

# CTO- Dialog

Prof. Dr.  
Christian Bauckhage  
über

## Quo Vadis Künstliche Intelligenz?

Das Thema künstliche Intelligenz (KI) erregt seit kurzem große öffentliche Aufmerksamkeit und wird teils begeistert, teils argwöhnisch diskutiert. In der Tat zeigen Systeme wie AlphaGO, Google Duplex oder IBM Debater, welchen sprunghaften Fortschritt die KI in den letzten Jahren gemacht hat, und deuten an, was nun alles in den Bereich des Möglichen rückt. Bei genauer Betrachtung der medialen Debatte fallen jedoch mehrere Aspekte auf. Einerseits hat vor allem maschinelles Lernen (ML) jüngst dazu geführt, dass KI in Form von Navigationssystemen, Sprachassistenten, oder Chatbots schon jetzt weitestgehend unbemerkt Bestandteil unseres Lebens- und Berufsalltags geworden ist. Andererseits werden die technischen Begriffe „künstliche Intelligenz“ und „maschinelles Lernen“ mit Assoziationen überfrachtet, die nicht angemessen sind. Wenn wir also über Chancen und Risiken der KI für Wirtschaft und Gesellschaft nachdenken, sollten wir zunächst klären, was diese eher unglücklich gewählten Fachbegriffe eigentlich bezeichnen.

„Künstliche Intelligenz“ umschreibt die Idee, kognitive Fähigkeiten des Gehirns (sehen, hören, planen,

entscheiden, etc.) auf Computern zu programmieren oder allgemeiner auf Maschinen nachzubilden. Eine Maschine gilt als schwach künstlich intelligent, wenn sie einzelne solcher Leistungen in einem fest umrissenen Anwendungsbereich robust erbringen kann. Schwache KI ist heutzutage bereits Realität; so gibt es etwa KI Systeme zur medizinischen Diagnostik oder industriellen Qualitätskontrolle, die in Bildern nach Anzeichen von Krankheiten oder Schäden suchen und diese oft schneller und verlässlicher erkennen als menschliche Experten. Starke künstliche Intelligenz würde einer Maschine zugeschrieben, deren Fähigkeiten deutlich über solche spezielle Expertise hinausgehen. Sie müsste z.B. in der Lage sein, auch jenseits ihres eigentlichen Anwendungsbereichs verlässlich zu funktionieren, was eine generelle Form künstlicher Kreativität erfordern würde, die beim heutigen Stand der Technik nach wie vor Fiktion ist. Schwachen KI wurde zuletzt vor allem durch lernende Algorithmen beflügelt, die anhand von Big Data trainiert werden, spezifische Aufgaben zu lösen. Aber auch die Begriffe „lernen“ und „trainieren“ sind vor allem technisch zu verstehen. „Maschinelles Lernen“ beschreibt nichts anderes als die Idee, die Parameter eines mathematischen Modells

**HR. PROF. DR. BAUCKHAGE**

Christian Bauckhage ist Professor für Informatik (Mustererkennung) an der Universität Bonn und wissenschaftlicher Direktor des Fraunhofer Zentrums für maschinelles Lernen am Fraunhofer IAIS in Sankt Augustin. Nach Studium der Informatik und theoretischen Physik in Bielefeld und Forschungssemestern am Institut National de Recherche en Informatique et en Automatique (INRIA) in Grenoble promovierte er 2002 in Bielefeld in Informatik. Anschließend forschte er am Centre for Vision Research in Toronto und war Senior Research Scientist bei den Deutschen Telekom Laboratories in Berlin, bevor er 2008 nach Bonn berufen wurde. In seiner Forschung widmet er sich Theorie und Praxis der künstlichen Intelligenz und des maschinellen Lernens und hat hierzu zahlreiche wissenschaftliche Beiträge veröffentlicht. Aktuelle Schwerpunkte seiner Forschung liegen auf hybriden kognitiven Systemen, die nicht nur anhand von Daten lernen, sondern auch Expertenwissen nutzen, um zu erklär- und nachvollziehbaren Entscheidungen zu kommen. In Projekten mit der Industrie werden derartige Lösungen etwa zur automatischen Qualitätskontrolle oder zur automatischen Analyse von Geschäftsberichten genutzt. Prof. Bauckhage ist überzeugter Proponent des Open Science und Open Education Gedankens; schon seit Jahren berät er Politik, Industrie und Wirtschaft zu diesen Themen und hält zahlreiche öffentliche Vorträge.

anhand vorhandener Daten automatisch so einzustellen, dass das Modell anschließend zur Analyse und Verarbeitung ähnlicher Daten eingesetzt werden kann. Dieses Prinzip kennen wir aus dem Mathematikunterricht, wo wir in der Kurvendiskussion die Parameter einer mathematischen Funktion berechnen mussten, so dass die Funktion vorgegebene Datenpunkte gut beschreibt. Allerdings sind die mathematischen Modelle, die in der KI benutzt werden, deutlich komplexer als diejenigen aus unserer Schulzeit.

Insbesondere so genannte künstliche neuronale Netze waren zuletzt höchst erfolgreich. Diese Modelle bestehen aus vernetzten Funktionen, die die Funktionsweise der Neuronen im Gehirn simulieren. Jedes einzelne Neuron berechnet eine einfache, mathematische Funktion seiner Eingaben und übermittelt die Ergebnisse an andere Neuronen, die damit weitere Berechnungen anstellen. Insgesamt ist ein neuronales Netz daher nichts Anderes als ein zugegebenermaßen komplizierter Algorithmus, der Eingabedaten in Ausgaben überführt. Soll ein neuronales Netz komplizierte Aufgaben lösen, müssen die Parameter der einzelnen Neuronen so eingestellt sein, dass die Berechnungen des gesamten Netzes der Aufgabe gerecht werden. In einer so genannten Trainingsphase werden dem Netz dazu Eingabedaten präsentiert, zu denen es Ausgaben berechnet. Weichen diese vom eigentlich gewünschten Ergebnis ab, werden die Parameter des Netzes durch einen Optimierungsalgorithmus so lange justiert, bis der durchschnittliche Ausgabefehler klein genug ist, um das Netz in der Praxis nutzen zu können.

Obwohl all diese Mechanismen gut verstanden sind, finden sich in den Medien dennoch regelmäßig Berichte über Fehlleistungen solcher KI System. Wie kann das sein? Ein fundamentales Problem besteht darin, dass Trainingsdaten fehlerhaft oder

verzerrt sein können. Selbst die Giganten des Silicon Valleys sind davor nicht gefeit. Legendär ist etwa ein Zwischenfall aus dem Jahr 2015 als Google Fotos ein Bild mit dunkelhäutigen Menschen als „Gorillas“ klassifizierte; die Entwickler des Systems hatten schlicht vergessen, nicht nur Bilder hellhäutiger Menschen in ihre Trainingsdaten aufzunehmen. Sobald das Problem erkannt war, ließ es sich durch weiteres Training einfach lösen. Generelle Methoden, die es erlauben würden, automatisch festzustellen ob Trainingsdaten verzerrt sind, gibt es allerdings noch nicht. KI Systeme haben eben keinen „gesunden Menschenverstand“, der es ihnen erlauben würde, ihre eigene Funktion kritisch zu reflektieren. Dies ist durchaus kritisch, denn datengetriebene KI ist immer nur so gut, wie die Daten, mit denen sie trainiert wurde. Ein weiteres Problem besteht darin, dass rein datengetriebene KI Systeme oft Black Boxes sind, deren interne Berechnungen nicht unmittelbar transparent sind. Anders gesagt ist zwar sehr gut verstanden, wie ein neuronales Netz funktioniert, welche seiner Millionen gekoppelten Berechnungen letztendlich dazu führen, dass eine gewisse Ausgabe produziert wird, ist im praktischen Einsatz aber meist nicht zu erkennen. Dies bereitet dort Probleme, wo KI Technologien Entscheidungen treffen sollen, die aufgrund regulatorischer Vorgaben zwingend nachvollziehbar sein müssen, etwa in der Kreditvergabe oder bei der Steuerung kritischer Anlagen oder Infrastrukturen. In der Tat rückt das Problem der erklärbaren KI (explainable AI) zunehmend in den Fokus und wird weltweit mit Hochdruck beforscht, in Deutschland etwa im Fraunhofer Zentrum für Maschinelles Lernen.

Zusammenfassend bleibt festzustellen, das schwache KI, die auf maschinellem Lernen beruht, mittlerweile zu erstaunlichen Leistungen fähig ist, enormes



wirtschaftliches Potential hat und vor allem Produktivitäts- und Effizienzsteigerungen ermöglicht. Vor diesem Hintergrund sollten Firmen, die noch nicht darüber nachgedacht haben, was KI für ihre Prozesse oder Geschäftsmodelle leisten kann, schnellstmöglich damit beginnen, entsprechende Überlegungen anzustellen und sich in Form kleiner, agiler Projekte mit den Möglichkeiten der KI vertraut machen.

Aber nicht nur die Wirtschaft, sondern auch Politik und Gesellschaft dies werden sich den Herausforderungen stellen müssen, die der zunehmende Praxiseinsatz von KI und die damit verbundenen Disruptionen mit sich bringen. Dabei muss aber immer bedacht werden, dass weder naiver Optimismus noch überkritischer Pessimismus angebracht sind. Insbesondere die Metaphern „Intelligenz“ und „lernen“ dürfen nicht falsch verstanden werden: Künstliche Intelligenz verhält sich zur natürlichen Intelligenz, so wie

sich Flugzeuge zu Vögeln verhalten. Aktuelle KI Systeme denken nicht, assoziieren nicht, haben keine Ideen und vor allem keine Selbsterkenntnis und keine eigenen Antriebe. Sie lernen weder von sich aus noch selbstständig, sondern sind nichts Anderes als komplexe Computerprogramme, die von Menschen entworfen und bedient werden. Daher gilt für Entwicklung und Anwendung des Werkzeugs KI das Gleiche, wie bei anderen Technologien auch: Der Mensch steht in der Verantwortung, sicher zu stellen, dass die Technologie adäquat und unfallfrei genutzt wird. Dort wo dies sichergestellt werden kann, bieten sich Chancen; dort, wo dies nicht gelingt, drohen negative Konsequenzen. Die gesellschaftliche Debatte zur KI sollten deren Potentiale und Probleme daher klar benennen und abwägen und vor dem Hintergrund der globalen Entwicklung in diesem Bereich Konsens darüber schaffen, welche Vor- und Nachteile wir in Zukunft bereit sind, zu akzeptieren.

