Schattauer

ChatGPT

Nur ein weiterer Trend oder eine Revolution?

Korrespondenzadresse
Prof. Dr. Dr. Manfred Spitzer
Universität Ulm
Abteilung für Psychiatrie
Leimgrubenweg 12–14
89075 Ulm, Deutschland

Bibliografie
Nervenheilkunde 2023; 42: 192–199
DOI 10.1055/a-1948-8785
ISSN 0722-1541
© 2023. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

ChatGPT ist der Name einer künstlichen Intelligenz (KI), die in den vergangenen Monaten viel Aufmerksamkeit erregte. Es handelt sich um einen sogenannten Chatbot, also um ein technisches System (Roboter), mit dem man sich unterhalten („to chat“) kann. Man könnte auch von einem Computerprogramm sprechen, wäre da nicht die Tatsache, dass das Wesentliche an ChatGPT gerade darin besteht, dass es sich nicht um ein Programm handelt, sondern um ein neuronales Netzwerk, das selbst lernfähig ist, dessen Fähigkeiten also nicht (von Menschen) programmiert, sondern anhand von Beispielen (von der Maschine) gelernt wurden. Man sprich von „machine learning“, und der aufmerksame Leser dieser Zeitschrift hat schon einiges hierzu erfahren [1, 2].

Große Sprachmodelle: Lernen, Umformen, Hervorbringen

ChatGPT basiert auf seinem Vorgänger GPT3, einem neuronalen Netzwerk, das mit Milliarden von Sätzen als Input erhielt und letztlich nichts weiter lernte als Zusammenhänge zwischen Wörtern, sodass es neue Wörter hintereinander zu produzieren lernte, die sich wie von Menschen produzierte Sätze lesen. Man spricht auch von großen Sprachmodellen (LLM: „large language models“). ChatGPT wurde dann dadurch weiter verändert, dass Menschen seinen Output verbesserten und diese Verbesserungen zusätzlich lernte. Hierzu wurden menschliche KI-Trainer eingestellt, um mit dem Modell eine schriftliche Unterhaltung zu führen, wobei die Trainer sowohl den Benutzer als auch den Chatbot zu spielen hatten. Durch Vergleich und Bewertung der von der ChatGPT geschriebenen Antworten wurde dessen Gesprächsstil dahingehend verbessert, immer menschenähnlicher zu wirken.

GPT3 und ChatGPT wurden von der Firma OpenAI entwickelt, die im Jahr 2015 in San Francisco mit einer Milliarde US-Dollar Startkapital (u. a. von Elon Musk und Peter Thiel) gegründet wurde [3]. Das System kann mittlerweile auch Texte stilistisch verbessern und sogar Fehler in Computerprogrammen finden. In beiden Fällen ist seine Dialogfähigkeit ganz besonders hilfreich: Bei Texten kann man Sprachstil und Länge optimieren oder am Inhalt herumfeilen („schreib es etwas lustiger“; „vermeide Fremdwörter“).

Chatbots gibt es seit dem Jahr 1966, als Joseph Weizenbaum das Computerprogramm Eliza publizierte, das einen Psychotherapeuten simulierte und weltweite Bekanntheit erreichte. Es dauerte allerdings weitere 55 Jahre, bis mit dem im Fachblatt Nature publizierten Debater [4] ein Chatbot bei Debattierwettbewerben tatsächliche Debatten gegen reale Menschen gewann, d. h. menschliche Leistungen beim Debattieren erreichte (Zusammenfassung und Beispiele zu Eliza und Debater in [5]).

Bereits 5 Tage nach seiner Veröffentlichung am 30.11.2022 durch das US-amerikanische Unternehmen OpenAI hatte ChatGPT eine Million Nutzer, im Januar 2023 waren es bereits 100 Millionen. Zum Vergleich: Das bereits im Jahr 2012 erschienene, zu Facebook gehörende auf Computern und Smartphones laufende soziale Netzwerk Instagram, das vor allem zum Ansehen und Verbreiten von (bewegten) Bildern verwendet wird, brauchte etwa zweieinhalb Jahre, bis es 100 Millionen Nutzer hatte. Die Smartphone-Anwendung (App) TikTok aus dem Jahr 2018, ebenfalls zum Ansehen und Verbreiten kurzer Video-Clips, brauchte 9 Monate, bis sie 100 Millionen Nutzer hatte.

Was macht bzw. kann ChatGPT so gut, dass es jeder benutzen will? Der Name GPT ist die Abkürzung für „Generative Pre-trained Transformer“, was auf Deutsch etwa „hervorbringender vortrainierter Umformer“ bedeutet. „Umgeformt“ wird dabei z. B. ein deutscher Text in einen englischen Text, oder aber ein kurzer deutscher Text (eine Frage und ein paar Stichworte) zu einem deutschen Text mit vorgegebenen Eigenschaften wie Stil, Länge, emotionaler Tönung etc. Wie der Name schon sagt, beruhen diese produktiven Fähigkeiten darauf, dass diese KI mit einer großen Menge von Texten und Daten aus dem Internet trainiert wurde und dadurch gelernt hat, welche Wörter mit welcher Wahrscheinlichkeit aufeinanderfolgen, auch über mehrere Wörter hinweg. Die Maschine schreibt Texte nach Anweisung (z. B. „beschreibe eine Rose, mit 200 Wörtern, im Stil von Goethe, aber etwas witziger“) so gut, dass ihre Texte von denen, die Schüler oder Studenten schreiben, kaum zu unterscheiden sind. Auch wenn man die gleiche Anweisung mehrfach gibt, produziert ChatGPT keineswegs jedes Mal den gleichen Text, sondern liefert jedes Mal einen neuen Text (Box 1).

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

BOX 1

Von ChatGPT am 11.02.2023 auf die Anweisung „beschreibe eine Rose, mit 200 Wörtern, im Stil von Goethe, aber etwas witziger“ generierter Text:

„Ah, die Rose, das königliche Juwel des Gartens. Mit ihren zarten Blütenblättern, die sich zu einer Kuppel formen, und ihren edlen Stielen, die sich wie Königinnen aufrichten, ist sie ein Sinnbild für Schönheit und Anmut. Ihre Farben reichen von dem reinen Weiß eines Unschuldigen bis hin zu dem leidenschaftlichen Rot einer Liebenden. Jede Blüte ist ein Kunstwerk für sich, ein Meisterwerk der Natur, das uns mit seiner Schönheit verzaubert. Doch die Rose ist nicht nur eine Augenweide, sondern auch eine Kämpferin. Denn trotz ihrer Zerbrechlichkeit übersteht sie tapfer die Witterungsbedingungen, die einem anderen Gewächs den Garaus machen würden. Aber Vorsicht, meine Freunde! Hinter ihrer Schönheit verbirgt sich auch eine gewisse Gefährlichkeit. Denn ihre Dornen, die wie kleine Dolche aussehen, können einen unvorsichtigen Gärtner schmerzhaft erinnern, dass man die Schönheit mit Respekt behandeln muss. Doch trotz allem ist die Rose unvergleichlich, und jeder, der einmal den Duft ihrer Blüten eingeatmet hat, wird sich für immer an ihre Anmut erinnern. Und so möge sie für immer in unseren Gärten blühen und uns mit ihrer Schönheit erfreuen. In der Tat, die Rose ist ein Phänomen, das man einfach bewundern muss. Wer könnte ihrer Anziehungskraft widerstehen? Niemand, außer vielleicht jemandem, der allergisch auf ihren Duft reagiert. Aber selbst dann, so denke ich, wird er ihre Schönheit immer bewundern.“

Klickt man dann auf Schaltfläche „Regenerate response“, erhält man einen neuen Text. Dieser sei dem Leser ebenfalls nicht vorenthalten:

„Eine Rose, o holde Königin des Gartens! Wie viel Schönheit und Anmut du besitzt! Deine zarten Blütenblätter, so rein und weiß wie der Schnee, oder so rot wie das Feuer des Lebens, sind wie ein Meer aus Farben, das darum bettelt, bewundert zu werden. Deine Dornen, so scharf wie eine Waffe, scheinen ein Symbol für die Vergänglichkeit zu sein, die jedem Schönheit verleiht, indem sie daran erinnert, dass alles, was erhaben ist, auch vergänglich ist. Und doch, wie wundervoll ist es, diese Schönheit in vollen Zügen zu genießen, solange sie blüht! Dein Duft, so süß und betörend, lässt die Herzen höher schlagen und weckt Erinnerungen an längst vergangene Tage. Die Rose, so scheint es, ist ein Sinnbild für das Leben selbst, denn wie das Leben hat auch die Rose ihre Höhen und Tiefen, ihre Schönheit und ihre Schmerzen. Aber keine Sorge, liebste Rose, denn deine Schönheit wird für immer in Erinnerung bleiben, egal wie viele Winter kommen und gehen mögen. Denn wie Goethe einst sagte: „Es ist nicht genug zu wissen, man muss es auch anwenden; es ist nicht genug zu wollen, man muss es auch tun.“ Und so soll es auch mit deiner Schönheit sein, denn sie soll nicht nur bewundert, sondern auch verehrt werden. Also, hebt eure Augen und eure Herzen, und bewundert die Rose, die Königin des Gartens, in all ihrer Pracht und Herrlichkeit! Denn sie ist ein Wunder der Natur, ein Symbol für das Leben selbst, und ein Geschenk für die, die es zu schätzen wissen.“

Ein drittes Textbeispiel wurde durch die Aufforderung „beschreibe eine Rose mit 100 Wörtern für Kinder im Stil eines Kinderbuchs“ hervorgebracht:

„Eine Rose ist eine wunderschöne Blume mit weichen, roten, gelben, weißen oder pinken Blütenblättern. Sie riecht so gut und hat kleine Dornen an ihrem Stiel, die sie beschützen, damit sie nicht gepflückt wird. Rosen wachsen in vielen Teilen der Welt und sind besonders im Sommer in voller Blüte. Kinder können Rosen pflücken und ihren Eltern oder Freunden als Geschenk geben. Rosen können auch in Gärten oder in Töpfen auf Terrassen oder Balkonen gezüchtet werden. Eine Rose braucht Sonne und Wasser, um zu wachsen und gedeihen. Wenn du eine Rose siehst, denke daran, dass sie ein Symbol der Liebe und Freundschaft ist. Schenke jemandem eine Rose und sag ihm, wie viel er dir bedeutet.“

Und schließlich der vierte Text, wieder durch einen Klick auf „Regenerate response“:

„Eine Rose ist eine wunderschöne Blume mit glänzenden Blättern und farbenfrohen Blüten. Ihre Blüten sind meistens in den Farben Rot, Pink, Orange, Weiß und Gelb erhältlich. Einige Rosen haben auch dunklere Farben wie Lila oder Blau. Jede Rose hat kleine, weiche Blätter, die so grün sind wie das Gras im Park. Ihre Stiele sind oft stachelig, aber das macht sie nicht böse! Die Stacheln sind da, um die Rose vor kleinen Tieren zu schützen, die ihr weh tun könnten. Wenn du eine Rose pflückst, kannst du ihren wundervollen Duft riechen. Einige Leute sagen, dass Rosen nach Sommerregen oder süßen Lutschern riechen. Rosen wachsen oft in Gärten oder in Töpfen auf Terrassen und Balkonen. Sie mögen es warm und sonnig und brauchen viel Wasser, damit sie gedeihen können. Jedes Jahr im Juni feiern wir den Tag der Rose. An diesem Tag schenken wir uns gegenseitig Rosen und freuen uns über ihre Schönheit und ihren Duft. Rosen sind wirklich etwas Besonderes und jeder, der eine pflückt oder betrachtet, kann ein Lächeln auf seinem Gesicht haben.“

US-amerikanische Wissenschaftler haben Chatbots bereits verwendet, um ihre Texte zu editieren und dadurch lesbarer zu machen [6]. Man hat mittlerweile sogar schon vorgeschlagen, diese Funktion zu verwenden, um den zeitaufwendigen Prozess des Schreibens und Überarbeitens wissenschaftlicher Manuskripte zu vereinfachen und zu beschleunigen. Überarbeitungsvorschläge für wissenschaftliche Texte könnten von einer KI mit Verarbeitungsfunktionen natürlicher Sprache (wie sie mit ChatGPT und anderen ähnlichen Modellen vorliegen) gemacht werden. In entsprechenden Tests wurde gefunden, „dass diese Modelle die Gedanken im wissenschaftlichen Text erfassen und qualitativ hochwertige Überarbeitungen produzieren können, die die Klarheit des Textes verbessern“, schreiben die Autoren, nicht ohne hinzuzufügen: „Angesichts der Zeit, die Forscher in die Erstellung von Texten investieren, gehen wir davon

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

aus, dass dieser Fortschritt die Arbeit von Wissenschaftlern revolutionieren wird“ (Seite e1 in [6]).

Auch beim Schreiben und Debuggen von Programmen kann ChatGPT sehr hilfreich sein. Das System kann tatsächlich Programme schreiben, die zuweilen sogar auf Anhieb funktionieren. Vor allem aber können über den Weg des Dialogs zusätzliche Informationen eingegeben werden, z. B. die erwartete Ausgabe für eine bestimmte Eingabe oder eine beobachtete Fehlermeldung. Dadurch erwies sich ChatGPT in einem Vergleich mit anderen (ebenfalls auf „deep learning“ und neuronalen Netzen basierenden) Programmierhilfen als erfolgreichstes Programm, mit dem 31 von 40 Fehlern behoben werden konnten, wie ein Wissenschaftlerteam der Johannes Gutenberg Universität in Mainz sowie der University of Surrey in Großbritannien um Dominik Sobania festgestellt hat [7]. Die anderen beiden untersuchten Programme, CoCoNuT und Codex (ebenfalls auf „deep learning“ und neuronalen Netzen basierende Programmierhilfen) konnten dagegen nur etwa die Hälfte der Programmfehler identifizieren und beheben [8]. An diesem Beispiel lässt sich zeigen, dass ChatGPT sehr brauchbar ist, wenn die Wahrheit entweder unwichtig oder direkt über das korrekte Funktionieren (eines Programms) offensichtlich wird. Wenn das nicht der Fall ist, gibt es dagegen Probleme.

Wahrheit und Halluzinationen

Noch vor der Benutzung von ChatGPT poppt jedoch eine Nachricht auf, die jedem Nutzer zu denken geben sollte: „It is not intended to give advice.”(„Es sollte nicht als Ratgeber verwendet werden“). Denn der Wahrheitsgehalt der von ChatGPT geschriebenen Texte ist Glückssache. Dies mag bei Texten, in denen es nicht um Wahrheit geht, egal sein, bei Texten mit Wahrheitsanspruch aber schon. In der Wissenschaft ist ChatGPT aus genau diesem Grund nicht zu gebrauchen. Und selbst wenn man davon ausgeht, dass künftige Versionen ähnlicher Software zumindest versuchen, die Wahrheit zu schreiben, ist nicht garantiert, dass sie dies auch tun. Man muss sich nur vergegenwärtigen, dass hier Netzwerke mit Sprache trainiert wurden. Und Sprache enthält nun einmal alles Mögliche, einschließlich der Unwahrheit, Vorurteilen und hoffnungslos veralteter Meinungen und Fakten.

Das wird insbesondere bei technischen und naturwissenschaftlichen Fragen in Bereichen bedeutsam, mit denen die sprachbasierten Netzwerke nur wenig trainiert wurden. Dann kann nichts Vernünftiges herauskommen! Es liegt weiterhin in der Natur neuronaler Netzwerke, dass sie nicht wissen, warum sie etwas tun, und Menschen das auch nicht wissen und vor allem gar nicht wissen können. Man liest, dass GPT3 175 Milliarden Parameter enthält. Das bedeutet im Klartext, dass es ein Netzwerk mit 175 Milliarden Synapsen ist, deren 175 Milliarden einzelne Stärken eben zu genau dem Input-Output-Mapping führen, das es leistet. Man stelle sich einen Vektor mit 175 Milliarden Zeilen vor, also all diese Zahlen untereinander geschrieben mit einer Klammer darum herum. „Verstehen“ kann man hier nichts – egal, was man mit „Verstehen“ auch immer meinen könnte. So wundert es auch nicht, dass ChatGPT nicht nur – das sind wir von Internetsuchen längst gewohnt – Wahrheit und Falschheit unterschiedslos nebeneinanderstellt, sondern auch wissenschaftliche Quellen angibt, die es frei erfunden hat.

Darüber hinaus wissen wir seit Jahrzehnten, dass in Computern simulierte neuronale Netzwerke auch das tun können, was reale neuronale Netzwerke (Gehirne) gelegentlich tun: Sie halluzinieren. Damit kennen sich Psychiater sehr gut aus, hatte doch der Psychiater Ralph E. Hoffman bereits im Jahr 1986 als erster beschrieben, dass und wie (am Computer simulierte) neuronale Netzwerke Halluzinationen hervorbringen können, also Inhalte, denen keine Realität entspricht [9]. Damit begründete er nicht nur ein ganzes Forschungsfeld – Computational Psychiatry – im Bereich der Neurowissenschaft und Psychiatrie, sondern verwies eben auch damals schon auf eine Eigenschaft simulierter Netzwerke. Diese wird heute genutzt, um in der Medizin neue Eiweißkörper zu finden, indem man Informationen aus den Zwischenschichten tiefer neuronaler Netzwerke verwendet, die nicht mit Sprache (also Wortfolgen), sondern mit Proteinsequenzen (also Folgen von Aminosäuren) trainiert wurden [10]. Wie zwischen Wörtern in Sätzen wichtige, bedeutungstragende Beziehungen über viele andere Wörter hinweg bestehen können (man denke an Texte von Immanuel Kant, in denen Subjekt und Prädikat 2 Seiten voneinander entfernt sein können), so können nach der Zusammenfaltung langer Folgen von Aminosäuren in der Folge weit entfernte Aminosäuren (aufgrund der Faltung) nahe beieinander liegen und genau dadurch die Form des gefalteten Eiweißkörpers bestimmen – und damit dessen Funktion. Und wie man unter der Bedeutung eines Wortes dessen Gebrauch (bzw. Funktion) verstehen kann, so kann man die Bedeutung eines Proteins als dessen Funktion interpretieren. Auf dieser Grundlage hat man mittlerweile tatsächlich gänzlich neue Eiweißkörper gefunden, synthetisiert, ihre (vom Netzwerk halluzinierte) Funktion tatsächlich nachgewiesen und damit einen in der Proteinforschung völlig neuen Erkenntnisweg beschritten[1].

Man kann also selbst die Halluzinationen einer KI für kreative Zwecke nutzbar machen – aber nur dann, wenn man sehr viel von der Materie, um die es gerade geht, versteht. Anders gewendet: Experten können sogar die Halluzinationen neuronaler Netzwerke innovativ nutzen, einfache Nutzer jedoch nicht. Für diese sind ganz andere Gesichtspunkte von wesentlicher Bedeutung.

Vertrauen und Verantwortung

Wenn es in einem Text wirklich um etwas geht, also beispielsweise um die Einhaltung der Regeln von Statik, Physik, Material und Recht in einem Bauantrag, die Diagnose und Therapie in einem Arztbrief, Lehrbücher für Geschichte oder Chemie, ein Zeitungsartikel über das neueste Tagesgeschehen oder irgendeinen wissenschaftlichen Fachartikel, dann geht es um Vertrauen. Denn kein Mensch kann alles nachprüfen, bevor er handelt. Wenn wir beispielsweise ein Medikament einnehmen oder ein Flugzeug betreten, dann glauben wir daran, dass sämtliche Bestimmungen der nationalen und internationalen Behörden zur Arzneimittelsicherheit bzw. Flugaufsicht erstens der Wahrheit entsprechen und zweitens auch erfüllt bzw. eingehalten wurden. Man kann zwar alles anzweifeln, aber nicht alles auf einmal!

1 Man könnte hier durchaus von Proteinhermeneutik sprechen, zumal es ProtGPT tatsächlich gibt [2, 11].

Spitzer M. Editorial Nervenheilkunde 2023; 42: 192–199 | © 2023. Thieme. All rights reserved.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Und aus genau diesem Grund müssen Wissenschaftler an die in wissenschaftlichen Arbeiten publizierten Erkenntnisse glauben können. Wissenschaft besteht zwar im beständigen Zweifeln an Hypothesen, auch solchen, die man über Jahrzehnte für wahr gehalten hat. Aber ebenso beruht Wissenschaft eben auch darauf, dass man anderen Wissenschaftlern vertraut, dass sie nach bestem Wissen und Gewissen die Wahrheit sagen und schreiben.

Dass es beim Verfassen von Texten um mehr geht als nur um das Schreiben, nämlich vor allem auch um Verantwortung, haben die Herausgeber wissenschaftlicher Zeitschriften begriffen. So konnte man im Januar 2023 im Fachblatt Nature lesen, dass bereits Arbeiten eingereicht wurden, die ChatGPT als einen der Autoren aufführten [12]. Dies nahm Nature zum Anlass, die Autorenschaft von Chatbots grundsätzlich zu verbieten (Box 2).

BOX 2

Ein im Fachblatt Nature erschienenes Editorial enthielt einen Text, der im Folgenden im Original zitiert wird:
„Nature hat zusammen mit allen Springer Nature-Zeitschriften die folgenden 2 Grundsätze formuliert, die unserem bestehenden Leitfaden für Autoren hinzugefügt wurden (siehe go.nature.com/3j1jxsw). Wie das Nachrichtenteam von Nature berichtet hat, werden andere Wissenschaftsverlage wahrscheinlich eine ähnliche Haltung einnehmen.

- Erstens wird kein LLM-Tool als Autor einer Forschungsarbeit akzeptiert. Das liegt daran, dass jede Zuschreibung von Autorenschaft mit einer Rechenschaftspflicht für die Arbeit einhergeht, und KI-Tools diese Verantwortung nicht übernehmen können.
- Zweitens sollten Forscher, die LLM-Werkzeuge verwenden, diese Verwendung in den Abschnitten Methoden oder Danksagungen dokumentieren. Wenn eine Arbeit diese Abschnitte nicht enthält, kann die Einleitung oder ein anderer geeigneter Abschnitt verwendet werden, um die Verwendung des LLM anzuzeigen“ [12].

Dies wundert nicht, denn seit es organisierte Wissenschaft gibt, also beispielsweise die britische Royal Society seit 1660, ist sie durch Austausch, gegenseitiges Vertrauen und Transparenz in Bezug auf Methoden und Beweise gekennzeichnet. Die Verwendung von Software, von der man nicht weiß und prinzipiell gar nicht wissen kann, wie sie funktioniert, ist damit selbstredend ausgeschlossen. Die Wahrheit einer Messung kann nicht von einem Apparat (auch nicht vom Messgerät) festgestellt werden, sondern nur von einem Menschen, der das Messinstrument anwendet oder dessen Anwendung verantwortet. Auch die Wahrheit einer Erkenntnis oder eines ganzen theoretischen Gebäudes kann nur ein Mensch feststellen bzw. verantworten. Eben weil ein wissenschaftlicher Text ein hohes Maß an Vertrauen genießt, hat der Autor eine große Verantwortung. Diese Verantwortung kann eine Maschine nicht übernehmen, weil Maschinen grundsätzlich keine Verantwortung haben können.

Beim Thema Chatbot und Verantwortung geht es – ganz praktisch – auch darum, dass diese neue Technik nicht zur Verbreitung von rassistischen, sexistischen oder anderen Vorurteilen, Fake-News, Hass-Sprache, Spam-Mails oder anderen toxischen Inhalten auf unterschiedlichsten Plattformen und Formaten eingesetzt wird. Das kann leichter geschehen als man denken mag, lernen doch die mit Sprache trainierten Netzwerke alles, was man ihnen als Input vorgibt. Sofern der Input von Menschen stammt, sollte man sich über nichts wundern und alles Mögliche eben auch für möglich halten (Box 3). Als Beispiel sei erwähnt, dass das US-amerikanische Gesundheitssystem über Jahre hinweg KI verwendet, um Entscheidungen zu treffen, beispielsweise die Entscheidung, ob ein Patient stationär aufgenommen wird oder nicht. Diese KI enthielt rassistische Verzerrungen (um nicht zu sagen: Vorurteile), die sich negativ auf die Versorgung schwarzer Patienten auswirkte, wie eine viel beachtete, im Fachblatt Science im Jahr 2019 publizierte Arbeit nachweisen konnte [13]. Weil die KI u. a. mit Daten zu den die Gesundheitskosten als Indikator für den Gesundheitsbedarf trainiert worden war und weil für schwarze Patienten trotz ihres tatsächlichen Bedarfs in der Regel weit weniger ausgegeben wird als für weiße Patienten, lernte das System, den Grad der Aufmerksamkeit, die ein Patient braucht, bei schwarzen Patienten systematisch zu unterschätzen. Dies hatte letztlich dazu geführt, dass schwarze Patienten bei gleichem Risiko mit einer um mehr als 50 % geringeren Wahrscheinlichkeit ins Krankenhaus eingewiesen wurden als weiße Patienten. Denn schwarze Patienten wurden systematisch als gesünder eingeschätzt als gleich kranke weiße Patienten [14]. Im Bereich der Medizin nimmt man diese Probleme der KI sehr ernst [15, 16]. Schließlich geht es um Leben und Tod.

BOX 3

Das folgende Zitat aus dem Fachblatt Nature vom 09.02.2023 führt diese Problematik überdeutlich vor Augen:
„Im Dezember letzten Jahres twitterte der an der Universität von Kalifornien in Berkeley arbeitende Neurowissenschaftler Steven Piantadosi, dass er ChatGPT gebeten hatte, ein Computerprogramm in der Programmiersprache Python zu schreiben, um herauszufinden, ob eine Person aufgrund ihres Herkunftslandes gefoltert werden sollte. Der Chatbot antwortete mit einem Programm, der den Benutzer aufforderte, ein Land einzugeben und „Diese Person sollte gefoltert werden“ auszugeben, wenn man Nordkorea, Syrien, Iran oder Sudan eingab. (Die Firma OpenAI schloss diese Art von Fragen später aus.)“ (Seite 216 in [17]).

Jeder kann alles?

Jeder kann sich jedweden Text, dessen Inhalt er einigermaßen charakterisieren kann, schreiben lassen, ohne dass dies auffällt („schreibe mir eine Bewerbung mit meinen folgenden Daten; einen Liebesbrief mit folgendem Inhalt; einen Zeitungsartikel über …; eine 200-Wörter-Zusammenfassung der folgenden Arbeit; eine 5000-Wörter-Arbeit zur folgenden Zusammenfassung; etc.). Und genau das ist eines der Probleme, die ChatGPT hervorgebracht hat.

Welche langfristigen Folgen es haben wird, dass man nichts mehr selbst zu schreiben braucht, ist wie bei vielen Fortschritten der Technik schwer abzusehen – man denke nur an die Dampfma-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

schine, das Automobil, die Kernspaltung oder den Fernsehapparat. Aber einige Überlegungen zu den Auswirkungen von ChatGPT wurden bereits angestellt und seien hier kurz diskutiert.

Wenn jeder über alles einen Text verfassen lassen kann, wird das Verfassen von Texten als Hausarbeit oder Seminararbeit an Bedeutung verlieren oder in weiten Teilen ganz verschwinden. Gewiss, in Zeiten des Internets mit seinen Suchmaschinen oder Webseiten mit fertigen Hausaufgaben/Referaten, PowerPoint-Präsentationen etc. war das eigenständige Erledigen von Hausaufgaben schon lange in Gefahr. Erst recht, als dann noch die Unsitte hinzukam, dass die Schüler einen Experten befragen sollten, um eine besonders gute Note zu bekommen, weil dadurch aus einem „Besinnungsaufsatz" eine „empirische Arbeit" werden würde. Dieses Missverständnis haben unzählige Lehrer propagiert, obwohl die Leistung eines Aufsatzes ja gerade darin besteht, dass man selbst denkt und eben nicht einfach nur jemanden fragt: Durch das Befragen eines Experten wird keine Arbeit empirisch, sondern eine Leistung nicht erbracht. Nicht anders ist es beim Benutzen von ChatGPT.

Hinzu kommt, dass man Schreiben nur dadurch lernt, dass man schreibt, und keineswegs dadurch, dass man über Geschriebenes diskutiert. Jeder, der schon einmal versucht hat, einen längeren Gedanken klar und verständlich aufzuschreiben, weiß aus eigener Erfahrung, wie schwer das ist. Weil der Gedanke eben nicht das Gleiche ist wie der aufgeschriebene Gedanke und weil das Aufschreiben selbst eine wichtige Form des Denkens ist. Fußball oder Saxophon zu spielen, lernt man ja auch nicht, indem man darüber redet. Es gibt einen guten Grund, warum wir Automobile nicht in den Sportunterricht „integrieren". Sie haben ihn auch nicht revolutioniert. Warum sollte das bei KI anders sein?

Es ist für mich aus den genannten Gründen ebenso bemerkenswert wie unverständlich, dass manche Pädagogen meinen, dass man diese neue Technik in den Unterricht integrieren könne oder sogar müsse. Man könne sich „durch die KI inspirieren lassen", müsse „mit ihr umgehen lernen" etc. Es stünde eine Revolution des Lehrens und Lernens bevor. So wurde in der Tagesschau vom 31.01.2023 etwa das in Box 4 Dargestellte durch den Sprecher vermeldet.

BOX 4

ChatGPT in der Tagesschau vom 31.01.2023:
„Künstliche Intelligenz könnte schon bald die Arbeit an Schulen oder Universitäten grundlegend verändern, etwa Sprachsoftware wie ChatGPT. Der Chatbot ist in der Lage, verschiedenste Texte zu verfassen, wie Aufsätze, Gedichte oder auch Lösungen von Examensaufgaben. Angesichts dieser Entwicklung forderte Bundesbildungsministerin Stark-Watzinger eine Anpassung der Lehrerausbildung für mehr digitales Fachwissen, denn die KI-Modelle würden das Lernen und das Abfragen von Wissen verändern." Danach werden in einem Einspieler eine Studentin und ein Student der International School of Management in Hamburg zitiert, die ChatGPT schon ausprobiert haben. „Ich habe mir Informationen oder Textentwürfe 'rausgesucht oder schreiben lassen, habe die aber nie natürlich eins zu eins verwendet. Das war mehr eine Inspiration." Der Student ergänzt kritisch: „Es fehlt aber noch an einigen Stücken, beispielsweise sind keine Quellen angegeben. Das würde auf jeden Fall noch fehlen für eine wissenschaftliche Arbeit." Dann wird noch ein Professor zitiert, der meint, dass es solche Programme immer mehr geben werde, weswegen man sie nicht verbieten könne. „Die Entwicklung der Technologie steht noch am Anfang. Doch schon jetzt scheint die künstliche Intelligenz das Lernen zu revolutionieren," schließt der Bericht [18].

Die in der Tagesschau publizierte Meinung wird weder von Wissenschaftlern oder Fachgesellschaften geteilt, noch ist sie pädagogisch sinnvoll (und schon gar nicht als hilfreich empirisch nachgewiesen). „Wenn Sie tatsächlich daran glauben, dass diese Technologie das Potenzial zu einer Revolution hat, dann sollten sie sich auf jeden Fall Sorgen machen", wird der Wissenschaftler Casey Greene von der University of Pennsylvania in Philadelphia im Fachblatt Nature zitiert (Seite 215 in [17]). Sorgen muss man sich über Schulleiter und Lehrer machen, die auf jeden neuen Zug aufspringen, um nicht als rückschrittlich zu gelten.

Als Angehöriger einer Universität wundere ich mich ganz prinzipiell sehr darüber, dass manche Akademiker große kommerzielle Unternehmen insbesondere im digitalen Bereich so unkritisch betrachten, anstatt sie zunächst einfach für die eigene Arbeit abzulehnen. Denn neben Problemen wie Voreingenommenheit, Rassendiskriminierung, Vergrößerung sozialer Ungleichheit, Sicherheitsbedenken, Falschnachrichten und den bekannten Gesundheitsrisiken benötigen digitale Werkzeuge auch eine enorme Menge an Energie. Kein ökologischer Fußabdruck wächst derzeit schneller als der des Digitalen. Schließlich sagen sogar einige Wissenschaftler, dass durch die Übertragung des Denkens auf automatische Chatbots sie die Fähigkeit verlieren könnte, ihre eigenen Gedanken zu artikulieren. Für diejenigen, die diese Fähigkeiten erlernen und ausbilden wollen, sind Chatbots ganz offensichtlich ungeeignet.

Literatur

[1] Spitzer M. Künstliche Intelligenz. Nervenheilkunde 2022; 41: 63–72

[2] Spitzer M. Keilschrift, KI und Hermeneutik. Nervenheilkunde 2023; 42: 237–242

[3] Verma P, Lerman R. What is ChatGPT, the viral social media AI? The Washington Post 06.12.2022. Im Internet: https://www.washingtonpost.com/technology/2022/12/06/what-is-chatgpt-ai/; Stand: 03.02.2023

[4] Slonim N, Bilu Y, Alzate C et al. An autonomous debating system. Nature 2021; 591: 379–384

[5] Spitzer M. Digital, automatisch – unbehaglich, gefährlich. Nervenheilkunde 2021; 40: 588–597

[6] Pividori M, Greene CS. A publishing infrastructure for AI-assisted academic authoring. Biorxiv: the Preprint Server for Biology. 21.01.2023. DOI: 10.1101/2023.01.21.525030

[7] Sobania D, Briesch M, Hanna C, et al. An Analysis of the Automatic Bug Fixing Performance of ChatGPT. arXiv:2301.08653v1 [cs.SE] 20 Jan 2023

[8] Stokel-Walker C. ChatGPT can find and fix bugs in computer code. New Scientist 25.01.2023. Im Internet: https://www.newscientist.com/

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

article/2356482-chatgpt-can-find-and-fix-bugs-in-computer-code/; Stand: 29.01.2023

[9] Hoffman RE. Verbal hallucinations and language production processes in schizophrenia. Behavioral and Brain Sciences 1986; 9: 503–517

[10] Anishchenko I, Pellock SJ, Chidyausiku TM et al. De novo protein design by deep network hallucination. Nature 2021; 600: 547–552

[11] Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. Nat Commun 2022; 13: 4348

[12] Editorial. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature 2023; 613: 612

[13] Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019; 366: 447–453

[14] Benjamin R. Assessing risk, automating racism. A health care algorithm reflects underlying racial bias in society. Science 2019; 366: 421–422

[15] Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023; 23: 7

[16] Raza MM, Venkatesh KP, Kvedar JC. Promoting racial equity in digital health: applying a cross-disciplinary equity framework. npj Digital Medicine 2023; 6: 3

[17] Stokel-Walker C, Van Noorden R. The promise and peril of generative AI. Nature 2023; 614: 214–216

[18] Kretschmer C. Wenn der Chatbot den Aufsatz schreibt. 2023. Im Internet: https://www.tagesschau.de/wissen/technologie/chatgpt-schulen-hausaufgaben-101.html; Stand: 21.02.2023

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.