

FEUILLETON

Würdig

Der Schauspieler Jens Harzer bekommt den Iffland-Ring – eine weise Entscheidung > Seite 15



PHOTO: IMAGO

Der Mensch: ein Fehler

Welches Weltbild steckt in künstlicher Intelligenz? Die Antwort auf diese Frage wird das Verhältnis von Mensch und Maschine bestimmen

Von Sarah Spiekermann

Künstliche Intelligenz (KI) ist ein epochaler Technologiesprung, der die Menschheit vor ganz neue Fragen stellt, die keine Disziplin alleine beantworten kann. John Brockman, Agent für Wissenschaftsliteratur und Gründer des Debattenforums edge.org, hat das „Possible Minds“-Projekt ins Leben gerufen, das Natur- und Geisteswissenschaften zusammenführt, um KI und deren wahrscheinliche Ausformungen und Folgen zu ergründen. Das Feuilleton der SZ druckt Texte aus dem Projekt sowie europäische Reaktionen wie diese hier als Serie.

DER GEIST IN DER MASCHINE

Was bedeutet künstliche Intelligenz? Eine Serie von Essays sucht Antworten. Teil 3

Die Autorin ist Expertin für KI-Ethik und leitet das Institute for Management Information Systems der Wirtschaftsuniversität Wien. Im April erscheint ihr Buch „Digitale Ethik – Ein Wertesystem für das 21. Jahrhundert“ bei Droemer.

Künstliche Intelligenzen haben ein Gegenüber: die Menschen. Mit ihnen soll die KI kooperieren. Sie soll sie unterstützen. Dafür braucht die KI eine Vorstellung von den Menschen. Wenn man jedoch in die Informatikwelt hineinhört und fragt, wie dieses Menschenbild aussieht, dann steht ein Fragezeichen im Raum. Ein grundlegendes Bild vom Menschen ist nicht bekannt. Ist es vielleicht der „Dau“? Das ist die noch gängigste Bezeichnung und steht für „dümmsten anzunehmenden User“, für den eine Anwendung möglichst einfach konstruiert werden muss. Oder ist es eines der ebenso unerfreulichen Synonyme, die auf Wikipedia zu finden sind, wie etwa „FSVG“, „Fehler sitzt vor Gerät“?

Charlie Chaplins Film „Modern Times“ fasst das Menschenbild unserer ersten Maschinenparks zusammen, die damals nur in Firmen zu finden waren. Bei Chaplin war der Mensch Diener der Maschine, ein Vereinfachter und Ausgelieferter. Sind diese Bilder veraltet in der heutigen Welt, wo sich uns Technik doch als super fortschrittliche, bonbonsüße Community-Oase voller Freiheiten darstellt?

Dass die Frage nach dem Menschenbild aktueller ist denn je, hat technische, ökonomische und gesellschaftliche Gründe. Die Digitalisierung umfängt uns immer stärker. Bisher waren es die Bildschirmoberflächen unserer Computer und Smartphones, die man ausschalten kann, zumindest zu Hause. Sollten Betreiber und Entwickler uns also als „dümmste anzunehmende User“ ansehen, dann sei es so. Aber mit der „digitalen Transformation“ und dem Eintritt künstlicher Intelligenzen in unser tägliches Leben, mit durchdigitalisierten Autos, persönlichen Kredit-Scores, Alexas im Wohnzimmer, Bots im Call Center und dergleichen kommt uns Menschen dieses Digitale zunehmend so nahe, dass es nicht mehr egal ist, ob wir Nutzer als die Dümmeren angesehen werden.

Im Zentrum steht – technisch gesehen – die Frage der „Funktionsallokation“: Wie viel Kontrolle bekommt der Mensch zugewiesen und wie viel die Maschine? Geisteswissenschaftler würden von Machtverteilung sprechen. Gibt es einen „Aus“-Knopf für den Roboterpolizisten, der vielleicht schon bald an unserer Haustür klingeln wird?

könnte? Oder gibt es diesen „Ausknopf“ nicht mehr, ganz im Sinne eines mehr oder weniger gut gemeinten Technologiepaternalismus? Nach welchen Regeln bestellt Alexa Vorräte für den Kühlschrank? Vollautomatisch? Mit dem Menschen als letzter Entscheidungsinstanz? Oder hält sie sich ganz zurück, uns überhaupt zu beeinflussen? Solche granular abgestuften Fragen der Kontrollallokation zwischen Mensch und Maschine sind zentral für die langfristige Freiheit und Würde von uns Menschen. Und wie sie beantwortet werden, hängt davon ab, was für ein Menschenbild die Entwickler haben, beziehungsweise die Institutionen, die diese beauftragen.

Hat man ein vertrauensvolles Menschenbild, so lässt man Homo sapiens auch die Kontrolle. Die Bedeutung der Kontrollallokation wird dabei leicht unterschätzt. Ökonomisch könnte man argumentieren, dass die unsichtbare Hand des Marktes es richten wird. Wer will schon für paternalistische Maschinen im Wohnzimmer bezahlen? Was ist aber, wenn ein krankhaft übergewichtiger Mensch einen staatlich geförderten Vertrag mit seiner Versicherung abgeschlossen hat, Amazons Sprachassistenten Alexa so zu nutzen, dass diese ihm nur Gemüse nach Hause liefert, wofür es Premiennachzahlung gibt? Was wenn der Vertrag mit der Autoversicherung beinhaltet, dass weniger Prämie anfällt, wenn der Fahrer sich nicht übernimmt? Solche Deals, die die Freiheit von Menschen gegen Geldvorteile eintauschen und uns in ein eigenartiges Maschinenkorsett einspannen, sind nicht so abwegig, sondern gängig werdende Praxis, wie die Harvard-Professorin Shoshana Zuboff in ihrem Buch „Das Zeitalter des Überwachungskapitalismus“ darlegt.

Denn in der Digitalökonomie haben sich Geschäftsmodelle herausgebildet, denen kein gutes Menschenbild zugrunde liegt. Hier wird der Mensch zwar nicht mehr ausschließlich auf den „Homo oeconomicus“ reduziert, also einen egoistischen Präferenzoptimierer. Dafür ist man aber zu einem ebenso fragwürdigen Menschenbild übergegangen, das uns als vorhersehbar irrational darstellt.

Insbesondere dieses letztere Bild des Menschen aus der Verhaltensökonomie paart sich nun fruchtbar mit der Digitalisierung. Was wäre besser, als die digitalen Dienste, Sprachassistenten, KIs, Autos, et cetera so zu bauen, dass sie uns vorhersehbar irrationale Selbstoptimierer in ein „richtiges“ Verhalten hineinlotsen? Die Verhaltensökonom Cass Sunstein und Richard Thaler sprechen von „Stupsen“ (Englisch: Nudge). Also einer digitalen Transformation des Alltags und des Wirtschaftens, in denen Maschinen Menschen in ihr fortschrittliches Dasein hineinstupsen. Das bringt uns zurück zu Charlie Chaplin und dem Bild des ausgelieferten, fremdbestimmten Menschen. In der Schule wurde diese Form von „Modern Times“ als überkommener Auswuchs früher Industrialisierungsphasen dargestellt. Sind wir nun mit der Digitalisierung schnurstracks auf dem Weg in diese veraltete geglaubte Vergangenheit?

Goethe hat geschrieben: „Wenn wir Menschen behandeln, wie sie (scheinbar) sind, so machen wir sie schlechter; wenn wir sie so behandeln, wie sie sein sollten (wenn wir das Gute in ihnen bejahen), so machen wir sie zu dem, was sie werden können.“ Betrachtet man aktuelle Forschungsergebnisse zur menschlichen Intelligenz, wie sie etwa vom Max-Planck-Direktor Gerd Gigerenzer als Antwort auf die klassische Verhaltensökonomie zusammengefasst werden, dann wird klar, dass Menschen trotz vieler kleiner Fehler insgesamt eine hervorragende ausgeprägte „ökologische“ Intelligenz haben.

Dem erstaunten Philosophen oder Technikhistoriker bleibt bei der naiven Vergleichslogik nur ein Blick in die Vergangenheit: Bereits im 17. und 18. Jahrhundert lieferte die damals vorherrschende Technik das Modell zur Beschreibung des Menschen. Damals wurden die ersten Uhren erfunden und pneumatische Orgeln gebaut, und so beschrieb man den Menschen nach dem Modell einer Uhr oder einer Orgel. Leider haben Bilder aber die Macht, die Real-

genz und Kooperationsfähigkeit haben. Wäre es nicht weise, bei der Digitalisierung diese Erkenntnisse zu berücksichtigen? Unsere digitale Infrastruktur sollte doch auf einem vertrauensvollen Menschenbild aufgebaut sein. Sie sollte uns Mut machen, mehr Kontrolle und soziale Verantwortung zu übernehmen. Bei so einem positiven Menschenbild würden freilich weniger bevormundende Systeme gebaut werden und gestupst würde nur noch auf Wunsch eines als mündig respektierten Nutzers.

Leider stößt ein positives Menschenbild bei Digitalisierungsfans und Zynikern auf Granit. Der Mensch, so wird oft insistiert, ist nun mal ein egoistischer, unberechenbarer, geiziger Präferenzoptimierer. Er sei vor allem ein suboptimales „System“, das im Vergleich zu Computerschaltkreisen unauweilich dümmer sein muss als die ach so „intelligenten“ Maschinen.

Eine Beobachtung der Sprache über uns Menschen unterstreicht diese Haltung. Da spricht der ehemalige Chefentwickler der Google-Autos, Chris Urmson, in einem Vortrag bei der Ted Conference vom Menschen als letzten „Bug“ im System (das englische Wort für Käfer bedeutet im Jargon der Informatik „Fehler“). Transhumanistische „Eliten“ beschreiben den Menschen als grundsätzlich eher unglücklich. Von einigen werden Menschen sogar als „Wetware“ bezeichnet, also als „Nassware“, weil unsere Körper aus Wasser bestehen, im Gegensatz zur „Software“ der Computer.

Vielen dieser „Experten“ ist gemein, dass sie sich Menschen irrtümlich als „Informationsobjekte“ vorstellen, also als Wesen, die Computern und ihrer Datenlogik erstaunlich ähnlich sehen. Dass Menschen weder Objekte sind noch stringent miteinander interagierende Dateneinheiten und auch keine rein kognitiv arbeitenden Festplattengehirne haben, sondern integrierte Körper-Geist-Systeme sind, das wird ausgeklammert.

Dem erstaunten Philosophen oder Technikhistoriker bleibt bei der naiven Vergleichslogik nur ein Blick in die Vergangenheit: Bereits im 17. und 18. Jahrhundert lieferte die damals vorherrschende Technik das Modell zur Beschreibung des Menschen. Damals wurden die ersten Uhren erfunden und pneumatische Orgeln gebaut, und so beschrieb man den Menschen nach dem Modell einer Uhr oder einer Orgel. Leider haben Bilder aber die Macht, die Real-

tät zu bestimmen. Wenn der Mensch als suboptimales Gehirn im Tank, als schwach und dumm im Vergleich zu KIs dargestellt wird, ist es nicht verwunderlich, dass Technik dazu dienen soll, ihn oder sie entsprechend zu verbessern oder gleich zu ersetzen. Besonders hartnäckig hält sich die Vorstellung, Menschen hätten eine eingeschränkte Entscheidungsfähigkeit. So findet sich bereits auf der ersten Seite der „Ethikempfehlungen für Vertrauen stiftende Künstliche Intelligenz“ der EU Kommission (beziehungsweise ihrer „High Level Expert Group“) der Satz: „Es ist bekannt, dass Menschen in ihren Entscheidungen voreingenommen sind.“

Stimmt das? Oder treffen wir Entscheidungen vor dem Hintergrund komplexer kulturell gewachsener Normen und Erfahrungen, die in einem ständigen Fluss sind? Was ist mit den weisen Menschen unter uns? Oder gibt es die nicht mehr?

Auf einer Tagung der Eidgenössischen Technischen Hochschule Zürich betitelte kürzlich ein dort berufener Professor seine Vortragsfolien damit, dass es nun 1:0 gegen den Menschen stünde, dann 2:0, 3:0 und so fort. Die Folien enthielten dann Beispiele, wie sogenannte narrow AIs, also „enge“ künstliche Intelligenzen für Spezialaufgaben, die in geschlossenen Kontexten wie Brettspielen oder Bilderkennungssystemen (Schach, Go, Krebsdiagnose), Menschen überlegen sind. Dass solche geschlossenen Kontexte in einer offenen Alltagswelt eher selten vorkommen, wird ausgeblendet, weil dann der Mensch vielleicht gewinnen würde?

Als Vorbild für die KI-Entwicklung wurde in demselben Vortrag Spike Jonzes Film „Her“ beschrieben. Dort treibt die Sprachassistentin Samantha mit ihrer attraktiven Stimme den erbärmlich traurigen Menschen Theodor in eine emotionale Abhängigkeit, die sich wirklich niemand als Nebenwirkung im Umgang mit dem eigenen Sprachassistenten wünscht. Leider sind solche Bilder in ihrer Wirkung nicht zu unterschätzen. Wenn Wissenschaftler, Entwickler und Firmen sich dunkle Science-Fiction-Märchen zum Vorbild nehmen, in denen das skizzierte Menschenbild eines der Schwäche, der Abhängigkeit und der Entscheidungsunfähigkeit ist, dann werden sie Maschinen schaffen, die genau diese traurige Seite unserer Natur bestätigen und befördern.

Doch wie entkommen wir diesem traurigen Menschenbild? Wir müssen entlarven,

wie es in unserer humanistisch geprägten Welt zu solchen Negativbildern über den Menschen kommen konnte und warum sie sich so hartnäckig halten. Hier lohnt sich ein Blick in die Geschichte.

Möglicherweise hat die oft falsch verstandene Rhetorik in modernen christlichen Kirchen dazu beigetragen, die uns Menschen seit Jahrhunderten als „sündig“ bezeichnet. Obwohl hier vor allem die Demut gemeint ist, so war das Bild des Sünders doch machtpolitisch opportun und paarte sich schon im 17. Jahrhundert mit dem politisch umarmten Bild des einsamen „Wolfs“, das der Philosoph Thomas Hobbes prägte und das bis heute den bereits erwähnten Homo oeconomicus unterfüttert. Der Mensch erscheint in diesen Darstellungen nicht mehr so wie in der aristotelischen Tradition, also als ein politisches oder soziales Tier, das fähig ist, Tugenden zu kultivieren, sondern als autonomer Egoist.

Selbst der zentrale Philosoph der Aufklärung, John Locke, blieb bei diesem negati-

ven Menschenbild, wenn er die wichtigste Funktion des staatlichen Gemeinwesens darin sah, uns gegen Übergriffe auf unser Privateigentum zu schützen.

Der Philosoph Jean-Jacques Rousseau schließlich sah den Menschen zwar als von Natur aus gut an, war allerdings davon überzeugt, dass unser soziales Zusammenleben und zumal unsere moderne Lebensform in städtischen Umfeldern unseren Charakter pervertieren, was wiederum den Staat zwingt, korrigierend einzugreifen.

Kurz: Eine ganze Reihe von Vorvätern unseres heutigen Denkens haben schon ein fragwürdiges Menschenbild entwickelt. Vor allem aber haben Institutionen der Macht solche Theorien dankend aufgegriffen, ob kirchlich oder staatlich, ob ökonomisch oder technisch. Seit Jahrhunderten haben sich diese auf das negative Menschenbild gestützt. Nur wenn der Mensch schlecht ist, gibt es nämlich eine Legitimation ihn einzunorden, zu verwalten und ihn vor sich selbst zu schützen.



COLLAGE: STEFAN DIMITROV

Die Männer, die unser Denken prägten, sahen das Böse im Menschen

KUNSTHALLE MÜNCHEN

SAMURAI
PRACHT DES JAPANISCHEN RITTERTUMS
Die Sammlung Ann und Gabriel Barbier-Mueller
1.2. – 30.6.2019

TÄGLICH 10-20 UHR
17.4., 15.5., 19.6.: 10-22 UHR

Die Kunsthalles München ist ein Engagement der HypoVereinsbank