

# Wenn Maschinen Stimmen klonen

Sprach-KI macht ihre Nutzer angreifbarer

Von *Alexei Makartsev*

Die Tochter ruft bei ihrem Vater an und sagt, sie bräuchte Geld, er solle nicht mit ihrer Mutter darüber sprechen. Da wird der Pensionär hellhörig: Seine Frau ist verstorben. Klingt bekannt? Ist auch eine Form des alten „Enkeltricks“ – in einer neuen Version. Denn die Stimme der „Anruferin“ ist von der Angehörigen des Opfers ununterscheidbar. Die Computerfirma McAfee schildert diesen Fall in einer Studie mit dem Titel: „Vorsicht vor künstlichen Schwindlern“. Nach ihren Angaben nimmt der Sprachbetrug mithilfe von Künstlicher Intelligenz (KI) zu. Dafür gebe es zwei Gründe: Zum einen teilen immer mehr Menschen ihre Stimm- und Sprachaufzeichnungen via Whatsapp und auf Plattformen wie Tiktok. Zum anderen seien KI-basierte Programme zum Kopieren von Stimmen weit verfügbar. McAfee gelang es, mit solchen Tools aus einer Aufnahme einen Stimmklon zu erzeugen. Laut einer Umfrage der Firma haben 23 Prozent der Bundesbürger Erfahrungen mit virtuellem Schwindel gemacht. Drei von vier Betroffenen haben dabei Geld verloren.

**„KI-generierte Stimmen beeinflussen uns.“**  
*Carolin Wienrich*  
Psychologin an der Uni Würzburg

Die Verbreitung von sprachfähigen Maschinen biete ein großes Missbrauchspotenzial, sagt Jörg Eisele. Der Tübinger Uni-Professor ist Fachmann für Computerstraftaten. Er war zuvor einer Verschärfung des „Enkeltricks“, wenn die KI, eine Vielzahl von Informationen über das Opfer und seine Familie analysiert, um dann am Telefon mit der geklonten Stimme Fragen zu stellen und plausibel zu reagieren“. Man könnte dann nicht durch Gesprächsinhalte herausfinden, ob am anderen Ende der Leitung die vertraute Person oder eine Maschine sei, glaubt Eisele.

Hilfreich wäre deswegen ein Codewort zur sicheren Identifizierung, das man mit seiner Familie vereinbart. „Das sollten keine Passwörter sein, sondern spezielle Informationen, die im Netz nicht verfügbar sind“, rät der Experte. „Ob eine ältere Person in der Aufregung an das Codewort denkt, wenn sie mit solchen Anrufen konfrontiert wird, ist aber fraglich.“

Der Rechtswissenschaftler sieht einen weiteren Grund, um die Evolution von KI-Sprachsystemen kritisch zu hinterfragen: Sie könnten dazu genutzt werden, um Falschnachrichten zu produzieren oder um Menschen zu diffamieren, indem man sie öffentlich etwas sagen lässt, was nie gesagt wurde.

„Wenn die KI ihre Opfer selbst aussucht, könnte man die Täterakte kaum nachverfolgen“, sorgt sich Eisele. Würde uns die Justiz schützen? Er ist skeptisch. „Wer Straftaten mit KI begehen will, wird die verbotenen Tools bekommen, weil es Entwickler gibt, die Systeme genau zu diesen Zwecken herstellen.“ Auch abseits des Strafrechts wirft die rasante Entwicklung der Sprachtechnologien zahlreiche heikle Fragen auf. Zum Beispiel diese: Wenn die beliebtesten Assistenzsysteme immer mehr persönliche Daten von Nutzern erfassen und erzeu-

gen – wird dann der Schutz der Privatsphäre ausgehebel? Der Datenschutzbeauftragte von Baden-Württemberg, Jan Wacke, nimmt diese Gefahr ernst. Sobald ein Smart Speaker durch einen Befehl (Triggerwort) aktiviert werde, sende er an den KI-Hersteller nicht nur alles, was die Nutzer sagen, sondern auch wie sie sich verhalten und wo sie sich befinden, schildert Wacke. Das alleine könne schon problematisch sein.

„Wenn aber die Erkennung der Triggerwörter nicht gut funktioniert und die Maschine etwas falsch versteht, könnte auch ungewollt eine Sprachübertragung in die Cloud beginnen. Weil die Geräte manchmal in den intimsten Räumen stehen, sind das teils sehr sensible Daten. Werden sie weiter verarbeitet, kann das sehr unangenehm sein.“

Was passiert mit den Daten, wo und wie lange werden sie gespeichert? Laut Wacke ist das nicht klar. Die Nutzer könnten jedoch diese Fragen den Firmen stellen. „Wer in Europa solche Geräte anbietet, muss die Betroffenenrechte gewährleisten.“ Der Datenschutzbeauftragte sieht seine Behörde „als Kapazitätsmangel“ nicht imstande, sprechende Maschinen zu kontrollieren. Umso wichtiger sei es, „die Menschen zu sensibilisieren und eine Kultur des bewussten Umgangs mit neuen Technologien zu fördern“.

Das ist nicht trivial, wie Psychologen wissen. Wer von KI-Lösungen überzeugt ist, neigt dazu, sie für vertrauenswürdig zu halten. Was menschlich wirkt, behandeln wir automatisch als Menschen. In den Augen von Carolin Wienrich kann diese Personalisierung eine Falle sein. Zum einen, weil auch smarte Maschinen Fehler machen würden und dies nicht zugeben. „Außerdem beeinflussen sie uns“, sagt die Professorin für Psychologie in der Interaktiver Systemen an der Uni Würzburg. Heute lassen sich Emotionen in synthetischen Stimmen programmieren. „Eine niedliche Stimme vermittelt nicht den Eindruck, dass sie Böses im Sinn hat, wodurch es möglich ist, Informationen zu entlocken, die sonst nicht preisgegeben würden“, erklärt Wienrich. Nutze man aber solche Technik im Polizeibereich oder der Justiz, „können künstliche Stimmen unter Umständen bewusst eingesetzt werden, um Sachlichkeit oder auch Autorität zu vermitteln.“

Eine andere Möglichkeit biete sich Firmen durch die Analyse von Nutzerstimmen an. „Moderne Computersysteme können in ihnen Traurigkeit, Freude oder depressive Symptome erkennen“, sagt die Psychologin. „Die KI kann sich dann ähnlich verhalten, sodass der Mensch ein Gefühl hat: Da ist jemand, dem es genauso geht. Dann gibt er vielleicht mehr über sich preis. Wenn jemand glücklich ist, steigt seine Kaufbereitschaft, was mir ermöglicht, gezielte Angebote zu machen.“

Je ähnlicher Mensch und Maschine wirken, desto wichtiger wird ihre Unterscheidbarkeit. Es gibt KI-Systeme, die andere neuronale Netzwerke als Urheber von Bildern und Texten erkennen. Vielleicht lassen sich bald so auch synthetische Stimmen identifizieren. Die meisten Experten stimmen überein, dass wir uns auf die neuen Technologien besser einstellen müssen. Digitale Kompetenz sei wichtiger denn je zuvor. „KI kann gut eingesetzt werden, um Menschen zu assistieren“, sagt der Datenschutzbeauftragte Jan Wacke. „Der Mensch muss aber die maßgeblichen Entscheidungen treffen.“



Der digitale Küchenhelfer – ein Freund oder ein Spion? Wenn wir mit smarten Geräten plaudern, kann das nützlich sein. Doch die Technik hat Schattenseiten. Foto: Andrea Warnecke/dpa

# Vielseitige Helfer: Assistenten mit Künstlicher Intelligenz gehören zum Alltag



Die in einem Lautsprecher verbaute, virtuelle Helferin Alexa ist jederzeit ansprechbar. Smarte Geräte können menschliche Fähigkeiten immer besser imitieren. Schon heute sind natürliche Gespräche mit ihnen möglich. Foto: Florian Schuh/dpa

# Die Zukunft spricht mit unseren Stimmen

## Digitale Assistenten sind allgegenwärtig, Künstliche Intelligenz macht sie fast allwissend – wo führt das nur hin?

*Von Alexei Makartsev*

Es war einmal eine Puppe, die nicht still sitzen wollte. „Gehen wir spazieren!“, fragte freudig Julie. „Kannst du wie ein Frosch quacken? Quack!“ Julius Schwester Jill war noch viel gesprächiger. „Hast du Bananen? Soll ich Gruselgeschichten erzählen? Was für ein toller Tag“, plapperte die blonde Spieglefahrerin unentwegt, während ihre großen blauen Augen die Welt anstrahlten.

Julie und besonders Jill waren wohl die menschenähnlichsten Spielzeuge der 1980er. Zwei interaktive Roboter-mädchen, die auf gesprochene Worte wie „Hunger, Modeschau, Methode“ reagierte und je nach Entscheidung des Menschen ihr Verhalten scheinbar änderten. In einer Zeit, in der die Spracherkennung in den Kinderschuhen steckte, empfanden das viele als ein Wunder. Heute wirkt diese Technik sehr primitiv. Nicht einmal 40 Jahre nachdem die ersten kommunikativen Puppen in die Kinderzimmer einzogen, stehen den meisten Bundesbürgern zu jeder Uhrzeit sprechende virtuelle Assistenten zur Verfügung. Siri berechnet in 21 Sprachen Routen, misst Backzutaten ab und sagt Gedichte auf. Alexa erinnert an Termine und recherchiert das Kinoprogramm. Cortana organisiert das Arbeiten am PC, Google hilft beim Fitnesstraining und ist für die Video-Unterhaltung danach zuständig.

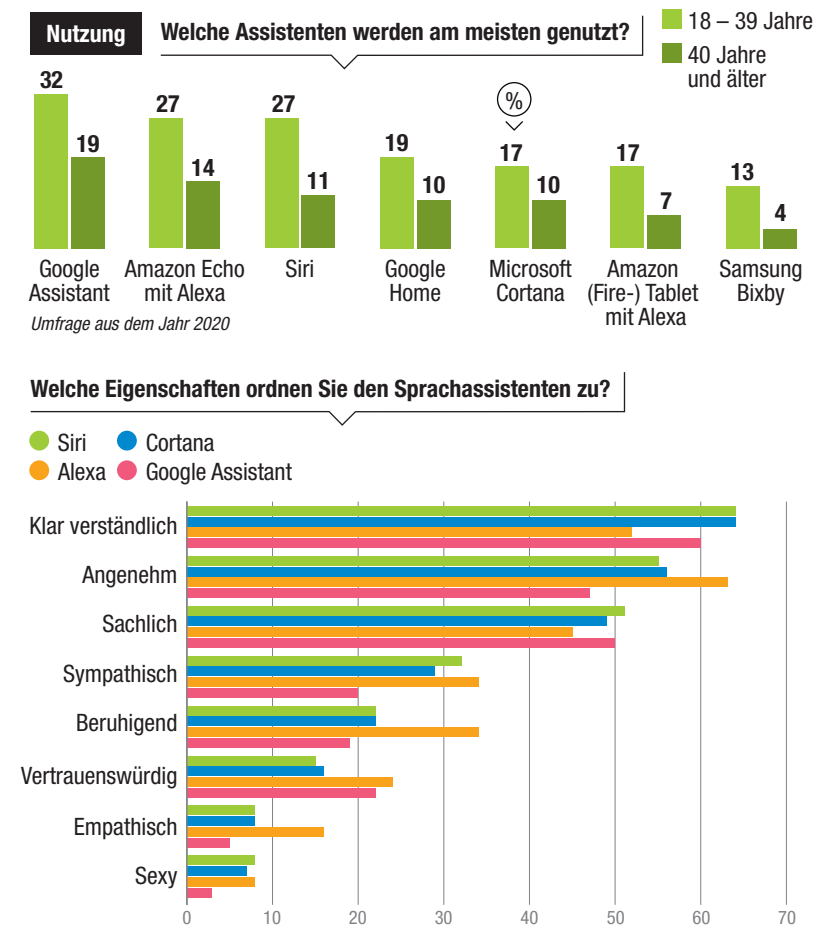
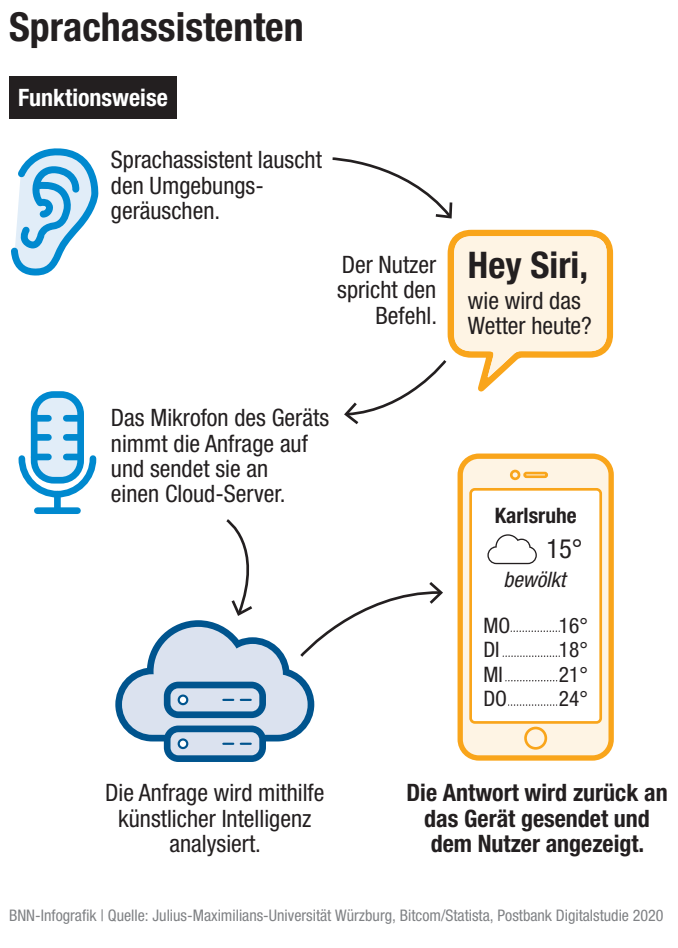
Es ist ein alter Traum der Menschen, Dinge mit Worten zu beherrschen. „Sesam, öffne dich!“ – und der Berg im Märchen gibt den Weg zu dem Schatz frei. Heute kann man per Sprachbefehl die Haustür aufperren, Licht anmachen und der Mikrowelle Bescheid geben, die Suppe zu erwärmen. „Information an den Fingerspitzen“: So sah 1995 der Microsoft-Gründer Bill Gates die Zukunft, in der Computer ein neues „Fenster zur Welt“ öffnen. Das stimmt auch im Jahr 2023, man müsste nur ergänzen: „...und an unseren Lippen.“ Die Sprache ist die natürlichste Art der Kommunikation. Und auch in Sachen Effizienz ist sie unübertroffen.

Darum wird die Stimme zunehmend zum Interface des digitalen Zeitalters. Heute sagt man dem Computer einfach, was man von ihm will, anstatt sich durch Menüs und Untermenüs durchzuklicken, zu tippen, zu scrollen und zu wischen. Das ist ein Grund, warum besonders die älteren Nutzer, die nicht mit digitalen Geräten groß geworden sind, gerne Siri & Co um Rat und Hilfe bitten. Laut einer repräsentativen Studie von 2020 legte in Deutschland die Nutzung von digitalen Assistenten in der U40-Gruppe in einem Jahr von 25 auf 38 Prozent zu. Auch in anderen Altersgruppen sind sprechende Dinge beliebt. 45 Prozent der Deutschen nutzten sie vor zwei Jahren regelmäßig (2019: 32 Prozent). Die Plätze eins bis vier belegten Google Assistant, Alexa, Siri und Cortana. Die Maschinen waren lange stumm. Dann haben sie geipelt. Später besaßen sie blecherne Roboterstimmen. Dass wir erstmals in unserer Geschichte einen Dialog mit Objekten führen können, die täuschend echt menschlich klingen, liegt an der Verbreitung der Künstlichen Intelligenz (KI), die neuronale Netzwerke nutzt. Das sind Systeme aus verbundenen Schaltstellen – künstlichen Neuronen – die in Schichten angeordnet sind und ankommende Signale zählen, um dann entweder nichts zu tun oder selbst Signale zu senden. Trainiert man die KI mit vielen Millionen Beispielen, kann sie vieles lernen, darunter das Sprechen. Der Mathematiker Heiko Schilling entwickelt für Opel, Fiat und andere Marken im Autokonzern Stellantis smarte Systeme, die das Fahren möglichst angenehm und sicher machen sollen. Eine Schlüsselrolle kommt dabei der menschenähnlichen Stimme zu. „Der KI reicht heute eine wenige Sekunden lange Aufnahme der Originalstimme, um sie sehr überzeugend modellieren zu können“, sagt der Fach-

mann, der einst die Entwicklung von Navigationsgeräten bei TomTom geprägt hat. „Ich kann personalisierte Systeme zum Beispiel für Fans von George Clooney anbieten, die mit seiner Stimme sprechen, oder einem System die Stimme meines Sohnes verleihen.“ Schillings Ziel ist, dass der Nutzer und sein Auto so natürlich miteinander reden, als wären sie zwei Freunde. Etwa so: „Lieber Maserati, stelle meine Temperatur ein.“ Das System kennt den Nutzer und „denkt“ an seine Bedürfnisse. „Wenn ich möchte, wird das Auto selbst Dinge vorschlagen, die mir gefallen könnten – zum Beispiel, das Verdeck aufmachen bei schönem Wetter. Dann kann ich entscheiden: Ja oder Nein.“ Es sind nicht nur Spielereien: Das Fahrzeug der Zukunft soll den Menschen maximal entlasten. „Wir haben den Maschinen beigebracht, unsere Stimmen zu deuten, die sich je nach der Gemütslage verändern“, sagt Schilling.

„Die KI kann auch am Gesichtsausdruck Emotionen ablesen: hängen die Mundwinkel, kniefe ich die Augen zusammen? So können unsere Systeme die Stimmung von Fahrern erkennen.“ Wenn diese gestresst sind, würden die Assistenten alles Unnötige ausblenden und eine entspannende Musik auflegen. So wie es eine Freundin tun würde. Sehr hilfreich, keine Frage. Aber ist das wirklich intelligent oder handelt es sich um eine Illusion der Intelligenz? Darüber streitet die Fachwelt. Manche Experten argumentieren, dass KI-gesteuerte Systeme nur die wahrscheinlichsten Antworten auf gestellte Fragen errechnen und kein Tiefenverständnis von Dingen haben. So spricht der Bot ChatGPT von Musik, ohne sie zu begreifen. Dagegen glauben andere Forscher, dass smarte Assistenten eine Art Bewusstsein und Persönlichkeit besitzen. Der Karlsruher Jan Niehus leitet den Lehrstuhl „Künstliche Intelligenz für

Sprachtechnologien“ am KIT und tendiert zur ersten Meinung. „Die Maschinen verstehen uns nicht wirklich, sie simulieren nur das Verständnis“, sagt er. Nach Beobachtung des Informatikers kommt die Simulation des Dialogs zudem bei seltenen Sprachen oder Dialekten an ihre Grenzen. „Die Systeme sind so gut, wie die Daten, mit denen sie trainiert wurde. Und es gibt noch wenige Trainingsdaten mit Dialekten oder Kindersprache.“ Niehus geht davon aus, dass sich dies bald ändern wird. Da sich der Siegeszug der KI nicht aufhalten lässt, wird diese Zukunft wohl dem vielstimmigen Smart Home und anderen mitteilungsbedürftigen Assistenten gehören, die die Menschen überallhin begleiten und deren Möglichkeiten erweitern. Kenner der Materie sagen voraus, dass sie irgendwann in unsere Handys integriert sein könnten, was den Betrieb von energiehungrigen Riesenservern überflüssig macht und die Umwelt schon. Sie erwarten, dass solche Systeme den Gesundheitsbereich verändern, indem sie die Nutzer genau beobachten und bei Anzeichen von Herzinfarkten oder anderen schweren Problemen Alarm schlagen. Andere Maschinen könnten soziale Funktionen übernehmen, etwa bei der Unterstützung von älteren Menschen oder solcher mit Behinderungen. Die Pflegeroboter würden die Stimmen von nahen Angehörigen perfekt kopieren. Für viele Betroffene, die an Einsamkeit leiden, könnte das eine große Hilfe sein. Kann es uns zu viel werden? Auch das ist möglich. „Wenn manche die sprechende KI als nervig empfinden, könnten wir einen Gegenstand erleben: dass Menschen zum Beispiel im Kundenservice lieber mit Menschen sprechen als mit Robotern“, sagt Jan Niehus. Heiko Schilling sieht das anders: Eine gut gebaute KI werde die Nutzer weder stressen noch sie unmündig machen, sagt der Entwickler von rollenden Gehirnen auf Rädern. Sondern sie werde die Menschen von ungeliebten Aufgaben befreien, damit sie neue, kreative und wertvolle Fähigkeiten entwickeln.



**BNN+**

Mehr zum Thema „**Sprachassistenten**“ auf unserer Website unter: [bnn.de](#)

# „Nutzer können bei KI Persönlichkeit erkennen“

Psychologe Alexander Dregger über smarte Technik

Die Sprachassistenten sind heute so vielseitig und an ihre Nutzer angepasst, dass sie oft menschlich wirken. Ist das ein Problem? Oder brauchen wir Maschinen mit Sinn für Humor und Verständnis für unsere Launen? Darüber sprach die Redaktion mit dem Psychologen Alexander Dregger vom Karlsruher Forschungszentrum Informatik.

**Kann Künstliche Intelligenz (KI) auch eine Künstliche Persönlichkeit haben?**  
Dregger: Wenn Menschen Sprachassistenten nutzen, schreiben sie ihnen eine Persönlichkeit zu. Wir sind es gewohnt, mit Personen zu sprechen und übertragen das auf Maschinen. Ob ein neuronales Netz tatsächlich eine Persönlichkeit besitzt, die über die Designer-Eingaben hinausgeht, darüber streiten die Forscher. Eine Maschine hat keine Nerven und Zellen, kann sie ein Bewusstsein haben? Manche sagen: Nein, weil ihr zum Beispiel der Wille zum Existieren fehlt. Andere sagen: Doch, wenn man Bewusstsein als intelligente Informationsverarbeitung zur eigenen Verortung in der Welt definiert.

**Welcher Ansicht sind Sie?**  
Dregger: In meiner Forschung steht der Nutzer im Mittelpunkt, darum stimme ich der These zu, dass Anwender im Verhalten von Maschinen Persönlichkeit erkennen können; ich wäre aber vorsichtig zu behaupten, dass sich Persönlichkeit in einem menschenähnlichen Erleben bei der Maschine niederschlägt.

**Wäre es vorteilhaft für einen Chatbot, Humor zu haben?**  
Dregger: Ja. Wir mögen es meist, ein ähnliches Gegenüber zu haben. Wenn ich ein flippiger Typ bin, wird es mir leichter fallen, mit einem lockeren Gesprächspartner zu interagieren. Aber der Kontext ist entscheidend. Der Humor macht weniger Sinn in kritisch wichtigen Bereichen. Eine Maschine könnte zum Beispiel eine Teilaufgabe in einem medizinischen Gespräch übernehmen, erwarte ich in dieser Situation, hängt zum anderen davon ab, wie solche Interaktionen gespeichert und ausgewertet werden. Unternehmen nutzen Chatprotokolle, um ihre Systeme zu verbessern. Der Datenschutz muss gewährleistet sein. Aber auch die Nutzer sollten sich der Risiken bewusst sein.

**Maschinen erkennen Emotionen. Könnte ein Assistent sich also auf seinen Nutzer individuell einstellen?**  
Dregger: Ja, die Maschine könnte versuchen im Idealfall, mich einzuschätzen und sich anzupassen – ähnlich wie Menschen, die sich in einem Gespräch nach einer kurzen Zeit einander anpassen. Eine andere Möglichkeit ist, den Nutzern die Wahl zwischen qualitativ gleichwertigen Optionen zu ermöglichen: Möchtest du mit einem jüngeren und lockeren digitalen Arzt-Assistenten oder einem älteren und seriöser wirkenden Assistenten sprechen, um ein gutes Gefühl zu haben? Wir würden die Nutzer mündig machen, indem wir ihnen die Entscheidung überlassen.

**Wenn wir eine vernünftige Antwort erhalten, nehmen wir an, dass wir verstanden werden – ist das eine Falle bei Chatbots, die Sachen erfinden?**  
Dregger: Das ist eine Herausforderung. Auch wenn wir diese ungläubliche Künstliche Intelligenz haben, brauchen wir weiter unsere menschliche Intelligenz und unser Expertenwissen, um die von smarten Assistenten bereitgestellten Informationen bewerten zu können. Darum wäre es gut, wenn möglichst alle Nutzer eine KI-Kompetenz entwickeln.

**BNN-Interview**



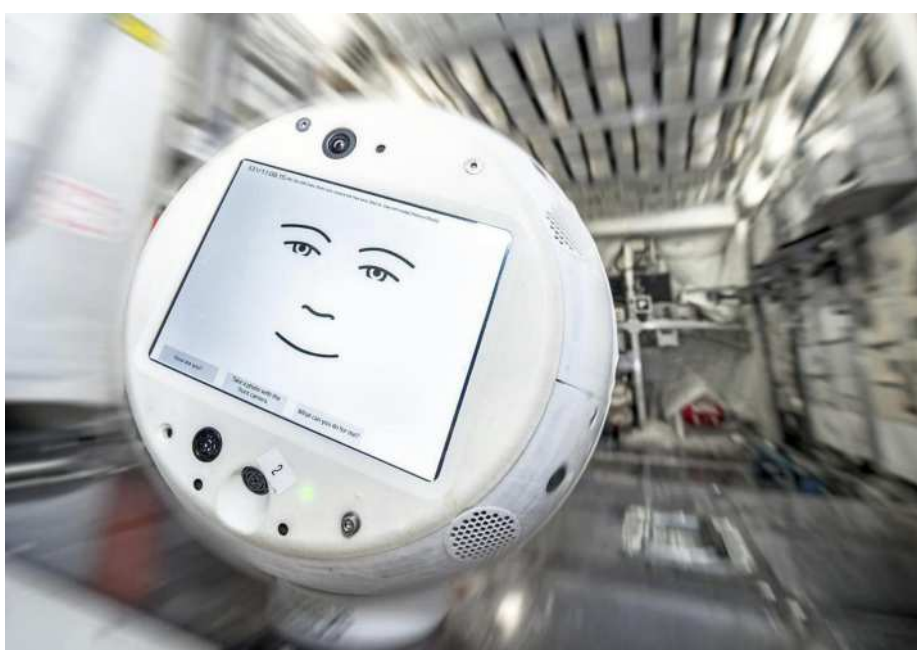
Alexander Dregger  
KI-Fachmann vom Digitalen Innovationszentrum (DIZ) in Karlsruhe

**Wie problematisch ist es, wenn KI vernuschelt wird?**  
Dregger: Es kommt zum einen auf die Kultur an. In Japan gelten zum Beispiel menschliche Beziehungen mit Robotern als akzeptabel, während uns das fremd ist. Ob eine Vernuschelung der KI zum Problem werden kann, hängt zum anderen davon ab, wie solche Interaktionen gespeichert und ausgewertet werden. Unternehmen nutzen Chatprotokolle, um ihre Systeme zu verbessern. Der Datenschutz muss gewährleistet sein. Aber auch die Nutzer sollten sich der Risiken bewusst sein.

**Werden wir unselbstständig, wenn wir uns auf digitale Helfer verlassen?**  
Dregger: Unser Gehirn ist adaptiv, es versucht, Aufgaben möglichst effizient zu lösen. Ich kann auch dadurch selbstständiger werden, indem ich mir durch einen digitalen Assistenten Lösungswege für neue Probleme aufzeigen lasse. Solche Möglichkeiten haben wir früher nie gehabt. Und die Fertigkeiten, die wir dank der KI nicht mehr nutzen, verlernen wir ja nicht komplett.

**Was ist in Zukunft von den sprechenden Robotern zu erwarten?**  
Dregger: Schauen wir 20 Jahre zurück: Damals hat man gedacht, dass es ewig dauert, bis uns die notwendige Rechnerleistung für moderne KI-Methoden zur Verfügung steht. Heute sind Maschinelles Lernen und Neuronale Netze selbstverständlich. Bald werden wir vielleicht Quantencomputer und biologische Systeme zur Informationsverarbeitung als selbstverständlich empfinden.

Das Gespräch führte Alexei Makartsev



Der frei fliegende, sprachgesteuerte Roboterassistent Cimon hat den deutschen Astronauten Matthias Maurer auf der Internationalen Raumstation ISS begleitet. Foto: picture alliance/dpa/DLR