

Disruption, Evolution oder viel heiße Luft – was ist dran an „künstlicher Intelligenz“?

Stefan Oppl

Als am 30. November 2022 ChatGPT online der weltweiten Öffentlichkeit zugänglich gemacht wurde, war noch nicht abzusehen, welche großen Wellen in allen Bereichen der Gesellschaft diese Veröffentlichung auslösen würde. Nach fünf Tagen hatten sich über eine Million Benutzerinnen und Benutzer angemeldet, nach nicht einmal zwei Monaten wurde die 100-Millionen-Marke erreicht. ChatGPT hielt damit den Rekord für die am schnellsten wachsende Internet-Anwendung.

Diese rasche Verbreitung führte auch zu einem breiten öffentlichen Diskurs über die Potentiale und Risiken jener Technologien, die oft unter dem Begriff „Künstliche Intelligenz“ zusammengefasst werden. Dieser Diskurs ist bis heute nicht abgeflaut und oft von einem gewissen Unbehagen und von Unsicherheit geprägt. „Künstliche Intelligenz“ wird als etwas Vages, nicht Greifbares, aber doch im weitesten Sinn wie ein Mensch handelndes Etwas wahrgenommen – „die KI“ kann auf unseren Wunsch hin Texte schreiben, Bilder malen oder ein Kochrezept empfehlen und dabei noch sehr eloquent mit uns kommunizieren. Gleichzeitig bleibt unklar, wie „die KI“ zu ihren Fähigkeiten kommt und wo ihre Grenzen liegen ... und dadurch ist auch schwer einschätzbar, ob daraus uns als Menschen eine Konkurrenz in unserer beruflichen Tätigkeit, unserer Kreativität und unserem sozialen Miteinander erwächst.

Dieser Unsicherheit und diesem Unwohlsein entgegenzuwirken und in der Breite der Bevölkerung eine realistische Einschätzung der Möglichkeiten und Grenzen von künstlicher Intelligenz zu verankern, ist eine der größten Bildungsaufgaben der nächsten Jahre, die in der ganzen Gesellschaft und insbesondere auch in den Schulen adressiert werden muss. Dieser Text soll eine erste Annäherung an diese Herausforderung ermöglichen und Anknüpfungspunkte für die weitere Beschäftigung mit dem Thema Künstliche Intelligenz im eigenen Handlungsfeld bieten.

Künstliche(?) Intelligenz(?)

Der Begriff „Künstliche Intelligenz“ wurde bereits vor fast 70 Jahren, im Jahr 1955, in einem Antrag für ein Forschungsprojekt geprägt. Der Anspruch war damals wie heute, mithilfe von Informationstechnologie ein System zu erstellen, das „von außen“ betrachtet menschenähnliche Verhaltensweisen zeigt. Die dazu gewählten Herangehensweisen unterschieden sich über die Jahrzehnte sowohl in ihren technischen Grundlagen als auch im jeweiligen Anspruch, wie weitgehend die Imitation von menschlichem Verhalten gehen sollte. Bereits in den 1960er-Jahren gelangen rasche Fortschritte im Bereich der sogenannten „Expertensysteme“, in denen das Wissen von Expertinnen und Experten über bestimmte Zusammenhänge in spezifischen Bereichen in Form von Regeln gespeichert wurden - etwa im Bereich der medizinischen Diagnose das Wissen über das Auftreten von bestimmten Symptomen im Zusammenhang mit gewissen Krankheiten. Diese Regeln werden in der Folge durch ein Computerprogramm verwendet, um Anfragen von Benutzerinnen und Benutzern zu beantworten - etwa eben nach möglichen Krankheiten, wenn eine bestimmte Kombination von Symptomen beobachtet wurde. Derartige Systeme können tatsächlich in ihrem jeweiligen Anwendungsbereich

menschenähnliches Verhalten zeigen und bestimmte Aufgaben effektiver unterstützen, als Menschen das tun könnten. Fragen von außerhalb des Anwendungsbereichs können solche Systeme aber nicht beantworten und nicht einmal interpretieren. Eine Vision, oder - je nach Sichtweise - ein Schreckensszenario, der Forschung im Bereich der künstlichen Intelligenz ist die Erstellung einer „künstlichen allgemeinen Intelligenz“, die die Fähigkeit besitzt, jede beliebige Aufgabe zu verstehen und deren Bearbeitung zu erlernen, die auch ein Mensch ausführen oder erlernen könnte. Derartige Systeme existieren heute nicht - ob sie jemals erzeugt werden können und ob der Zeithorizont einige Jahre oder Jahrzehnte in der Zukunft liegt, ist umstritten.

Im Bereich der aufgabenspezifischen künstlichen Intelligenz wurden jedoch über die letzten Jahrzehnte große Fortschritte gemacht, so dass heute Systeme, wie das oben erwähnte ChatGPT, existieren, die zumindest in Teilbereichen bereits sehr weitgehend menschenähnliche Fähigkeiten imitieren können. Wesentliche Schritte dahin fanden ab den 1990er-Jahren statt, als mit sogenannten „Neuronalen Netzen“ eine neue technische Grundlage für die Entwicklung von künstlicher Intelligenz gelegt wurde. Neuronale Netze sind Computerprogramme, deren Struktur dem menschlichen Gehirn nachempfunden ist. Im Gegensatz zu den Expertensystemen, die auf Basis von transparenten und nachvollziehbaren Regeln operieren, werden neuronale Netze mit Hilfe von vorgefertigten Datensätzen für ihre jeweilige Aufgabe „trainiert“ - sie „lernen“ aus diesen Trainingsdaten, welche Muster in den Eingabedaten zu welcher Ausgabe führen soll.

So könnte etwa ein neuronales Netz darauf trainiert werden zu erkennen, ob ein Bild eine Katze zeigt. Zu diesem Zweck würde man das neuronale Netz mit einer großen Anzahl an Bildern speisen, auf denen Katzen zu sehen sind und ihm dabei mitteilen, dass die erwartete Ausgabe für die Bilder der Text „Es handelt sich um eine Katze“ ist. Daneben würde man eine ebenso hohe Anzahl von Bildern, die keine Katzen zeigen, einspeisen und dort als erwartete Ausgabe „Es handelt sich um keine Katze“ angeben. Wenn dies mit einer ausreichend hohen Anzahl an Bildern (üblicherweise mehrere hunderte bis tausende) durchgeführt wird, wird das neuronale Netz „gelernt“ haben, eine Katze auf einem Bild zu erkennen. Wenn es nun mit einem bislang unbekanntem Bild konfrontiert wird, wird es eine Aussage darüber treffen können, ob auf diesem Bild eine Katze zu sehen ist oder nicht. Es kann jedoch keine anderen Tiere unterscheiden und wird - je nachdem, wie die Trainingsdaten ausgestaltet waren - vielleicht auch ähnlich aussehende Bildmuster (wie etwa kleine Hunde) als Katzen identifizieren. Für eine exaktere Identifikation oder eine Unterscheidung mehrerer Tierarten müsste man das neuronale Netz mit entsprechenden Datensätzen anders und neu trainieren.

An dieser Stelle ist es wichtig, sich zwei Grundeigenschaften dieser Form von „künstlicher Intelligenz“ bewusst zu machen: Ein neuronales Netz kann zum einen Aussagen über neue Eingaben immer nur auf Basis jener Daten ableiten, mit denen es trainiert wurde - wenn in diesen Daten fehlerhafte Zuordnungen oder Lücken enthalten waren, wird das neuronale Netz diese Fehler und blinden Flecken fortschreiben. Zum anderen ist für niemanden, auch nicht für die Erstellerinnen und Ersteller des neuronalen Netzes nachvollziehbar, wie dieses bei bestimmten Eingabedaten (etwa ein Bild eines Blumenstocks) zu einer Aussage über diese Daten (ob etwa das Bild von vornhin eine Katze zeigt) kommt. Das Ergebniss des „Lernprozesses“ auf Basis der Trainingsdaten bildet sich im neuronalen Netz in Form einer riesig großen Anzahl an numerischen Parametern ab,

die für sich genommen einzeln nicht interpretiert werden können - erst aus deren Zusammenspiel im neuronalen Netz ergibt sich letztlich die Funktion. Dies ist ein fundamentaler Unterschied zu den "Expertensystemen", deren Regelbasis überprüft und – wenn notwendig – nachvollziehbar angepasst werden kann.

Die möglichen Verzerrungen und Lücken in den Fähigkeiten eines neuronalen Netzes und das Faktum, dass nicht direkt überprüfbar ist, ob solche vorhanden sind, sind Eigenschaften dieser Art von „künstlicher Intelligenz“, deren man sich für die kompetente Verwendung zumindest bewusst sein muss - der Einsatz einer „künstlichen Intelligenz“, deren Trainingsdaten unbekannt oder nicht nachvollziehbar sind für kritische Einsatzzwecke (wie etwa im Bereich der medizinischen Diagnose) ist fragwürdig, wenn nicht sogar fahrlässig.

Sprich mit mir

Es handelt sich bei der eben beschriebenen Problematik um kein Nischenthema mehr - auch die "künstlichen Intelligenzen", die Anwendungen wie ChatGPT zugrunde liegen, beruhen auf den gleichen Prinzipien. Diese Art von „künstlicher Intelligenz“ wird als „großes Sprachmodell“ („large language model“) bezeichnet. Große Sprachmodelle werden - wie im Namen schon abgebildet - mit riesig großen Datenmengen trainiert. Der dazu notwendige Rechen- und Ressourcenaufwand ist enorm, als Quelle für die Trainingsdaten wird üblicherweise das World Wide Web herangezogen - alles, was auf Webseiten als Text oder auch in anderen Medienformaten veröffentlicht ist, fließt so in den Trainingsprozess ein. Daraus ergibt sich eine erstaunliche Eigenschaft: das große Sprachmodell "lernt" aus diesen unglaublich großen Datenmengen, wie Sprache grundsätzlich funktioniert und kann auch unterschiedliche konkrete Sprachen, die in den Trainingsdaten in ausreichend großem Umfang enthalten sind, reproduzieren - deshalb kann ChatGPT nicht nur auf Englisch, Deutsch oder Spanisch und auch in anderen Sprachen kommunizieren, sondern etwa auch Programmcode in den Programmiersprachen Python oder Javascript schreiben. Gleichzeitig führt diese Art von Trainingsprozess auf einer inhaltlichen Ebene aber wieder zur gleichen Problematik wie oben beschrieben: wenn in den Trainingsdaten problematische Inhalte enthalten sind (und diese sind im World Wide Web zuhauf vorhanden), wird auch das große Sprachmodell diese problematischen Inhalte reproduzieren. Dort, wo in den Trainingsdaten Lücken vorhanden sind, wird das große Sprachmodell einen sehr eloquent formulierten Text produzieren (es hat ja gelernt, wie Sprache funktioniert), der vermeintlich wie eine Antwort auf die gestellt Anfrage aussieht, mit der Realität aber nichts zu tun hat - man spricht hier von „Halluzinationen“, für die große Sprachmodelle anfällig sind. Die großen Anbieter aktueller Systeme wie ChatGPT betreiben großen Aufwand in manueller Nachkorrektur des Verhaltens ihrer Systeme, um derartige problematische Antworten oder Halluzinationen zu vermeiden - letztlich ändern diese Korrekturmaßnahmen aber nichts an der grundlegenden Funktionsweise von großen Sprachmodellen.

Trotz dieser Einschränkungen bergen große Sprachmodelle enormes Potential in der Art, wie wir als Menschen mit Computern interagieren können und wie wir bei der Aufgabenerfüllung unterstützt werden können. Durch die Möglichkeit, einem Computer in natürlicher Sprache ein Anweisung zu geben, und dann schrittweise in einer natürlich

ablaufenden Interaktion mit dem System etwaige Unklarheiten auszuräumen oder Nachfragen zu beantworten, revolutioniert die Art und Weise, wie Computer im täglichen Leben genutzt werden können. Die Nutzbarkeit und Zugänglichkeit neuer, mächtiger und damit oft komplexe IT-Systeme im Alltag und im Beruf kann so massiv gesteigert werden. Personengruppen, denen durch die Digitalisierung in ihrem Alltag bislang eine Anpassung ihrer gewohnten Abläufe abverlangt wurde (man denke nur an den Kauf eines Bahntickets am Automaten und die damit verbundenen Hürden für digital nicht affine Personen), können so die Vorteile dieser neuen Systeme verfügbar gemacht werden. Auch neue Möglichkeiten der Digitalisierung im Berufsleben, die oft den komplexen Umgang mit großen Datenbeständen und deren Auswertung und Interpretation erfordern, können durch die neuen Möglichkeiten der großen Sprachmodelle auch für Personen ohne technische Spezialausbildung zugänglich gemacht werden.

Die Hebung dieser Potentiale steckt noch in den Kinderschuhen - sie werden jedoch schnell an Verbreitung gewinnen und Einzug in den Alltag finden. Erste Eindrücke von den Möglichkeiten zeigen sich, wenn man in aktuellen Versionen von ChatGPT nicht nur mit dem System chatten kann, sondern es auch auffordern kann, einen bereitgestellten Datensatz auszuwerten und zu interpretieren oder im Web auf die Suche nach aktuellen Informationen zu einer bestimmten Frage zu gehen. Wenn sich „die KI“ als Vermittlerin zwischen uns Menschen und anderen Computersystemen etablieren kann und in unserem Auftrag Werkzeuge (wie eine Websuche oder ein Ticketbuchungssystem) benutzen kann, können viele bislang aufwändige Aufgaben rascher, zugänglicher und weniger anstrengend erfüllt werden.

Auch die Nutzung von großen Sprachmodellen als Assistenzsysteme in ihre ureigenen Domäne - dem Umgang mit Sprache - birgt großes Potential. Die Fähigkeiten von großen Sprachmodellen in diesem Bereich machen deren Einsatz als Unterstützer beim Schreiben und Editieren von Texten zu einem naheliegenden Anwendungsszenario - sie sind letztlich ein logischer Evolutionsschritt zu den Funktionen zur Rechtschreib- und Grammatikprüfung, die wir aus unseren Textverarbeitungsprogrammen kennen. In diesem Bereich wird „die KI“ nicht zur kreativen Texterstellung, sondern zur Prüfung und Generierung von Verbesserungsvorschlägen eingesetzt - die Kreativität liegt bei den Autorinnen und Autoren. Die anderen Aufgaben im Schreibprozess, die eher mühsam und mechanistisch sind, können effektiv unterstützt werden. Auch hier gilt: die ersten Systeme dieser Art sind bereits verfügbar, deren weite Verbreitung liegt durch die bereits angekündigte Integration in die Textverarbeitungssysteme der großen Softwareanbieter in naher Zukunft.

Alles gut?

Die Nutzung großer Sprachmodelle als Vermittler zu anderen IT-Werkzeugen und als Assistenzsysteme in der Erstellung von Texten (und auch anderer Medienarten - trotz ihres Namens können große Sprachmodelle durchaus auch für den Umgang mit visuellen und auditiven Medien verwendet werden) birgt also bereits kurzfristig große Potentiale in der Unterstützung von bislang mühsamen oder aufwändigen Aufgaben. Ein weiteres Anwendungsfeld, das von den anderen Einsatzgebieten abzugrenzen ist, ist die Nutzung dieser Systeme als Kreativwerkzeug - also etwa zur Erstellung von Texten oder Bildern auf Basis von in natürlicher Sprache formulierter Anforderungen. Dieser Bereich ist

differenzierter zu betrachten, weil hier die negativen Eigenschaften großer Sprachmodelle - also deren Tendenz zum „Halluzinieren“ oder der Reproduktion von problematischen Inhalten - viel stärker und umfassender zum Tragen kommen kann als in den anderen Anwendungsfeldern.

Die Funktion eines großen Sprachmodell basiert vereinfacht gesagt letztlich darauf, auf Basis der zur Verfügung gestellten Eingabe und des bislang generierten Textes das jeweils wahrscheinlichste nächste Wort vorherzusagen - so ergibt sich überraschenderweise ein üblicherweise sinnvoll klingender Text. Die heute verfügbaren System unterscheiden sich darin, wie lang der Eingabetext sein kann, der in der Generierung der Ausgabe berücksichtigt wird und wie umfassend die Ausgabe ausfallen kann, bevor die Generierung abgebrochen wird. Welches Wort jeweils im Generierungsprozess als nächstes Wort gewählt wird, hängt so wiederum wesentlich von den Trainingsdaten des großen Sprachmodells ab. Wenn nun Anfragen gestellt werden, deren Beantwortung die Reproduktion, Zusammenfassung oder Interpretation von Fakten bedingen würde, dann kann "die KI" diese nur aus ihrer verfügbaren Wissensbasis, also den Trainingsdaten, ableiten, solange sie nicht selbst Zugriff auf Recherchewerkzeuge hat, die sie zu Beantwortung einsetzen kann. Während derartige werkzeugnutzende KI-Systeme bereits existieren, bietet etwa die weit verbreitete kostenfreie Version von ChatGPT diese Funktionalität nicht - diese greift ausschließlich auf ihre Wissensbasis zum Stand des Trainingszeitpunkts zu. Das Risiko, dass es hier also zu oberflächlichen, lückenhaften oder sogar falschen Antworten kommt, ist - je nach Art der Anfrage und Umfang des generierten Textes - nicht unerheblich. Trotzdem gibt es sinnvolle Einsatzszenarien für diesen Anwendungsfall, etwa zur Generierung von Schreibideen oder zur Erstellung einer Inhaltsskizze zur Strukturierung eines längeren Textes. Die Anforderungen an die Medienkompetenz der Nutzerinnen und Nutzer ist jedoch ungleich höher als in den vorhergehenden Fällen - die Prüfung der Ergebnisse auf seine Stimmigkeit und Korrektheit liegt bei diesen. Dazu muss neben dem Bewusstsein für die grundlegende Problematik der Textgenerierung auch die inhaltliche Kompetenz vorhanden sein, diese Prüfung durchzuführen. Angesichts der auf den ersten Blick erstaunlichen Fähigkeiten der großen Sprachmodelle in diesem Bereich ist es umso wichtiger, das Bewusstsein zu schaffen, dass es sich dabei letztlich auch „nur“ um eine Werkzeug handelt, dessen kompetente Nutzung erlernt werden muss und das auch verantwortungsvoll genutzt werden muss - wie ein Taschenrechner, der Berechnungen massiv beschleunigt und vereinfacht, bei dessen Nutzung aber dennoch die Kompetenz vorhanden sein sollte, einzuschätzen, ob das Ergebnis mathematisch überhaupt korrekt sein kann. Dazu ist es notwendig, die zu unterstützende Aufgabe prinzipiell auch selbst durchführen zu können. Im Fall des Taschenrechners wäre das also auch händisch rechnen zu können - im Fall der Nutzung von Systemen wie ChatGPT bedingt dies, ein inhaltliches Grundverständnis über den Gegenstandsbereich der Anfrage und die Recherchekompetenz zur Prüfung des Ergebnisses zu haben. Wenn es gelingt, bei den Benutzerinnen und Benutzern das Bewusstsein für die Grenzen und die notwendigen Kompetenzen für den Einsatz von „künstlicher Intelligenz“ auch im kreativen Bereich zu schaffen, kann sie auch in diesem Bereich eine wertvolle Ergänzung zu den bisher verfügbaren Kreativtechniken und -werkzeugen zu schaffen.

Gerade dieser letzte Aspekt ist eine große Herausforderung für das Bildungssystem im Allgemeinen und die Schulen im Speziellen. Hier ein differenziertes Verständnis für die

Potentiale und Grenzen von „künstlicher Intelligenz“ zu vermitteln und den Schülerinnen und Schülern jene Kompetenzen mit auf den Weg zu geben, die sie zu einem verantwortungsvollen Umgang mit dem Werkzeug „KI“ befähigen, ist eine wesentlicher Schritt hin zu einer zukunftsfähigen Bildung, die die Orientierung auch in einer Welt ermöglichen wird, in der der Umgang mit „künstlicher Intelligenz“ so alltäglich sein wird wie der Umgang mit einem Lichtschalter.

Kontakt

Univ.-Prof. DI Dr. Stefan Oppl
Universität für Weiterbildung Krems
Dr.-Karl-Dorrek-Straße 30
3500 Krems
stefan.oppl@donau-uni.ac.at