KI-basierte Sprachassistenten im Alltag

Forschungsbedarf aus informatischer, psychologischer, ethischer und rechtlicher Sicht

The implications of conversing with intelligent machines in everyday life for people´s beliefs about algorithms, their communication behavior and their relationship building

Ein Projekt gefördert durch

VolkswagenStiftung

Impressum

Autoren:

Nicole Krämer, André Artelt, Christian Geminn, Barbara Hammer, Stefan Kopp, Arne Manzeschke, Alexander Rossnagel, Pauline Slawik, Jessica Szczuka, Lina Varonina, Carina Weber

http://www.impact-projekt.de/

Kontakt:

Prof. Dr. Nicole Krämer

Telefon: +49 203 379 - 2482
E-Mail nicole.kraemer@uni-due.de

Universität Duisburg-Essen
Forsthausweg 2
47057 Duisburg

Details zur Publikation:

DOI: 10.17185/duepublico/70571
ISBN (Print): 978-3-940402-24-0

Verlag: Universität Duisburg-Essen, Universitätsbibliothek
Verlagsort: Essen

1. Auflage, Oktober 2019

Sprachassistenten auf dem Vormarsch

Ein Lautsprecher, der dem Nutzer tröstende Worte zuflüstert, der Mut zuspricht, der auf die eigenen Stimmungen Rücksicht nimmt, den man alles fragen kann und der um keine Antwort verlegen ist, der einen an alle Termine und Vereinbarungen erinnert und in schwierigen Lebenssituationen Empfehlungen gibt oder mit dem man einfach nur Konversation betreiben kann – all das ist keine ferne Vision mehr. Sprachassistenten, die aus ihrer Kommunikation ständig lernen und sich verbessern, sind auf dem Vormarsch. Sie versprechen die jeweils passende Unterrichtung, Unterredung, Unterstützung und Unterhaltung. Sie sind aber auch mit Risiken verbunden.

Diese Risiken betreffen einerseits die eingesetzten Algorithmen und Logiken, die dem technischen System zugrunde liegen. Sie können beispielsweise unerwünschte Haltungen und Wertungen vermitteln und schädliche Verhaltensanreize setzen. Andererseits liegen auch in der Qualität und Quantität der im Zuge der Nutzung des Sprachassistenten anfallenden Daten erhebliche Risiken. Sie geben Dritten einen tiefen Einblick in Gewohnheiten, Einstellungen, Vorlieben und Abneigungen, Kommunikationen und Beziehungen und können Selbstbestimmung und Entscheidungsfreiheit gefährden. Die Risiken wirken sowohl individuell als auch gesamtgesellschaftlich.

Gleichzeitig liegen in dieser Technik auch Chancen für die Selbstentfaltung – gerade für Nutzergruppen, die auf Assistenz angewiesen sind. Sprachassistenten können helfen, ältere Menschen und Menschen mit Einschränkungen im Alltag unabhängiger zu machen und ihre Selbstbestimmung zu steigern; Kinder können den Umgang mit Technologie erlernen.

Diese und weitere Chancen werden sich aber nicht von selbst realisieren, sondern es muss aktiv daran gearbeitet werden, durch Gestaltung der Technik die Risiken zu minimieren und die Chancen zu nutzen. Welche individuellen und sozialen Risiken bei lernfähigen Assistenzsystemen bestehen oder entstehen können und welche Chancen diese Techniken bieten, erörtert der folgende Text thesenartig und will damit die Aufmerksamkeit auf den Gestaltungsbedarf solcher Techniken lenken. Der Text gibt somit auch einen Ausblick darauf, welche Problemstellungen im von der Volkswagenstiftung geförderten Projekt IMPACT („The implications of conversing with intelligent machines in everyday life for people´s beliefs about algorithms, their communication behavior and their relationship building“, 2019-2024) behandelt werden.

Technische Ausgestaltung

Die modernen Sprachassistenten stellen eine Vielzahl an Funktionen zur Verfügung: Kalenderplanung, Geräte- und Smart-Home-Steuerung, Navigation, Informationsauskunft, Nachrichtenverfassung, Erledigen der Aufgaben in der realen Welt (z.B. Anrufe zur Terminvereinbarung über Google Duplex) u.v.m. Diese Assistenten funktionieren meistens nach dem gleichen Grundmodell, das im Folgenden beschrieben wird, und sind auf vielen Geräten präsent: Smartphones, Tablets, Laptops und PCs, smarte Lautsprecher, Uhren und Module in Smart Home-Systemen. Diese Geräte besitzen verschiedene technische Eigenschaften jenseits der Sprachein- und -ausgabe, wie Bildschirme unterschiedlicher Auflösung und Farbqualität, LEDs oder Webbrowser.

Dementsprechend müssen die über die Assistenten zur Verfügung gestellten Funktionen flexibel mit dieser Diversität umgehen können.
Im inaktiven Zustand erwartet der Sprachassistent ein bestimmtes vordefiniertes Weckwort. Sobald es in der Spracheingabe erkannt wurde, ist der Assistent aktiviert und reagiert auf die darauffolgenden Aussagen. Jetzt kann die Nutzer*in dem Assistenten einen Befehl erteilen oder eine Frage stellen. Moderne Assistenten erkennen menschliche Sprache oft mithilfe von tiefen neuronalen Netzen [1]. Dieses sind Black-Box Verfahren, die eine hochgradig nichtlineare Funktionalität der Eingabe (Signale) zur Ausgabe (Text) auf der Basis großer Datenmengen lernen können. Zumeist kommen in diesem Gebiet rekurrente Netze zum Einsatz, die intrinsisch die Tatsache, dass die Signale eine zeitliche Komponente haben, berücksichtigen können. Da tiefe Netze anspruchsvolle Rechenkapazitäten selbst für die Anwendung der gelernten nichtlinearen Funktionen, mehr noch für ihr Training, benötigen, läuft die Spracherkennung oft nicht lokal auf dem Gerät, mit dem die Nutzer*innen interagieren, sondern wird als Cloud Service vom Assistenten in Anspruch genommen.
Nachdem die Spracheingabe verstanden wurde, wird vom Assistenten ein Service ausgewählt, der für die Erfüllung der gestellten Aufgabe passend ist. Die Services können sowohl von dem Unternehmen, dem der Sprachassistent gehört, als auch von Drittanbietern zur Verfügung gestellt werden und laufen meistens auf einer Cloud, deren Server von dem Anbieter des Sprachassistenten verwaltet werden. Die Nutzer*innen können sich aktiv einen bestimmten Service wünschen (z.B. „Spiele die Playlist ‚Tanzmusik' vom Streamingdienst X"), wenn aber in der Spracheingabe derartige Informationen nicht explizit vorkommen, kann das System selbst einen Service vorschlagen, mit dem die Aufgabe der Nutzer*in erfüllt werden kann.
Die Spracherkennungskomponente liefert eine strukturierte Repräsentation der aktuellen Anfrage und sendet diese über ein Internetprotokoll in die Cloud an eine Instanz des ausgewählten Services. Dort wird die Anfrage nach der vom Serviceanbieter hinterlegten Logik verarbeitet, von der Nutzer*in gewünschte Aktionen werden ausgeführt und eine textuelle Antwort wird generiert, die ggf. durch visuelles Material ergänzt wird (z.B. Bilder, Videos, Tabellen). Diese Antwort wird dann an den Sprachassistenten zurückgesendet, der Sprache aus dem gelieferten Text synthetisiert. Als Gegenpol zu den klassischen Methoden wie konkatenative Sprachsynthese, bei der aus vorher aufgenommenen Sprachsegmenten Aussagen kombiniert werden, oder parametrische Sprachsynthese, bei der Sprache mithilfe von stochastischen Modellen erzeugt wird [2], werden in Sprachassistenten zunehmend neue Methoden verwendet, die auf tiefem Lernen basieren, bspw. WaveNet beim Google Assistenten [3]. Die Besonderheit dieser tiefen Lerner ist ihre 'end-to-end'-Verarbeitung: Statt der Modellierung einer komplexen Funktion unter Ausnutzen von Expertenwissen und modularer Zerlegung der Aufgabe in ihre einzelnen Bestandteile, werden komplexe Zusammenhänge als Ganzes betrachtet und anhand vieler Beispieldaten direkt gelernt.
Nach der Sprachsynthese bekommt die Nutzer*in die Antwort via Audioausgabe des Sprachassistenten und eventuelle ergänzende Informationen via weitere Ausgabemodalitäten, die auf dem aktuellen Gerät vorhanden sind.

Technische Ausgestaltung

Methoden der künstlichen Intelligenz können an allen Stellen in einem solchen Assistentensystem Verwendung finden:
(i) Die Erkennung eines Weckwortes ist ein sogenanntes Klassifikationsproblem: ein Audiosignal muss in die Kategorie „Weckwort" versus „kein Weckwort" unterteilt werden. Dieses Problem ist relativ einfach zu realisieren, da es sehr speziell ist, und kann daher lokal auf dem Gerät realisiert werden.
(ii) Die Transkription von gesprochener Sprache in geschriebenen Text findet, wie bereits erwähnt, in der Regel durch sogenannte rekurrente Netze statt [4]. Diese Aufgabe ist schwierig, da verschiedene Sprecher*innen unter realistischen Bedingungen, insbesondere bei Hintergrundgeräuschen, erkannt werden müssen. Zudem betrifft die Transkription den Gebrauch beliebiger natürlichsprachlicher Äußerungen und unterschiedlicher Sprachen. Große Internetfirmen können hier in der Regel auf eine große Anzahl von Daten zurückgreifen, da entsprechende Daten von den Nutzer*innen solcher Geräte gesammelt werden. Auf dieser Basis können dann leistungsfähige neuronale Netze trainiert werden. Das Training erfordert oft spezielle Hardware (GPUs oder TPUs) und erhebliche Trainingszeit, und wird offline bei den Firmen realisiert. Auch die anschließende Verwendung der Netze ist derzeit aufgrund der erforderlichen Ressourcen nicht lokal realisiert, sondern wird als Cloud-Service angeboten.
(iii) Der umgekehrte Prozess, die Synthetisierung gesprochener Sprache aus einer generierten textuellen Antwort, findet meistens ebenfalls durch rekurrente Netze statt [3]. Diese Aufgabe ist sehr viel einfacher als die Spracherkennung, da hier weniger Variationen und Störungen der Signale vorliegen; daher wird dieses in der Regel durch lokale, auf dem Gerät installierte Komponenten realisiert. Eine Herausforderung ist dabei, wie es erreicht werden kann, dass die generierte Sprache nicht nur verständlich ist, sondern natürlich wirkt; dieses betrifft neben dem Inhalt und der stilistischen Realisierung auch die Intonation des Gesprochenen.
(iv) Ein Kernstück von Assistenzsystemen, in denen künstliche Intelligenz eingesetzt wird, ist die Generierung einer sinnvollen Antwort oder Aktion auf der Basis der Spracheingabe der Nutzer*in. Derzeit ist dieses auf spezifische Funktionalitäten beschränkt wie Terminplanung, Bedienung von Hardware in Smart Home Settings, Internetabfragen oder e-Commerce. Dialoge finden daher in sehr klar begrenzten Gebieten statt und die Realisierung erfolgt entlang einer Anzahl vordefinierter Funktionalitäten und Logiken: es muss erkannt werden, um welches Szenario es gehen soll, und es müssen die wesentlichen Parameter zur Realisierung des Szenarios (z.B. welches Lied soll abgespielt werden) aus dem Text extrahiert oder vom Assistenten erfragt werden. Hier wird in der Regel auf Zusatzwissen zurückgegriffen und die wahrscheinlichsten Aktionen unter Einsatz von tiefem Lernen oder statistischen Modelle generiert. Typischerweise wird dabei Bedeutungswissen nicht explizit in Datenbanken gespeichert, sondern durch geeignete semantische Einbettungen als Vektoren repräsentiert, die sich leicht verrechnen lassen (z.B. für die automatisierte Übersetzung, bei der Sätze aus verschiedenen Sprachen in denselben Vektorraum eingebettet und einander zugeordnet werden können). Die Zuordnung einer so verarbeiteten Eingabe auf eine Systemantwort wird oftmals durch sogenanntes „sequence-to-sequence"-Lernen oder statistische Modelle realisiert und folgt damit den in großen Datenmengen vorhandenen Statistiken, nicht unbedingt aber expliziten Regeln – hierdurch entsteht für den Menschen oft eine Black Box-Charakteristik der Funktionsweise, die eine Nachvollziehbarkeit der Vorgänge erschwert [5].

Leistungsfähigkeit und -potentiale von Sprachassistenten

Derzeit sind die realisierten Funktionalitäten der Assistenzsysteme auf spezielle Bereiche limitiert. Die Reichhaltigkeit dieser Kern-KI wird allerdings stetig erweitert, etwa um die Möglichkeit der Diagnose des Gesundheitszustands der Nutzer*innen auf der Basis der Charakteristik der Spracheingabe [6]. Auch Einsichten in den aktuellen emotionalen Zustand und die Persönlichkeitsstruktur der Nutzer*innen können aus der Sprachintonation und den Stimmdaten gewonnen werden und somit in absehbarer Zukunft die Funktionalität der Sprachassistenten bereichern [7]. Emotionserkennung ist eine anspruchsvolle Aufgabe aus mehreren Gründen, u.a. da es bei den Ausdrucksformen zwischenmenschliche und kulturelle Unterschiede gibt. Außerdem ist es schwierig zu bestimmen, welche Eigenschaften der Sprache für die Emotionserkennung am relevantesten sind, z.B. akustische Merkmale wie Tonhöhe [8]. Das Problem der Merkmalsselektion kann auch mithilfe von tiefen neuronalen Netzen dank ihrer 'end-to-end'-Verarbeitung gelöst werden [9]. Zunehmende Möglichkeiten, verschiedene Modalitäten und individuelle Aspekte zu berücksichtigen, versprechen in diesem Gebiet dabei erhebliche Fortschritte [10].
Eine weitere Forschungsthematik besteht in der Fragestellung, wie solche Assistenzsysteme so gestaltet werden können, dass der Mensch den Dialog als natürlich wahrnimmt und letztendlich nicht mehr unterscheiden kann, ob er mit einem Menschen oder einer Maschine kommuniziert. Diese Bestrebungen wurden etwa durch Google Duplex, einer auf KI basierenden Technologie, die am Telefon natürliche, in ihrer Form oder Struktur nicht vorgegebene Dialoge zu Themen wie Terminvereinbarung realisiert, eindrucksvoll demonstriert, wobei diese Menschenähnlichkeit durch das Erlernen des Verhaltens auf der Basis von beobachteten Beispielen erreicht wurde [11]. Die immer reichhaltigeren Funktionalitäten von Assistenten werden in Wettbewerben evaluiert: IQ Tests für Sprachassistenten adressieren die Fähigkeit der Systeme, unterschiedlichste Fragen zu beantworten; einige Systeme erlangen inzwischen über 90% Erfolgsgüte [12]. Es wird dabei in der Wissenschaft kontrovers diskutiert, ob Assistenten in der Zukunft in der Lage sein werden, nicht nur spezifische Aufgaben zu lösen, sondern auch die Funktion einer universellen KI, die beliebige ihr vorab unbekannte Probleme lösen kann, übernehmen können.
Mit diesen Entwicklungen einhergehend entsteht eine Reihe von Herausforderungen für die Systeme: Wie bereits erwähnt, sind viele Services nur in der Cloud realisierbar. Dieses bedeutet aber, dass Funktionen nicht notwendig überall und zu jeder Zeit zur Verfügung stehen. Zudem müssen eventuell sensible Daten in die Cloud übertragen werden, d.h. es muss etwa durch Verschlüsselung sichergestellt werden, dass die übertragenen Daten nicht von unbefugten Personen abgefangen werden. Es ist Gegenstand der Forschung, wie Funktionalitäten schlanker und weniger aufwändig gestaltet werden können, so dass sie direkt auf den Endgeräten realisiert werden können, sogenanntes Edge-Computing [13].
Oft treten die Nutzer*innen dabei den Serviceanbietern die Hoheit über die gesammelten Daten ab, und das Geschäftsmodell großer Internetfirmen beruht oft darauf, die in den massiven Datenmengen verfügbare Information für Werbezwecke zu verwerten.

Die Tatsache, dass nur einige Monopolisten so Zugang zu diesen Datenmengen erhalten und nur sie die entsprechenden neuronalen Netze bereitstellen können, bietet erhebliche Risiken durch die dadurch entstehende Machtstellung etwa im wirtschaftlichen Bereich, insbesondere sobald diese Services für den Alltag notwendige Funktionalitäten bereitstellen. Es wird untersucht, inwieweit neuronale Netze oder Methoden des maschinellen Lernens auch anhand weniger Daten zuverlässig trainiert werden können – etwa Prototyp-basierte Verfahren erlauben in gewissem Rahmen diese Funktionalität [14].

Im Fall zentral gesammelter Daten und dadurch trainierter Modelle sind Anforderungen an die Privatheit zu berücksichtigen. Neben dem Schutz der Rohdaten selber muss sichergestellt werden, dass die neuronalen Netze, sofern öffentlich verfügbar, sowie deren Nutzung keine Informationen über Einzelpersonen beinhalten. Es darf Dritten nicht gelingen, individuelle Informationen über Personen herauszufinden, deren Daten zum Training der Netze verwandt wurden. Um diese Anforderungen zu erfüllen, wurden Verfahren entwickelt, die das Training leicht abändern und beweisbar die Möglichkeit, private Information zu erfahren, einschränken; bekannt sind etwa Methoden der sogenannten differenziellen Privacy, die bei großen Datenmengen die Information einzelner Individuen schützen kann [15].

Da die Logik von Assistenten oft auf erlerntem Verhalten beruht, ist die Sicherheit und Transparenz der Verfahren eine große Herausforderung: Sie können sich jederzeit unerwartet verhalten. Es muss etwa sichergestellt werden, dass nur befugte Nutzer*innen Aktionen wie das Bestellen von Ware im Internet initiieren können – dieses erfordert eine zuverlässige Identifikation befugter Nutzer*innen aufgrund ihrer Stimmen. Eine erhebliche Herausforderung ist dabei, dass es möglich ist, tiefe Netze durch für diesen Zweck gezielt optimierte Daten anzugreifen und für den Menschen unerwartetes Verhalten zu provozieren, so genannte adversariale Beispiele. Diese sind sehr bekannt etwa im Bereich der Bilderkennung, wo zum Teil bereits die Änderung einzelner Pixel ausreicht, die prognostizierte Klasse in unvorhersehbarer Weise abzuwandeln. Obschon es hier eine sehr große Forschungsaktivität gibt, ist es im Allgemeinen ungelöst, wie diese Angreifbarkeit von tiefen Netzen verhindert werden kann [16].

Es muss bei statistischen Verfahren und neuronalen Netzen dabei immer mit der Möglichkeit gerechnet werden, dass die Systeme auf die jeweilige Situation nicht angemessen reagieren können, etwa weil die derzeitige Situation unbekannt und nicht durch Trainingsdaten abgedeckt ist. Um hier zuverlässige Interaktionen zu gestalten, ist es zwingend nötig, KI-Verfahren immer zusammen mit ihren Beschränkungen zu modellieren und Fallback-Optionen zu realisieren. Eine Herausforderung ist dabei, dass viele Verfahren keine validen Wahrscheinlichkeiten berechnen und es in der Praxis unklar ist, wie genaue solche Rückweisungsoptionen optimiert werden sollen [17].

Die Tatsache, dass neuronale Netze oft als Black Box agieren, kann auch weiter reichende Konsequenzen haben: Netze spiegeln die in Daten gefundenen statistischen Zusammenhänge wider ohne hier auf Kausalitäten und Erklärungen zu achten. Insbesondere können so in den Modellen Zusammenhänge abgebildet werden, die nicht durch den Sachverhalt selber begründet sind, sondern auf der Statistik historischer Daten beruhen - Modelle sind dann unter Umständen "unfair" bzw. besitzen einen Bias.

Ein Beispiel ist etwa, wenn Modelle die Ethnie für die Entscheidungsfindung, ob eine Person wahrscheinlich kriminell wird, berücksichtigen, alleine auf der Basis der Tatsache, dass – auch verursacht durch menschliche Biases - in historischen Daten Ethnien bezogen auf das Kriterium der Kriminalität nicht gleichverteilt sind [18]. Um solche Implikationen zu verhindern, wird versucht, formale Konzepte von Fairness zu modellieren und während des Trainings zu berücksichtigen [19]. Ein weiteres zentrales Konzept, das solche Problematiken verhindert und gleichzeitig die Sicherheit und Benutzbarkeit erhöht, ist der Versuch, durch Black-Box-Verfahren gemachte Entscheidungen zu erklären. Die Frage der Erklärbarkeit von Modellen ist daher ein in mehrfacher Hinsicht hoch relevantes, allerdings noch nicht zufriedenstellend erforschtes Gebiet. Besonders wichtig ist bei dieser Fragestellung, sich darüber im Klaren zu sein, an welche Zielgruppen solche Erklärungen adressiert sind. Es besteht ein Unterschied dazwischen, (Fach-)Expert*innen mithilfe von Tools und mathematischer Methoden eine Möglichkeit zur Interpretation des Entscheidungsmodells zu verschaffen [20] und einzelne Entscheidungen im Dialog mit einem Sprachassistenten den Nutzer*innen zu erklären, die über kein weiterführendes technisches und domänenbezogenes Wissen verfügen. Im letzteren Fall sollten Erkenntnisse aus den Sozial- und Kognitionswissenschaften miteinbezogen werden, um zu verhindern, dass die Intuition der Entwickler*innen zum Hauptmaß der Güte der Erklärungen wird [21]. Ein idealer Sprachassistent soll imstande sein, unter Berücksichtigung der Besonderheiten der Nutzer*in (bspw. kognitive Fähigkeiten, Wissensstand) passende Erklärungen zu generieren und diese in einem interaktiven Dialog zusammen mit der Nutzer*in zu explorieren, z.B. indem weiterführende Fragen beantwortet und fehlende Hintergrundinformationen mitgeliefert werden. Um dies zu erreichen, sollte das System über weitreichende Kommunikationsfähigkeiten und Dialogmanagement verfügen und von einem Theory-of-Mind Modell Gebrauch machen, das ermöglicht, Annahmen über Bewusstseinsvorgänge in einer anderen Person zu treffen und zu erkennen, dass diese Person über eigene Perspektive, Glauben, Intentionen etc. verfügt. Mithilfe von diesen und weiteren Techniken sollen sich die Assistenzsysteme dynamisch an die Nutzer*innen in unterschiedlichen Dialogkontexten anpassen können [22].

Kommunikationskultur und Beziehungsbildung

Die neue Generation von Sprachdialogsystemen ist sicherlich auch deshalb so erfolgreich, weil erstmals ein zumindest ansatzweise natürlichsprachlicher Dialog mit Maschinen ermöglicht wird. Psychologisch bedeutet dies, dass die Maschinenhaftigkeit in den Hintergrund rückt, zum Beispiel dadurch, dass die dahinterliegenden Algorithmen und Funktionsweisen im Gespräch mit einem Sprachdialogsystem für den/die Nutzer*in der Regel wenig salient sind. Die Interaktion wird durch verbale Kommunikation gesteuert, so wie es der Mensch aus der Kommunikation mit anderen Menschen gewohnt ist. Zahlreiche Studien mit prototypischen Sprachdialog-, Roboter- oder Agentensystemen zeigen, dass eine solche natürlichsprachliche Interaktionsmöglichkeit unwillkürliche soziale Verhaltensweisen und Reaktionen gegenüber einer Maschine anregt, die man sonst nur anderen Menschen gegenüber erwarten würde (z.B. höfliche Ansprache, Anwendung von Stereotypen, sozial erwünschtes Verhalten).

Ein besonderes Merkmal dieser sozialen Reaktionen ist, dass diese unbewusst erfolgen und ihre Berechtigung von den Nutzer*innen sogar bewusst verneint wird (im Sinne der Aussage, dass man sich einer Maschine gegenüber nicht sozial verhalten sollte). Im Rahmen der Media Equation Theory nach Reeves und Nass [23] wurde in diesem Zusammenhang herausgestellt, dass eine Interaktion mit einer künstlichen Entität drei wichtige Aspekte beinhalten kann, welche dazu führen, dass Menschen unbewusst und unmittelbar sozial auf eben diese reagieren.

Zum einen wird der Dialog mit dem System in der *natürlichen Sprache* des Nutzers bzw. der Nutzerin geführt. So muss nicht extra ein komplizierter Code eingegeben werden, der einen bestimmten Output zur Folge hat. Es wird vielmehr ermöglicht, dass wir in den Kommunikationsmustern bleiben können, die wir bereits aus der zwischenmenschlichen Kommunikation kennen.

Zweitens ermöglicht die *Interaktivität des Systems* einen dynamischen Austausch zwischen Mensch und Maschine. Hierbei sei zu erwähnen, dass die Qualität der Interaktivität variieren kann. Während einige Systeme vor allem darauf ausgelegt sind, bestimmte Anfragen zu verarbeiten, betonen andere besonders die soziale Komponente der Interaktion und können eine distinkte Persona darstellen [24]. Verschiedene empirische Studien zu Interaktionen mit künstlichen Entitäten konnten bereits zeigen, dass sozialere Interaktionen (beispielsweise wenn Selbstauskunft betrieben wird oder Witze gemacht werden) im Vergleich zu funktionaler Ausführung von Befehlen zu stärkeren sozialen Reaktionen bei Nutzer*innen führten. Hierbei gelten oft ähnliche kommunikationspsychologische Gesetzmäßigkeiten wie bei zwischenmenschlichen Interaktionen. So konnte in mehreren Studien bereits gezeigt werden, dass beispielsweise Reziprozität und Selbstoffenbarung im Dialog mit künstlichen Entitäten bei Menschen zu einer gesteigerten Gesprächsbereitschaft und inhaltlich fundierteren Interaktionen führen (u.a. [25]; [26]). Wie in zwischenmenschlicher Interaktion wurde auch in Mensch-Maschine Dialogen sogenanntes Alignment beobachtet: Die Tendenz, in einer Unterhaltung die verbalen [27] und nonverbalen [28] Verhaltensweisen des*der Interaktionspartner*in zu übernehmen, ist auch bei einem maschinellen Gesprächspartner gegeben.

Der letzte Aspekt, der im Rahmen der Media Equation Theory für soziale Reaktionen im Hinblick auf artifizielle Interaktionspartner spricht, ist die *soziale Rolle*, welche von dem Dialogsystem ausgefüllt wird. Zum jetzigen Zeitpunkt haben die gängigen Sprachdialogsysteme vor allem die Aufgabe, bei bestimmten Anfragen zu assistieren (z.B. Fragen zu Weltwissen beantworten, Musik spielen). Mit einer gesteigerten sozialen Kompetenz wechselt die potentielle Rolle des Dialogsystems jedoch stärker zu einer Art Begleiter des Alltags („Companion"). Im Hinblick auf die soziale Rolle, welche ein Sprachdialogsystem ausfüllt, sollte reflektiert werden, dass die Systeme zum jetzigen Zeitpunkt prädominant mit weiblichen Stimmen und Namen (z.B. Alexa, Cortana, Siri) versehen werden. Dies könnte bezüglich einer potentiellen Beziehungsbildung ebenfalls mit Implikationen einhergehen, da verschiedene Studien bereits zeigen konnten, dass sich Geschlechterstereotype (z.B. in Bezug auf die Kompetenz oder Attraktivität eines Systems) ebenfalls auf künstliche Entitäten übertragen lassen (u.a., [29]). Dies wurde bereits von verschiedenen Autoren und Institutionen durchaus kritisch beleuchtet, da beispielsweise die Gefahr besteht, dass hierarchische (Befehls-)Strukturen im Mensch-Technik-Bereich repliziert werden könnten (u.a. [30]; [31]).

Zusammenfassend lässt sich also festhalten, dass künstliche Entitäten soziale Reaktionen hervorrufen können, sofern bestimmte soziale Hinweisreize gegeben sind. Dies konnte in zahlreichen empirischen Studien gezeigt werden, selbst wenn den Menschen bewusst gemacht wurde, dass es sich um einen nichtmenschlichen Hinweisreiz handelte.
Ein wichtiger Erklärungsansatz der Theorie besagt, dass Interaktivität im Lauf der Evolution ein Anzeichen für Leben war und dass sich unser Perzeptionsprozess noch nicht daran angepasst hatte, dass Intelligenz und Interaktionsfähigkeit nun auch von artifiziellen Interaktionspartnern ausgehen kann.
Diesbezüglich ist es wichtig zu untersuchen, welche Vorstellungen Menschen von der Funktionsweise maschineller Interaktionspartner*innen haben – gegebenenfalls auch weil ein Wissen über das Funktionieren die unmittelbaren sozialen Reaktionen abschwächen könnte. Zusätzlich konnte gezeigt werden, dass es positive Effekte auf die Akzeptanz haben kann, wenn Nutzer*innen eine klare Vorstellung von der Funktionsweise interagierender Maschinen haben [32]. Vor dem Hintergrund der Tatsache, dass mehr und mehr Systeme auf maschinellem Lernen beruhen, wird das Funktionieren allerdings eher weniger transparent. Erste Versuche, zu verstehen, wie Menschen mentale Modelle von künstlichen Entitäten entwickeln, zeigen, dass sich Menschen bislang vor allem an fiktionalen und nicht-fiktionalen Medienbeiträgen orientieren, um zum Beispiel zu verstehen, was ein Algorithmus ist und was er macht [33].
Auch hinsichtlich der Beziehungsbildung mit künstlichen Entitäten stellen sich noch zahlreiche Fragen. Vor dem Hintergrund, dass es kaum Langzeitstudien über die Beziehungsentwicklung zwischen Menschen und Maschinen gibt – auch verursacht durch die bislang fehlende Verfügbarkeit von langfristig in Interaktion stehenden Systemen – ist noch nicht ausreichend geklärt, ob maschinelle Systeme auf Dauer tatsächlich als „Companion", „Assistent" oder gar „Freund" empfunden werden. Weiter gefragt: Ob maschinelle Systeme so konstruiert werden sollten, dass sie als „Companion", „Assistent" oder gar „Freund" empfunden werden – was eine ethische Frage ist.

Interaktion mit Sprachdialogsystemen als Herausforderung für die Ethik

Die im Projekt IMPACT zu untersuchenden technischen Systeme führen kein Eigenleben – die Nutzer*innen sind vielmehr gefordert, in den Möglichkeitsräumen der Technik kooperative Handlungen zu vollziehen, die wiederum nicht für sich stehen können, sondern neben einer ontologischen Taxonomie (Womit habe ich es hier zu tun? Was heißt es, als Mensch mit einer Maschine zu interagieren?) der ethischen Bewertung bedürfen. Besonders im Hinblick auf den ontologisch, sozial wie moralisch unklaren Status der technischen Systeme, deren Beziehung zum Menschen sowie die Transparenz der Systeme und ihre Anwendung lassen sich einige Bedenken und Herausforderungen formulieren.

Moralisch unklarer Status des Sprachassistenzsystems

Der moralisch unklare Status des Systems ist in zwei Richtungen zu denken. Erstens kann das technische System per se nicht als explizit moralischer Akteur gelten, da es zur Selbstreflexion nicht in der Lage ist – was bisher als Merkmal eines moralischen Subjektes gilt.

Allerdings muss bedacht werden, dass es bereits Bestrebungen gibt, den technischen Systemen ein moralanaloges Kalkül zu implementieren, was diese Systeme zu quasi-moralischen Akteuren machen könnte. Mit der Vorstellung, dass in Maschinen die moralischen Regeln der Menschen einprogrammiert werden und diese so zu moralischen Akteuren werden, beschäftigt sich in erster Linie die Maschinenethik. An dieser Stelle muss weitergehend beleuchtet werden, inwiefern sich diese anfangs implementierte Moral durch Künstliche Intelligenz und maschinelles Lernen verändern könnte.

Die mögliche Entwicklung der Moral über ein regelbasiertes System hinaus ist, vor allem für Laien, weder einzuschätzen noch abzusehen. Doch auch schon zum Zeitpunkt der Programmierung können moralisch strittige Absichten in die Reaktionsregister implementiert werden, da Institutionen mit explizitem oder auch implizitem moralischen Anspruch die Sprachdialogsysteme aufsetzen. Moralavers wäre an dieser Stelle beispielsweise eine Art der Datenverarbeitung, die monetäre Interessen über die Privatheit stellt und nicht sensibel genug mit Nutzerdaten verfährt. Hier ist die Systemtransparenz entscheidend, jedoch nicht nur, was die Nutzer*innendaten, sondern auch was die Funktionsweise des Systems und die Explainability betrifft. Die Nutzer*innen sollten im Vorfeld ausreichend über die Datenverarbeitung aufgeklärt werden und dabei in der Lage sein, diese prinzipiell auch nachzuvollziehen. Es lässt sich beobachten, dass Algorithmen und die Funktionsweise künstlicher Intelligenz zwar im Groben bekannt, aber nur selten detaillierter beschreibbar sind. Zumindest die Logik der Entscheidungsfindung solcher Systeme sollte den Nutzer*innen verständlich gemacht werden, um falsche Erwartungen und negative Implikationen zu vermeiden. Auch die Transparenz in Bezug auf die prinzipielle Logik der Entscheidungsfunktionen technischer Systeme (wie kommt welche Aktion des technischen Systems zustande?), spielt eine nicht von der Hand zu weisende Rolle.

Es kommt hinzu, dass ein praktisches Eingreifen der Nutzer*innen im Sinne einer tatsächlichen Handlungsmöglichkeit gegenüber dem System in vielen Fälle gar nicht vorgesehen ist.

Technikphilosophisch liegt hier ein fundamentales Problem zugrunde, das Hubig in [34] als »Indisponibilität der Schnittstellen« markiert hat: 1) Der Mensch wisse nicht genau, was da im maschinellen Gegenüber in der Interaktion vor sich gehe. Das ist letztlich ein programmiertechnisches und kalkulatorisches Problem. Die einzelnen Schritte eines Algorithmus nachzuvollziehen, ist für Menschen extrem schwierig und vor allem nicht in einer sinnvollen Zeit zu realisieren. Fast unmöglich wird das bei selbstlernenden Algorithmen, bei denen sich der Programmcode selbst modifiziert. 2) verfügt der nutzende Mensch nicht oder nur sehr eingeschränkt über die Maschine; er kann und soll sie gerade nicht manipulieren, weil das ihre Funktionalität beeinträchtigen könnte. Also soll er es auch nicht tun. Diese doppelte Indisponibilität gilt auch für den Umgang mit Sprachassistenten; sie verschärft sich noch einmal bei vulnerablen Nutzergruppen.

Die Unternehmen, die Sprachassistenten in Umlauf bringen, verfolgen damit verschiedene Interessen. Man tut ihnen sicher kein Unrecht, wenn man ihnen ökonomische unterstellt. Daneben mögen aber auch moralische Interessen oder zumindest Intuitionen eine Rolle spielen. Soweit sie explizit sind, mögen sie sich u. U. in einprogrammierten Regeln wiederfinden. So ist der häufig vorgetragene Anspruch, mit einem Assistenten die Selbstbestimmung und gesellschaftliche Teilhabe von Menschen mit Hilfebedarf zu stärken zunächst einmal ein moralischer, der grundsätzlich bejaht wird: Es ist (moralisch) gut, Menschen zu helfen.

Wenn nun aber der Assistent die zu unterstützende Person immer zu einer bestimmten Apotheke lotst, weil diese für die Applikation entsprechend bezahlt, so kann nicht grundsätzlich von der moralischen Vorzugswürdigkeit des technischen Systems ausgegangen werden. Darüber hinaus haben die Hersteller von solchen Systemen oder Applikationen bestimmte Vorstellung davon, was gut oder richtig ist. Diese Vorstellungen werden aber nicht unbedingt in einem breiteren ethischen Diskurs überprüft, sondern gehen in den Programmcode ein.
Darüber hinaus ist davon auszugehen, dass auch moralische Hintergrundannahmen, die zumeist implizit sind, in den Programmcode (»Code is Law«, Lawrence Lessig [35]) eingehen. Da sie aber für die Interaktion bedeutsam und folgenreich sind, ist es Aufgabe der Ethik, diese moralischen Implikationen zu explizieren.
Ist es also moralisch richtig, Systeme mit moralischen Fähigkeiten und Fähigkeiten zur Entscheidung auszustatten? Wenn ja, wie sollten diese Systeme beschaffen sein? Auch stellt sich hinsichtlich der Mensch-Technik-Interaktion die daran anschließende Frage, welche Verpflichtungen die Nutzer*innen gegenüber der KI als moralischer und auch sozialer Entität zu erfüllen haben, sofern sie diese als solche betrachten. Vor allem in den vulnerablen Nutzergruppen wird dies aufgrund der fehlenden Systemtransparenz sowie der spezifischen mentalen Modelle der Nutzer*innengruppen der Fall sein. Dies wiederum lässt sich nur adäquat beleuchten, wenn der Beziehungsaufbau im sozio-technischen Arrangement beleuchtet und zu jeder Zeit mitbedacht wird.

Kommunikation, Interaktion und Beziehungsaufbau

Die Kommunikation und Interaktion und somit auch die Beziehung zwischen Mensch und KI ist etwas wesenhaft anderes als das, was allgemeinhin unter dem Begriff *Beziehung* gefasst wird. Bereits beim Beziehungsaufbau beider Entitäten spielt es, so die These, eine nicht von der Hand zu weisende Rolle, ob dem sozio-technischen Arrangement ein Name oder eine schlichte Typenbezeichnung verliehen wird. Ersteres fördert die Individualisierung des Systems und stiftet Beziehung und Dialog. Dem unbelebten Objekt wird so ein Platz im sozialen Gefüge zugewiesen und somit wird auch die Empathie, die die Nutzer*innen gegenüber dem System zu empfinden in der Lage sein können, gefördert. Dass solche technischen Systeme Gefühle und Emotionen, Stimmungen und Sorgen darstellen können, muss als problematisch erachtet werden. Nutzer*innen der Technologie zumal wenn es sich um vulnerable Nutzergruppen handelt – aufgrund fehlender Transparenz und ihrem spezifischen mentalen Modell – könnten dazu neigen, solche Systeme zu vermenschlichen, sie mit unangemessenen Erwartungen zu befrachten, was zu Enttäuschungen führen muss. Das Menschen nicht menschliche Wesen und unbelebte Entitäten immer wieder einmal emotional und sozial besetzen, sie in gewisser Weise vermenschlichen, ist nicht grundsätzlich problematisch, solange es gewissen Entwicklungsschritten entspricht oder als soziale Spiel bewusst inszeniert wird. Moralisch problematisch wird es dort, wo Menschen in ihrer sozialen Orientierung verunsichert oder gar irregeführt werden.
In diesem Zusammenhang sollte beobachtet werden, ob und inwieweit die sprachliche Interaktion zwischen Mensch und Sprachassistenz die menschliche Kommunikation beeinflusst, besonders im Hinblick auf die vulnerablen Nutzergruppen. Wie lassen sich an dieser Stelle negative Folgen vermeiden – und was kann mit welchen Gründen als negative Folge gewertet werden?

Sprachassistenten als Herausforderung für das Recht

Sprachassistenten stellen das Recht vor zahlreiche Herausforderungen – insbesondere das Datenschutzrecht. Das Datenschutzrecht ist für die Verarbeitung von Daten mit Hilfe von Algorithmen, wie sie in Sprachassistenten zum Einsatz kommen, immer dann relevant, wenn die verarbeiteten Daten auf eine bestimmte oder bestimmbare Person zurückzuführen sind. Von besonderer Bedeutung ist die konkrete Ausgestaltung des Assistenten.
Wo befinden sich die gespeicherten Daten und wie werden sie verwendet? Welche Kontrollmöglichkeiten haben die Nutzer?

Massenhafte Datennutzung als Grundlage

Sprachassistenten interagieren mit dem Nutzer und sind potenziell in einem ständigen Datenaustausch mit anderen smarten Geräten. Dies führt zu einer enormen Informationsfülle über den*die Nutzer*in, auf die der Anbieter Zugriff erhält. Gespeichert werden nicht nur Sprachbefehle, sondern unter anderem auch IP-Adresse, Daten über das verwendete Endgerät wie auch Informationen über Shopping- und Suchanfragen [36]. Zudem werden biometrische und damit besonders schützenswerte Daten verarbeitet: die Stimme des*der Nutzer*in. Je mehr Daten vorliegen, desto besser können die Geräte durch Algorithmen lernen und den größten Mehrwert durch bestmögliche Unterstützung des Nutzers schaffen. Werden Sprachassistenten häufig genutzt, sind mehr Daten im Umlauf. Das erhöht wiederum die Wahrscheinlichkeit, dass auch weitere sensible Daten darunter sind oder auf diese geschlossen werden kann, wie etwa Religionszugehörigkeit oder sexuelle Vorlieben. Das Erfordernis großer Datenmengen stellt zudem einen Konflikt mit dem Grundsatz der Datenminimierung gem. Art. 5 Abs. 1 lit. c DS-GVO dar, der eigentlich Datenverarbeitung auf die Zweckerfüllung begrenzen soll. Liegt der Zweck aber in der umfassenden Unterstützung des*der Nutzer*in, so ist gerade keine Begrenzung, sondern eine möglichst breite Datenbasis erforderlich. Der Zweck legitimiert dann die Schaffung dieser Datenbasis.

Vereinbarkeit mit den Grundprinzipien des Datenschutzrechts

Bei der Nutzung von Sprachassistenten handelt es sich im Kern um die Verarbeitung von personenbezogenen Daten in Form von Tonaufzeichnungen. Darunter fallen alle von Personen gesprochenen Befehle und Inhalte sowie alle lautlichen Äußerungen oder verursachten Geräusche. Laut Anbieter werden die Sprachbefehle zwar erst aufgezeichnet und an den Server der Anbieter gesendet, wenn das Signalwort fällt. Da die Sprachassistenten jedoch aus dem Stand-by-Modus mit dem Signalwort aktiviert werden müssen, muss das Gerät stets mithören, um zu erkennen, wann das Signalwort fällt.
Das Datenschutzrecht fordert unter anderem Transparenz der Datenverarbeitung. Derjenige, der personenbezogene Daten erhebt, ist deshalb verpflichtet, den Betroffenen zum Zeitpunkt der Erhebung über die Datenverarbeitung, den Zweck dessen, den Verantwortlichen und wie der*die Betroffene diesen erreichen kann zu informieren. Die Herstellung von Transparenz gestaltet sich insbesondere bei komplexen Systemen aber schwierig und ist regelmäßig mangelbehaftet [37].

So ist oftmals unter anderem aufgrund der Angabe weiter Zwecke der Verarbeitung (etwa „Verbessern des Nutzererlebnis“) nicht eindeutig erkennbar, wie und wo die Daten verwendet werden, zumal regelmäßig neben der Verarbeitung im Gerät über die Internetanbindung eine Verarbeitung auf den Servern des Anbieters stattfindet.
Liegen die Daten einmal auf den Servern der Anbieter vor, so können diese auch mit Dritten geteilt werden oder mit Daten aus anderen Quellen zusammengeführt werden. Beispielsweise können Daten aus der Nutzung einer Suchmaschine mit denen aus der Nutzung eines Sprachassistenten zusammengeführt werden. Die Hersteller haben ein großes Interesse an den hochsensiblen Daten und den Möglichkeiten zur Wertschöpfung aus diesen Daten. Diese werden anschließend von ihnen selbst oder von Dritten beispielsweise für Marketingzwecke oder zum Schalten personalisierter Werbung mittels individualisierter Nutzerprofile genutzt.
Transparenz bedeutet aber auch, dass sich ein Sprachassistent etwa in einem Telefongespräch mit einem Menschen als Maschine identifizieren müsste.
Da die großen Anbieter von Sprachassistenten vornehmlich aus den USA stammen, haben die dortigen Behörden wiederum Zugang auf die von den Anbietern gesammelten Daten. Europäische Nutzer*innen unterliegen hinsichtlich des Zugriffs lediglich dem Schutz durch das umstrittene EU-US-Privacy-Shield-Abkommen ([38] Art. 44 Rn. 14f.). Daten aus Sprachassistenten wurden in den USA bereits in Strafprozessen als Beweismittel eingebracht. Gleichzeitig fordern auch die deutschen Innenminister eine Stärkung von Ermittlungsbefugnissen [39].

Weitere Rechtsprobleme

Probleme für das Recht ergeben sich aber auch jenseits des Datenschutzrechts. So werden etwa Haftungsfragen aufgeworfen, die sich etwa aus einem Fehlgebrauch durch den*die Nutzer*in, aus Programmierfehlern oder aus dem Lernprozess des Systems ergeben können. Ein Beispiel ist ein Assistent, der laut Musik abzuspielen beginnt, obwohl niemand in der Wohnung ist, und dadurch einen Polizeieinsatz auslöst.
Auch die Nutzung von Sprachassistenten im E-Commerce wirft Fragen auf; so etwa die Frage, wie der*die potenzielle Käufer*in unmittelbar vor dem Kauf mit den gesetzlich vorgeschriebenen Informationen versorgt werden kann. In Deutschland muss der*die Verkäufer*in trotz reduzierter Anforderungen für Einkäufe per Sprachbefehl Angaben zum Produkt, zum Preis, zur Identität des*der Verkäufer*in und zum Widerrufsrecht machen. In einem ähnlich gelagerten Fall wurde bereits der Amazon Dash Button von deutschen Gerichten für illegal erklärt. Die Schlussfolgerung ist, dass gesetzliche Anforderungen eine reibungslose Bestellung per Sprachbefehl verhindern.
Auch das Telekommunikationsrecht kann mit bestimmten Formen von Sprachassistenten in Konflikt geraten. So wurde etwa eine interaktive Puppe als verbotene Sendeeinrichtung eingestuft. Hier wurde unter anderem bemängelt, dass sich die Übertragung der Daten aus der Puppe leicht hacken ließe. Vor allem sei aber nicht ersichtlich, dass die Puppe ein Mikrofon in sich trage und diese deshalb zur heimlichen Überwachung geeignet [40].

Vulnerable Nutzergruppen – Kinder und Senioren

Wie bei den bisherigen Betrachtungen bereits angeklungen ist, gilt es ein besonderes Augenmerk auf Nutzergruppen zu legen, die potenziell besonders von negativen Folgen betroffen sein könnten. Dazu zählen ältere Personen und Kinder, da für beide Gruppen ein eher gering ausgeprägtes Verständnis von technologischen Geräten und KI angenommen werden kann. Es kann daher stärker als bei technikversierteren Personen zu Fehleinschätzungen kommen, die nicht nur für einen unbedachteren Umgang mit der Technologie sorgen können, sondern auch psychologische Folgen haben können, zum Beispiel für die Beziehungsbildung.

Kinder in Interaktion mit Sprachassistenten

Kinder, die mit im Haushalt leben, können ebenfalls mit dem Sprachassistenten in Berührung kommen. Sie sind aus rechtlicher Sicht eine besonders schutzwürdige Gruppe, da sie, je nach Alter und Stand der Entwicklung, die Folgen ihres Handelns nicht vollständig verstehen und abschätzen können (BMJV, 2017). Hieraus ergeben sich besondere rechtliche Anforderungen. Probleme beginnen bereits bei der Altersfeststellung und reichen hin zur Nutzung von Sprachassistenten durch Kinder ohne Einwilligung der Eltern.

für Werbezwecke oder
für die Erstellung von Persönlichkeits- oder Nutzerprofilen

Die Datenschutz-Grundverordnung enthält an zahlreichen Stellen besondere Vorgaben für die Verarbeitung personenbezogener Daten von Kindern oder zumindest Aufforderungen zur Beachtung der besonderen Bedürfnisse von Kindern. Insbesondere bei der Verwendung solcher Daten für Werbezwecke oder für die Erstellung von Persönlichkeits- oder Nutzerprofilen solle für einen Schutz von Kindern gesorgt werden. Zudem soll die Sprache bei Informationen zur Datenverarbeitung kindgerecht gestaltet sein.

Erste Sprachassistenten speziell für Kinder als Zielgruppe sind bereits am Markt erhältlich, so etwa Amazons Echo Dot Kids Edition. Das Gerät soll mit kindgerechten Anwendungen auf die Bedürfnisse von fünf- bis zwölfjährigen Nutzer*innen zugeschnitten sein. Außerdem kann es höfliche Umgangsformen in der Kommunikation des Kindes mit dem Gerät anmahnen. Weitere Einstellmöglichkeiten geben den Eltern Kontrollmöglichkeiten, insbesondere auch zur Überwachung der Aktivitäten des Kindes.

Eine Studie der EU-Kommission zu Smart Toys (JRC Technical Reports: Kaleidoscope on the Internet of Toys, 2017) ergab signifikante Mängel im Bereich des Datenschutzes für Kinder, insbesondere bezogen auf Transparenz der Datenverarbeitung. Auch der Wissenschaftliche Dienst des Bundestages äußerte sich kritisch. Es bleibe offen, „wie unbeteiligte Dritte und Minderjährige von der Datensammlung ausgeschlossen werden können".

Im Hinblick auf die Anwendungsbereiche bei Kindern ist davon auszugehen, dass interaktive Sprachsysteme vor allem Entertainmentzwecken dienen. Da liegt nahe, dass aus psychologischer Sicht vor allem die potenzielle Beziehungsbildung betrachtet werden muss. Kinder haben ohnehin eine stärkere Tendenz dazu, unbelebten Objekten (wie beispielsweise Spielzeugen) einen Charakter zuzuschreiben und diese zu anthropomorphisieren, welches im Laufe des Erwachsenwerdens in der Regel abnimmt [41]. Dementsprechend ist davon auszugehen, dass die sozialen Hinweisreize, die durch eine künstliche Entität gesendet werden, in verschiedenen entwicklungspsychologischen Phasen zu unterschiedlichen Annahmen über das Wesen einer künstlichen Entität führen.

Erste Studien, in denen Kinder mit Robotern konfrontiert worden sind, deuten auf Unterschiede in der Wahrnehmung des Roboters basierend auf dem Alter der Kinder hin. So konnten Kahn et al. in [42] zeigen, dass 9-12 jährige Kinder, im Gegensatz zu 15 jährigen, der künstlichen Entität signifikant stärker mentale Zustände (z.B. in Bezug auf die Fähigkeit, Gefühle zu haben) zugesprochen haben und stärker davon überzeugt waren, dass der Roboter eine Art soziales Wesen ist.
Die jüngeren Kinder waren ebenfalls stärker überzeugt, dass der Roboter beispielsweise als eine Art Freund angesehen werden kann und ihm damit einhergehend auch Geheimnisse anvertraut werden können. Hierbei ist es wichtig zu erwähnen, dass es sich um ein physisch präsentes Objekt handelte, welches über anthropomorphe Merkmale verfügte (u. a. menschenähnliche Statur mit „künstlichen Augen").
Verschiedene Studien konnten zudem zeigen, dass sich das Verständnis einer künstlichen Entität (Computer, Roboter etc.) mit dem Alter wandelt und schließlich zu einer Wahrnehmung als intelligentes Objekt führt, dass es jedoch an einem konkreten Verständnis im Hinblick auf Programmierung fehlt, welches wiederrum mit Problemen bei der Kategorisierung der Objekte einhergeht (z.B. [43]; [44]).
Zusammenfassend lässt sich somit schließen, dass soziale Hinweisreize einer Maschine gerade bei jüngeren Kindern dazu führen können, dass nicht nur unmittelbare soziale Reaktionen erfolgen, sondern einer künstlichen Entität auch bewusst Attribute zugeschrieben werden, welche sonst menschlichen Interaktionspartnern vorbehalten sind und beispielsweise „Freundschaften" angestrebt werden können. Gerade das im jungen Alter noch gering ausgeprägte Verständnis über dahinterliegende technologische Prozesse wird eine wichtige Rolle im Hinblick auf eine adäquate Kategorisierung künstlicher Interaktionspartner spielen.
Zusätzlich kann und muss jedoch auch reflektiert werden, dass eine Beziehungsbildung nicht immer negativ bewertet werden muss. Der Einsatz künstlicher Interaktionspartner in Therapiesettings lebt beispielsweise vom Aufbau einer tragfähigen sozio-emotionalen Beziehung.
Diese und weitere Themen müssen vor dem Hintergrund ethischer Expertise diskutiert werden. Zu den weiteren Themen die aus ethischer Sicht zentral sind, gehören der Einfluss auf die Moralentwicklung von Kindern, das Kindeswohl oder die Grenze zwischen Erziehung und Nudging. Auch hinsichtlich der Transparenz der algorithmischen Prozesse und einer entsprechenden kindgerechten Vermittlung müssen ethische Betrachtungen erfolgen. Auch die Rolle der Eltern ist aus ethischer Sicht thematisierbar: Welche Verantwortung haben diese bzw. welche Interessen verfolgen sie, wenn ein Gerät angeschafft wird, das auch die Kinder nutzen?
Es stellen sich somit noch zahlreiche Fragen, die durch ein Zusammenspiel rechtlicher, ethischer und psychologischer Forschung beantwortet werden müssen.

Senioren in Interaktion mit AAL

Eine weitere Nutzergruppe, in welcher die Nutzungsszenarien und damit auch die soziale Bedeutung von jener der Erwachsenen mittleren Alters abweichen könnte, sind Senioren (mit und ohne kognitive Einschränkungen).

Vulnerable Nutzergruppen – Kinder und Senioren

So ist es beispielsweise denkbar, dass interaktive Sprachassistenten den Alltag von Menschen im fortgeschrittenen Alter bzw. bei kognitiven Einschränkungen insofern erleichtern, als dass diese an die Einnahme von Tabletten erinnern oder Termine koordinieren (z.B. [45]). Jedoch können auch die sozialen Eigenschaften des Systems für Senioren oder kognitiv eingeschränkte Personen von Relevanz sein, da besonders in diesen Gruppen Menschen häufiger an Einsamkeit oder auch gesundheitsbedingter Isolation leiden [46].

Kopp und Kollegen [45] konnten beispielsweise zeigen, dass virtuelle Agenten in ersten Langzeiteinsätzen von zwei Wochen durchaus als soziale Interaktionspartner genutzt wurden. So wurde das System vor allem dazu eingesetzt Termine zu koordinieren, wurde aber von einigen Nutzer*innen auch in eine Art morgendliches Begrüßungsritual einbezogen.

Die soziale Kategorisierung einer künstlichen Entität wurde bisher vor allem im Hinblick auf den Roboter Paro untersucht.

Hierbei war bei Nutzer*innen, die unter Demenz leiden, scheinbar durchaus nicht immer deutlich, ob es sich bei dem Roboter um einen künstlichen oder lebenden Interaktionspartner handelte [47]. Gerade im Zusammenspiel mit ersten kognitiven Einschränkungen lassen sich somit Unsicherheiten im Hinblick auf die soziale Kategorisierung auch von künstlichen Interaktionspartnern wie Sprachassistenten erwarten. Aus ethischer Perspektive ist hier zu erforschen, ob und ab welchem Grad die Vermenschlichung des Systems zum Beispiel zu emotionalem Missbrauch führt, der die Menschenwürde untergräbt.

Hinzu kommt, dass weitere Forschung zum Wissen von Senioren über neue Technologien zeigt, dass sich viele ältere Nutzer*innen in Bezug auf ihr Verständnis über Prozesse innerhalb eines Computers nicht selbstsicher fühlen [48]. Es kann also davon ausgegangen werden, dass eine Transparenz über Algorithmen bei älteren Nutzer*innen schwieriger zu erreichen ist. Hier muss aus ethischer Sicht diskutiert werden, wieviel Aufklärung über die Mechanismen notwendig ist und welcher Grad an Autonomie des Nutzenden nicht unterschritten werden darf.

Diese Fragen sind letztlich verwandt mit denen, die sich aus rechtlicher Sicht stellen: Es muss ein grundsätzliches Verständnis für die Sprachassistenten vorliegen, um eine selbstbestimmte Nutzung zu ermöglichen. Weitere wichtige rechtliche Themen beziehen sich auf den Datenschutz, da die erhobenen Daten häufig Rückschlüsse auf die persönliche Gesundheit zulassen, weshalb diese einen besonderen Schutz erfordern. Ein Problem stellt zudem dar, dass die Einwilligungsfähigkeit von älteren Menschen in bestimmten Fällen unklar sein kann.

Ältere Menschen sind rechtlich in der Regel nicht anders gefasst als andere Erwachsene. Insbesondere Art. 25 der Charta der Grundrechte der Europäischen Union, der ein Recht älterer Menschen auf ein würdiges und unabhängiges Leben und auf Teilhabe am sozialen und kulturellen Leben enthält, kann aber durchaus als Gestaltungsauftrag an Hersteller auch von Sprachassistenten im Sinne eines Rechts auf Teilhabe am technischen Fortschritt verstanden werden.

Auch für in der Regel technikferne Nutzergruppen wie Senioren muss die Nutzung von Sprachassistenten demnach transparent, leicht verständlich sowie nachvollziehbar gestaltet werden.
Bei der Gruppe der Älteren muss also sowohl aus rechtlicher als auch aus ethischer und psychologischer Perspektive zunächst besonders sorgfältig abgewogen werden, unter welchen Umständen ältere Personen als besonders vulnerable Personengruppe verstanden werden können.
Beide herausgehobenen Nutzergruppen könnten (in einem nutzergruppengerechten Rahmen) von der sogenannten Explainable AI profitieren. Hierbei handelt es sich um Erklärungen, welche ein System über die Verarbeitung eines Inputs macht. Durch diese Erklärungen könnte zudem der Unterschied zu zwischenmenschlichen Interaktionen deutlich gemacht werden, welches wiederum eine eindeutige soziale Kategorisierung des Interaktionspartners zulassen würde.

Ausblick

Der Überblick über die informatischen, psychologischen, ethischen und rechtlichen Dimensionen der Nutzung von Sprachassistenten zeigt einerseits, dass einige Effekte und Implikationen bekannt, dass aber andererseits zahlreiche wichtige Fragen noch unbeantwortet sind. Es herrscht also Gestaltungsbedarf und durchaus auch Gestaltungsmöglichkeit von Sprachassistenten. Da es sich um ein sozio-technischem System handelt, bei dem weitergehende Effekte betrachtet werden müssen als zum Beispiel bloße Akzeptanz und Benutzerfreundlichkeit des Systems, ist die Liste an Forschungsdesiderata komplex.

- Allgemein formuliert müssen die Auswirkungen von sozio-technischen KI Systemen besser verstanden werden, um nachteilige Effekte verhindern zu können.
- Das Problem, dass die Prozesse und Mechanismen mindestens für technische Laien nicht durchschaubar sind (Transparenzproblem), muss gelöst werden.
- Implikationen für Privatheit und Datenschutz müssen nicht nur aus rechtlicher, ethischer und psychologischer Perspektive betrachtet werden, sondern auch Niederschlag in technischen Gestaltungen finden.
- Die Auswirkungen der Systeme im Hinblick auf soziale Kategorisierung als „Companion“ und hinsichtlich der Beziehungsentwicklung müssen erforscht und hinterfragt werden.
- Die Effekte auf das zwischenmenschliche Kommunikationsverhalten durch verstärkte Kommunikation mit Maschinen müssen analysiert und reflektiert werden.
- Vulnerable Gruppen (wie Kinder oder ältere Personen) müssen in der Technikentwicklung besonders berücksichtigt werden.
- Geschäftsmodelle müssen hinterfragt werden.

Literatur

Zur Lösung der aufgeworfenen Probleme ist eine interdisziplinäre Technikgestaltung erforderlich. Die hier untersuchten Thesen und die aufgeworfenen Fragestellungen untersucht das Forschungsprojekt IMPACT („The implications of conversing with intelligent machines in everyday life for people´s beliefs about algorithms, their communication behavior and their relationship building"). Das Projekt wird von der Universität Duisburg-Essen (Sozialpsychologie - Medien und Kommunikation), der Universität Bielefeld (Social Cognitive Systems Group und Machine Learning Group), der Evangelischen Hochschule Nürnberg (Anthropologie und Ethik) und der Universität Kassel (Projektgruppe verfassungsverträgliche Technikgestaltung - provet) durchgeführt. Das Projekt ist Teil der Förderinitiative „Künstliche Intelligenz – Ihre Auswirkungen auf die Gesellschaft von morgen" der VolkswagenStiftung und wird von 2019 bis 2024 gefördert.

Literatur

[1] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.

[2] Y. Tabet and M. Boughazi, "Speech synthesis techniques. A survey," in *Proc. 7th Int. Workshop Systems, Signal Process.and their Appl., WoSSPA 2011*, 2011, pp. 67–70.

[3] A. Van Den Oord *et al.*, "Parallel WaveNet: Fast high-fidelity speech synthesis," in *Proc. 35th Int. Conf. Mach. Learning, ICML 2018*, 2018, vol. 80, pp. 3918–3926.

[4] A. Hannun *et al.*, "Deep Speech: Scaling up end-to-end speech recognition," *ArXiv e-prints*, 2014.

[5] Y. Wu *et al.*, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *ArXiv e-prints*, 2016.

[6] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, no. July, pp. 1–13, 2018.

[7] H. Jin and S. Wang, "Voice-based determination of physical and emotional characteristics of users," U.S. Patent US10096319B1, 2017.

[8] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition,* vol. 44, pp. 572–587, Mar. 2011.

[9] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for Speech Emotion Recognition," *Neural Netw., vol. 92, Aug. 2017.*

[10] P. Barros, G. I. Parisi, and S. Wermter, "A Personalized Affective Memory Neural Model for Improving Emotion Recognition," *ArXiv e-prints*, 2019.

[11] Y. Leviathan and Y. Matias, "Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone," *Google AI Blog*, 2018. [Online]. Available: https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html. [Accessed: 19-Aug-2019].

[12] G. Munster and W. Thompson, "Annual Smart Speaker IQ Test," 2018. [Online]. Available: https://loupventures.com/annual-smart-speaker-iq-test/. [Accessed: 19-Aug-2019].
[13] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016.
[14] M. Biehl, B. Hammer, and T. Villmann, "Prototype-based models in machine learning," *Wiley Interdiscip. Rev. Cognitive Sci., vol. 7, no. 2, pp. 92–111, 2016*.
[15] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.
[16] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Appl. Sci.*, vol. 9, no. 5, 2019.
[17] Y. Geifman and R. El-Yaniv, "SelectiveNet: A Deep Neural Network with an Integrated Reject Option," *ArXiv e-prints*, 2019.
[18] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine Bias," *ProPublica*, 2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [Accessed: 19-Aug-2019].
[19] P. Hacker, "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law," *Common Market Law Rev.*, vol. 55, pp. 1143–1186, 2018.
[20] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," *ACM Comput. Surveys.*, vol. 51, no. 5, Aug. 2018.
[21] T. Miller, P. Howe, and L. Sonenberg, "Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences," in *Proc. IJCAI 2017 Workshop Explainable Artificial Intell.*, 2017.
[22] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *ArXiv e-prints*, 2017.
[23] B. Reeves and C. I. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press, 1996.
[24] K. Dautenhahn, "Socially intelligent robots: Dimensions of human-robot interaction," *Philosoph. Trans. Roy. Soc. B: Biol. Sci.*, vol. 362, no. 1480, pp. 679–704, 2007.
[25] T. Bickmore, D. Schulman, and L. Yin, "Engagement vs. deceit: Virtual humans with human autobiographies," in *Proc. Int. Workshop Intell. Virtual Agents*, 2009, pp. 6–19.
[26] A. M. Rosenthal-von der Pütten, L. Hoffmann, J. Klatt, and N. Krämer, "Quid Pro Quo? Reciprocal Self-disclosure and Communicative Accomodation towards a Virtual Interviewer," in *Proc. Int. Workshop Intell. Virtual Agents*, 2011, vol. LNAI 6895, pp. 183–194.
[27] A. M. Rosenthal-von der Pütten, N. C. Krämer, C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro, "The Uncanny in the Wild. Analysis of Unscripted Human-Android Interaction in the Field," *Int. J. Social Robot.*, vol. 6, no. 1, pp. 67–83, 2014.

Literatur

[28] N. Krämer, S. Kopp, C. Becker-Asano, and N. Sommer, "Smile and the world will smile with you - The effects of a virtual agent's smile on users' evaluation and behavior," *Int. J. Human-Comput. Stud.*, vol. 71, pp. 335–349, 2013.

[29] S. Lee, R. Ratan, and T. Park, "The Voice Makes the Car: Enhancing Autonomous Vehicle Perceptions and Adoption Intention through Voice Agent Gender and Style," *Multimodal Technol. Interact.*, vol. 3, no. 1, p. 20, 2019.

[30] S. Alesich and M. Rigby, "Gendered Robots: Implications for Our Humanoid Future," *IEEE Technol. Soc. Mag.*, vol. 36, no. 2, pp. 50–59, 2017.

[31] M. West, R. Kraut, and H. Ei Chew, "I'd blush if I could: closing gender divides in digital skills through education," EQUALS Skills Coalition, 2019.

[32] B. Benninghoff, P. Kulms, L. Hoffmann, and N. Krämer, "Theory of mind in human-robot-communication: Appreciated or not?," in *Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten*, 2013, vol. 1.

[33] M. A. DeVito, J. Birnholtz, J. T. Hancock, M. French, and S. Liu, "How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation," in *Proc. ACM CHI Conf. Human Factors Comput. Systems*, 2018.

[34] C. Hubig, "Der technisch aufgerüstete Mensch – Auswirkungen auf unser Menschenbild," in *Digitale Visionen*, A. Roßnagel, T. Sommerlatte, and U. Winand, Eds. Berlin, Heidelberg: Springer, 2008, pp. 165–176.

[35] L. Lessig, *Code: And Other Laws of Cyberspace*. New York, NY: Basic Books, 1999.

[36] J. Heidrich and N. Maekeler, "Alexa, darfst du das?," *c't magazin für computertechnik*, vol. 22, pp. 86–87, 2017.

[37] A. Roßnagel, C. Geminn, S. Jandt, and P. Richter, *Datenschutzrecht 2016 - "Smart" genug für die Zukunft?* Kassel: Kassel University Press, 2016.

[38] G. Sydow, Ed., *Europäische Datenschutzgrundverordnung. Handkommentar*, 2nd ed. Nomos, 2018.

[39] "Innenminister wollen Daten von Smart-Home-Geräten nutzen," *ZD-Aktuell*, p. 06669, 2019.

[40] "Faktenblatt Smartes Spielzeug," Bundesministerium der Justiz und für Verbraucherschutz, 2017.

[41] M. S. Geerdts, "(Un)Real Animals: Anthropomorphism and Early Learning About Animals," *Child Dev. Perspectives*, vol. 10, no. 1, pp. 10–14, 2016.

[42] P. H. Kahn *et al.*, "'Robovie, you'll have to go into the closet now': Children's social and moral relationships with a humanoid robot," *Dev. Psychol.*, vol. 48, no. 2, pp. 303–314, 2012.

[43] M. Van Duuren, B. Dossett, and D. Robinson, "Gauging Children's Understanding of Artificially Intelligent Objects: A Presentation of 'Counterfactuals,'" *Int. J. Behavior Dev.*, vol. 22, no. 4, pp. 871–889, 1998.

[44] M. T. Rücker and N. Pinkwart, "Review and Discussion of Children's Conceptions of Computers," *J. Sci. Educ. Technol.*, vol. 25, no. 2, pp. 274–283, 2016.
[45] S. Kopp *et al.*, "Conversational assistants for elderly users – The importance of socially cooperative dialogue," in *Proc. CEUR Workshop*, 2018, vol. 2338, pp. 10–17.
[46] A. P. Dickens, S. H. Richards, C. J. Greaves, and J. L. Campbell, "Interventions targeting social isolation in older people: A systematic review," *BMC Public Health*, vol. 11, no. 1, p. 647, 2011.
[47] K. Wada, T. Shibata, T. Musha, and S. Kimura, "Robot therapy for elders affected by dementia," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 4, pp. 53–60, 2008.
[48] J. C. Marquié, L. Jourdan-Boddaert, and N. Huet, "Do older adults underestimate their actual computer knowledge?," *Behav. Inform. Technol.*, vol. 21, no. 4, pp. 273–280, 2002.

IMPACT

Ein gemeinsames Projekt von

UNIVERSITÄT DUISBURG ESSEN

UNIVERSITÄT BIELEFELD

UNIKASSEL VERSITÄT

Evangelische Hochschule Nürnberg

Gefördert durch

VolkswagenStiftung

DOI: 10.17185/duepublico/70571
URN: urn:nbn:de:hbz:464-20191004-091327-7