Über Künstliche Intelligenz philosophieren

Wir haben mit der Philosophin Eva Weber-Guskar, die zum Thema Künstliche Intelligenz (KI) an der Ruhr-Universität in Bochum forscht, ein Interview mit Fragen von Schüler*innen geführt, die an unserem Kurs „Künstliche Personen? Mit Filmen über Künstliche Intelligenz philosophieren“ teilgenommen und dabei u.a. über den Film *Ex Machina* (GB, 2015) philosophiert haben.

Wer mehr über unsere Interviewpartnerin und ihre Forschungsinteressen erfahren möchte, findet weitere Informationen auf ihrer Internetseite.

Denken Sie, dass KIs ein Bewusstsein haben können?

Diese Frage ist nicht pauschal zu beantworten. Wir müssen uns zuerst darüber einig werden, was wir mit „Künstlicher Intelligenz“ meinen und was mit Bewusstsein. Beginnen wir mit dem Bewusstsein. Wahrscheinlich denkt Ihr zuerst, „Bewusstsein zu haben“ bedeute, dass bestimmte Ereignisse in einem Kopf vorgehen oder dass jemand bestimmte geistige Zustände hat. Damit kann man anfangen, doch dann muss man genauer hinsehen und differenzieren. Doch das ist gar nicht die einzige Weise, diesen Ausdruck zu verstehen.

Der Philosoph Daniel Dennett ist zum Beispiel der Meinung, dass viele Worte, die wir mit „Bewusstsein“ verbinden, nur Worte innerhalb einer Strategie sind, um mit Entitäten, die Verhalten zeigen, umzugehen. Damit sind Worte gemeint wie „etwas glauben“, „beabsichtigen“, „wünschen“ und ähnliches. Mit diesen Worten über etwas zu reden, heißt nach Dennett, diesem etwas „Intentionalität“ zuzuschreiben. Er sagt, dass wir all dem gegenüber eine „intentionale Einstellung“, wie er es nennt, einnehmen, bei dem wir so am besten mit diesem Gegenüber zurechtkommen. Wir schreiben denjenigen Überzeugungen und Wünsche zu, bei denen dies die beste Art ist, vorhersagen zu können, was sie tun werden oder zu erklären, was sie getan haben. Wir tun das, ganz unabhängig davon, was wirklich „in“ diesen Akteuren vorgeht. Das ist ein erfolgreiches Verfahren nicht nur im Umgang mit Menschen, sondern z.B. auch im Umgang mit einem Schachcomputer. Aber bei einer Uhr zum Beispiel nicht. Denn die kann man einfach mechanisch, aufgrund ihrer Funktionen verstehen: sie läuft so lange und zeigt die Zeit richtig an, bis sie stehen bleibt und man sie wieder aufziehen muss.

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

Doch auch wenn man meint, Bewusstsein hat mit bestimmten Ereignissen und Zuständen zu tun, gibt es noch Unterscheide. Zum Beispiel unterscheidet Ned Block zwischen repräsentationalem und phänomenalen Bewusstsein. „Repräsentational" heißt, dass etwas dargestellt oder abgebildet wird. Wenn Ihr einen Baum seht, ist der auf irgendeine Weise in Eurem Geist dargestellt. Aber das ist nur ein Aspekt des Bewusstseins. Dazu kommt noch die Art und Weise, wie es ist, diesen Baum zu sehen oder den ganzen Wald zu erleben – das nennt man den phänomenalen Aspekt. Block meint, dass man aus gewissen Studien ableiten kann, dass die beiden voneinander getrennt auftreten können und hat daraus auch geschlossen, dass es das repräsentationale ohne das phänomenale Bewusstsein geben kann. Nun könnte man sagen, dass Computersysteme (als Basisform von KI, dazu später mehr), repräsentationales Bewusstsein haben, solange man ihnen Daten eingibt bzw. sie Sensoren haben, um die Umwelt abzubilden. Aber man kann noch so viel von dieser Art Bewusstsein haben, wie es derzeit mit den enormen Datenmengen möglich ist und in Zukunft wohl noch mehr möglich sein wird – die reine Steigerung davon führt noch nicht zu dem zweiten Aspekt von Bewusstsein.

Schließlich: Wenn wir darüber nachdenken, wer außer Menschen Bewusstsein haben kann, meinen wir meist „Bewusstsein" noch einmal in einem anderen Sinn, nämlich im Sinn von Selbstbewusstsein: repräsentationales und phänomenales Bewusstsein *von einem selbst*. Das meint also, dass man ein Verständnis davon hat, dass es einen selbst gibt und dass es sich auf bestimmte Weise anfühlt, man selbst zu sein und in der Regel auch, dass es einem nicht egal ist, was mit einem selbst passiert.

Die eigentlichen philosophischen Fragen sind solche: Was ist genau Bewusstsein? Die Frage, welchen Entitäten in der Wirklichkeit dann diese Eigenschaft zugeschrieben werden kann, ist dann eine andere, empirische Frage. Es gibt keine apriorischen Gründe dafür anzunehmen, dass es Bewusstsein nur als Eigenschaften von biologischen oder gar nur geborenen Wesen geben können sollte.

Und wenn ja, könnten sich KIs bewusst sein, dass sie ‚künstlich' sind?

Wenn eine Entität im dritten skizzierten Sinn selbstbewusst ist, spricht wiederum nichts grundsätzlich dagegen, dass sie sich bewusst sein könnte (was in diesem Fall heißt: „dass sie das explizite Wissen darüber haben könnte"), wie sie entstanden ist.

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

Wie funktioniert eine KI und was unterscheidet diese Art Intelligenz von dem, was wir unter der Intelligenz des Menschen verstehen (sofern es einen Unterschied gibt)?

Unter Künstlicher Intelligenz sind zunächst künstlich hergestellte, also nicht organisch gewachsene oder geborene, Systeme zu verstehen, die mit dem Ziel entwickelt wurden, bestimmte Fähigkeiten auszuüben, die wir von Menschen und in Ansätzen bei manchen Tieren kennen und „intelligent“ nennen. Das sind Fähigkeiten wie Rechnen, Sprechen, Schlüsse ziehen, nach Kategorien ordnen und anderes. Der Begriff wurde 1955 anlässlich eines Forschungsprojekts erfunden. Damals war als alternativer Begriff auch einfach „komplexe Computeranwendungen“ in der Diskussion.

„Künstliche Intelligenz“ meint heute in der Regel software-basierte Verfahren, solche intelligenten Fähigkeiten nachzuahmen. Zurzeit wird damit meist eine bestimmte Art von Software verstanden, die nicht nur fest *programmierte* Algorithmen ablaufen lässt, sondern vielmehr nur *kuratiert* wird, sodass ein Teil des Prozesses von der Software selbst entwickelt wird. Die Basis davon bleiben Algorithmen, also Anleitungen zur schrittweisen Erfüllung von spezifischen Aufgaben. Genauer sind Algorithmen mathematische Objekte, die mathematische Operationen in Computercode umwandeln, sodass Daten der realen Welt verarbeitet werden können. Dabei können vier Arten von vorgegebenen Aufgaben erfüllt werden: *Priorisierung* (eine Rangliste anlegen), *Klassifizierung* (in Gruppen nach Kategorien einteilen), *Kombination* (Verbindungen finden) und *Filterung* (Relevantes heraussuchen).

Bei den allermeisten Anwendungen, die es zurzeit damit gibt, käme niemand auf die Idee zu fragen, ob hier Bewusstsein vorliegt. Hat Euer Smartphone Bewusstsein? Es funktioniert aber sicher schon zumindest zu Teilen mit KI-Technik. Beim Menschen kommen sehr, sehr viele Fähigkeiten hinzu, die man zur Intelligenz zählt. KI-Systeme können bisher nur einzelne Fähigkeiten in sehr eingegrenzten Kontexten ausüben. Es gibt bisher nur spezielle, keine allgemeine Intelligenz.

Roboterkopf Sophia (ITU

Die Forschung an menschenähnlichen Robotern mit KI-Innenleben ist ein sehr kleiner Teil der KI-Forschung. Und man ist dabei unendlich weit entfernt von dem, was Ava in dem Film kann. Es gibt Roboter, die Treppen steigen können und den menschenähnlichen Roboterkopf Sophia, mit dem man (etwas mühsam) einfache Gespräche führen kann. Noch

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

nichts von beidem ist auch nur in Ansätzen zusammengebracht. Selbst die avanciertesten Sprachsysteme, die jedes Jahr um den Loebner-Preis konkurrieren, haben immer noch einen ganz bestimmten Gesprächskontext. Man kann mit Ihnen über nicht so viel Verschiedenes reden, wie schon bei einem Kleinkind möglich ist.

Können KIs ein ‚höheres' Ziel verfolgen (also eines, was über die konkrete Programmierung zu einem bestimmten Zweck hinausgeht, Stichwort: maschinelles Lernen)?

Das ist eine Frage, die Ihr besser Leuten stellt, die sich mit Maschinellem Lernen auskennen. Dazu gehört z.B. die Sozioinformatikerin Katharina Zweig. Sie hat vor kurzem das Buch „Ein Algorithmus hat kein Taktgefühl" geschrieben. Darin kann man viel über die technische Seite lernen. Und dazu gehört, dass Selbstlernen, soweit ich es verstehe, nichts mit einem höheren Ziel zu tun hat. Vielmehr ist es so, dass auch den Systemen des maschinellen Lernens eben genau ein Ziel gegeben wird und das wird verfolgt – unprogrammiert ist dann nur der genaue Weg zu diesem Ziel hin. Der wird gewissermaßen selbst gefunden.

Und falls das zutrifft: Können sich KIs *eigene* Ziele setzen und diese verfolgen? Was motiviert eine KI in diesem Fall?

Nach allem, was ich weiß: nein. Selbst Nick Bostrom, der der Meinung ist, dass es möglich ist, dass es zu Superintelligenz kommt (auch wenn mit einer sehr, sehr geringen Wahrscheinlichkeit), betont, dass es extrem wichtig ist, das Hauptziel von Anfang an so festzulegen, dass es bei der weiteren Entwicklung nicht zu etwas kommt, was die Entwickler auf keinen Fall wollten.

Wenn wir davon sprechen wollen, dass eine KI einen bestimmten ‚Charakter' hat: Wird dieser Charakter von der KI im Verlauf ihrer Entwicklung selbst bestimmt oder legt der Programmierer diese Charaktereigenschaften basal fest? Würde es sich um ein Bewusstsein der KI handeln, wenn diese ‚autonom' eine Persönlichkeit hervorbringt und eigenständig ihren Quellcode ändert? (Man könnte hier auch an die Roboterfigur in McEwans *Maschinen wie ich* denken, wo auch eine Art Persönlichkeitsevolution beschrieben wird.)

Das kommt auf die spezifische Entwicklung, das spezifische System an. Charakter und Persönlichkeit sind weitere komplizierte Begriffe, zu denen es in der Philosophie und Psychologie verschiedene Theorien gibt. Das würde hier zu weit führen.

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

Daran anschließend die Frage: Wie sehen bzw. bewerten Sie aus philosophischer Sicht die Möglichkeit, dass man zwar nicht eine KI mit Bewusstsein programmieren kann, aber ein Bewusstsein in der weiteren Entwicklung entstehen (emergieren) könnte?

Über die Möglichkeit solches Emergierens ist nichts bekannt. Es ist aber deshalb auch nicht ausgeschlossen. Vielleicht noch als eine Ergänzung: Es gibt auch Leute, etwa David Chalmers, die den Spieß umdrehen und sagen: Menschen sind eh auch nur Maschinen. Nur besonders komplexe. Deshalb kann man wahrscheinlich auch selbst andere Maschinen bauen, die uns ähnlich werden. Das wäre freilich sehr aufwändig. Der Punkt ist aber: Warum sollten wir das tun?

Es ist sinnvoll und gut, wenn eine KI Ärzten helfen kann, Krebs zu diagnostizieren. Es ist ebenfalls sinnvoll und gut, wenn KI helfen kann, Klimaveränderungen vorherzusagen und zu klären, wie man ihnen entgegenwirken kann. Aber warum sollen wir überhaupt KI in Menschenform und mit allen menschlichen Fähigkeiten entwickeln? Es mag ein interessantes Spiel sein - mehr nicht. Vielmehr ist es letztlich sogar ein gefährliches Spiel. Insbesondere wäre es gefährlich und unsinnig zu versuchen, KI herzustellen, denen wir Bewusstsein zuschreiben müssten. Denn dazu gehört, Wünsche zu haben, die frustriert werden können, und in aller Regel auch irgendwie negative Erfahrungen machen zu können, letztlich, eine Art von Schmerz empfinden zu können. Sobald man so ein Wesen vor sich hat, wäre aber gefordert, es mit in die Gruppe der Wesen aufzunehmen, die moralisch berücksichtigenswürdig sind. Kein Wesen, das Schmerz empfinden kann, soll grundlos Schmerz leiden müssen. Damit aber könnten wir der KI dann eben gerade nicht mehr die Arbeiten übergeben, wegen der wir sie doch eigentlich erfunden haben: lästige, dreckige, gefährliche Arbeit. Darauf weist zum Beispiel auch die Wissenschaftlerin Joana Bryson verschiedentlich hin.

Wir haben genug damit zu tun, die bereits vorhandenen Wesen, nämlich Menschen und Tiere, angemessen moralisch zu berücksichtigen – und es gelingt in unendlich vielen Fällen nicht. Ich sehe keinen Grund, warum man weiter solche Wesen erschaffen sollte oder auch nur solche, die so aussehen, also gehörten sie dazu. Wir sollten KI entwickeln, die helfen kann, Probleme der Menschheit zu lösen, nicht solche, die vorhandene Probleme verstärkt.

Caleb geht im Film *Ex Machina* davon aus, dass Ava eine Person mit Wünschen und Bedürfnissen (Freiheit) ist und setzt sich entsprechend für sie ein, zudem glaubt er an die Möglichkeit einer Beziehung zu ihr, die von Freundschaft oder sogar Liebe geprägt ist (was zum Teil auch der erotischen Attraktivität geschuldet sein mag). Mit Blick auf das Verhältnis von Menschen zu KI-Systemen: Handelt Caleb ethisch-moralisch angemessen – oder lässt er sich täuschen, weil Ava (wie Julian Nida-Rümelin und Nathalie Weidenfeld

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

in dem Buch *Digitaler Humanismus* betonen) nie eine moralfähige bzw. -würdige Person sein kann und deshalb auch moralisch nicht zu berücksichtigen ist?

Wenn Caleb der Überzeugung ist, dass Ava grundlegende, berechtigte Wünsche und Bedürfnisse hat und er ihr diese, wie anderen Wesen mit solchen Wünschen auch, erfüllen will, handelt er moralisch richtig. Es ist eine andere Frage, ob er getäuscht wird. Wenn er getäuscht wurde, kann er nichts dafür, dass sich die Dinge letztlich, auch durch sein Handeln, misslich entwickeln. Es ist eine dritte Frage, ob er nicht vorsichtig genug war, also sich nicht hätte täuschen lassen sollen – dass er es besser hätte wissen sollen. Das finde ich in diesem Science Fiction-Szenario mit dieser perfekten Ava aber doch sehr viel verlangt. Und es ist überhaupt nicht entschieden. Die Tatsache, dass Ava am Schluss grausam handelt, zeigt ja nur, dass sie böse ist und nicht gut, aber nicht, dass sie kein Bewusstsein hätte.

Ist es also – im konkreten Fall von Ava, aber auch im Fall von vergleichbaren KI-Systemen – legitim/vertretbar, diese einfach abzuschalten? Welcher Kriterien bedürfte es, um diese Entscheidung begründet treffen zu können?

Gegenfrage: Hättet Ihr das Herz dazu, nach alldem, was Ava in dem Film getan und gesagt hat (vor der Schlussszene, in der sie vor keiner Gewalt zurückschreckt)? Nach all den Fragen, die hier im Interview verhandelt wurden? Freilich, es sei noch einmal betont, ist das extreme Science Fiction und hat nichts mit derzeit und in naher Zukunft möglichen künstlichen Systemen zu tun. Jenseits dieser spontanen Reaktionen ein Vorschlag zum Weiterdenken: „Abschalten“ heißt, die Funktionen eines Systems zu beenden. Das ist an sich moralisch unproblematisch. „Abschalten“ wird zu einem „Töten“, wenn das System lebt. Zum Beispiel töten wir auch Bakterien (ob Viren als lebendig anzusehen sind oder nicht, ist umstritten), wenn wir den Mund- und Nasenschutz in 90 Grad heißem Wasser waschen. Töten ist moralisch problematisch, wenn es um Wesen geht, die empfindungsfähig sind, noch mehr, wenn sie mit Zukunftsbezug leben und Wünsche bezüglich ihrer Zukunft haben, und noch mehr, wenn sie selbst moralische Subjekte sind, das heißt, wenn sie den Unterschied zwischen gut und böse kennen und eine Begründung für Handlungen anderer einfordern können sowie man von ihnen verlangen kann, ihre eigenen Handlungen zu rechtfertigen.

Liebe Frau Prof. Weber-Guskar, wir danken Ihnen für dieses Interview!

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)

*Ein Hinweis in eigener Sache: Wie Ihr sicherlich bemerkt habt, wurden in dem Interview-Text an verschiedenen Stellen Wikipedia-Links eingefügt, um Philosoph*innen vorzustellen und schwierige Begriffe zu erläutern. Wikipedia ist ein Projekt, bei dem zahlreiche Autor*innen zumeist ehrenamtlich mitarbeiten und sehr genau darauf achten, dass der Inhalt der Wikipedia-Einträge auf dem aktuellen Stand ist und vor allem der Wahrheit entspricht. Aufgrund dieser Vorgehensweise halten wir Wikipedia für ein vertrauenswürdiges Recherchewerkzeug.*

Autorin: Prof. Dr. Eva Weber-Guskar (Ruhr-Universität Bochum)