Künstliche Intelligenz trifft „künstliche Dummheit“

INTERVIEW. Mehr Daten führen nicht immer zu besseren Erkenntnissen, betont Professor Gerd Antes, bis 2018 Direktor des deutschen Cochrane Zentrums. Für den Mathematiker und Biometriker sind Digitalisierung und künstliche Intelligenz inzwischen zur Ideologie verkommen, für die wissenschaftliche Qualitätsstandards über Bord geworfen werden. Antes warnt, die Korrelation von Daten sage alleine überhaupt nichts aus.

Im Personalbereich gibt es immer mehr Start-ups, deren Programme mithilfe künstlicher Intelligenz (KI) die Personalauswahl oder Mitarbeiterbeurteilung vereinfachen und verbessern sollen. Dabei fällt auf, wie unkritisch die neuen Angebote oftmals hochgejubelt werden. Wie erklären Sie sich das?

Professor Dr. Gerd Antes: Digitalisierung ist derzeit ein absoluter Hype, der die normale Kritikfähigkeit bei vielen einfach ausschaltet. Das ist natürlich gerade bei Journalisten verheerend. Das Ganze lässt sich auch nicht mehr rational erklären. Das ist schon fast so etwas wie eine Gehirnwäsche. Digitalisierung ist inzwischen eine oft sehr stark vom Marketing getriebene Ideologie. Dabei werden manchmal Dinge behauptet, die völlig absurd und mit theoretischen Grundlagen völlig unvereinbar sind.

Ideologie? Aber gerade datenbasierte Verfahren sollen doch in Wirklichkeit vorurteilsfrei sein.

Antes: Wir haben doch längst eine religiös anmutende Verehrung von Daten. Man kann zwar lange darüber philosophieren, ob Religion eine Ideologie ist. Auf jeden Fall ist es etwas, was nicht hinterfragt werden kann. Und das ist hier genauso. Was heißt denn eigentlich künstliche Intelligenz? Wenn man nachfragt, hört man, dass die Daten mit Methoden der künstlichen Intelligenz ausgewertet werden. Und wenn man weiter fragt, wie das funktioniert, bekommt man das nächste Schlagwort um die Ohren gehauen: Deep Learning. Und weil der Computer beim Deep Learning alleine lernt, wissen wir nicht, wie es geht. Daher gibt es leider keine Transparenz. Da steht man dauernd irgendwo im definitorischen Niemandsland. Ich habe das schon mal als Trumpismus in der Wissenschaft bezeichnet. Das, was wir derzeit verstärkt in der Gesellschaft sehen wie Aberglaube, Fake-News, Verschwörungstheorien, also alles, bei dem die Ratio ausgeschaltet wird, ist zumindest an den Rändern auch schon in die Wissenschaft eingedrungen. Vieles von dem, was heute versprochen wird, kann einfach nicht funktionieren.

Warum nicht?

Antes: Der fundamentale Fehler ist die Annahme, man brauche nur riesige Datenmengen und der Rest ergäbe sich dann von allein. Das ist einfach nur unglaublich dümmlich. Es gibt Arbeiten, die in jüngster Zeit eindrücklich bestätigt wurden und die zeigen, dass mehr Daten sogar ungünstiger sein können als weniger Daten. Wenn ich zum Beispiel eine Befragung vor einer Wahl mache, hängt das Ergebnis eng damit zusammen, wer überhaupt mitmacht. Es gibt also eine Selbstselektion. Das ist zum Beispiel bei der Wahl von Donald Trump der Fall gewesen, deren Ergebnis keiner vorhergesagt hat. Inzwischen gibt es Studien der Harvard University, die zeigen, dass die Ergebnisse einer Befragung nicht besser sind, wenn man die Daten von einem Prozent der US-Bevölkerung, also ungefähr 2,3 Millionen Menschen, verwendet als wenn man eine sorgfältig ausgewählte, repräsentative Stichprobe von 500 Men-

„Wir haben doch längst eine religiös anmutende Verehrung von Daten.“

Foto: Monty Rakusen / gettyimages

schen nutzt. Man braucht eine sorgfältig geplante, repräsentative Zufallsstichprobe und muss eine Befragung durchführen, bei der auch alle mitmachen und bei der größten Wert auf Qualität gelegt wird. Sonst bekommt man immer systematische Verzerrungen. Aber darum bemüht sich heute keiner mehr, weil alle dem Mythos Big Data verfallen sind.

Mehr Daten bedeuten also letztlich mehr Fehler?

Antes: Wer über künstliche Intelligenz spricht, muss auch über künstliche Dummheit reden. Denn unter der Menge der Daten leidet die Qualität der Ergebnisse, wenn ich naiv vorgehe. Und daher bedarf es größerer menschlicher Anstrengungen, um die Qualität sicherzustellen. Die Tragödie von Big Data ist: Je mehr Variablen ich habe, desto mehr signifikante Korrelationen habe ich. Damit wächst jedoch auch die Zahl der unechten oder „falschen" Korrelationen. Dafür gibt es unzählige Beispiele. Je größer die Datenmenge, umso mehr Korrelationen gibt es, die jedoch inhaltlich völlig schwachsinnig sind. Zum Beispiel, dass Menschen mit kleineren Füßen weniger verdienen. Hier ist es das Geschlecht, das das Ergebnis erklärt. Frauen haben kleinere Füße und verdienen weniger. Das ist hier leicht zu erkennen, ist jedoch üblicherweise viel komplexer. Im Englischen heißen diese Korrelationen „Spurious Correlations".

Sie meinen Scheinkorrelation?

Antes: Das ist der übliche deutsche Begriff, aber er ist eigentlich falsch. Denn die Zahlen sind tatsächlich korreliert. Der Schein kommt durch die falsche Interpretation, die das Ganze zu einem irreführenden kausalen Zusammenhang verfälscht. Darum muss man unbedingt Mechanismen entwickeln, um all diese irreführenden Korrelationen und fälschlich hineingedeuteten Kausalitäten zu entlarven und möglichst rasch auszuschalten. So kann man zum Beispiel mit Daten belegen, dass sich der Käsekonsum in den USA von 2000 bis 2010 parallel zur Anzahl der Menschen entwickelt hat, die auf tragische Weise unter ihrer Bettdecke erstickt sind. Wenn diese Zusammenhänge kausal wären, sollte man durch die Reduktion des Käsekonsums die Gefahr des Todes im Bettlaken reduzieren können – was natürlich blanker Unsinn ist. Durch Big Data kommt es zu einem starken Anstieg falsch positiver Ergebnisse und damit ist Big Data vor allem ein Bullshit-Generator.

Was bedeuten „falsch positive Ergebnisse"?

Antes: Es gibt allgemeingültige Regeln, die für jedes diagnostische Verfahren gelten. Wenn nach etwas gesucht wird, findet man einen Teil dessen, wonach man sucht, aber fast nie alles und unvermeidlich auch einen Teil von dem, wonach man nicht sucht. Das erlebt jeder, der sich durch den Berg der Ergebnisse einer Google-Suche quält, um die wenigen Treffer zu finden, die tatsächlich Antworten auf seine Frage liefern. Das Gleiche gilt für jede wissenschaftliche Untersuchung. Bei klinischen Untersuchungen, Massenscreenings, Tests oder dem Datensammeln ist es immer so: Es gibt richtig positive und negative sowie falsch positive und negative Ergebnisse.
Nehmen Sie eine Laboruntersuchung auf eine Infektion. Positiv heißt, dass eine Probe auf eine Infektion hinweist. Richtig positiv, dass die Analyse die Infektion bestätigt, falsch positiv, dass die Probe fälschlicherweise eine Infektion anzeigt. Darin liegt ein großes Schadenspotenzial, vor allem wenn die Diagnoseverfahren auf Gesunde oder Unverdächtige angewendet werden. Selbst wenn in Deutschland niemand einen Terroranschlag im Sinn hätte, würde jedes Verfahren trotzdem Verdächtige produzieren. Damit wären die aber alle falsch positiv und jeder Verdacht wäre zu hundert Prozent ein Fehlalarm. Wenn wir also mehr Polizei einsetzen, haben wir auch mehr falsch positive Ergebnisse. Dasselbe gilt natürlich für falsch negative Ergebnisse. Wer als nicht verdächtig eingestuft wurde, ist es vielleicht doch.

Der Geschäftsführer der Bertelsmann Stiftung, Jörg Dräger, der sich für den gesellschaftlich verantwortungsvollen Einsatz von Algorithmen einsetzt, hat →

→ **vor kurzem in einem Interview gesagt: Algorithmen sind gut im Verstehen von Mustern und Korrelationen. Wenn ein erkanntes Muster stabil ist und nicht diskriminiert, braucht es auch nicht immer eine erklärbare Kausalität.**

Antes: Das ist einer der dümmsten Sätze, die ich seit langem gehört habe. Was heißt Algorithmen verstehen Korrelationen? Das ist inhaltsleer und zeigt, dass er nicht verstanden hat, wovon er redet. Das kann ich nur so hart sagen. Denn Korrelationen allein sagen gar nichts aus, weil es eben so viele und oftmals schwer erkennbare „Spurious Correlations" gibt. Und dagegen sind die Maschinen völlig hilflos.

Eine weitere Aussage von ihm ist, dass man nicht unbedingt eine Theorie braucht, sondern den Erfolg eines Verfahrens – zum Beispiel bei der Personalauswahl – am Ergebnis messen kann. Man schaut also zum Beispiel, ob die mit Algorithmen ausgewählten Bewerber schneller Karriere machen.

Antes: Dann muss man auch Studien haben, die genau das sauber zeigen. Da kommt man stets schnell auf den wunden Punkt: Wie sind die Gruppen gebildet worden, die man vergleicht. Denn die müssten möglichst randomisiert sein und man müsste das zeitgleich und parallel untersuchen, um die massive Verfälschung durch andere Faktoren zu reduzieren oder idealerweise sogar auszuschließen. Wenn man mal vor drei Jahren eine Studie ohne maschinelle Selektionsmechanismen gemacht hat, kann man das nicht mit einer aktuellen Stichprobe von Bewerbern vergleichen, die mit Algorithmen ausgewählt wurden. Der Selektionsbias ist der Killer vieler Studien.

Es gibt bereits Anbieter, die behaupten, ihre Programme können errechnen, welche Mitarbeiter demnächst kündigen werden.

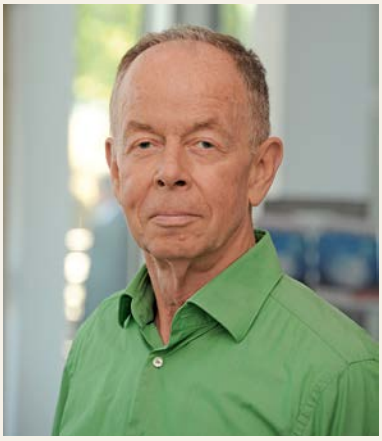
Antes: Das sind Prognosen. Dazu gibt es in der Wissenschaft einen bösen Spruch: von Daten ein sehr viel bedeutenderes Problem stehe: die Annahme, dass große Datenmengen stets besser seien als wenige Daten und dass Datenanalysen letztlich zur Lösung jedes Problems führten. Im „Ärzteblatt Baden-Württemberg" prangerte er im Juni 2019 an: „Die fundierte Technikfolgenabschätzung ist mutiert zu einer Ideologie, die nicht hinterfragt, sondern umgesetzt werden muss." Auf der „Skepton", einer Konferenz für kritisches Denken, referierte Antes im Mai 2019 zum Thema „Gute wissenschaftliche Praxis und Evidenz versus Big Data und künstliche Intelligenz – Partner oder Gegner?" Die Konferenz wurde von der Gesellschaft zur wissenschaftlichen Untersuchung von Parawissenschaften (GWUP e. V.) veranstaltet.

Wer ist Professor Gerd Antes?

Hintergrund. Professor Gerd Antes war bis 2018 Direktor des deutschen Cochrane Zentrums am Institut für Medizinische Biometrie und Medizinische Informatik am Universitätsklinikum in Freiburg im Breisgau.

Das Cochrane-Netzwerk ist benannt nach dem britischen Arzt und Epidemiologen, Sir Archibald Leman Cochrane. Dessen Überlegungen zur Überprüfung von Therapien in Studien und die Aufbereitung dieser Ergebnisse in systematischen Übersichten waren der Ausgangspunkt für die Gründung von Cochrane im Jahr 1993. In den vergangenen Jahrzehnten ist daraus ein globales unabhängiges Netzwerk von klinischen Forschern, Ärzten, Methodikern, Angehörigen der Gesundheitsfachberufe und Patienten entstanden, das die wissenschaftlichen Grundlagen für Entscheidungen im Gesundheitssystem verbessern will. Antes ist Mitinitiator des 1998 gegründeten „Deutschen Netzwerks Evidenzbasierte Medizin", für das er später auch als Vorstand fungierte. Das Leitbild des Vereins ist geprägt durch ein kritisch-wissenschaftliches Denken und die Orientierung am Patientennutzen.

Im Jahr 2012 ernannte ihn die Medizinische Fakultät der Universität Freiburg zum Honorarprofessor. Heute ist der 70-Jährige als Gastwissenschaftler am Institut Evidenz in der Medizin der Universität Freiburg und am Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Universität München tätig. In seinen Vorträgen und Artikeln prangert Antes vor allem den unkritischen Umgang mit Big Data im Gesundheitswesen an und plädiert dafür, die Grundsätze guter Wissenschaft nicht über Bord zu werfen. Antes gilt unter Experten als ein Freund der klaren Worte. In einem Gastbeitrag für die „Süddeutsche Zeitung" schrieb er, dass hinter der missbräuchlichen Nutzung



Gerd Antes. Der 70-Jährige arbeitet noch als Gastwissenschaftler an der Uni Freiburg.

Prognosen sind das Schwierigste, vor allem, wenn sie die Zukunft betreffen. Wenn ich so etwas mache, muss ich eine Studie durchführen, um herauszufinden, ob meine Vorhersageregel tatsächlich richtige Prognosen garantiert. Kündigt der wirklich? Und ist das nicht vielleicht ein ganz normales Geschehen und ich hätte auch genauso gut einen Würfel werfen können. Dafür brauche ich Gütekriterien, was jedoch fast nie gemacht wird. Das ist ja auch die Hauptkritik bei der Testbewertung der Sprachanalyse „Precire“, die vor kurzem vom Testkuratorium quasi für unbrauchbar erklärt wurde. Das Ganze ist schlecht dokumentiert, die Kriterien sind nicht offengelegt und man weiß eigentlich gar nicht, was da passiert. Und der Rest ist Behauptung. Um Persönlichkeit zu bestimmen, muss ich erst einmal wissen, wie ich sie definiere. Und dann muss ich prüfen, ob meine Ziele mit dem Verfahren überhaupt erreicht werden. Wenn ich zum Beispiel diejenigen ausschalten will, die zum Diebstahl neigen, stimmt dann meine Prognose? Stehlen diese Menschen wirklich mehr? Dafür gibt es eine ausgefeilte Methodik, um das zu überprüfen. Aber viele Start-ups machen sich nicht einmal die Mühe und meist sind die auch gar nicht dafür ausgebildet, methodisch sauber den Nachweis zu erbringen, dass ihr Verfahren funktioniert. Aber das sollte eigentlich ihre Pflicht sein wie bei jedem Arzneimittelhersteller. Der muss auch den Beweis erbringen, dass sein Mittel einen Nutzen bringt und höchstens einen begrenzten Schaden.

Wie würde so ein sauberer Nachweis aussehen?

Antes: Ich muss erstens das Ziel definieren: Was will ich eigentlich? Das fehlt zum großen Teil oder es ist so schwammig formuliert, dass man danach eigentlich immer recht hat. Wenn ich das Ziel habe, brauche ich Kriterien, um zu beurteilen, wie gut das Ziel erreicht wurde. Dafür wurde in den letzten Jahrzehnten eine Fülle von Methoden entwickelt und angewendet. Das ist während des gegenwärtigen Hypes alles über Bord gekippt worden. Mit der Versprechung, mit genug Daten bräuchte ich diese Anstrengungen nicht mehr – das wird aber auch nicht so deutlich gesagt, sonst würde klar, dass es absurd ist –, ich bräuchte nur möglichst viele Daten und das erledige sich quasi von allein. Das ist blanker Unfug.

Gerade in der Medizin setzt man bei der Diagnose stark auf Big Data, weil Computer angeblich besser und schneller sind.

Antes: Auch hier fehlt mir der Qualitätsbegriff. Es gibt keine Beweise, dass Big Data besser ist. Ein besonders schockierendes Beispiel ist IBM mit seinem Supercomputer Watson. IBM behauptete im Jahr 2011, dass Watson dank künstlicher Intelligenz die Krebstherapie revolutionieren werde. Das renommierte Krebszentrum MD Anderson der University Texas investierte über 60 Millionen Dollar für den Einsatz von Watson und brach das Experiment nach 2017 nach drei Jahren ab, weil sie gemerkt haben, dass die Behandlungsempfehlungen der Maschine einfach schlecht sind. Als eine der letzten Institutionen in Deutschland ist Watson auch bei den Rhönkliniken rausgeflogen. Der CEO sagte gegenüber dem „Spiegel“: „Ich dachte mir: Wenn wir da weitermachen, investieren wir in eine Las-Vegas-Show.“ IBMs Vorgehen war vor allem marketinggetrieben.

„Es gibt keine Beweise, dass Big Data (zum Beispiel bei medizinischen Diagnosen) besser ist.“

In der Wissenschaft gibt es derzeit ein großes Problem mit der Reproduzierbarkeit von wissenschaftlichen Studien. Hilft da Big Data?

Antes: Auch das ist wieder so ein Taschenspielertrick. Big Data kann Analysen nicht reproduzieren, weil sich die reale Welt und damit die Daten ständig ändern und vor allem das Volumen rasant zunimmt. Damit wird das Problem der Reproduzierbarkeit quasi ausgeschaltet. Man braucht sie nicht mehr. Aber so renne ich in alle bekannten Fallen wie zum Beispiel die unechten Korrelationen, die sich gerade bei anderen Daten nicht reproduzieren lassen. Die zeigen mir dann, dass heute etwas zusammenhängt, morgen aber nicht mehr. Leider schweigen die Methodiker dazu, statt einmal klarzustellen, dass Big Data dieses Problem nicht lösen wird.

Wie konnte es überhaupt dazu kommen, dass man jetzt plötzlich alle wissenschaftlichen Grundregeln über Bord schmeißt?

Antes: Da gibt es für mich drei Triebkräfte: erstens grenzenlose Naivität. Das beträfe dann die Leute, die die Zukunft als Datenparadies beschreiben, aber selbst noch nie Daten in die Hand genommen oder ein Computerprogramm geschrieben haben; zweitens eine unglaubliche Inkompetenz, mit der man sich die Ignoranz gegenüber der Fehlinterpretation von Korrelationen erklären muss, obwohl diese Zusammenhänge schon für Anfänger gelehrt werden; und drittens massivste Interessenkonflikte. Das trifft auch gerade in der Wissenschaft zu. Obwohl viele wissen oder wissen müssten, dass vieles ohne jedes theoretische Fundament ist und so nicht funktionieren kann, schweigen sie, weil sie sich nicht selbst von ihren Drittmittelströmen abkoppeln wollen.

Wie ist Ihre Prognose? Wird sich der Hype wieder legen und bald wieder mehr Wert auf Rationalität gelegt werden?

Antes: Die Antwort ist sehr zynisch. Wir brauchen eine Katastrophe und eigentlich haben wir die schon. Die Abstürze der beiden Boeing 737 Max sind das beste Beispiel. Da wurde dem Menschen die Entscheidungsfindung abgenommen. In der Medizin gab es den größten Schub für Qualitätsstandards nach dem Contergan-Skandal. Manchmal geht das ganz schnell. In Arizona hat der Gouverneur einen Tag nach einem tödlichen Unfall mit einem selbstfahrenden Auto diesen Fahrzeugen die Nutzung der Straßen verboten. Bei den Firmen ist die Frage, wie schnell sie begreifen, dass sie mit den KI-Verfahren vielleicht doch das schlechtere Personal auswählen – so wie es mit Watson in Houston passiert ist.

Interview: Bärbel Schwertfeger