

# Von Big Data zur Künstlichen Intelligenz – Maschinelles Lernen verändert Wissenschaft und Gesellschaft

Der schnelle Fortschritt in Informatik, Messtechnik und Sensorik versetzt uns heute in die Lage, immer vielfältigere, umfangreichere und genauere Daten immer schneller und günstiger zu erheben. Gleichzeitig wächst unser Verständnis dafür, wie solche Datenmengen intelligent verknüpft und durch Verfahren des Maschinellen Lernens so analysiert werden können, dass daraus neues Wissen und intelligente Leistungen entstehen können, deren Automatisierung noch vor wenigen Jahren undenkbar schien. Der Vortrag beleuchtete die Grundlagen dieser Entwicklung und erläuterte die Prinzipien, nach denen Verfahren des Maschinellen Lernens und der intelligenten Datenverknüpfung ihre Leistungen erbringen.



Prof. Dr. **Stefan Wrobel** ist Professor für Informatik an der Universität Bonn und Leiter des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS. Er studierte Informatik in Bonn und Atlanta, Georgia, USA (M.S., Georgia

Institute of Technology) mit Schwerpunkt Künstliche Intelligenz und promovierte an der Universität Dortmund. Nach Stationen in Berlin und Sankt Augustin wurde er Professor für Informatik an der Universität Magdeburg, bevor er im Jahr 2002 den Ruf auf seine aktuelle Position annahm. Seit 2014 ist er zudem einer der Direktoren des Bonn-Aachen International Center for Information Technology (b-it).

Professor Wrobel beschäftigt sich seit vielen Jahren mit Aspekten der Digitalisierung, insbesondere mit intelligenten Algorithmen und Systemen zur Analyse großer Datenmengen und dem Einfluss von Big Data/Smart Data auf die Nutzung von Informationen in Unternehmen und der Gesellschaft. Er ist Autor einer großen Zahl von Publikationen in den Gebieten des Data Mining und des Maschinellen Lernens, Mitglied des Herausgeber-Gremiums mehrerer führender Fachzeitschriften und Gründungsmitglied der International Machine Learning Society.

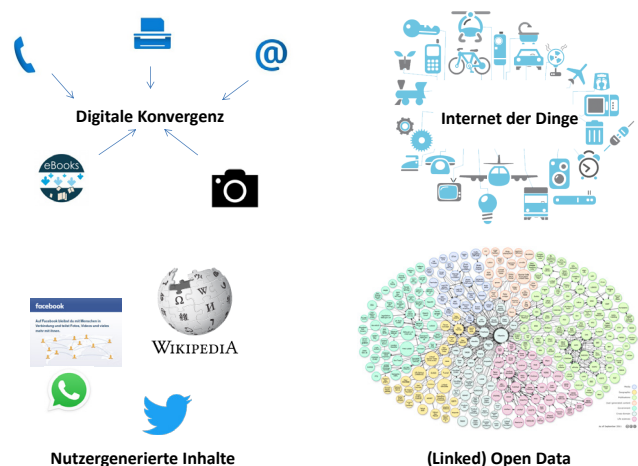
*Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS  
Schloss Birlinghoven, 53757 Sankt Augustin und Rheinische Friedrich-Wilhelm-Universität Bonn, Institut für Informatik, Endenischer Allee 19a, 53115 Bonn*

„Wir machen heute im Alltag viele Dinge digital, die vor zehn Jahren noch meist analog abliefen. Musik hören, online Geld überweisen oder online bestellen und einkaufen sind nur einige dieser Alltagstätigkeiten, die überall auf der Welt Daten generieren.“

Mit diesem Satz knüpfte Stefan Wrobel an den Vortrag seines Vorredners an, der bereits auf die Bedeutung von Big Data im industriellen Bereich hingewiesen hatte. Die eingangs genannten Beispiele, die man unter dem Stichwort „Digitale Konvergenz“ zusammenfasst, zeigen, dass das Thema umfassend betrachtet werden muss. Der Trend geht dahin, für ursprünglich unterschiedlich aufgestellte Systeme eine gemeinsame – digitale – Plattform zu schaffen, die Vernetzungsmöglichkeiten und Zuwachs an Daten ermöglicht.

## Digitalisierung und Big-Data-Trends

Die digitale Konvergenz ist einer von vier digitalen Trends, die „Big Data“ ausmachen. (Abb. 1).



**Abb. 1.** Digitalisierung und die vier Big-Data-Trends. [Abb. www.free-pik.com, LOD cloud: Richard Cyganiak Anja Jentszsch CC BY-SA 3.0]

„Im Internet und in den sozialen Netzwerken produziert jeder von uns täglich große Mengen an Daten als nutzergenerierte Inhalte, die verarbeitet werden können und wertvolle Informationen enthalten“, zählte Wrobel weiter auf. Hinzu kommt, dass die weltweit verknüpften intelligenten Maschinen im „Internet der Dinge“ (Internet of Things) enorme Datenströme erzeugen. Schließlich nannte Wrobel noch die Öffnung der Zugänge zu Daten – Unternehmen und Organisationen können zunehmend auf Daten zugreifen und sie für ihre Geschäftsmodelle oder Projekte nutzen: „Institute, Forschergruppen und Regierungen stellen immer mehr Daten öffentlich und kostenfrei zur Verfügung, mit denen wir etwas tun und die wir nutzen können.“

Das Datenwachstum erzeugt ein riesiges Universum an potentiell nutzbaren Daten. Für das Jahr 2025 sind Datenmengen im Bereich von Zettabyte zu erwarten (Abb. 2).

„Das Datenwachstum ist eine große Herausforderung. Es ist aber auch eine Chance etwas zu tun, wovon der Mathematiker Alan Turing bereits 1950 träumte. Er überlegte sich, wie man intelligente Maschinen realisieren könnte. Ihm war klar, dass wir durch Programmieren kein intelligentes Verhalten erzeugen können. Mittlerweile wissen wir, dass wir aus den vielen Daten, die wir sammeln, lernen können“, so Wrobel.

Auf diese Weise wurden beachtliche Erfolge erzielt (Abb. 3). Populär wurde Künstliche Intelligenz und Maschinelles Lernen unter anderem durch die ersten Computer, die in den USA bei der Quizshow Jeopardy gewannen (2011) – hierfür waren in natürlicher Sprache gestellte Fragen zu beantworten – und eine

Dezimalpräfixe		
Name	Symbol	Anzahl Bytes
Kilobyte	KB	1 000 = 10 <sup>3</sup>
Megabyte	MB	1 000 000 = 10 <sup>6</sup>
Gigabyte	GB	1 000 000 000 = 10 <sup>9</sup>
Terabyte	TB	1 000 000 000 000 = 10 <sup>12</sup>
Petabyte	PB	1 000 000 000 000 000 = 10 <sup>15</sup>
Exabyte	EB	1 000 000 000 000 000 000 = 10 <sup>18</sup>
Zettabyte	ZB	1 000 000 000 000 000 000 000 = 10 <sup>21</sup>
Yottabyte	YB	1 000 000 000 000 000 000 000 000 = 10 <sup>24</sup>

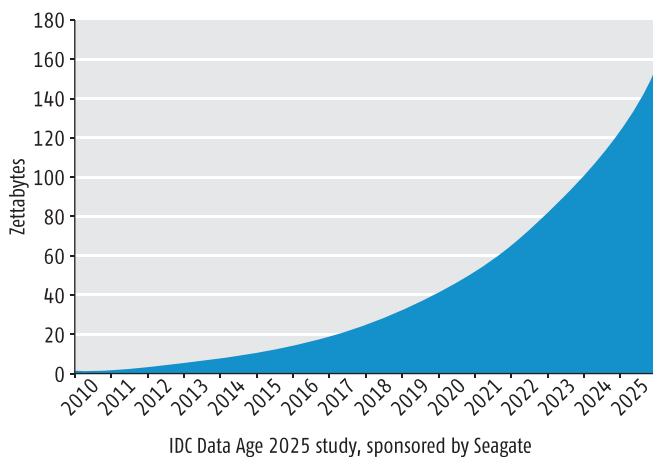


Abb. 2. Wachsende Datenmengen stellen eine Herausforderung dar, sie zu verwalten, bieten aber zugleich die Chance, sie zu nutzen.

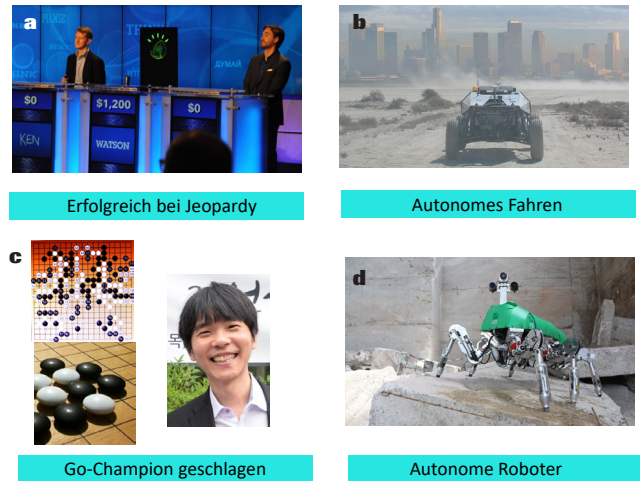


Abb. 3. Vier populäre Beispiele für Künstliche Intelligenz: – a. 2011 schlägt der IBM Watson Computer zwei der erfolgreichsten Wettkämpfer in der US-amerikanischen Quizz-Show Jeopardy. – b. Roboterfahrzeug bei der DARPA Grand Challenge, an der autonome Fahrzeuge teilnahmen. – c. 2016 schlägt AlphaGo den Go-Meister Lee Se-Dol (der eine von fünf Partien gewann). Das Match wurde von Science als „Breakthrough of the Year“ angesehen. – d. Autonome, sechsbeinige Roboter, der sich in unbekanntem Terrain fortbewegen kann. [Photos von oben links im Uhrzeigersinn: IBM | Spaceape/Wikimedia CC-BY-SA 3.0 | KIT | LG Electronics CC-BY SA 2.0]

KI, die 2017 Go-Champion wurde. Auch die Entwicklung von autonomen Fahrzeugen und autonom sich bewegenden und orientierenden Robotern sorgten für Aufmerksamkeit. Die Prognose von Alan Turing wird langsam Realität.

Damit hat eine neue Ära begonnen, vergleichbar mit der Entwicklung intelligenter cyber-physischer Systeme (vgl. Vortrag Liggesmeyer). Der Weg geht von programmierten Systemen zu lernenden KI-Systemen:

Programmierte Systeme	→	Lernende KI-Systeme
nutzen vordefinierte Regeln (Programme)		werden trainiert, nicht programmiert
verfolgen vordefinierte Regeln (Programme)		lernen und verbessern sich auf Basis von Erfahrung
stellen allen Nutzern den gleichen Output bereit		interpretieren Sensordaten und unkonventionelle Daten

Vom händischen Programmieren zum Lernen

„Wir sind jetzt dabei, uns weiter zu entwickeln von programmierten Systemen, deren Verhalten wir genau vorgegeben haben, hin zu lernenden Systemen, die auf ihre Umgebung reagieren und die die gesamten Daten, die ihnen zur Verfügung stehen, verwenden, um daraus Verhalten zu lernen. All dies trägt dazu bei, die Leistungsfähigkeit solcher Systeme enorm zu steigern. Computer beginnen jetzt, Bilder, Töne und Texte zu ‚verstehen‘ und zu interpretieren. Sie können schon bestimmte Informationen aus Texten extrahieren, Objekte erkennen und klassifizieren. Zwar noch nicht wie wir Menschen, aber schon mit beeindruckenden Ergebnissen.“

Es handelt sich für Wrobel um einen wahrhaft fundamentalen Sprung: „Die Fähigkeit der Computer war zu Beginn

Kein Nachdruck, keine Veröffentlichung im Internet oder einem Intranet ohne Genehmigung des Verlags und der GDNÄ

begrenzt durch die Fähigkeit der Menschen, die sie programmieren. Dann lernten Computer aus Datenbanken. Wir werden zukünftig sehen, dass Computer überall auf der Welt, auch hier im Vortragssaal, aus der natürlichen Umgebung lernen können.“

Für Wrobel ist klar, dass diese neue Ära, die von Künstlicher Intelligenz geprägt ist, unsere Zukunft bestimmen wird.

„Künstliche Intelligenz begegnet uns heute schon überall im Alltag, wie einige Beispiele aus unserer Arbeit am Fraunhofer IAIS zeigen: Lernende Systeme analysieren zum Beispiel die Inhalte von Medienarchiven, verschlagworten die Inhalte von Radio- und Fernsehbeiträgen und ermöglichen so Redakteuren eine gezielte Suche in riesigen audiovisuellen Datenmengen, die Menschen alleine nicht mehr erfassen könnten.

In der Medizin ermöglicht KI das Erkennen und Interpretieren von Krankheitsherden auf Röntgenaufnahmen oder CT-Bildern – dank des Maschinellen Lernens können sie dies teils schon besser und schneller als das geschulte Spezialistenauge, welches schnell ermüdet. Intelligente Algorithmen helfen beim Autokauf und empfehlen auf Basis von persönlichen Präferenzen im Bereich Reisen, Architektur oder Essen das zum Lebensstil passende Auto zu finden. Beim Autofahren geben autonome Systeme dem Fahrer eine Orientierung im Schilderwald, weil sie in der Lage sind, die Straßenschilder zu erkennen und zu interpretieren.“

Die Ära der Künstlichen Intelligenz setzt nicht schlagartig ein, sie hat schon längst begonnen und erlangt immer größere Bedeutung. Zentral ist die Verbindung von den Daten über ihre Verknüpfung bis hin zum Lernen. Aus diesem Grund hat Fraunhofer das Forschungscluster „Cognitive Internet Technologies (CCIT)“ ins Leben gerufen: Mit der gebündelten Kompetenz von insgesamt 13 Fraunhofer-Instituten sollen kognitive Technologien für das industrielle Internet entwickelt werden. Forscher aus unterschiedlichen Disziplinen erarbeiten ein ganzheitliches Lösungsangebot entlang der Wertschöpfungskette vom Sensor über intelligente Lernverfahren bei der Datenverarbeitung bis hin zur Cloud. Hier entstehen bereits heute Schlüsseltechnologien für vertrauenswürdige Echtzeit-Kommunikation, für Datenräume zur kontrollierbaren unternehmensübergreifenden Kollaboration sowie nachvollziehbare Lernverfahren, basierend auf domänenspezifischem Expertenwissen.

**Was braucht man, um Künstliche Intelligenz zu realisieren?**

1. Die Fähigkeit zum Maschinellen Lernen durch geeignete Algorithmen, z. B. zum Training neuronaler Netze, von denen es unterschiedliche Varianten gibt, von denen die aktuell bedeutendste das Deep Learning ist,
2. eine große Rechenleistung,
3. ausreichend viele Trainingsdaten zum Lernen und schließlich,
4. strukturiertes Wissen.

„Was einfach klingt, ist aber eine fundamentale Herausforderung für die Wissenschaft. Maschinelles Lernen und die Nutzung von Wissen müssen sich für langfristig erfolgreiche KI unmittelbar miteinander verbinden zur sogenannten Hybriden KI“.

Beispielhaft ging Wrobel auf das Maschinelle Lernen ein und illustrierte an einem einfachen Beispiel die Heraus-

**„Lerne“ eine mathematische Funktion aus Beispielen (ein Polynom)**

x	f(x)	x	f(x)
1	1	1	1
2	4	2	1
3	9	3	1
4	16	4	1
5	25	5	1 121

---

$f(x) = x$

$f(x) = (x-4)(x-3)(x-2)(x-1)x+1$

**Abb. 4.** Was macht ein Lernverfahren? – Eine Aufgabe zum Selbsttest.

forderung, die damit verbunden ist: Während sich aus der Zahlenreihe 1, 4, 9, 16 als nächstes leicht 25 vorhersagen lasse und damit die Funktion  $f(x) = x^2$ , wird wohl jeder bei die Zahlenreihe 1, 1, 1, 1 als nächstes mit 1 fortsetzen und doch danebenliegen. Denn es gibt im Prinzip eine unendliche Zahl von Funktionen, mit denen man eine endliche Zahl von Punkten korrekt wiedergeben kann (Abb. 4). Es handelt sich um eine fundamentale Herausforderung! Fazit: Die Datenmenge alleine reicht nicht aus. Man braucht Wissen als „Leitplanken“.

**Warum ist KI so wichtig für Wirtschaft und Wissenschaft?**

„Verschiedene Studien zeigen, dass allein in Deutschland das Bruttoinlandsprodukt bis 2030 mit KI um 160 Milliarden Euro höher sein wird also ohne. Wobei dieser Wert noch sehr konservativ ist“, erklärte Wrobel, „denn KI ist keine Spartentechnologie, sondern wird alle Bereiche unseres Alltags beeinflussen. Da verhält es sich ähnlich wie mit der Elektrizität und dem Strom: Die werden auch überall gebraucht und genutzt. KI ist eine flächen-deckende Technologie und sie wird die Wettbewerbsfähigkeit quer über alle Branchen beeinflussen und stärken.“

Der Durchbruch der KI hänge daher auch von unserem Technologieverständnis bzw. von unserer Offenheit für neue Technologien ab. Die Technologieentwicklung träfe aber mittlerweile auf veränderte, offenere Kunden mit weniger Technikfeindlichkeit und weniger Kundenvorbehalten, dies könne den Durchsatz von KI in unserem Alltag erheblich beschleunigen.

Auch für die Wissenschaft sieht Wrobel ein großes Potential durch die KI. Als Beispiel nannte er „Robot Scientists“, die im Labor von Ross King in Manchester Genfunktionen von Hefebakterien (*Saccharomyces cerevisiae*) und vielversprechende Kandidaten-Wirkstoffe gegen Krebs und Malaria entdeckt haben, also echtes wissenschaftliches Wissen generiert haben.

„Der Roboter hat selbst in die Welt geschaut und sich Experimente ‚ausgedacht‘. Man kann nun diskutieren, ob das nun schon Wissenschaft ist oder doch nur Kärrnerarbeit. Aber Kärrnerarbeit ist in der Wissenschaft sehr wichtig und eröffnet nun die Chance auf höhere Effizienz und Beschleunigung des wissenschaftlichen Fortschritts – und das bei niedrigen Kosten. Die Verbesserung der Reproduzierbarkeit von wissenschaftlichen Ergebnissen ist ein weiterer positiver Aspekt; denn Reproduzierbarkeit ist eine

Kein Nachdruck, keine Veröffentlichung im Internet oder einem Intranet ohne Genehmigung des Verlags und der GDNÄ

große Herausforderung in allen Wissenschaften – selbst in der Informatik stellt sich die Frage, ob andere Labore zum selben Ergebnis kommen“, sagte Wrobel. In der Pharmaindustrie wird dieser Trend mit großem Interesse verfolgt, hier liegt sicher ein großes Potential. 90% der Experten in der Pharmaindustrie nahmen 2017 an, dass Robotik im Jahr 2020 eine bedeutende Rolle spielen wird.

Ein weiterer Aspekt, um die in der Wissenschaft erforderliche Reproduzierbarkeit zu gewährleisten, ist die Vernetzung von Datenbeständen, um sie anderen Forschern zugänglich zu machen. Aktuell wird dies über die *Open Access Infrastructure for Research in Europe (OpenAIRE)* praktiziert und erprobt.

### Wo liegen die Herausforderungen für die Gesellschaft?

- **Arbeitsplätze**

Die Arbeitswelt wird sich durch KI ebenfalls entscheidend wandeln. „Arbeitsplätze werden durch KI substituiert, an anderen Stellen entstehen neue Arbeitsformen. Auch hochwertige Tätigkeiten können rationalisiert werden, z. B. bei Beraterfirmen, die Geschäftsberichte analysieren und Kennzahlen interpretieren“, nannte Wrobel ein Beispiel. Dabei wies er nachdrücklich darauf hin, dass sich entsprechend auch die Ausbildung ändern müsse, um den Erfordernissen gerecht zu werden.

- **Digitale Souveränität**

„Datensicherheit und digitale Souveränität sind eine weitere Herausforderung für unsere Gesellschaft.“ Hier gehe es vor allem um die Anonymisierung und Pseudonymisierung zum Schutz der Privatsphäre bei sensiblen Daten. Welche tiefen Einblicke KI selbst mit unverfänglich erscheinenden Daten gewinnen kann, demonstrierte Wrobel anhand einer Untersuchung zur Auswertung von Facebook-„Likes“. Bereits 2014 waren mit hoher Sicherheit Aussagen zu ethnischer Zugehörigkeit (95%), Geschlecht (93%), sexueller Orientierung (75 bzw. 88%) oder politischer Einstellung (85%) der Nutzer möglich. Dies bedeutet eine Herausforderung für die ganze Gesellschaft. Aber es gibt nicht nur regulatorische Antworten, sondern auch technische, die die Wissenschaft bieten kann, wie Wrobel betonte. Tatsächlich können (wünschenswerte) KI-Analysen durchgeführt werden, ohne dass personengebundene Daten notwendig sind. Als Beispiel nannte Wrobel Fraunhofer-Algorithmen zur Datenanalyse, die nicht zentralisiert ausgewertet werden, sondern lokal und geschützt bleiben und dennoch eine hohe Analysequalität bei medizinischen Untersuchungen garantieren:

„Für das Erkennen bestimmter Krankheiten ist es nämlich nicht wichtig zu wissen, welche Krankheiten Frau Müller oder Herr Meier vorher hatten, sondern es kommt nur auf das Wissen an, wie sich bestimmte Krankheiten entwickeln“, so Wrobel.

Das Problem der digitalen Souveränität stellt sich insbesondere auch für die Industrie, deren Entwicklung und Konkurrenzfähigkeit ja entscheidend von der Anwendung der KI abhängig ist. Eine Antwort ist hier die *International Data Space Initiative*, in der souveräne Datenräume für Wirtschaft, Wissenschaft und Gesellschaft geschaffen werden. Ursprünglich ein vom Bundesministerium für Bildung und Forschung (BMBF) geför-

deres Forschungsprojekt der Fraunhofer-Gesellschaft, hat die Initiative großen Widerhall gefunden und firmiert nun unter der Bezeichnung *International Data Space Association IDSA*. An ihr sind rund 100 international tätige Unternehmen und Organisationen beteiligt, auch aus den USA und dem asiatischen Raum. Das Data-Space-Referenzarchitekturmodell liegt mittlerweile in der dritten Version vor und die Pilotierung in ausgewählten Anwendungsfällen läuft. Der International Data Space ist ein Daten-Ökosystem, in dem unterschiedliche Akteure den sicheren Austausch von Daten und die einfache Verknüpfung von Daten in Geschäftsökosystemen auf Basis von Standards und mit Hilfe gemeinschaftlicher Governance-Modelle sicher und nach marktwirtschaftlichen Prinzipien durchführen können. Die Idee ist: Daten zu teilen, breit verfügbar zu machen, aber doch souverän über sie bestimmen zu können.

- **Verlässlichkeit**

Eine große Herausforderung sieht Wrobel nicht zuletzt in der Verlässlichkeit von KI Systemen.

„KI agiert anders als der Mensch und wird auch weiterhin nicht perfekt sein können, denn die aus Trainingsdaten gelernten Modelle haben in der Regel kein allgemeines Verständnis der zu Grunde liegenden Phänomene.“ Beispielsweise wurden für uns eindeutig als Stoppschilder erkennbare Bilder von Computern komplett fehlinterpretiert, weil die trainierten Modelle auf manipulierte Bilder nicht vorbereitet waren. „Wenn dem Rechner zu viele falsche oder nicht zielführende Informationen zum Lernen eingetrichtert werden, können diese später zu fehlerhaften Interpretationen führen“, warnte der KI-Experte.

Ein anderes Beispiel aus der Medizin zeigte ein Röntgenbild einer Kardiomegalie (krankhafte Herzvergrößerung), das vom Computer zwar richtig diagnostiziert wurde, aber aus den falschen Gründen: Das Modell orientierte sich zu stark an einem eingblendeten Geräteschriftzug eines portablen Röntgengeräts. Dieses Gerät wird häufig bei tatsächlich schwer Erkrankten eingesetzt, die nicht zu einem stationären Gerät transportiert werden können, so dass der indirekte Schluss auf eine Erkrankung das richtige Ergebnis hat, aber nicht aus medizinischen Gründen erfolgte.

Es wird zukünftig also entscheidend sein, weiter an vertrauenswürdiger und abgesicherter Künstlicher Intelligenz zu arbeiten, wie das bei Fraunhofer und vielen anderen Forschungseinrichtungen bereits passiert. Denn der Trend zur KI wird weitergehen. In der breiten Öffentlichkeit wird sie zum Beispiel im Bereich der Spracherkennungssysteme wahrnehmbar sein. Wrobel ist überzeugt, dass immer mehr Benutzerschnittstellen durch Sprache ersetzt oder ergänzt werden. „Alexa ist erst der Anfang. Natürlichsprachliche Dialogassistenzen werden zunehmend unseren Alltag bestimmen. Die Interaktion mit Computern wird sich stark verändern. In ferner Zukunft wird der Mensch natürlich auch jenseits der Sprache mit Computern interagieren. Irgendwann werden wir auch Roboter mit unseren Gedanken steuern können. Das ist noch Zukunftsmusik, wird aber bereits jetzt erforscht.“