

Klüger als wir?

Über kognitive Robotik

MICHAEL DECKER

Geboren 1965 in Ludwigshafen am Rhein, Physiker, Professor für Technikfolgenabschätzung und Leiter des Bereichs Informatik, Wirtschaft und Gesellschaft am Karlsruher Institut für Technologie (KIT).

Im März 2016 unterlag der weltbeste Go-Spieler Lee Sedol einem Computer mit der Software AlphaGo, die von Google entwickelt wurde. Ähnlich wie bei dem Sieg von Deep Blue gegen den damaligen Schachweltmeister Garri Kasparow wurde in der Berichterstattung der Eindruck erweckt, als ginge es um mehr als nur eine „technisch-sportliche“ Nach-

richt: Die Vorherrschaft der Menschen als der intelligentesten Spezies mit den höchsten kognitiven Fähigkeiten sei in Gefahr. Warum nehmen wir diese Nachrichten so interessiert auf? Weil sie unser Menschenbild infrage stellen? Oder weil wir uns unmittelbar die technischen Veränderungen vor Augen führen, die durch kognitive Systeme möglich werden?

Während Roboter in Produktionshallen etabliert sind, gestaltet sich die Entwicklung von Service-Robotern schwieriger. In der Produktionshalle können alle Vorkehrungen für den optimalen Einsatz eines Roboters getroffen werden. In der realen Welt ändern sich jedoch sowohl die Umgebung als auch die Handlungskontexte ständig. Um diese Veränderungen rechtzeitig wahrnehmen und dann auch adäquate Aktionen ausführen zu können, sind „menschliche Fähigkeiten“ wie Sehen, Hören, Sprechen, Fühlen, Lernen, Anpassen und Schlüsse ziehen gefragt. Die klassischen Stärken der Industrierobotik, präzise, reproduzierbar und unermüdlich programmierte Bewegungen auszuführen, treten in den Hintergrund.

Wenn das die Herausforderungen sind, dann ist es nicht verwunderlich, dass seitens der technischen Entwickler ebenfalls auf menschliche Fähigkeiten Bezug genommen wird, wenn von künstlicher Intelligenz, kognitiver oder sozialer Robotik, maschinellem Lernen oder autonomer Technik gesprochen wird. Es hilft bei der Beschreibung der Technik, sich dieser allgemein verständlichen Begrifflichkeiten zu bedienen. Gleichzeitig ist diese Begriffsnutzung auch der Ursprung vieler Missverständnisse, weil autonomes, kognitives, soziales Verhalten bei Menschen und Maschinen prinzipiell unterschiedlich sind. Diese Missverständnisse hätte man vermeiden können, wenn man Mitte des letzten Jahrhunderts „künstliche Intelligenz“ zum Beispiel „alternatives Programmieren“ genannt hätte. Wobei unbenommen ist – was von Kritikern im Wettbewerb um die Gewinnung von Fördermitteln beklagt wird –, dass die „anthropomorphe“ Rede deutlich mehr Neugier weckt.

DEM MENSCHEN MÖGLICHT ÄHNLICH

Humanoide Roboter üben einen besonderen Reiz aus, gerade dann, wenn sie nicht nur eine am Menschen orientierte Körperform haben, sondern auch durch künstliche Haut, Haare, Augen et cetera dem Menschen zum Verwechseln ähnlich gestaltet werden. Bei der CeBIT 2017 konnte das beobachtet werden, als Starwissenschaftler Hiroshi Ishiguro zum Thema „Androids, Robots, and Our Future Life“ vortrug und seine Roboter vorstellte. Sein erklärtes Ziel ist es, Roboter zu bauen, die dem echten Menschen möglichst ähnlich sind. Im europäischen Raum und in Deutschland werden humanoide Roboter als Mittel zum Zweck gebaut: Dabei liegt die Annahme zugrunde, die humanoide Körperform ermögliche es dem Roboter besser, seine Aktionen in einer Welt, die auf die Nutzung durch Menschen optimiert ist, durchzuführen.

Technik als Mittel zum Zweck zu analysieren, ist auch die Perspektive der Technikfolgenforschung. Menschen setzen Technik als Werkzeug ein, um ihre Handlungsziele besser erreichen zu können. In der Fertigung sind das etwa Schweißen, Bohren, Werkstücke positionieren, im Servicebereich

Putzen, Haare waschen, Pflegen und so weiter. Natürlich verändern sich mit vorhandenen Mitteln auch die Zwecke, die Menschen verfolgen.

Zunächst soll der Blick auf die (nicht) intendierten, die (un)erwünschten Folgen und allgemein auf die Chancen und Risiken moderner Robotersysteme in Bezug auf die Ersetzbarkeit des Menschen gerichtet werden. Danach wird spezieller auf lernende Systeme eingegangen.

WO SIND WIR ERSETZBAR?

Dass Roboter Menschen ersetzen – und auch deren Arbeitsplätze gefährden –, ist seit Beginn der Automatisierung ein wichtiges Thema. Seit 2013 wird es auch mit Blick auf Dienstleistungen verstärkt diskutiert, nachdem Carl Benedikt Frey, Professor für Volkswirtschaftslehre an der Universität Oxford, gemeinsam mit seinem Kollegen Michael Osborne ihre Studie „Die Zukunft der Beschäftigung“ publiziert und darauf hinwiesen hatten, welche Berufe besonders durch Computer und Roboter gefährdet sind. Seitens der technischen Entwicklung ist weniger die Ersetzung als die Kooperation – das geschickte Zusammenwirken von Mensch und Roboter – das Paradigma, wobei letztendlich der Roboter den Menschen in den Teilen der Kooperation ersetzt, die er durchführt. Denn bevor es Roboter gab, wurden diese Anteile von Menschen ausgeführt.

Bei der *technischen* Ersetzbarkeit muss zunächst analysiert werden, welche Tätigkeiten Roboter übernehmen können. Ein Staubsaugroboter muss Staub saugen können. Aber er sollte auch Teppichkanten überwinden können, auf „Kontakt“ an die Möbel heranfahren, ohne diese zu beschädigen, eigenständig seine Ladestation aufsuchen und keine Stelle im Wohnzimmer ungesaugt lassen. Für den eigenen privaten Haushalt stellt man sofort die Kostenfrage. Diese *ökonomische* Ersetzbarkeit wird nach der eigenen Präferenz beurteilt, das heißt: Wenn man Staubsaugen als entspannende Haushaltstätigkeit einstuft, wird die Zahlungsbereitschaft geringer sein, als wenn man Staubsaugen hasst oder gar unter einer Hausstauballergie leidet. Diese eher betriebswirtschaftliche Perspektive muss für eine umfassende Analyse um eine volkswirtschaftliche Sicht ergänzt werden, weil mit einem großflächigen Einsatz von Putzrobotern auch entsprechende Arbeitsmarkteffekte verbunden sein können.

Die *rechtliche* Ersetzbarkeit wird aktuell besonders prominent beim autonomen Fahren diskutiert, wenn die Frage gestellt wird, wer bei einem Unfall verantwortlich ist. Nach wie vor der Fahrer des Fahrzeugs, der Produzent oder der Programmierer der Software? Während Rechtsexperten darauf hinweisen, dass man mit der etablierten Aufteilung zwischen Produzenten- und Halterhaftung heute schon verfügbare Robotersysteme sehr gut beurteilen kann, stellen adaptive Systeme, also Systeme, die sich durch maschinelles

Lernen selbst verändern und so an neue Gegebenheiten anpassen, möglicherweise eine Herausforderung aus rechtlicher Sicht dar. Diese Technik verändert sich in der Nutzung selbst, sodass der Roboterproduzent nach einiger Zeit des Einsatzes nicht mehr vorhersagen kann, was das System in einer bestimmten Situation tun wird. Damit wird er auch nicht mehr für mögliche Schäden haften wollen. Der Roboterhalter respektive -nutzer ist aber im Allgemeinen kein Robotikexperte, er wird den Lernprozess nicht beurteilen können, vielleicht sogar über Veränderungen des Roboters überrascht sein und daher ebenfalls nicht für mögliche Fehler haften wollen.

„SOZIALE ROBOTIK“

Die Frage nach Handlungskontexten, in denen wir auch als moderne Gesellschaft nicht wollen, dass menschliche Akteure durch Roboter ersetzt werden, bezieht sich auf die *ethische* Ersetzbarkeit. Gerade in Zusammenhängen, in denen die zwischenmenschliche Interaktion einen besonders hohen Stellenwert hat, wie beispielsweise in der Pflege, wird diese Frage relevant.

Bei Pflegehandlungen spielen neben den technischen Aspekten immer auch sogenannte Soft Skills wie Freundlichkeit, Zuvorkommenheit, Zugewandtheit eine Rolle, denen auch ein entsprechender Anteil am Wohlbefinden der zu Pflegenden und damit am Heilungserfolg zugeschrieben wird. Seitens der Robotik werden diese Aspekte der Interaktion unter dem Stichwort „Soziale Robotik“ diskutiert, die auch mit der humanoiden Robotik verknüpft ist, weil die Darstellung von Emotionen typischerweise über ein künstliches Gesicht umgesetzt wird. Einerseits gibt es bereits Forschungsergebnisse, die bestätigen, dass die Interaktion zwischen Roboter und Mensch besser funktioniert, wenn der Roboter auch diesen Kommunikationszweig nutzt. Andererseits wird darauf hingewiesen, dass in den Interaktionen zwischen Pflegendem und Pflegebedürftigen Letztere auch nachdrücklich aufgefordert werden müssen, beispielsweise die Medikamente einzunehmen. Hier muss geklärt werden, inwieweit auch derartige Aufgaben der Technik übertragen werden sollten.

Eigenständiges maschinelles Lernen kann als eine besondere kognitive Leistung angesehen werden. Es stellt gleichermaßen eine große Chance dar, denn zu lernen und sich selbst über das Gelernte an neue Handlungskontexte anzupassen, sind Schlüsselemente für den erfolgreichen Einsatz von Robotern in ihnen bisher unbekanntem Umgebungen (Bahnhöfe, Krankenhäuser, Bürogebäude, private Wohnungen). Gleichzeitig sind damit Risiken verbunden, wie am Beispiel des Go-Spielens erläutert werden soll.

AlphaGo verwendet eine Kombination von *Deep Learning* in Künstlichen Neuronen Netzen (KNN) und Monte-Carlo-Methoden („taktische Suche“). Dafür wurde es mit dreißig Millionen Spielzügen „trainiert“, die

zwischen Menschen gespielt wurden. Das System konnte am Ende des Trainings zu 57 Prozent den menschlichen Spielzug vorhersagen.

„WUNDERSCHÖNE KREATIVE ZÜGE“

Um auch eigenständige Spielzüge zu generieren, spielte AlphaGo Tausende Spiele gegen sich selbst, die nach Trial and Error bewertet wurden („reinforcement learning“). „Dabei ist offenbar eine neue Spielweise entstanden, die die Go-Elite der Welt nachhaltig verstört hat. AlphaGo machte Züge, die menschliche Profis zunächst für Fehler hielten. Tatsächlich hatte die Maschine aber einen Weg gefunden, das Spiel siegreich zu spielen, auf den Menschen in Tausenden Jahren Go-Geschichte nicht gekommen sind“, berichtet Christian Stöcker auf *SPIEGEL ONLINE* über das eingangs genannte Spiel zwischen AlphaGo und Lee Sedol. Die Profi-Spielerin Young Sun Yoon ist der Ansicht, dass AlphaGo Züge macht, die ein Mensch nie machen würde. Demis Hassabis, Mitbegründer der britischen Firma DeepMind (heute Google), die AlphaGo entwickelt hatte, twitterte, „der zweite Sieg sei für ihn selbst schwer zu fassen. AlphaGo hat in diesem Spiel einige wunderschöne kreative Züge gespielt.“ Auch andere Experten sprachen von einem Meilenstein in der Entwicklung des maschinellen Lernens.

Zusammenfassend¹ kann man festhalten, dass sowohl die Software-Entwickler als auch insbesondere die Go-Spieler-Community überraschend kreative Spielzüge identifiziert haben. Die „Überraschung“ der Community bezog sich nicht nur auf einzelne Züge, die nach den Regeln der heutigen Spielkunst als falsch beurteilt worden wären, sondern auch auf die Gesamtleistung, in der durchaus „menschliche Spielformen“ wahrgenommen wurden. Die KNN-Funktionsweise ist natürlichen Gehirnen nachempfunden. Es gibt künstliche Neuronen, die miteinander Signale austauschen, wobei das einzelne Neuron durch Gewichtung ein Eingangssignal zu einem modifizierten Ausgangssignal umwandelt. Die Veränderung der Gewichtungen in allen Neuronen im Training oder auch im Einsatz wird dann als Lernen des gesamten KNN beschrieben. Eine klassische Programmierung lässt sich im Wesentlichen über „Wenn-dann“-Beziehungen beschreiben, sodass ein Informatikexperte in einem konkreten Fall die „Wenn“-Situation nachvollziehen kann, die ein entsprechendes „Dann“ auslöst. Im Gegensatz dazu findet man in einem KNN nur die Neuronen und deren (möglicherweise veränderte) Gewichtungen. Hinweise zu einer konkreten Handlung, etwa warum ein vermeintlich falscher Spielzug gespielt wurde, lassen sich nicht ableiten. Das macht die Zuschreibung von Verantwortung für diese Handlung problematisch.

Wenn sowohl die Software-Entwickler als auch die Experten in dem jeweiligen Handlungskontext von einer Aktion des lernenden Systems „überrascht“ sind und die Experten die Aktion sogar für einen Fehler halten,

dann bringt das besondere Herausforderungen bei der Einführung der Systeme mit sich. Wir müssen die Frage stellen, wie viel „Überraschung“ wir tolerieren möchten. Das hängt offensichtlich vom Handlungskontext ab: Beim Spielen sind diese Überraschungen geradezu erwünscht. Auch bei einem Museumsroboter, der durch Ausstellungen führt, sind sie sicher tolerabel. Beim autonomen Fahren oder in Pflegezusammenhängen ist es dagegen schwer vorstellbar, dass der Roboter Aktionen durchführt, die etwa vom Pflegepersonal als Fehler eingestuft werden.

Gleichzeitig wäre es aber eine große vergebene Chance, wenn diese Systeme sich nicht durch Lernen an ihre Handlungsumgebungen anpassen könnten. Hier gilt es, geschützte Möglichkeiten des Lernens zu schaffen, in denen die Systeme lernen, ohne das Gelernte unmittelbar in Handlungen umzusetzen. Eine Prüfinstanz liest dann regelmäßig das Gelernte aus und implementiert nur die Veränderungen, die wünschenswerte Aktionen veranlassen. Sie stellt damit in verantwortlicher Weise sicher, dass das neu Gelernte tatsächlich funktioniert und – was ebenso beim Lernen in KNN vorkommen kann – dass das System nichts Wichtiges verlernt hat. Prozeduren dieser Art würden es uns erlauben, die Chancen des maschinellen Lernens zu nutzen und dabei die Risiken gering und die Verantwortung zurechenbar zu halten.

¹ Diese Argumentation ist ausführlicher nachzulesen in dem Beitrag von Michael Decker, „Adaptive Robotik und Verantwortung“, in: Gless, Sabine / Seelmann, Kurt (Hrsg.): *Intelligente Agenten und das Recht*. Nomos Verlag, Baden-Baden 2016, S. 23–44.