Steuerungen kaum Anforderungen, die Menschen für den Begriff Intelligenz haben, erfüllen. Trotzdem gilt KI – nicht nur in Medien, sondern auch auf höchsten Ebenen der bundesdeutschen und europäischen Politik – als die Schlüsseltechnologie der Zukunft: KI soll wirtschaftliche Prosperität sichern, für effizientes staatliches Handeln sorgen, die Wissenschaft beschleunigen und das

Klima retten.

Künstliche Intelligenz [https://www.zeit.de/gesundheit/2021-07/ki-psychotherapie-kuenstliche-intelligenz-therapeut-algorithmen-gpt-3] ist allerdings keine Technologie, sondern ein Narrativ, das seit seiner Prägung in der Mitte des 20. Jahrhunderts immer wieder mit neuen Technologien (und Versprechungen) gefüllt wurde. Wenn wir den Begriff heute verwenden, beziehen wir uns allermeistens auf Systeme, die mittels statistischer Verfahren, dem sogenannten Machine Learning, bestimmte Verhalten gelernt beziehungsweise besser gesagt antrainiert bekommen haben.

Die Annahme und das Versprechen dieser Systeme ist, dass, wenn man nur genug sogenannter Trainingsdaten hat, man ein System trainieren kann, das die in diesen Daten steckenden Strukturen lernt und danach effizient und vorurteilsfrei anwendet. Der Wahrheitsanspruch dieser Systeme leitet sich dabei direkt aus einer Art maschinellen Neoplatonismus ab, der behauptet, dass Daten die Welt korrekt abbilden und dass man deshalb aus "echten" Daten aus der realen Welt auf die inneren Strukturen dieser zurückschließen kann.

Kein Abbild der Realität

Es ist nun keine neue Erkenntnis, dass diese Annahme etwas naiv ist. Machine-Learning-Systeme tun leider vor allem genau das, was man denkt: Sie lernen irgendwelche Strukturen aus Eingabedaten. Aber leider sind große Mengen von möglichen Trainingsdaten keineswegs fair und ausgewogen oder gar Abbild einer objektiven Realität, sondern voller Diskriminierung [https://www.zeit.de/2020/32/kuenstliche-intelligenz-diskriminierung-hautfarbe-gesichtserkennung], Exklusion und sachlicher Fehler. Man spricht hier – besonders, wenn es um Diskriminierung und Exklusion geht – von sogenanntem Bias, einer Voreingenommenheit in den Daten. Dieser Kombination von Bias und Wahrheitsversprechen wegen prägte der bekannte polnische Softwareentwickler Maciej Cegłowski schon 2016 den Satz, Machine Learning sei "money laundering for bias" – wie Geldwäsche für Vorurteile.

Die Situation ist leider aber noch weit schlimmer: Was genau, welche Strukturen diese Systeme "gelernt" haben, ist oft kaum nachvollziehbar. Man kann testen, ob sich das trainierte System in etwa so verhält, wie man sich das wünscht, aber reicht das? Eine Software von Amazon namens "Rekognition", die Menschen mit Vorstrafen auf Basis von Polizeifotos erkennen sollte, erkannte fälschlicherweise 28 US-Abgeordnete als Kriminelle, tat das aber mit auffallender Häufigkeit bei Abgeordneten mit schwarzer Hautfarbe. Erkannte Amazons Produkt einfach schwarze Gesichter schlechter? Hatte es einen impliziten Bias gegen Schwarze? Hatte es einen impliziten Bias, der weiße Gesichter bevorteilte? Unklar. Das System war bei Amazon durch die technischen Tests gegangen und erst die Kontrolle durch Nichtregierungsorganisationen und Watchdogs brachte das Problem zutage.

Aus soziotechnischer Perspektive ist also der Einsatz von KI-Systemen, insbesondere wenn es um die Leben von Menschen geht, sehr problematisch und schließt sich teilweise sogar aus: Nicht nur lagert man extreme Macht an einfache Automatisierungssysteme aus, man kann diese Systeme sogar nur ungenügend auditieren, um sicherzustellen, dass sie sich korrekt und nicht bezogen auf einzelne Gruppen diskriminierend verhalten. Im Zweifel zeigen sich solche Verfehlungen erst im Betrieb und die Betroffenen müssen neben dem Schaden, der ihnen entsteht, auch noch den Nachweis der Verfehlung des Systems bringen.

Auch darf man nicht vergessen, dass Machine-Learning-Systeme qua Definition strukturkonservativ sind: Das Wahrheitsversprechen leitet sich daraus ab, aus echten Daten, das heißt, aus einem Abbild der Vergangenheit, zu lernen, wie Dinge sein sollen. Die Vergangenheit normiert damit die Gegenwart und die Zukunft und gibt die Strukturen und den Raum der Möglichkeiten vor. Gerade im Hinblick auf die anstehenden politischen Herausforderungen wie zum Beispiel der Klimakatastrophe [https://www.zeit.de/politik/ausland/2021-07/klimakatastrophe-g20-treffen-nepal-umweltminister-erderwaermung] oder sozialen Ungleichheit [https://www.zeit.de/2021/20/soziale-ungleichheit-corona-krise-einkommen-armut-reichtum-wirtschaft] erscheint dieser Ansatz bestenfalls zynisch zu sein. Wie geht noch dieses fälschlicherweise Einstein zugeschriebene Zitat? "Wahnsinn ist, immer wieder das Gleiche zu tun und andere Ergebnisse zu erwarten." Das ändert sich auch nicht durch die Automatisierung des Prozesses.

Manchmal ist "Nein" die einzig richtige Antwort

Als Gegengift zu diesen toxischen Handlungsweisen von KI-Systemen etabliert sich seit einiger Zeit sowohl im Bereich von Nichtregierungsorganisationen und Forschung aber auch in der Politik ein Diskurs, der "KI der Werte"

beziehungsweise "KI für das Gemeinwohl" fordert und umsetzen will: Dabei sollen zum Beispiel Trainingsdatensätze überprüft oder der Entwicklungsprozess durch Teilhabe von Betroffenen fairer gestaltet werden. Und all diese Ansätze haben im konkreten Projekt einen Wert und sicher auch oft einen positiven Beitrag. Wobei man nicht vergessen darf, dass die Antwort auf die Frage, was genau denn nun einem Gemeinwohl zuträglich ist und wer zur Gemeinschaft gehört, der wohlgetan werden soll, ebenso von Vorurteilen und Diskriminierungen geprägt ist wie unsere gesamte Gesellschaft – denke man nur daran, wie oft das Gemeinwohl nur von den Bürgerinnen und Bürgern her gedacht wird, ohne an die Menschen mit anderen Pässen zu denken, die ebenso Teil unserer Gemeinschaften sind.

Eine "KI der Werte" ist eine Kapitulation vor den Systemen

Aber der Ansatz der "KI der Werte" ist vor allem eine Kapitulation. Der Einsatz von künstlich intelligenten Systemen wird damit als quasi alternativlos akzeptiert. Das kann man als eine Geste der Unterordnung unter das wirtschafts- und politikgetriebene Versprechen von Wohlstand und Glück durch künstliche Intelligenz interpretieren – niemand ist gerne Verhinderer. Doch gerade wegen dieses Dranges diverser einflussreicher Gruppen, künstliche Intelligenz als fast magisches Heilsversprechen zu propagieren, wird es immer wichtiger, den harten politischen Fragen nicht auszuweichen und Feigenblätter für technische Systeme zu suchen. Die Umsetzung solcher Systeme kann sich nicht beschränken auf die Frage: "Wie können wir ein besseres KI-System bauen?" Sie muss – gerade in Anbetracht der immer wieder untersuchten und bestätigten Charakteristika dieser Systeme, wie sie auch dieser Text referenziert – das "Nein" erlauben, wenn nicht gar von ihm ausgehen: Es reicht nicht, zu evaluieren, ob man ein KI-System verbessern kann, oft ist die richtige Lösung, das System abzuschalten. Diese Entscheidung wird aber oft in einem Versuch, die Trainingsdaten oder die Auswertung irgendwie zu optimieren, vergessen.

Medienethik

KI-Regeln der EU

"Würden Sie bei einem Atomkraftwerk erst einmal gucken, was passiert?"

[https://www.zeit.de/digital/internet/2021-04/kuenstliche-intelligenz-regeln-eu-kommission-medienethik-jessica-heesen]

Zu Beginn des 19. Jahrhunderts organisierten sich in England gut ausgebildete Textilhandwerker in einer Bewegung zum Schutz ihrer sozialen Stellung: Technologien wurden eingeführt, um Arbeitsstandards zu unterlaufen und die

Macht der Arbeiter zu untergraben. Die Ludditen zerstörten Webstühle in einem Protest nicht gegen die Technologie, sondern dagegen, dass Technologien eingesetzt wurden, um Menschen Mitbestimmung, einen halbwegs auskömmlichen Lebensunterhalt und Sicherheiten zu nehmen. Es ging nicht gegen die Technologie an sich, sondern darum, dass sie nur zur Unterdrückung eingesetzt wurde anstatt zum Beispiel für eine Reduktion der Arbeitszeit bei gleichem Lohn. Man wollte keine "automatisierten Webstühle der Werte", sondern hatte erkannt, dass das herrschende System diese Formen der Automatisierung immer zum Nachteil der Arbeiter einsetzen würde. Wir können – nicht nur in diesem Falle – viel aus der Geschichte der Ludditen, die tragischerweise meist zu einem Zerrbild verkürzt erzählt wird, lernen. Lernen, dass manchmal einfach "Nein" die einzig richtige Antwort ist.

KI-Systeme müssen weder grundsätzlich verboten werden noch sind sie niemals leistungsfähig oder einem Gemeinwohl zuträglich. Aber allein die strukturellen Probleme dieser Systeme, die eben keine Programmierfehler sind, die man nur reparieren muss, sollten den Einsatz von KI-Systemen an vielen Stellen grundsätzlich und kategorisch ausschließen: Wenn es zum Beispiel darum geht, ob Menschen notwendige Ressourcen wie Gesundheitsleistungen oder Sozialhilfe bereitgestellt werden, können automatische Systeme ganz grundlegend nicht eingesetzt werden, hier ist es essenziell, dass Menschen mit Blick auf die reale Situation entscheiden und nicht reine Statistik. In solchen Fällen ist es egal, wie viele Werte man glaubt, in die Trainingsdaten geschmuggelt zu haben: Das Ergebnis ist und bleibt intransparent, unüberprüfbar und im Zweifelsfalle unmenschlich.