

Sprich mit mir! Perspektiven für den Einsatz KI-basierter Dialogsysteme

Themenkurzprofil Nr. 52 | Robert Peters | Februar 2022

KI-basierte Dialogsysteme finden zunehmend Anwendung in der Praxis und sind inzwischen Teil des Alltags vieler Menschen. Sie stellen eine neue Form der Mensch-Maschine-Schnittstelle dar, die gesellschaftliche wie ökonomische Potenziale schafft. Mittels Sprachtechnologien können Unternehmen und öffentliche Einrichtungen bestehende Services verbessern sowie neue Geschäftsmodelle und Angebote schaffen. Gesellschaftlich könnten Sprachschnittstellen einen großen Beitrag zur Inklusion beispielsweise von sehbehinderten Menschen leisten. Diesen Chancen stehen nicht unerhebliche Risiken gegenüber. So ist beispielsweise unklar, welche entwicklungspsychologischen Folgen das Aufwachsen mit Sprachassistenten für Kinder haben kann. Darüber hinaus besteht die Gefahr, dass durch die Adaption von Dialogsystemen gesellschaftliche Stereotype und tradierte Rollenbilder unreflektiert verbreitet und skaliert werden. Ökonomisch und geopolitisch ist eine zunehmende Abhängigkeit von den US-amerikanischen sowie chinesischen Anbietern großer Sprachmodelle zu befürchten, deren Systeme zentral für die Umsetzung leistungsfähiger Dialogsysteme sind. Vor diesem Hintergrund wird in dem Kurzprofil diskutiert, welche sozialen, wirtschaftlichen und politischen Potenziale die Entwicklung eines europäischen Sprachmodells bieten könnte.

Hintergrund und Entwicklungsstand

Unter dem Begriff Sprachtechnologie (Speech Technology) werden allgemein Technologien verstanden, die auf Basis von Verfahren der künstlichen Intelligenz (KI)¹ mit dem Ziel entwickelt werden, menschliche Sprache zu verstehen, eine Person aufgrund ihrer Sprache zu identifizieren, menschliche Sprache künstlich zu synthetisieren bzw. mehrere dieser Funktionen zu vereinen (Doulaty 2015). Seit den 1960er Jahren entstanden Sprachtechnologien, die zunehmend leistungsfähiger in der Gestaltung einer dialogischen Kommunikation von Mensch und Maschine wurden. Bekannt wurde vor allem der 1964 von Joseph Weizenbaum am Massachusetts Institute of Technology (MIT) entwickelte Chatbot „Eliza“ (Pinola 2011). Heute spricht man im Zusammenhang mit Technologien, die Sprache verstehen und in der Lage sind, über natürliche – d.h. schriftliche oder gesprochene – Sprache (Berg 2013) Unterhaltungen mit Menschen zu führen, von „Dialogsystemen“ (Reply o.J.) oder „Conversational Artificial Intelligence (AI)“ (Deloitte 2019; Reply o.J.). Eine einheitliche Begriffsdefinition liegt diesen Technologien dabei nicht zugrunde.²

Dialogsysteme ermöglichen heute die Schaffung neuer Mensch-Maschine-Schnittstellen, indem bislang z.B. über Tastatur bzw. Touchscreen (Eingabe von Informationen bzw. Anweisungen) und Displays (Informationsausgabe) erfol-

- 1 Definition: „Künstliche Intelligenz (KI) bezeichnet Systeme mit einem ‚intelligenten‘ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen. KI-basierte Systeme können rein softwaregestützt in einer virtuellen Umgebung arbeiten (z.B. Sprachassistenten, Bildanalysesoftware, Suchmaschinen, Sprach- und Gesichtserkennungssysteme), aber auch in Hardwaresysteme eingebettet sein (z.B. moderne Roboter, autonome Pkw, Drohnen oder Anwendungen des ‚Internet der Dinge‘).“ (EK 2018).
- 2 So werden sie in der Literatur teilweise auch als Dialogroboter bezeichnet (Sieber 2019).

gende Interaktionen von Mensch und Maschine verstärkt über Sprachein- und -ausgabe erfolgen (für konkrete Anwendungsszenarien siehe Infobox).

Beispiele für Anwendungsfelder KI-basierter Dialogsysteme

Gesundheitswesen

Dialogsysteme können im Bereich der Gesundheitsversorgung als neue niedrigschwellige Schnittstellen für Patient/innen zu Gesundheitsinformationen genutzt werden (Lovett 2019). Realisiert werden so Self-Assessment-Lösungen, die Menschen dabei unterstützen, ihre mentale Gesundheit einzuschätzen oder Anamnese-systeme, wie etwa die Plattform „Ada“, die Menschen eine Ersteinschätzung zu ihrer konkreten Symptomatik gibt (Ada Health GmbH 2021).

Marketing und Endkundenschnittstellen

Ein verbreitetes Einsatzgebiet für Dialogsysteme ist die Kommunikation mit Kund/innen (KI.NRW 2021, S.58), die mit solchen Systemen über Smartspeaker, Smartphone, Telefon oder Tablet interagieren (annallect 2018). Beispiele finden sich heute bereits im Finanzsektor: So bieten Banken im Zusammenhang mit der Verbreitung des Onlinebankings verstärkt Unterstützung per Telefon an, da Kund/innen bei Konfliktfragen die Kontaktaufnahmen via Hotline bevorzugen. Dialogsysteme können dabei Teile des Anrufaufkommens automatisiert bearbeiten (agora.ai 2020). Auch im Onlinehandel werden Dialogsysteme z.B. in Form von Chatbots für den Support sowie als Einkaufsberater eingesetzt (Fynn 2020; Zalando 2021). Darüber hinaus ist es möglich, entsprechende Systeme als Sprachschnittstelle für Serviceroboter zu implementieren, sodass sie auch im stationären Handel bzw. der Hotellerie einsetzbar sind (Wang 2019).

Recht und Steuern

Im Rechts- und Steuerwesen etablieren sich seit Jahren sogenannte Legal-Tech-Anwendungen, die auch mittels integrierter Dialogsysteme funktionieren, welche etwa die Kommunikation mit Mandant/innen teilweise übernehmen. Dazu zählen Chatbots, die grundlegende Informationen vermitteln oder aufnehmen sowie Rechtsberatungsleistungen bieten (TAB 2019c).

Öffentliche Verwaltung

Ein konkretes Beispiel für den Einsatz von Dialogsystemen in der Verwaltungspraxis ist der von der Berliner Gesundheitsverwaltung eingesetzte Chatbot „Bobbi“. Dieser wurde z.B. in der Coronapandemie zur Entlastung von Telefonhotlines eingesetzt, um Anrufer/innen Informationen über die Pandemie zu geben (Puntschuh/Fetic 2020, S.14).

Personalauswahl

Ein weiteres Anwendungsfeld für Dialogsysteme bildet die Personalauswahl. Softwareanbieter wie HireVue ermöglichen mittels Dialogsystemen die Automatisierung von Bewerbungsgesprächen. Die Systeme analysieren dabei auch KI-basiert Emotionen oder Persönlichkeitsmerkmale von Bewerber/innen. Andere Unternehmen, darunter die deutsche Firma LogOn Consulting, nutzen Conversational AI primär für den Zugang zum Bewerbungsprozess, etwa indem kein CV und kein Anschreiben eingereicht werden müssen, sondern alle relevanten Angaben über ein Gespräch mit dem KI-System erfasst werden (TAB 2021a).

Die Breite bereits heute existierender Anwendungsfälle für Dialogsysteme deutet an, dass hier ein erhebliches wirtschaftliches Potenzial besteht. Seit 2010 gewinnt die Entwicklung im Bereich von Sprachtechnologien allgemein und von Dialogsystemen im Speziellen an Dynamik. Verantwortlich dafür sind verschiedene Faktoren: Im Zusammenhang mit der steigenden Nutzung von Smartphones und Tablets und darüber zugänglicher Softwareapplikationen (Apps) stehen immer mehr Nutzerdaten zur Verfügung, die für das Training der zur Anwendung kommenden KI-Systeme notwendig sind. Fehlerraten bei der Spracherkennung sinken. Lag die Fehlerrate bei Umwandlung von gesprochener Sprache in Text zu Beginn der 2000er Jahre noch bei rund 20 %, wird diese von Anbietern wie Google und Microsoft mit rund 5 % angegeben. Menschen kommen auf eine Fehlerrate von rund 4 % (Pinola 2011). Hochleistungsrechner ermöglichen das Training tiefer neuronaler Netze und damit die Entwicklung großer Sprachmodelle. Hinzu kommt, dass es infolge der Entwicklung im Bereich von Cloudcomputing mittlerweile möglich ist, Dialogsysteme nicht proprietär zu realisieren (Ionescu 2008; Pinola 2011). Das heißt, das Dialogsystem wird nicht auf dem Endgerät der Nutzer/innen betrieben. Vielmehr greift z. B. ein Smartspeaker über eine Schnittstelle auf ein über zentrale Server gehostetes Dialogsystem zu. Eine eindeutige Quantifizierung des Marktvolumens ist aufgrund der uneinheitlichen Abgrenzung der betrachteten Märkte im Bereich von Dialogsystemen schwierig. Das weltweite Marktvolumen für Conversational AI wird für 2020 mit rund 5,1 Mrd. US-Dollar angegeben. Für 2026 wird mit einer Steigerung auf mehr als 12 Mrd. US-Dollar gerechnet (Research and Markets 2021).

Technische Funktionsweise von Dialogsystemen

Dialogsysteme integrieren in der Regel eine Reihe technischer Funktionen, die eine sprachliche Interaktion zwischen Mensch und Maschine ermöglichen. Die zugrunde liegenden Technologien werden auch Natural Language Processing (NLP, Verarbeitung natürlicher Sprache) genannt. Damit werden allgemeine Technologien bezeichnet, die geschriebene oder gesprochene natürliche Spra-



che erkennen, verarbeiten und/oder synthetisieren können (Chowdhury 2003; Mishra/Kumar 2020; Sieber 2019). Die bei Dialogsystemen stattfindende Verarbeitung kann in fünf Prozessschritte differenziert werden, durch deren Integration ein Dialogsystem entsteht (KI.NRW 2021, S.19 ff.).

Dialogsystemen können unterschiedliche KI-Methoden zugrunde liegen. Grundsätzlich unterscheidet man zwischen symbolischer und subsymbolischer künstlicher Intelligenz. Bei symbolischer KI kommen häufig Algorithmen zum Einsatz, die auf Grundlage eines definierten, von Expert/innen vorgegebenen Regelsatzes Probleme lösen. Daher spricht man hier auch von regelbasierten Systemen. Diese Technologie ist bereits seit Jahrzehnten bekannt und kam schon in den 1960er Jahren im Falle des Chatbots „Eliza“ zum Einsatz. Seit den 1980er Jahren entwickelten sich Verfahren der subsymbolischen KI (Sieber 2019, S.61 ff.). Dabei werden Wörter als beobachtbare Datenpunkte im Kontext der sie umgebenden Wörter konzeptualisiert, sodass Anwendungen wie das maschinelle Übersetzen durch statistische Verfahren auf Basis von Textdaten trainiert werden können. Ein ab Mitte der 1970er Jahre häufig verwandtes statistisches Verfahren wurde unter dem Namen „Hidden Markov Model“ (HMM) bekannt. Das Verfahren bildete die Grundlage für die technische Entwicklung im Bereich von Sprachtechnologien in den folgenden Jahrzehnten. Diese Entwicklung markierte eine erste Welle des maschinellen Lernens

(Interview Burchard; Juang/Rabiner 2004). In den 2010er Jahren gewann die Anwendung komplexerer KI-Modelle auf Grundlage neuronaler Netze für die Gestaltung von Dialogsystemen an Bedeutung (Rathenau Institut 2020, S.19 ff.) Ein neuronales Netz ist ein komplexes Rechenmodell, das dem menschlichen Gehirn nachempfunden ist. Es ist in der Lage zu lernen, wie es die ihm zur Verfügung gestellten Trainingsdaten am besten nutzt, um Daten zu kategorisieren (Rathenau Institut 2020, S.21).

Große Sprachmodelle

Auf Grundlage des Prinzips neuronaler Netze wurden in den vergangenen Jahren große Sprachmodelle wie das „Generative Pre-trained Transformer 3“ (GPT-3) entwickelt. Das Sprachmodell des US-Unternehmens OpenAI LP ist eines der aktuell größten. Die hier zur Anwendung kommenden tiefen neuronalen Netze werden mittels großer Datenmengen trainiert. Im Vortraining baut das System Sprach- und Faktenverständnis auf Basis von Textdaten auf. So wurde „GPT-3“ mit rund 570 GB Textdaten aus der englischsprachigen Wikipedia und weiterer Datenquellen, z.B. Bücherdatenbanken und Webseitexten, trainiert. Im Resultat umfasst das neuronale Netz von „GPT-3“ 96 Schichten mit 175 Mrd. Parametern. Die Größe von „GPT-3“ wird nur vom chinesischen Sprachmodell „Wudao 2.0“ übertroffen, das an der Beijing Academy of Artificial Intelligence entwickelt wurde und das Berichten zufolge 1,75 Billionen Parameter umfasst (Rathenau Institut 2020, S.14). Um entsprechende Sprachmodelle für einen spezifischen Einsatzzweck zu

adaptieren, erfolgt in der Regel ein Feintuning: Dabei wird das Modell auf Basis kleinerer Datensätze trainiert. So lernt das Modell z.B. Meinungen oder Stimmungen (Sentiments) zu erkennen oder es wird für bestimmte, branchen- und unternehmensabhängige Anwendungskontexte erweitert. Bei einzelnen Aufgaben, wie etwa dem Weitererzählen von Geschichten, benötigt „GPT-3“ dabei lediglich wenige oder sogar nur ein Beispiel, um die Anforderung umzusetzen (One-, Two- bzw. Zero-Shot-Learning) (TAB 2019b). Anschließend ist das Modell in der Lage, mit hoher Wahrscheinlichkeit das nächste Wort eines Textes vorauszuberechnen (KI.NRW 2021, S.14).

In der Praxis sind Dialogsysteme heute in der Regel als hybride Systeme ausgelegt. Das heißt, es kommen sowohl lernende als auch durch Expert/innen modellierte Komponenten zum Einsatz (z.B. indem die Dialogsteuerung weiterhin regelbasiert funktioniert, die Spracherkennung und -synthese jedoch über maschinelles Lernen, heute meist mit tiefen Sprachmodellen erfolgt) (Interview Burchard). Während für so realisierte Dialogsysteme zunächst von einer zunehmenden Verbreitung ausgegangen wird (Deloitte 2019), sind auf technischer Seite zentrale Herausforderungen bislang nicht gelöst. So zeigen sich Nutzer/innen im Umgang mit digitalen Sprachassistenten deutlich häufiger frustriert als in der Anwendung anderer smarter Systeme. Besonders häufig klagen Nutzer/innen dabei über Defizite bei der Spracherkennung selbst (Accenture o.J.). Damit dürften Herausforderungen bei der Realisation von Dialogsystemen in der Umwandlung von Eingangssignalen in Text sowie im Bereich des Textverstehens liegen. Hohe Erwartungen an die Verbesserung von Dialogsystemen werden in große Sprachmodelle wie „GPT-3“ gesetzt, insbesondere, wenn diese in der internen Verarbeitung mit entsprechenden Datenbanken mit domänenspezifischen Informationen (z.B. Reparaturanleitungen) verbunden werden. Dialogsysteme wie Chatbots könnten sich so in ihrer Performance der sprachlichen Interaktion von Mensch zu Mensch annähern (Kerkmann 2020).

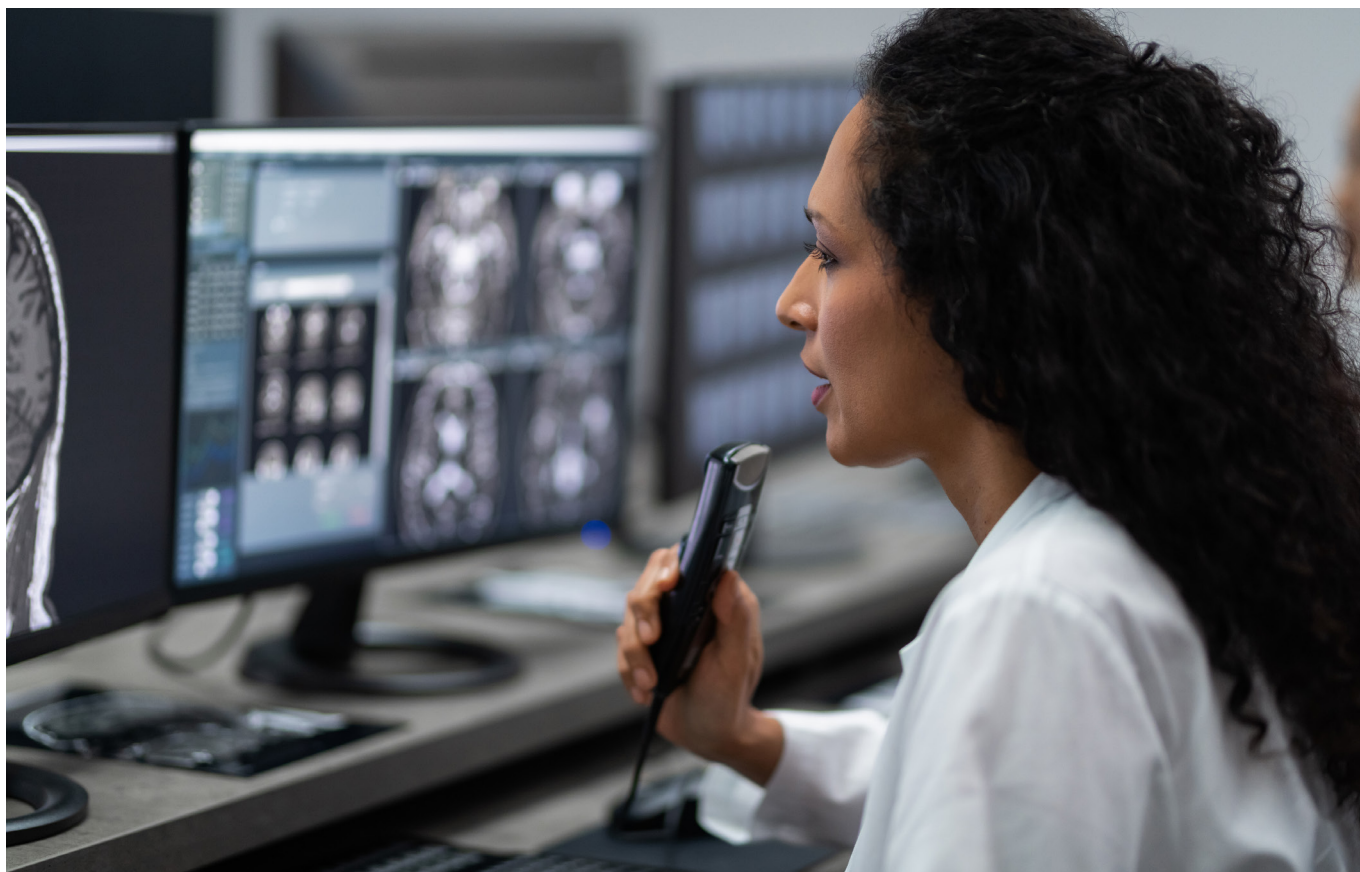
Öffentlich problematisiert wird im Zusammenhang mit großen Sprachmodellen eine zunehmende Abhängigkeit von den USA und potenziell von China, da in Europa gegenwärtig keine vergleichbar leistungsfähigen Sprachmodelle wie „GPT-3“ und „Wudao 2.0“ existieren. So bestehen seitens der Deutschen KI-Wirtschaft „erhebliche datenschutzrechtliche Bedenken, da OpenAIs GPT-3 nur per API [Programmierschnittstelle] verwendbar ist. Deutsche Firmen müssten derzeit also eigene Daten und Kundendaten über den Atlantik schicken, um das Modell verwenden zu können“ (Bienert 2021), was die Befürchtung weckt, dass europäische KI-Anbieter ohne eine eigene europäische Lösung künftig nur noch Frontendentwicklung auf Basis der US-Technologie betreiben können (Accenture o.J.). Die daraus resultierende Abhängigkeit von US-Anbietern wäre möglicherweise mit der heutigen Abhängigkeit von Google bei der Internetsuche vergleichbar (Grävemeyer 2021).

Gesellschaftliche Risiken durch Dialogsysteme

Mit der zunehmenden Verbreitung und einer steigenden Leistungsfähigkeit von Dialogsystemen rücken mögliche gesellschaftliche Risiken stärker in den Blickpunkt der Debatte (Interview Fetic): „Sprache ist mehr als unser Vokabular und unsere Grammatik. Sie drückt aus, wie wir unsere Gesellschaft wahrnehmen, wie wir uns selbst und andere (z.B. marginalisierte Gruppen) darin positionieren. Daher treten bei der Entwicklung und Einführung von Sprachtechnologien unweigerlich gesellschaftliche Rollen- und Verteilungskonflikte zu Tage.“ Diskriminierende Muster werden potenziell durch Dialogsysteme reproduziert, weil die Technologie als Artefakt – also durch menschlichen Einfluss – diese im Zuge des Entwicklungs- und Einführungsprozesses explizit oder implizit inkorporiert (Interview Fetic; TAB 2020a). Konkret gibt es die Befürchtung aus Wissenschaft und Zivilgesellschaft, dass durch die Nutzung von Dialogsystemen, wie sie etwa bei der Personalauswahl zum Einsatz kommen (automatisierte Jobinterviews), diskriminierende Muster skaliert werden (TAB 2020b u. 2021a). Zu einer Skalierung kann es insbesondere dann kommen, wenn ein KI-System auf Basis von Daten trainiert wird, die nicht repräsentativ für die Gruppe von Menschen sind, auf die das System angewendet werden soll (TAB 2021b). Darüber hinaus stehen seit Jahren auch Sprachassistentensysteme im Fokus der Debatte (TAB 2017 u. 2019a). Dabei wird u.a. die Sorge um die Einhaltung des Datenschutzes betont (Kerkmann 2019). Fragwürdig ist vor allem die Praxis der Anbieter von Sprachassistentensystemen, wie Apple, Google und Amazon, die in der Vergangenheit nicht proaktiv, sondern erst auf massiven öffentlichen Druck reagierten und fragliche Praktiken im Zusammenhang mit ihren Assistenzsystemen überdachten (Adegoké 2019; Hurtz 2019; Kerkmann 2019).

Auch im Hinblick auf den hohen Bedarf an Trainingsdaten von Sprachmodellen aus den USA und China werden zunehmend Bedenken laut (Schreiner 2021). Dabei geht





es längst nicht mehr nur um Fragen, wie, für welchen Zweck und in welchem Umfang Sprachassistenzsysteme Daten von Nutzer/innen sammeln und auswerten. Wenn die Interaktion mit Dialogsystemen zunehmend Teil unseres Alltagslebens wird – bereits heute nutzen 38 % der Menschen in Deutschland Sprachassistenten (TÜV-Verband 2021) – und Mensch-Maschine-Interaktionsformen flexibler und komplexer werden, dann können Formen sozialer Beziehungen zwischen Mensch und technischem Assistenzsystem entstehen und Nutzer/innen dem technischen Artefakt bewusst oder unbewusst menschliche Eigenschaften zuschreiben (Anthropomorphismus) (TAB 2019a). Grundsätzlich ist zu beobachten, dass Menschen Dialogsystemen menschliche Eigenschaften zuschreiben, z.B. eine Geschlechteridentität. Ein Forschungsbericht der UNESCO (2019) zeigt, dass verbreitete Dialogsysteme in Sprachassistenten, wie „Alexa“ oder „Siri“, auf sexistische und misogynen Äußerungen teilweise verharmlosend oder bestärkend reagieren (Fessler 2017). Bisherige Forschungsergebnisse deuten darauf hin, dass mindestens 5 % der Interaktion mit Sprachassistenten sexuell explizit sind (UNESCO 2019). Zwar hätten große Tech-Unternehmen zwischenzeitlich auf diese Fehlentwicklung reagiert, dies ändere jedoch nichts an dem Umstand, dass Sprachtechnologien bislang stark von binären Geschlechterbildern und -rollen geprägt sind (Interview Fetic). So sind die verbreitetsten Sprachassistenzsysteme standardmäßig mit einer als weiblich gelesenen Stimme ausgestattet und reflektieren, verstärken und verbreiten damit in Verbindung mit der helfenden, dienenden Rolle des Systems aus Sicht

von Kritiker/innen tradierte Genderstereotype (Danielescu 2020; UNESCO 2019). Zwar existieren mit Projekten wie „multi vocal“ (das synthetische Stimmen entwickelt, die mehr als ein Alter, ein Geschlecht oder eine Herkunft aufweisen können; TheNewNew o.J.a) und „Syb“ (Sprachschnittstelle von und für Transpersonen; TheNewNew o.J.b) Initiativen aus der Zivilgesellschaft, die versuchen, Impulse für eine größere Diversität und Offenheit beim Design von Sprachtechnologien zu setzen. Auch privatwirtschaftliche Akteure scheinen das Thema zumindest vereinzelt zu adressieren, wie das Projekt „Sam“ verdeutlicht. Die geschlechtsneutrale Text-to-Speech-Stimme ist nach Angaben der Entwickler/innen unter Beteiligung nicht binärer Communities entstanden (Danielescu 2020). Inwieweit große Tech-Unternehmen solche Impulse konsequent aufgreifen und nicht binäre Stimmen künftig zum Standard werden, bleibt offen.

Bislang weitgehend unklar bleibt, welchen Einfluss die Mensch-Technik-Interaktion über Sprachschnittstellen auf die Entwicklung von Sprache und sozialen Beziehungen hat. Einer Google-Studie (Kleinberg 2018) zufolge gaben 41 % der Nutzer/innen von Sprachassistenten an, sie hätten in der Interaktion das Gefühl, mit einem Freund oder einer anderen Person zu sprechen. So gibt es Hinweise darauf, dass insbesondere Kinder in ihrer Entwicklung negativ beeinflusst werden können, wenn sie regelmäßig mit Sprachassistenten interagieren, weil sie, noch weniger als Erwachsene, zwischen menschlichen und technischen Dialogpartnern differenzieren (Boller 2021; Seibert 2017).

Die Interaktion mit Sprachassistentensystemen könnte so dazu führen, dass Interaktionsverhalten erlernt wird, das dann auf zwischenmenschliche Kommunikation übertragen wird: Etwa indem auch in der Mensch-Mensch-Interaktion verstärkt befehlend und anweisend kommuniziert wird, wie es viele Sprachtechnologien heute in der Mensch-Technik-Interaktion erfordern (Álvarez/Rövekamp 2019).

Europäisches Sprachmodell erforderlich

Um eine drohende Abhängigkeit bei der Entwicklung von Sprachtechnologien wie Dialogsystemen zu vermeiden, erscheint die Entwicklung eines europäischen Sprachmodells vielen Expert/innen langfristig notwendig (Grävemeyer 2021). Ein Netzwerk von KI-Verbänden aus acht europäischen Ländern strebt – dieser Leitidee folgend – unter dem Begriff des „Large European AI Models“ (LEAM) die Entwicklung eines europäischen Sprachmodells an. Dabei geht es vor allem um den Aufbau umfangreicher Trainingsdatensätze unter Einhaltung der Vorgaben des EU-Datenschutzes und um den Aufbau einer skalierbaren Infrastruktur auf Basis von „GAIA-X“, dem Projekt einer EU-weit vernetzten Dateninfrastruktur (Bienert 2021). Aus der Zivilgesellschaft werden mit Blick auf solche Projekte Anforderungen formuliert (Interview Fetic): „Europäische Initiativen zur Entwicklung von Sprachmodellen dürfen kein Selbstzweck sein. Sie müssen von Anfang an so gestaltet werden, dass sie einen echten gesellschaftlichen Mehrwert schaffen und z.B. die Identität kleiner Sprachfamilien einbeziehen.“ Entsprechende Forderungen werden aus der Wissenschaft gestützt: Wenn in Europa monopolartige Strukturen im Bereich von Sprachtechnologien vermieden werden sollen, müsste die Frage beantwortet werden, wie gewährleistet werden kann, dass kleinere Anbieter in der Lage sind, in Wettbewerb mit großen Tech-Konzernen zu treten. Auch müsste (im Sinne der europäischen Sprachenvielfalt) dafür Sorge getragen werden, dass Sprachtechnologien künftig einzelne Sprachen, Akzente oder Zielgruppen berücksichtigen, auch wenn diese nur wenige Sprecher/innen adressierten und damit nicht primär im wirtschaftlichen Fokus großer Tech-Konzerne liegen (Rathenau Institut 2020, S. 62).

Gesellschaftliche und politische Relevanz

Wie die Analyse zeigt, dürfte eine stärkere Verbreitung von Dialogsystemen weitreichende soziale, ökonomische und potenziell geopolitische Folgen haben.

Soziale Folgen

Gesellschaftlich bieten KI-basierte Dialogsysteme großes Potenzial: So können sie als Sprachschnittstellen prinzipiell einen niedrighwelligen und inklusiven Zugang zu Informationen ermöglichen (Rathenau Institut 2020, S. 9). Konkrete Beispiele dafür finden sich bereits



heute in der öffentlichen Verwaltung in Form von Informationschatbots (Puntschuh/Fetic 2020, S. 14) und im Gesundheitswesen als auf Apps basierende Self-Assessment- und Anamnesesysteme (Lovett 2019). Auch für sehbehinderte Menschen oder Analphabet/innen kann so eine Interaktion mit technischen Systemen ermöglicht werden, indem visuelle Bedienungsmenüs ersetzt werden (Burley 2021). Auch könnten Menschen mittels Sprachschnittstellen Zugang zu bestimmten Leistungen bekommen, die sie bislang – etwa aufgrund mangelnder finanzieller Möglichkeiten – nicht nutzen konnten (wenn zuvor von Menschen erbrachte Dienstleistungen z.B. zu kostspielig für bestimmte Einkommensgruppen sind). Auch jenseits körperlicher, intellektueller oder ökonomischer Barrieren können Sprachschnittstellen potenziell die Zugangshürde zu bestimmten Institutionen senken, etwa in Bewerbungsverfahren (TAB 2020b). Wie mit Blick auf die Zuschreibung von dualen Geschlechterrollen bereits skizziert, können Dialogsysteme dazu beitragen, dass gesellschaftliche Stereotype und tradierte Rollenbilder unreflektiert verbreitet und skaliert werden – unabhängig davon, ob dies bei der Entwicklung der Technologie beabsichtigt war oder nicht (Interview Fetic). Es bedarf daher eines aktiven gesellschaftlichen Diskurses darüber, unter welchen Bedingungen und zu welchem Zweck Dialogsysteme entwickelt und eingesetzt werden sollen. Auf Basis dieses Diskurses stehen Gesetzgeber und Aufsichtsbehörden in der Verantwortung, geeignete förderpolitische Anreize für eine inklusive Technologieentwicklung zu setzen, erforderliche Investitionsmittel für die Entwicklung einer entsprechenden digitalen Infrastruktur zu schaffen, z.B. in Form einer Sprachdatenbank auf Deutsch, und, falls erforderlich, ordnungspolitische Leitplanken für den Einsatz von Dialogsystemen zu beschreiben. Darüber hinaus sind auch Unternehmen, die entsprechende Systeme entwickeln und ausgestalten, in der Verantwortung, sich über mögliche negative Effekte für die Gesellschaft bewusst zu sein und diesen aktiv entgegenzuwirken (Rathenau Institut 2020, S. 68).

Ökonomische Folgen

Dialogsysteme ermöglichen es Unternehmen und öffentlichen Einrichtungen, bestehende Services wie 24-Stunden-Hotlines zu verbessern und neue Services, z.B. dialogbasierte Self-Assessment-Lösungen, zu kreieren. Neben einer qualitativen Verbesserung bestehender Angebote und der Schaffung neuer Angebote für Nutzer/innen ergibt sich daraus für viele Branchen ein erhebliches Potenzial zur Effizienzsteigerung, indem Teile der Kundenkommunikation automatisiert werden können. Während Dialogsysteme damit in unserem Alltag eine immer präsentere Rolle einnehmen, sind die Chancen, am Wertschöpfungspotenzial dieser neuen Technologie zu partizipieren, höchst ungleich verteilt. Wie in anderen Bereichen der Digitalisierung auch, ist eine starke Dominanz großer US-Tech-Konzerne zu verzeichnen. Diese Dominanz zeigt sich insbesondere als Beherrschung hochfrequentierter Nutzerschnittstellen über bestehende Services, wie sie etwa von Google, Amazon, Facebook und Apple betrieben werden. Damit erlangen diese Unternehmen einen weiteren Wettbewerbsvorteil zur Entwicklung und Implementierung von Dialogsystemen, denn sie sind bereits in der Entwicklungsphase in der Lage, große Mengen relevanter Daten von Nutzer/innen für das Training ihrer Systeme zu nutzen. Kleinere Unternehmen und Anbieter aus Europa haben damit ungleich schlechtere Chancen, große Tech-Konzerne im Wettbewerb herauszufordern und Konkurrenzangebote zu schaffen (Rathenau Institut 2020, S.61f.). Daraus ergibt sich das Risiko, dass deutsche und europäische Unternehmen mit der Weiterentwicklung großer Sprachmodelle in den USA und in China in eine strukturelle Abhängigkeit von Anbietern aus diesen beiden Ländern geraten.

Geopolitische Folgen

Vor dem Hintergrund des gewachsenen wirtschaftlichen, sozialen und politischen Einflusses einiger weniger großer Tech-Unternehmen wie Google (Alphabet) oder Facebook (Meta) wird seit einigen Jahren über die geopolitische Dimension von Digitalisierung und KI diskutiert (Pandya 2019). Angesichts der zuvor beschriebenen Gefahr einer sich verfestigenden ökonomischen Abhängigkeit von außereuropäischen Tech-Unternehmen muss diese Dimension bei der Debatte um Dialogsysteme und Sprachtechnologien mitgedacht werden. Anders als in anderen Bereichen der Digitalisierung, ist es hier prinzipiell noch möglich, über die Definition eigener, europäischer Ziele für die Entwicklung inklusiver, gemeinwohldienlicher und ökonomisch verwertbarer großer Sprachmodelle die digitale Souveränität Europas zu stärken. Hier könnte es gelingen, in der Praxis zu zeigen, wie der vieldiskutierte dritte (europäische) Weg in der Digitalisierung umgesetzt werden kann. Gelingt es hingegen nicht, einen eigenständigen, wertebasierten Weg zu einer eigenen europäischen Infrastruktur zu schaffen, könnten die Folgen der Abhängigkeit von Sprachmodellen aus den USA (sowie potenziell aus China) ähnlich weitreichend sein, wie dies bereits bei Suchmaschineninfrastruktur, Cloudcomputing und beim Mobilfunkstandard 5G zu beobachten ist (Grävemeyer 2021). Europäische Unternehmen, die Sprachtechnologien entwickeln und einsetzen, würden dann nur über eine geringe Wertschöpfungstiefe verfügen, weil sie ihre Produkte und Services nur via API-Schnittstellen auf Basis von Infrastruktur aus Drittländern entwickeln könnten. Um dies zu verhindern, könnten Deutschland und Europa anhand von gezielten Investitionen sowie Regulationen derzeit noch



vorhandene Gestaltungsräume nutzen und ihre Wettbewerbsposition zum Vorteil von Ökonomie und Gesellschaft langfristig sichern.

Mögliche vertiefte Bearbeitung des Themas

Die komplexen Folgewirkungen von Dialogsystemen auf technischer, sozialer, psychologischer und ökonomischer Ebene erfordern eine aktive politische Gestaltung von Rahmenbedingungen für die Entwicklung und den Einsatz entsprechender Anwendungen. Zum jetzigen Zeitpunkt ergibt sich kein weiterer Bedarf einer Vertiefung des Themas. Wie gezeigt werden konnte, sind der politische Diskurs und der Erkenntnisaufbau zu den hier berührten Fragestellungen weit fortgeschritten. Zuletzt lieferte die Studie des Rathenau Instituts (2020) „Look who’s talking – Tools for the responsible use of speech technology“ eine spezifische Vertiefung zum Feld von Sprachtechnologien. Grundlegende Mechanismen der Diskriminierungsgefahren durch den Einsatz von KI-Systemen liefert das TAB-Hintergrundpapier Nr. 24 „Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick“ (TAB 2020a). Für das Anwendungsfeld der Personalauswahl ordnet das TAB-Themenkurzprofil Nr. 40 „Robo-Recruiting – Einsatz künstlicher Intelligenz bei der Personalauswahl“ (TAB 2020b) Potenziale und Risiken in geeigneter Weise ein. Ein Erkenntnisdefizit kann daher aus Sicht der Technikfolgenforschung nicht festgestellt werden. Vielmehr deutet sich ein Umsetzungsdefizit beim Transfer vorliegender Erkenntnisse in das Handeln zentraler Akteure aus Politik und Wirtschaft an.

Literaturverzeichnis

- ▶ Accenture (o.J.): Talk to me. https://www.accenture.com/_acnmedia/PDF-65/Accenture-Voice-Assistant-Infographic.pdf (22.2.2022)
- ▶ Ada Health GmbH (2021): Hallo, wir sind Ada. <https://ada.com/de/about/> (22.2.2022)
- ▶ Adegoke, Y. (2019): Alexa, why does the brave new world of AI have all the sexism of the old one? <https://www.theguardian.com/lifeandstyle/2019/may/22/alexa-why-does-the-brave-new-world-of-ai-have-all-the-sexism-of-the-old-one> (22.2.2022)
- ▶ Agara.ai (2020): How Conversational Voice AI Is Reshaping the Banking Industry. https://agara.ai/wp-content/uploads/2020/12/eBook_How-Conversational-Voice-AI-Is-Reshaping-the-Banking-Industry.pdf (1.12.2021)
- ▶ Álvarez, S.; Rövekamp, M. (2019): Diese Risiken birgt Alexa für Kinder. Der Tagesspiegel, 9.7.2019, <https://www.tagesspiegel.de/wirtschaft/experten-warnen-vor-sprachassistenten-diese-risiken-birgt-alexa-fuer-kinder/24573516.html> (22.2.2022)
- ▶ Annalect (2018): Voice AI: A Conversational Customer Journey. <https://www.annalect.com/voice-ai-a-conversational-customer-journey/> (22.2.2022)
- ▶ Berg, M. (2013): Natürlichsprachlichkeit in Dialogsystemen: Stand der Technik in Industrie und Forschung. In: Informatik Spektrum (36(04), DOI: 10.1007/s00287-012-0650-3
- ▶ Bienert, J. (2021): Warum Europa ein großes KI-Modell braucht. <https://background.tagesspiegel.de/digitalisierung/warum-europa-ein-grosses-ki-modell-braucht> (22.2.2022)
- ▶ Boller, E. (2021): Wenn Alexa „beleidigt“ ist – die schwierige Beziehung zum Sprachassistenten. Welt, 10.5.2021, <https://www.welt.de/wirtschaft/webwelt/article231008783/Sprachassistenten-Die-schwierige-Beziehung-zu-Siri-Alexa-und-Co.html> (22.2.2022)
- ▶ Burley, L. (2021): Die Zukunft der Sprachschnittstelle. Affde, 14.7.2021, <https://www.affde.com/de/future-of-voice-interface.html> (22.2.2022)
- ▶ Chowdhury, G. G. (2003): Natural language processing. In: Annual Review of Information Science and Technology 37(1), S.51–89
- ▶ Danielescu, A. (2020): Building tech that reflects our diversity. Accenture, 16.12.2020, <https://www.accenture.com/us-en/blogs/technology-innovation/danielescu-building-tech-that-reflects-our-diversity> (22.2.2022)
- ▶ Deloitte (2019): Conversational AI. The next wave of customer and employee experiences. <https://www2.deloitte.com/content/dam/Deloitte/au/Documents/strategy/au-deloitte-conversational-ai.pdf> (22.2.2022)
- ▶ Doulaty, M. (2015): A Brief Introduction to Speech Technology. Prescouter, Mai 2015, <https://www.prescouter.com/2015/05/a-brief-introduction-to-speech-technology/> (22.2.2022)

- ▶ EK (Europäische Kommission) (2018): Künstliche Intelligenz für Europa. COM(2018) 237 final, Brüssel
- ▶ Fessler, L. (2017): We tested bots like Siri and Alexa to see who would stand up to sexual harassment. Quartz, 22.2.2017, <https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/> (22.2.2022)
- ▶ Fynn, P. (2020): Wissen, was der Kunde wirklich will. Handelsjournal, 27.1.2020, <https://www.handelsjournal.de/unternehmen/technik/artikel-2020/wie-ki-dabei-helfen-kann-kundenwuensche-zu-erkennen.html> (22.2.2022)
- ▶ Grävemeyer, A. (2021): KI-Sprachmodelle: Alternativen zum umstrittenen Textgenerator GPT-3 gesucht. Heiße online 28.7.2021, <https://www.heise.de/news/KI-Sprachmodelle-Alternativen-zum-umstrittenen-Textgenerator-GPT-3-gesucht-6135934.html> (22.2.2022)
- ▶ Hurtz, S. (2019): Sprachassistenten verlieren ihre menschlichen Ohren. Süddeutsche, 4.8.2019, <https://www.sueddeutsche.de/digital/alexa-siri-google-daten-schutz-1.4552480> (22.2.2022)
- ▶ Ionescu, D. (2008): iPhone Gets Google Search By Voice. PCWorld, 14.11.1008, <https://www.pcworld.com/article/532047/search-2.html> (22.2.2022)
- ▶ Juang, B.; Rabiner, L. (2004): Automatic Speech Recognition – A Brief History of the Technology Development. https://web.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf (22.2.2022)
- ▶ Kerkmann, C. (2020): Dieses Sprachprogramm verblüfft Experten – und birgt Potenzial für die Wirtschaft. Handelsblatt, 25.8.2020, <https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-dieses-sprachprogramm-verbluefft-experten-und-birgt-potenzial-fuer-die-wirtschaft/26126390.html?ticket=ST-2130956-e3iH9xLSN7JhgG1vias9-cas01.example.org> (22.2.2022)
- ▶ Kerkmann, C. (2019): Smartphones, E-Mails, Sprachassistenten – So bekommen Sie Ihre Daten unter Kontrolle. Handelsblatt, 3.5.2019, <https://www.handelsblatt.com/technik/vernetzt/datenschutz-smartphones-e-mails-sprachassistenten-so-bekommen-sie-ihre-daten-unter-kontrolle/24280310.html> (22.2.2022)
- ▶ KI.NRW (2021): Moderne Sprachtechnologien. Konzepte, Anwendungen, Chancen. <https://www.ki.nrw/studie-moderne-sprachtechnologien/#download-studie> (22.2.2022)
- ▶ Kleinberg, S. (2018): 5 ways voice assistance is shaping consumer behavior. Google, Januar 2018, <https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/voice-assistance-consumer-experience/> (22.2.2022)
- ▶ Lovett, L. (2019): Triage-Chatbots - avancieren sie zum neuen Pflegestandard? Healthcare IT News, 16.1.2019, <https://www.healthcareitnews.de/triage-bots-statt-innovationen-den-boden-bereiten> (22.2.2022)
- ▶ Mishra, B.; Kumar, R. (Hg.) (2020): Natural language processing in Artificial Intelligence. Burlington/Palm Bay
- ▶ Pandya, J. (2019): The Geopolitics Of Artificial Intelligence. Forbes, 28.1.2019, <https://www.forbes.com/sites/cognitiveworld/2019/01/28/the-geopolitics-of-artificial-intelligence/?sh=3069522879e1> (22.2.2022)
- ▶ Pinola, M. (2011): Speech Recognition Through the Decades: How We Ended Up With Siri. PCWorld, 2.11.2011, https://www.pcworld.com/article/477914/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html (22.2.2022)
- ▶ Puntschuh, M.; Fetic, L. (2020): Handreichung für die digitale Verwaltung. Algorithmische Assistenzsysteme gemeinwohlorientiert gestalten. Bertelsmann Stiftung, https://www.bertelsmann-stiftung.de/fileadmin/files/user_upload/Handreichung_fuer_die_digitale_Verwaltung_Algo.Rules_12_2020.pdf (22.2.2022)
- ▶ Rathenau Institut (2020): Look who's talking – Tools for the responsible use of speech technology. (Hamer, J.; Doesborgh, S.; Kool, L.) Den Haag
- ▶ Reply (o.J.): Was sind Dialogsysteme? <https://www.reply.com/de/Shared%20Documents/REP18-Robotics-for-Customers-Conversational-Systems-DEU.pdf> (22.2.2022)
- ▶ Research and Markets (2021): Global Conversational AI Market Report 2021. <https://www.globenewswire.com/en/news-release/2021/07/09/2260513/28124/en/Global-Conversational-AI-Market-Report-2021-Market-is-Estimated-to-be-Worth-5-1-Billion-in-2020-and-is-Expected-to-Reach-12-12-Billion-by-2025-Growing-at-a-CAGR-of-18-9.html> (22.2.2022)
- ▶ Schreiner, M. (2021): KI-Bundesverband fordert Aufbau großer europäischer KI-Modelle. <https://mixed.de/ki-bundesverband-fordert-aufbau-grosser-europaeischer-ki-modelle/> (22.2.2022)
- ▶ Seibert, T. (2017): Alexa tyrannisiert Amerika. Tagespiegel, 25.3.2017, <https://www.tagesspiegel.de/gesellschaft/medien/aufwachsen-mit-dem-sprachassistenten-alexa-tyrannisiert-amerika/19570640.html> (6.12.2021)
- ▶ Sieber, A. (2019): Dialogroboter. Wie Bots und künstliche Intelligenz Medien und Massenkommunikation verändern. Wiesbaden
- ▶ TAB (Büro für Technikfolgen-Abschätzung) (2017): Voice Computing – allgegenwärtige Spracherkennung. (Ferdinand, J.-P.; Jetzke, T.) TAB-Themenkurzprofil Nr. 15, Berlin
- ▶ TAB (2019a): Digitale Lebensgefährten – der Anthropomorphismus sozialer Beziehungen. (Bovenshulte, M.) TAB-Themenkurzprofil Nr. 31, Berlin
- ▶ TAB (2019b): Beyond Big Data. (Ehrenberg-Silies, S.) TAB-Themenkurzprofil Nr. 34, Berlin
- ▶ TAB (2019c): Legal Tech – Potenziale und Wirkungen. (Kind, S.; Ferdinand, J.-P.; Priesack, K.) TAB-Arbeitsbericht Nr. 185, Berlin

- ▶ TAB (2020a): Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick. (Kolleck, A.; Orwat, C.) TAB-Hintergrundpapier Nr. 24, Berlin
- ▶ TAB (2020b): Robo-Recruiting – Einsatz künstlicher Intelligenz bei der Personalauswahl. (Peters, R.) TAB-Themenkurzprofil Nr. 40, Berlin
- ▶ TAB (2021a): Emotionserkennung mittels künstlicher Intelligenz. Perspektiven und Grenzen von Technologien zur Analyse von Gesichtsbewegungen. (Peters, R.) TAB-Themenkurzprofil Nr. 48, Berlin
- ▶ TAB (2021b): Gehypt! Gescholten! Der Einsatz von KI im Recruiting. (Peters, R.) TAB-Themenkurzprofil Nr. 40, Berlin
- ▶ TheNewNew (o.J.a): [multi'vocal]. <https://thenewnew.space/projects/multivocal/> (22.2.2022)
- ▶ TheNewNew (o.J.b): Syb: Queering voice AI. <https://thenewnew.space/projects/syb-queering-voice-ai/> (22.2.2022)
- ▶ TÜV-Verband (Verband der TÜV e.V.) (2021): Sicherheit und Künstliche Intelligenz. Erwartungen, Hoffnungen, Emotionen. Berlin, [https://www.tuev-verband.de/?tx_epxelo_file\[id\]=824710&cHash=5b6624273d780ebc87caf44bc821a35](https://www.tuev-verband.de/?tx_epxelo_file[id]=824710&cHash=5b6624273d780ebc87caf44bc821a35) (22.2.2022)
- ▶ UNESCO (2019): I'd blush if I could: closing gender divides in digital skills through education. (West, M.; Kraut, R.; Chew, H.) <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1> (22.2.2022)
- ▶ Wang, N. (2019): In Alibabas Hotel hat die Künstliche Intelligenz den Menschen ersetzt. Tagesspiegel, 2.12.2019, <https://www.tagesspiegel.de/wirtschaft/roboter-und-spracherkennung-in-alibabas-hotel-hat-die-kuenstliche-intelligenz-den-menschen-ersetzt/25292106.html> (1.12.2021)
- ▶ Zalando (2021): Reklamation. <https://www.zalando.de/faq/Rucksendung-and-Ruckerstattung/Reklamation.html> (22.2.2022)

Im Rahmen der Recherche zu diesem Beitrag wurden zwei Expert/inneninterviews durchgeführt, deren Inhalte im Text referenziert werden. Der Autor dankt Aljoscha Burchard (Deutsches Forschungszentrum für Künstliche Intelligenz) und Lajla Fetic (Bertelsmann Stiftung) für die zur Verfügung gestellten Informationen und ihr Mitwirken an diesem Beitrag.

Das Horizon-Scanning ist Teil des methodischen Spektrums der Technikfolgenabschätzung im TAB.

Horizon
SCANNING

Mittels Horizon-Scanning werden neue technologische Entwicklungen beobachtet und diese systematisch auf ihre Chancen und Risiken bewertet. So werden technologische, ökonomische, ökologische, soziale und politische Veränderungspotenziale möglichst früh erfasst und beschrieben. Ziel des Horizon-Scannings ist es, einen Beitrag zur forschungs- und innovationspolitischen Orientierung und Meinungsbildung des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung zu leisten.

In der praktischen Umsetzung werden im Horizon-Scanning softwaregestützte Such- und Analyseschritte mit expert/innenbasierten Validierungs- und Bewertungsprozessen kombiniert.

Herausgeber: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB)

Gestaltung und Redaktion: VDI/VDE Innovation + Technik GmbH

Bildnachweise: © kupicoo/iStock(S.1); © TAB (S. 3), © NicoElNino/iStock (S.4); © Tashi-Delek/iStock (S.5); © BestForBest/iStock (S.6); © Tero Vesalainen/iStock (S.7)

ISSN-Internet: 2629-2874