DGI-Forum Wittenberg 2019

Jürgen Buder*

Wieviel Mensch steckt in der Maschine?

Zur Leistungsfähigkeit und Wirkung von Künstlicher Intelligenz (KI)

<https://doi.org/10.1515/iwp-2019-2061>

Zusammenfassung: In vielen Bereichen ist die Leistung von Künstlicher Intelligenz (KI) der Leistung von Menschen mittlerweile ebenbürtig oder überlegen. Dies gilt vor allem in Bereichen, in denen Regeln angewendet oder erkannt werden müssen. Dennoch sollten diese Fortschritte nicht darüber hinwegtäuschen, dass es noch immer gravierende Defizite in der Leistungsfähigkeit von KI gibt. In diesem Beitrag werden sechs solcher Defizite identifiziert. Forschung aus der Psychologie legt allerdings umgekehrt den Schluss nahe, dass eine KI nicht besonders intelligent sein muss, um von Menschen als menschenähnlich wahrgenommen zu werden. Auf der Basis der Literatur zum Thema Anthropomorphismus werden drei Bausteine der Vermenschlichung von Technologien diskutiert.

Deskriptoren: Künstliche Intelligenz, Roboter, Mensch, Leistung, Psychologie

What is human about machines?

On the performance and effect of artificial intelligence (AI)

Abstract: In many areas, the performance of Artificial Intelligence (AI) is now equal to or superior to that of humans. This is especially true in areas where rules need to be applied or regularities need to be detected. Nevertheless, these advances should not obscure the fact that there are still serious deficits in the performance of AI. This paper identifies six such deficits. Conversely, research from psychology suggests that an AI does not have to be particularly intelligent in order to be perceived by humans as human-like. Based on the literature on anthropomorphism, three building blocks of the humanization of technologies are discussed.

Descriptors: Artificial intelligence, Robot, Human being, Performance, Psychology

Qu'y a-t-il d'humain dans les machines ?

Sur l'efficacité et les effets de l'intelligence artificielle (IA)

Résumée : Dans de nombreux domaines, la performance de l'intelligence artificielle (IA) est maintenant égale ou supérieure à celle des humains. C'est particulièrement vrai dans les domaines où des règles doivent être appliquées ou quand des régularités doivent être reconnues. Néanmoins, ces progrès ne doivent pas occulter le fait qu'il existe encore de graves déficits dans la performance de l'IA. Cet article identifie six de ces déficits. Inversement, la recherche en psychologie suggère qu'il n'est pas nécessaire qu'une IA soit particulièrement intelligente pour être perçue par l'être humain comme un être humain. Sur la base de la littérature sur l'anthropomorphisme, trois éléments constitutifs de l'humanisation des technologies sont discutés.

Descripteurs : Intelligence artificielle, Robot, être humain, Performance, Psychologie

Einleitung

Bei wenigen gesellschaftlichen Themen werden so stark Hoffnungen und Ängste geweckt wie beim Thema Künstlicher Intelligenz (KI). Der Hype um KI wurde in der Gesellschaft vor allem durch medienwirksam inszenierte Aktionen befeuert, in denen KI-Programme gegen die weltbesten Spieler im Schach (1996), Jeopardy (2010) und Go (2016) antraten und teils deutlich gewinnen konnten. Auch wurde die Mächtigkeit von KI und „big data“ durch Anekdoten unterstrichen, wie z. B. von dem Vater, der sich darüber beschwerte, dass eine Einzelhandelsfirma seiner minderjährigen Tochter Coupons für Babykleidung zugeschickt hatte. Die Tochter war tatsächlich schwanger, ohne dass der Vater davon wusste; der KI-Algorithmus hatte hingegen aus den Drogerie-Einkäufen der Tochter geschlossen, dass eine Geburt bevorsteht. Heutzutage ist KI in aller Munde, wobei sich neben den Verheißungen (z. B. selbstfahrende Autos) auch zunehmend Skepsis (z. B. gegenüber Pflegerobotern) und Sorgen (z. B. Weg-rationalisieren von Arbeitsplätzen, Einsatz von intelligen-

*Kontaktperson: Dr. Jürgen Buder, Leibniz-Institut für Wissensmedien (IWM), Schleichstraße 6, 72076 Tübingen, E-Mail: j.buder@iwm-tuebingen.de

ten Waffen) mischen. Während solche sichtbaren Themen breit diskutiert werden, hat KI schon längst – und oft unbemerkt – Einzug in den Alltag gehalten, sei es im Finanz- und Versicherungswesen, in der Werbebranche, oder bei der Nutzung des Internets (vor allem in der Auswahl der Inhalte, die Usern angezeigt werden).

Vor diesem Hintergrund sollen in diesem Beitrag zunächst die Stärken und (derzeitigen) Schwächen von KI aus einer psychologischen Sichtweise heraus skizziert werden. Nachdem vor allem die Schwächen dargestellt wurden, an deren Behebung natürlich fieberhaft geforscht wird, erörtert dieser Beitrag die Frage, wie hoch die Leistungsfähigkeit und wie nuanciert das „Verhaltensrepertoire“ von KI eigentlich sein muss, um von Menschen als menschenähnlich wahrgenommen zu werden.

Stärken von KI

Zunächst einmal basieren die Stärken von KI natürlich auf der generellen Fähigkeit von Computern, sehr große Datenmengen in sehr kurzer Zeit zu verarbeiten. Bei KI geht es zusätzlich aber um die Frage, wie solche Datenmengen verarbeitet werden. Hier lassen sich vereinfacht zwei verschiedene Ansätze unterscheiden. Der erste Ansatz dominierte die KI-Forschung im 20. Jahrhundert und basierte auf dem Grundgedanken, dass sowohl Menschen als auch Computer letztlich informationsverarbeitende Systeme sind (Newell & Simon, 1972). Sowohl Menschen als auch Computer verarbeiten Informationen, indem sie bestimmte Wenn-dann-Regeln anwenden, beispielsweise beim schriftlichen Addieren mit Übertrag („Wenn die Summe zweier Ziffern größer neun ist, dann bilde einen Übertrag in der nächsten Spalte“). Solche regelbasierten KI-Systeme wurden daher mit einem reichen Repertoire an Wenn-Dann-Regeln versehen, die jeweils in einer gegebenen Situation („Wenn“) das Ausführen einer Aktion („Dann“) nach sich zogen (Anderson, 1983). Dies setzte voraus, dass eine Domäne (z. B. Algebra) möglichst vollständig beschrieben und modelliert wurde. Im Bildungskontext wurden in den 1980er Jahren beispielsweise sogenannte intelligente tutorielle Systeme entwickelt, die in der Lage waren, Schülerinnen und Schüler in Algebra oder Geometrie zu unterrichten. Diese Systeme waren in der Lage, aus den schriftlichen Antworten der Schüler Rückschlüsse über deren Vorwissen oder über vorhandene Fehlkonzepte zu ziehen, und auf dieser Basis den Unterricht (z. B. das Wiederholen einer Lehreinheit oder das Steigern der Aufgabenschwierigkeit) anzupassen. Für gut modellierbare Domänen wie Algebra zeigte sich tatsächlich, dass intelligente Tutoren zu ähnlichen guten Lernleistungen bei

Schülern führten wie Einzelunterricht durch eine Lehrkraft. Allerdings scheiterten intelligente Tutoren – wie generell regelbasierte Systeme – wenn es um komplexere und weniger gut modellierbare Domänen wie Chemie oder gar Geschichte ging.

Ein zweiter Ansatz dominiert bislang die KI-Forschung im 21. Jahrhundert. Er wurde dadurch motiviert, dass die Funktionsweise von regelbasierten Systemen keinerlei Ähnlichkeit damit hat, wie Informationen in einem menschlichen Gehirn neuronal codiert und verarbeitet werden. Außerdem sind die alten, regelbasierten Systeme nicht von sich aus lernfähig, sondern können immer nur die vorhandenen Regeln anwenden. Analog zur Struktur von Neuronen im Gehirn modellieren neuere KI-Systeme das Zusammenspiel von einer sehr großen Anzahl von „Knoten“, die miteinander zu neuronalen Netzen verschaltet sind. Die Knoten selber haben keine semantische Bedeutung, sondern haben lediglich ein Aktivationsniveau, welches zum „Feuern“ des Knoten führt, wenn ein bestimmter Schwellenwert überschritten wird. Im Gegensatz zu regelbasierten Systemen wenden neuronale Netze keine vorgegebenen Regeln an – stattdessen sind sie selber in der Lage, Regelmäßigkeiten in Daten zu erkennen und somit auch zu lernen. Um ein solches maschinelles Lernen zu ermöglichen, benötigen diese Systeme sehr große Datenmengen. Als Gegenstück zu intelligenten tutoriellen Systemen hat sich beim maschinellen Lernen im Bildungsbereich das Konzept der Learning Analytics (Siemens, 2005) etabliert. Hier werden beispielsweise Millionen von Daten aus Interaktionen von Tausenden von Schülerinnen und Schülern mit einer Lernumgebung nach statistischen Verfahren ausgewertet, um dabei Regelmäßigkeiten zu entdecken. Eine entsprechende Software würde dann bei einer Interaktion einer Schülerin mit dem System beispielsweise berechnen können, dass 53 Prozent aller vergleichbaren Schüler eine bessere Note im abschließenden Geometrie-Test erhielten, wenn ihnen in der aktuellen Situation Lernaufgabe X präsentiert wurde, aber nur 49 Prozent, wenn Aufgabe Y dargeboten wird. Mit solchen Ansätzen maschinellen Lernens, die nicht vorgegebene Regeln anwenden, sondern bestehende Regelmäßigkeiten in Daten erkennen können, ist es der KI-Forschung gelungen, erhebliche Fortschritte zu erzielen. Dies betrifft vor allem auch Bereiche, die viel zu komplex sind, als dass sie vollständig regelbasiert modelliert werden könnten, wie z. B. das Erkennen gesprochener Sprache, das Erkennen von Gesichtern, oder das Erkennen von Verkehrssituationen (selbstfahrende Autos).

Schwächen von KI

Die erstaunlichen Leistungen, die mit KI auf der Basis maschinellen Lernens verzeichnet werden können, sollten aber nicht darüber hinwegtäuschen, dass diese Systeme derzeit auch einige fundamentale Schwächen aufweisen. Im Folgenden sollen sechs solcher Schwächen skizziert werden.

Verzerrungsproblem. Eine KI, die auf maschinellem Lernen beruht, ist nur so gut wie die Daten, anhand derer sie trainiert wurde. Die Datengrundlage kann aber fehlerhaft oder nicht-repräsentativ sein. Beispielsweise wird vereinzelt KI von Strafverfolgungsbehörden eingesetzt, um besondere Brennpunkte zu identifizieren (Predictive Policing). Bestimmte Delikte wie beispielsweise häusliche Gewalt werden allerdings häufig durch Nachbarn zur Anzeige gebracht; infolgedessen neigt eine solche KI dazu, die Fälle häuslicher Gewalt in dünner besiedelten Gebieten oder in Regionen mit vielen allein stehenden Häusern zu unterschätzen. Auf diese Art und Weise sind die Vorhersagen einer KI oft verzerrt. Dadurch, dass KI immer auf bestehenden Daten agiert, ist sie auch rein deskriptiv und somit wenig geeignet, Entscheidungen zu fällen, die eine präskriptive Komponente beinhalten. Im Jahr 2015 musste Amazon eine KI-Entwicklung zur eigenen Personalauswahl einstellen, weil das System fast ausschließlich Männer für Posten im Management vorschlug – dass es präskriptiv wünschenswert sein mag, den Frauenanteil in einer Branche zu erhöhen, kann eine KI, die nur auf maschinellem Lernen beruht, nicht berücksichtigen. Ähnliche Verzerrungen können in allen Bereichen zum Tragen kommen, sei es im Versicherungswesen, oder auch im Bereich der Learning Analytics.

Transparenzproblem. Dadurch, dass maschinelles Lernen meist auf hochkomplexen und dynamischen Mustern in neuronalen Netzen basiert, sind die Entscheidungen oder Vorhersagen der KI nicht rückführbar. Eine KI kann beispielsweise auf der Grundlage von Millionen von einzelnen Datenpunkten zum Schluss kommen, dass eine Person nicht kreditwürdig ist, aber es ist nicht möglich, einzelne Parameter zu bestimmen, die zu dieser Entscheidung geführt haben. Je folgenreicher die Entscheidung einer KI ist, umso größer ist das Problem einer solchen Intransparenz. Derzeit wird unter dem Stichwort „Explainable AI“ daran geforscht, die Transparenz der maschinellen Lernprozesse zu erhöhen.

Das Problem des Alltagsverständnisses. Neuere KI, die auf maschinellem Lernen beruht, ist rein datengetrieben – es gibt keine vorgegebenen Regeln, mit denen die KI Umweltinformationen verarbeitet. Im Gegensatz dazu besitzen Menschen bereits in sehr früher Kindheit intuitive

Theorien über physikalische oder psychologische Begebenheiten, die es ihnen erlauben, die Welt zu verstehen (Lake, Ullman, Tenenbaum, & Gershman, 2017). Ein Kind versteht daher schon sehr früh, dass Objekte nicht aufhören zu existieren, wenn sie von anderen Objekten verdeckt werden, oder dass Personen üblicherweise Ziele erreichen wollen, wenn sie sich von einem Ort zum anderen bewegen. Zwar kann eine KI ebenfalls solche Zusammenhänge auf statistischem Weg lernen, doch was ein Kind mit einer intuitiven Theorie in Minuten lernen kann, benötigt oft mehr als 100.000 Trainingsdurchgänge für eine KI. Alltagsverständnis ist es auch, Informationen kategorial interpretieren zu können: ein Mensch erkennt sofort, dass die Antwort „20 Zoll“ auf die Frage „Wann wurde Abraham Lincoln geboren?“ nicht richtig sein kann. Eine KI ohne Alltagsverständnis würde hingegen versuchen, die Güte der Antwort zu ermitteln.

Das Problem der Eventualität. Menschen sind sehr gut darin, sich an veränderte Situationen anzupassen – sie können somit auch sehr flexibel mit Änderungen oder Ausnahmen umgehen. Für ein Kind macht es keinen besonderen Unterschied, wenn beim Spielen rote Klötzchen durch blaue ersetzt werden. Bei einer KI, die Klötzchen erkennen kann, aber nur anhand von roten Klötzchen trainiert wurde, wäre es prinzipiell denkbar, dass sie ein blaues Klötzchen gar nicht als solches erkennen könnte. Zudem sind Menschen in der Lage, bei der Beurteilung von Situationen (z. B. bei der ärztlichen Erstellung von Therapieplänen) Ausnahmen zu berücksichtigen. Eine KI im medizinischen Kontext würde aufgrund einer spezifischen Befundlage möglicherweise vorschlagen, dass eine Dialyse einer Nierentransplantation vorzuziehen sei. Eine Ärztin würde hingegen weitere Faktoren berücksichtigen (z. B. Anfahrtswege zur nächsten Dialyse-Station) und in ihre ärztliche Empfehlung mit einbinden. Derzeit besteht der einzige Lösungsansatz für das Eventualitätsproblem darin, eine KI möglichst umfassend zu trainieren, so dass auch Daten zu Ausnahmesituationen vorliegen. Doch dies wirft nur weitere Probleme auf (s. nächstes Problem).

Das Problem des Datenhungers. Um möglichst gut auf Änderungen oder Ausnahmen reagieren zu können, benötigt eine KI möglichst viele Daten. Big Data wird als eine Verheißung gepriesen, ist aber eine schiere Notwendigkeit, um eine KI lernfähig zu machen. Dass KI so viele Daten (z. B. über die User-Navigation im Internet oder in der physikalischen Umwelt) sammelt, schafft einerseits massive ethische und datenschutzrechtliche Probleme. Andererseits zeigt es auch, dass die Mechanismen, die KI intelligent erscheinen lassen, fundamental anders strukturiert sind als beim Menschen. Menschen sind in der Lage, aufgrund von „Small Data“ gute und begründbare Entschei-

dungen zu treffen. Die Güte der Entscheidungen steigt nicht einmal notwendigerweise mit einer größeren Datenbasis (less-is-more-Effekte; Gigerenzer & Gaissmaier, 2011). Im Gegensatz dazu benötigt maschinelles Lernen so viele Daten wie möglich, um menschenähnliche Performanz zu erreichen. In der Forschung und Entwicklung ist man sich dieser Probleme durchaus bewusst, und so hat die US-amerikanische Militärbehörde DARPA, die maßgeblich zur Entwicklung des Internets beigetragen hat, im Herbst 2018 eine „dritte Welle der KI“ ausgerufen, bei der es darum gehen soll, KI menschenähnlicher zu machen: ein Alltagsverständnis zu entwickeln, mit Ausnahmen und einer geringen Datenbasis umgehen zu können (DARPA, 2018).

Das Problem des Sozialen. Wie der Begriff „Künstliche Intelligenz“ bereits andeutet, geht es bei KI vor allem darum, Maschinen oder Computer zum Erbringen kognitiver Leistungen zu befähigen. Im Vergleich dazu sind Menschen aber nicht nur erstaunlich gut darin, kognitive Leistungen zu vollbringen (angesichts ihrer vergleichsweise geringen „Rechenpower“), sondern vor allen Dingen exzellent darin, sozial zu interagieren – die Emotionen und Motivation anderer Menschen zu erkennen oder selber Emotionen und Motivationen auszudrücken. Die meisten KI-Produkte muten in dieser Hinsicht geradezu autistisch an. Zwar gibt es KI-Programme, die versuchen, soziale Information zu erkennen (z. B. den emotionalen Gehalt von Texten zu bestimmen), doch gibt es zu weiten Bereichen zwischenmenschlichen Zusammenlebens (z. B. Erkennen von politischen Einstellungen; von Machtstrukturen; von sozialem Status; von sozialer Identität) bisher kaum Anstrengungen, diese maschinell codierbar zu machen.

Trotz der beeindruckenden Leistungen, die sich mit KI erzielen lassen, gibt es auch noch erhebliche Lücken oder Schwächen: abhängig von den Trainingsdaten können die Vorhersagen von KI stark verzerrt sein. Die Gründe, die zu einer Vorhersage geführt haben, sind häufig intransparent. KI mangelt es zudem insofern an Menschenähnlichkeit, als Stärken der menschlichen Informationsverarbeitung (Nutzung intuitiver Theorien, flexibler Umgang mit Daten, Rückgriff auf „Small Data“, soziale Informationsverarbeitung) derzeit nicht implementiert oder implementierbar sind. Man könnte daher zu dem Schluss kommen, dass eine sinnvolle und menschenähnliche Interaktion zwischen Mensch und Maschine gar nicht möglich sei. Dies würde aber voraussetzen, dass Menschen sehr hohe Standards ansetzen, wenn es um die Beurteilung von Menschenähnlichkeit geht. Forschung zum Anthropomorphismus legt aber den Schluss nahe, dass dies nicht unbedingt der Fall ist.

Anthropomorphismus allgemein

Der Hang zum Anthropomorphismus, also zur Vermenschlichung von nicht-menschlichen Entitäten, ist in vielen Bereichen gut dokumentiert: Menschen schimpfen auf ihre Autos, sie unterstellen Tieren (insbesondere ihren Haustieren) menschenähnliches Erleben und sie versehen ihre Gottheiten mit menschlichen Attributen und Emotionen. Vor diesem Hintergrund liegt der Schluss nahe, dass auch Computer anthropomorphisiert werden. Dies machen sich auch Designer von Computersystemen oder KI zunutze, wobei zwischen zwei Formen von Anthropomorphismus unterschieden werden kann: bei physikalischem Anthropomorphismus geht es darum, einem Computer oder Roboter ein menschenähnliches Aussehen zu verleihen; beim psychologischen Anthropomorphismus geht es darum, dass Computer oder Roboter ein möglichst menschenähnliches Verhalten zeigen.

Physikalischer Anthropomorphismus

Psychologische Forschung hat gezeigt, dass Menschen eine Information anders verarbeiten, wenn diese Information von einem sogenannten Agenten präsentiert wird, also einem Computerprogramm, welches menschenähnlich aussieht und agiert. In einer Studie konnte ein Agent, der über menschenähnliche Gestik und Gesichtsausdrücke verfügte, einen Lerninhalt über Solarzellen besser an Schülerinnen und Schüler vermitteln als ein statischer Agent und führte somit zu erhöhten Leistungen in Wissenstests (Mayer & DaPra, 2012). In einer anderen Studie führte Feedback über die Leistung von Versuchspersonen bei diesen zu höherer Motivation, wenn das Feedback von einem Agenten dargeboten wurde, als wenn das Feedback nur in Textform dargestellt war (Mumm & Mutlu, 2011). Wenn ein Computerprogramm oder Roboter mit einer menschenähnlichen Stimme spricht, führt dies stärker dazu, den Computer als menschenähnlich wahrzunehmen, als wenn derselbe Output schriftlich erfolgt (Schroeder & Epley, 2016). Umstritten ist allerdings die Annahme, dass es bei physikalischem Anthropomorphismus zu einem sogenannten „uncanny valley“-Effekt kommt, bei dem eine fast perfekte Menschengestalt als weniger menschlich und „unheimlicher“ wahrgenommen wird als entfernte Menschenähnlichkeit (z. B. ein Cartoon-Charakter). Insgesamt spricht die Befundlage somit dafür, dass menschenähnliches Aussehen von Computern zu menschenähnlichen Zuschreibungen bei Usern führt.

Psychologischer Anthropomorphismus

Bereits in den 1990er Jahren hatte Clifford Nass die sogenannte CASA-Hypothese aufgestellt (ein Akronym für „Computers Are Social Actors“), in der postuliert wurde, dass Menschen mit Computern oft nicht anders umgehen als mit anderen Menschen. So zeigten Studien, dass a) Menschen Geschlechterstereotypen anwenden, wenn ein Computer mit einer männlichen oder weiblichen Stimme sprach; b) dass Menschen gegenüber Computern Höflichkeitsformen wahren, die sie aus dem interpersonalen Kontext übertragen; und c) dass Menschen Computer nach ähnlichen Kriterien (z.B. vermeintliche Expertise) bewerten wie Mitmenschen (Nass & Moon, 2000). Laut einer anderen Theorie wird der Hang zur Vermenschlichung größer, je mehr eine Person über andere Menschen und je weniger die Person über Computer weiß (Epley, Waytz, & Cacioppo, 2007). Zudem ist die Anthropomorphisierung verstärkt, wenn Menschen sich einsam fühlen oder wenn sie in einer bestimmten Situation verunsichert sind. Beispielsweise ist ein unerwartetes Verhalten eines Computers leichter „erklärbar“, wenn man dem Computer einen eigenen Willen unterstellt.

Wenig darüber bekannt ist allerdings, wie sich ein Computer verhalten muss, um als menschenähnlich wahrgenommen zu werden – also gewissermaßen die Bausteine, die zum Anthropomorphismus führen. Auf der Basis der aktuellen wissenschaftlichen Befundlage werden im Folgenden drei solcher Bausteine postuliert:

Interaktivität. Zwischenmenschliche Kommunikation ist von Interaktivität geprägt, einem Wechsel von Rede und Gegenrede, bei der der bisherige Verlauf der Kommunikation berücksichtigt wird. In dem Maße, in dem ein Computer diese Verhaltensmuster imitieren kann, wird er vermenschlicht werden.

Unvorhersagbarkeit. Ein weiteres Kennzeichen von zwischenmenschlicher Kommunikation besteht darin, dass das Verhalten des Gegenübers nicht exakt vorhersehbar ist – man weiß nicht, ob und wann man angesprochen wird, und wenn man etwas äußert, ist die Reaktion des Gegenübers nicht antizipierbar. Wenn ein Computerprogramm in der Lage ist, eine Interaktion mit einem Menschen zu initiieren (z.B. an einen anstehenden Termin zu erinnern), und wenn der Output des Computers nicht genau vorhersagbar ist, wird das Programm als menschlicher angenommen.

Geteiltes Wissen. Zwischenmenschliche Kommunikation basiert auf geteiltem Wissen. Wenn Menschen mit einem Gegenüber „face-to-face“ sprechen, werden dabei

zahlreiche identifizierende Merkmale übermittelt (z.B. Alter, Geschlecht). Bei einem vertrauten Gegenüber ist es zudem normal, dass die andere Person etwas über einen selber weiß (z.B. den Namen, bestimmte Vorlieben). In dem Maße, in dem ein Computer den Anschein entwickeln kann, selbst über identifizierbare Merkmale zu verfügen (z.B. Vorlieben, Meinungen) oder in dem der Computer über Wissen zum Interaktionspartner zu verfügen scheint (z.B. Kenntnis des Namens, Leistungsfeedback), wird die Vermenschlichung zunehmen.

Wenn diese theoriegeleiteten Annahmen stimmen, müsste sich empirisch zeigen, dass bestimmte Standardtechnologien, mit denen Menschen interagieren, verschieden anthropomorphisiert werden. Eine Textverarbeitung ist zwar interaktiv, aber vorhersagbar, und sie verfügt zudem über nur geringes Wissen über die Person, die sie nutzt. Der Output einer Suchmaschine ist zwar unvorhersagbar, aber von geringer Interaktivität und geringem Maß an geteiltem Wissen gekennzeichnet. Höher dürfte die Anthropomorphisierung bei jenen Technologien ausfallen, deren Verhalten zwei der drei Bausteine adressiert: so ist ein Empfehlungsalgorithmus zwar vergleichsweise wenig interaktiv, doch die Empfehlungen sind unvorhersagbar und zeugen – wenn es personalisierte Empfehlungen sind – von Wissen über die Person. Das Verhalten einer KI in einem Computerspiel ist interaktiv und unvorhersagbar, aber nicht durch geteiltes Wissen geprägt. Schließlich gibt es Technologien, die sich durch Interaktivität, Unvorhersagbarkeit und geteiltes Wissen auszeichnen. Dazu gehören z.B. intelligente Chatbots, aber auch personalisierte Lernumgebungen, in denen Lernende Feedback und persönlich angepasste Instruktionen erhalten.

Fazit

Die Erwartungshaltung, die mit KI verknüpft wird, ist enorm – sei es in der Politik, der Industrie, in der Wissenschaft, oder in der Gesellschaft. Dies ist einerseits nachvollziehbar, da KI bereits in einigen Bereichen enorme Leistungen erbringen kann. Das sollte aber nicht darüber hinwegtäuschen, dass es eine Reihe von fundamentalen Problemen gibt, die mit KI derzeitigen Zuschnitts nicht adressierbar sind, sei es im Alltagsverständnis oder in der Fähigkeit, Sozialverhalten zu erkennen, und auf der Grundlage weniger Daten gute und nachvollziehbare Schlüsse zu ziehen. Dass solche Probleme in Diskursen wenig aufgegriffen werden, mag das Ergebnis ausgeklügelter Marketingstrategien von KI-Entwicklern sein. Es mag aber auch damit zusammenhängen, dass der Glaube

an die Segnungen moderner Technologien nur eine weitere Spielart des Anthropomorphismus ist. Aus psychologischer Warte betrachtet spricht viel dafür, dass Technologie gar nicht so viel leisten können muss, um von Menschen als menschenähnlich wahrgenommen zu werden.

Literatur

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press
- DARPA (2018). AI Next Campaign. <https://www.darpa.mil/work-with-us/ai-next-campaign> [24.10.2019]
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, e253.
- Mayer, R. E., & DaPra, S. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied*, 18(3), 239–252.
- Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, 27, 1643–1650.
- Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Upper Saddle River, NJ: Prentice-Hall.
- Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology: General*, 145(11), 1427–1437.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1).

Dr. Jürgen Buder

Leibniz-Institut für Wissensmedien (IWM)
Schleichstraße 6
72076 Tübingen
www.iwm-tuebingen.de/www/en/personen/ma.html?uid=jbuder
j.buder@iwm-tuebingen.de

Dr. Jürgen Buder ist kommissarischer Leiter der Arbeitsgruppe Wissensaustausch am Leibniz-Institut für Wissensmedien (IWM) in Tübingen. Außerdem bekleidet er die Stabsstelle Strategie im Leibniz-WissenschaftsCampus Tübingen „Cognitive Interfaces“. Nach dem Studium der Psychologie wurde er 2002 über Prozesse des Wissensaustauschs via Computer promoviert. Seine Forschungsschwerpunkte sind das computergestützte Lernen in Gruppen, der Umgang mit Konflikten in Internet-Diskussionen, sowie die Auswirkungen von Künstlicher Intelligenz auf die menschliche Informationsverarbeitung.