

WIE INTELLIGENT IST DIE KÜNSTLICHE INTELLIGENZ?

Dritter Teil: Ausblick

Das eigentliche Problem

Was Künstliche Intelligenz (KI) sein könnte, was nicht und wo sie beginnt, ist beschreib- und diskutierbar. Der wirtschaftliche und gesellschaftliche Nutzen einzelner Anwendungen ebenfalls. Die einsetzende Entwicklung »echter« oder »starker« KI fordert aber vor allem die menschliche Selbstwahrnehmung heraus. Fragen nach dem Beginn und Ende von Bewusstsein, Leben und menschlicher Autonomie drängen sich auf. Die Dringlichkeit dieser Fragen nimmt nochmals dramatisch zu, wenn zukünftige KI-Entwicklungen möglicherweise nicht mehr vom Menschen abhängig sein werden, weil einmal angestoßene Prozesse eigenständig weiterlaufen.

Angesichts der Vielzahl ungeklärter Grundfragen bezüglich der Funktion von KI und ihrer Auswirkungen auf den Menschen ist es verantwortungslos, die weitere Entwicklung auf gut Glück laufen zu lassen. Es gibt keine historische Parallele für einen so tiefen Eingriff in die Verfasstheit und das Selbst des Menschen. Alle Analogien zu den Neuerungen der industriellen Revolutionen verfehlen die Lage, denn nicht »nur« die Lebenswelt wird nun radikal verändert. Vielmehr wird die Prägung der Welt durch den Menschen selbst infrage gestellt. Das ist nur vergleichbar mit der Unterwerfung der Natur durch den Menschen im Neolithikum – aber die Veränderung erfolgt jetzt mit viel höherer Geschwindigkeit. Deswegen sind KI-Entwicklung und Fragen des Transhumanismus das eigentliche Zukunftsthema, und nicht die eher werkeuhafte und vergleichsweise harmlose Digitalisierung.

Werden durch KI-Anwendungen wesentliche Kulturtechniken auf Maschinen übertragen, mag das noch als gewöhnungsfähige Alltagsentlastung gelten. Wo aber die Funktionsmechanismen der Menschenwelt nicht mehr von Menschen bestimmt, sondern großmaßstäbliche Entscheidungen im Selbstvollzug der

Maschinenintelligenz getroffen werden, da »gelten anthropologische Grundkonstanten nicht mehr«. ¹ Dem ist, wenn überhaupt, nur durch Besinnung auf die Grundlagen menschlichen Daseins zu begegnen.

Offene Fragen der Philosophie

Die großen Fragen des Menschseins und von dessen Verhältnis zur Umwelt werden sich nicht beantworten lassen. Das gehört zum Wesen der Philosophie. Wir wollen aber zeigen, dass die Fortentwicklung von KI und deren Einbindung in alle Lebensbezüge auch drängende konkrete Fragen stellt.

Fehlende Analysewerkzeuge – Die zur Analyse berufenen Gelehrten sollten aus der Erfahrung früherer technologieinduzierter Umstürze gelernt haben, dass man offenen ethischen und politischen Fragen nicht durch vorgeblichen »Pragmatismus« oder Berufung auf die »normative Kraft des Faktischen« aus dem Weg gehen kann. Der düstere, weil Ratlosigkeit verdeckende Zweckoptimismus unserer Tage und die überall waltende irrationale Zuversicht, dass es diesmal schon klappen werde mit der Weltrettung, lässt Böses ahnen.

Zukunftsoffenheit – Harald Seubert weist auf die verantwortungsethische Frage der »vergangenen Zukunft« hin, »des Erwartungshorizontes, den eine Gegenwart von der bevorstehenden Zukunft hat und haben kann. Erst die Nachwelt wird wissen können, welche Zukunftserwartungen sich bewährt haben [...]. Zu diesen Perspektiven kann der Zeitgenosse selbst nur im probabilistischen Sinn Stellung nehmen.« ² Seubert möchte dieses Problem durch diagnostisches Denken und die Suche nach einer Mitte zwischen dem best- und dem schlimmstmöglichen Verlauf lösen. Es

¹ Harald Seubert: Digitalisierung. Die Revolution von Seele und Polis. Baden-Baden 2019, S. 151.

² Ebenda, S. 30, vgl. bereits S. 19.

scheint aber noch wichtiger, dass ein Zahlen-Denken, das aus inhärenter Logik heraus nach einem mathematisch definierten Optimum sucht, die zur Verfügung stehenden Problemlösungswege so sehr verengt, dass differierende »Zukunftserwartungen« gar nicht mehr vorstellbar sind bzw. von der KI nicht berücksichtigt werden (können). So erhebt sich die Frage, ob mechanisches Analysieren überhaupt zukunfts offen sein kann – ob es nicht vielmehr in einem einmal festgeschriebenen Lösungsraum gleichsam eingesperrt bleibt. Dies wird zum Problem, wenn die möglichen Lösungen nicht dicht³ im tatsächlichen Lösungsraum liegen, man also nicht für jedes reale Problem eine hinreichend nahe Approximation durch das KI-System erreichen kann.

Kausalität – Die Entscheidungen von Algorithmen sind nicht transparent, oft nicht einmal für die Entwickler. Häufig sind gut funktionierende Algorithmen komplex und die Menge der pro Zeiteinheit zu verarbeiteten Daten zu hoch, als dass die Prozesse von Menschen nachvollzogen werden könnten. Wo Entscheidungen in einer Geschwindigkeit unterhalb der menschlichen Wahrnehmungsschwelle getroffen werden und die Entscheidungssysteme sogar bei Experten als »Black Boxes« zu gelten beginnen, wird Kausalität fremd. Unklarheit über Ursache-Wirkung-Zusammenhänge aber verhindert strukturiertes Denken. Bevor Menschenhirne sich über Folgen moderner Technologien Gedanken machen können, sehen sie sich vor das Problem gestellt zu erfassen, was eigentlich geschieht. Nach Ansicht von Judea Pearl etwa werden Ursache-Wirkung-Mechanismen beim maschinellen Lernen sträflich vernachlässigt.⁴ Immerhin aber versucht man, diese Schwäche zu beheben.

Offene Fragen der Technik

Fehleranfälligkeit – Im Mai 2016 stieß ein autonom fahrendes Auto in die Seite eines Lkw-Anhängers. Der Lkw-Fahrer hatte die Vorfahrt missachtet und die Pkw-Software die Situation falsch eingeschätzt. Sie klassifizierte den hellfarbigen Lkw-Anhänger bei klarem Himmel als Autobahnschild und fuhr ungebremst in ihn hinein; der Passagier des autonomen Fahrzeugs starb. Ein menschlicher Fahrer hätte die Situation vermutlich erkannt und eine Notbremsung eingeleitet. Auch in einem KI-gesteuerten System können also noch Fehler passieren; diese unterscheiden sich aber teilweise fundamental von menschlichen Fehlern. Die *Bayes-Fehlerrate* ist in der Regel nicht null, das heißt, dass es Situationen gibt, die der Algorithmus nicht

korrekt einschätzt. Der Wirkungsbereich der (anderen, neuen) Fehler kann unter Umständen für einen anderen, nicht einbezogenen Personenkreis verhängnisvoll sein.

Stochastik und Folgenabschätzung – Die Technikkritik des 20. Jahrhunderts beschäftigte sich überwiegend mit der Frage, wie die vom Menschen geschaffenen Mächte beherrschbar bleiben könnten.⁵ KI und Digitalisierung hingegen bringen andere Herausforderungen mit sich. Heutige Problemlagen haben sich von der Distribution von Potenzialen längst hin zur Wahrscheinlichkeitsrechnung verschoben. Von einer einfachen Wenn-dann-Logik ist somit nicht mehr auszugehen, obwohl das Wort »Programmierung« dieses Missverständnis nahelegt.

Moderne Steuerungsprogramme haben einen Code von mehreren hunderttausend Zeilen. Geistig erfassen kann ihn kein Programmierer. Auch Funktionalität absichernde Tests können niemals alle möglichen Szenarien einbeziehen, was die Handhabung in der Praxis teils bedenklich unabsehbar macht. Dass dies kein spekulativer Pessimismus ist, zeigen drei Beispiele aus Bereichen, die als höchst gesichert gelten und unter intensiver Beobachtung stehen:

Der Flash Cash Crash: Am 6. Mai 2010 um 2.45 Uhr kam ein gewaltiger Kurseinbruch über den US-amerikanischen Aktienmarkt. Dabei verlor u. a. der US-amerikanische Leitindex S&P 500 innerhalb von Minuten sechs Prozent seines Wertes. Hochautomatisierte und auf Echtzeit getrimmte Algorithmen hatten eine Börsenmanipulation als echtes Marktsignal gewertet und massive Verkäufe ausgelöst, die zu den Wertverlusten führten. Es dauerte mehrere Jahre, bis dieser Fall vollständig aufgeklärt und verstanden war.

Der Cambridge-Analytica-Skandal: In der Folge der 2018 aufgedeckten Verwendung persönlicher Daten von Millionen Facebook-Nutzern zu Wahlkampfzwecken ohne deren Wissen traten die Fragen der Datensicherheit und Anonymisierung in den Blickpunkt der Öffentlichkeit. Insbesondere Facebook sah sich zu hohen Investitionen in diesen Bereich gezwungen. Allerdings zeigen neue Publikationen⁶, dass die bislang zur Anonymisierung eingesetzten Methoden die Privatsphäre nicht ausreichend schützen. Mit nur wenigen demographischen Daten lässt sich selbst in stark fragmentierten Datensätzen wieder ein Personenbezug herstellen.

Boing 737 Max 8: Eine fehlerhafte Software brachte 2018 und 2019 zwei solcher Maschinen zum Absturz.

Der Autopilot hatte aktiv in die Flugzeugsteuerung eingegriffen und aufgrund fehlerhafter Sensordaten zu diesen Abstürzen beigetragen.

Erst extreme Ereignisse erzwangen hinreichend präzise technologische Handlungsanweisungen (z.B. im Contergan-Skandal). Offenbar führen erst schmerzhaft spürbare Fehlentwicklungen zu einem Umdenken beim vertrauensseligen Einsatz neuer Technologien. Wir gehen somit gefährlichen Zeiten entgegen, denn viele Technologien werden den Alltag beherrschen, ohne dass je eine Risikofolgenabschätzung vorgenommen worden wäre.

Irrationale Erwartungen und irrationale Enttäuschungen – Von Fehldiagnosen durch Algorithmen abgesehen, kann die Erwartungshaltung vieler Anwender zu Enttäuschungen führen. Wo sich falsche Vorstellungen von der Wirkungsweise und den Anwendungsmöglichkeiten statistischer Methodik und der KI verfestigen, bleiben unvermeidlich viele Hoffnungen unerfüllt. Das stärkt wiederum eine Abwehrhaltung gegenüber dieser Problemlösungsweise. Die durchaus vorhandenen positiven Potenziale werden dann, allem Technikoptimismus zum Trotz, nicht genutzt.⁷ Insbesondere in der medizinischen Forschung wird seit Längerem über die falsche Nutzung statistischer und maschineller Lernmethoden diskutiert, nachdem sich herausgestellt hat, dass ein großer Teil der modernen datengetriebenen (scheinbar höchst rationalen) Forschung auf falschem Methodenverständnis und naiver Gläubigkeit basierte.⁸

Mehr Spiel als Regel – Dass KI sich durch aktives Lernen/Adaption gegenüber dem Ausgangszustand verselbständigen und weiterentwickeln kann, ist ein technologiehistorisches Novum. Werkzeuge erhalten die Fähigkeit, sich an neue Situationen anzupassen. So wird bei *OpenAI*⁹ ein einfaches Such-und-Versteckspiel simuliert. Interagierende (KI-gesteuerte) Agenten lernen dabei allmählich, eine Umwelt und zur Verfügung gestellte Werkzeuge für ihre Ziele zu nutzen, und entwickeln selbst unerwartete, nicht beabsichtigte Strategien und brechen damit die unerklärten »Spielregeln«.

Offene Fragen menschlichen Zusammenlebens

Krise des Vertrauens? – Durch das Internet verbreitete Informationen haben die Glaubwürdigkeit verloren, die sie einst hatten, als man ihre Quellen für unabhängig von klassischen Meinungsmanipulatoren hielt. In jüngster Zeit werden nun aus Bildern Videos gemacht, Bildausschnitte mit zunehmender Präzision

auf bewegte Bilder montiert und mit täuschend echt erscheinenden »Aussagen« unterlegt. Die neuartige Technik des »Deepfake« wird bald ununterscheidbare Abbilder einer bestimmten Person alles Mögliche sagen lassen können.¹⁰ Bilder sind nicht mehr verlässlich; wir können unseren Augen und Ohren nicht mehr trauen. Klassische bildgestützte Nachrichten werden allmählich wertlos, es sei denn, sie kommen von jenen Medien, denen man selbst gerade vertraut. Jeder lebt dann – aus guten Gründen – in seiner »Filterblase«. Bedeutet dies, dass die vielbeschworene, meist überschätzte »kritische Öffentlichkeit« nun endgültig vor ihrem Ende steht?

Vermutlich gewinnen vor diesem Hintergrund persönliche Bindungen wieder an Bedeutung. Vertrauen gegenüber Informationsquellen wird zu einer knappen Ressource, die wir – je nach Gruppenzugehörigkeit – gleichsam spekulativ und »politisch« nutzen. Aber müssen sich die »Nutzer« nicht bereits heute meist blindlings auf Technik verlassen, deren Funktionsweise sie nicht ansatzweise verstehen? Das wachsende Misstrauen den Informationsmedien gegenüber wird jedenfalls das Unsicherheitsgefühl im Umgang mit der ubiquitären Technologie noch verstärken.

Krise der Arbeit? – Schon seit Jahrzehnten wird die Vernichtung klassischer Industriearbeitsplätze durch Automatisierung vorhergesagt.¹¹ Bevor »der Mensch zur

3 Zur Dichtheit im mathematischen Sinn: <https://www.spektrum.de/lexikon/mathematik/dichtheit/1830>.

4 Judea Pearl: *Causality – Models, Reasoning, and Inference*. Cambridge (Mass.) 1999. Pearl ist von den aktuellen Entwicklungen um das »Deep Learning« unbeeindruckt: »All the impressive achievements of deep learning amount to just curve fitting.«

5 Tiefgehend, nämlich anthropologisch nachfragend bereits Günther Anders: *Die Antiquiertheit des Menschen*. München 1956.

6 Z. B. Luc Rocher / Julien M. Hendrickx / Yves-Alexandre de Montjoye: »Estimating the success of re-identifications in incomplete datasets using generative models«, in: *nature communications*, 23.07.2019. <https://www.nature.com/articles/s41467-019-10933-3>.

7 Sich prinzipiell zu verschließen, wie es in Deutschland als Tendenz zu beobachten ist, bedeutet angesichts des weltweiten Wettbewerbs um technologische Innovationen auf dem noch nicht verteilten Markt aber, sich selbst die Möglichkeit zu nehmen, der Hammer zu sein statt der Amboss (siehe dazu Simon Müller / Yannic Weber: »Wie intelligent ist die Künstliche Intelligenz? Zweiter Teil: Wirtschaft«, in: *TUMULT*, Sommer 2019, S. 95–98).

8 John P. A. Ioannides: »Why most research findings are false«, in: *PLOS Medicine*, Vol. 2/8 (30.08.2005). Spyros M. Makridakis et al.: »Predicting Medical Risk and Appreciating Uncertainty«, in: *The International Journal of Applied Forecasting*, Nr. 52/2019, S. 28–35. John P.A. Ioannides: »Medical Science and Practice. Does Anyone Want to Fix Them?«, in: *The National Journal of Applied Forecasting*, Nr. 53/2019, S. 28–30. <https://replicationindex.com/2019/01/08/psychologists-p-values/>.

9 <https://openai.com/blog/emergent-tool-use/>.

10 <https://www.youtube.com/watch?v=T76bK2t2r8g> (Phänomen); <https://tinyurl.com/y4vkg4vr> (Gegenmaßnahmen); <https://tinyurl.com/y6gg5vml> (Stimmkopien).

11 Vielbeachtet zu seiner Zeit Jeremy Rifkin: *The End of Work*. New York 1995.

Funktion der Maschinen wird« (Harald Seubert) – ist dann eine Welt ohne Arbeit nicht wirklich die humane Alternative?¹²

Aber es ist ja gar nicht abzusehen, wie die Arbeitswelt sich verändern wird. Ein paar Hunderttausende von sich selbst überzeugten jungen Innovatoren weltweit sind nicht repräsentativ für die Heere der Arbeitsfähigen. Technologiebrüche, beispielsweise die für die Sozialordnung in bestimmten Gebieten so folgenreiche Ablösung der Handweberei durch Webmaschinen, hatten andernorts keine oder kaum negative Auswirkungen. Außerdem haben während der letzten einhundertfünfzig Jahre neue Technologien die Zahl verfügbarer Arbeitskräfte gar nicht reduziert, sondern sie auf andere Sektoren verteilt. Weniger als fünf Prozent der Berufstätigen arbeiten heute in der Landwirtschaft; um 1850 herum waren es ca. 60 Prozent. Zugleich wird heute beträchtlich mehr Nahrung produziert als damals. Durch neue Technologien entstehen neue Berufszweige; Arbeitsplätze werden auch dort geschaffen, wo die neue Technologie nicht eingeführt, sondern ihre Verwendung nur vorausgesetzt wird.¹³

Viele KI-Technologien werden erst konzipiert oder beginnen gerade, Marktreife zu erlangen. KI-Ansätze führen bisweilen zu unerwarteten Lösungen, sodass konkrete Auswirkungen auf unser Leben nicht zuverlässig vorauszusagen sind.

Neue Verantwortung

Odo Marquard hat schon vor Jahrzehnten überzeugend dargestellt, wie der Einzelne von der Umwälzung aller Verhältnisse überfordert wird. Er kann sich nicht mehr auf eigene Erfahrungen verlassen, weil diese mit der Umwelt, der sie entstammen, veralten. Um sich noch in der Welt zurechtzufinden, muss der Mensch auf Fiktionen, Halbwahrheiten und Vermutungen bauen. Je schneller sich die Umwelt verändert, desto lieber greift er zu solchen Haltestangen – und damit auch zu Irrtum, Utopie, Lüge und Täuschung. Es entsteht eine »Flucht aus dem Erfahrungsverlust in den Erfahrungsverzicht«.¹⁴ Wo der Einzelne überfordert ist, hilft nur die gegenwartsunabhängige Kollektiverfahrung. Das ist die in Jahrhunderten gesammelte, geronnene Erfahrung vieler Generationen, mag man sie Tradition, Sitte oder historisches Bewusstsein nennen.¹⁵ Sie vermitteln nicht nur Einsichten, sondern wirken auch der zivilisatorischen Neigung zum Höher- und Weiterstreben entgegen. Sie sind dringend benötigte Gegenkräfte, »Ernüchterungsgrößen«.¹⁶ Übrigens war

dieser Appell für überkommene Leitplanken und die Wertschätzung der kulturellen Fundamente in den Selbstverständniskrisen der technologiegetriebenen Wachstumsgesellschaften regelmäßig zu hören, etwa im ausgehenden 19. Jahrhundert und in den Jahren nach den beiden Weltkriegen. Dass diese Rufe stets leiser waren als die tosenden Verheißungen von Fortschritt und einer Abschaffung aller menschlichen Defizite, darf uns nicht dazu verleiten, in der heute anschwellenden Technologiewelle stumm zu bleiben.

Gerade die Künstlichkeit der KI erfordert es, ihr ein Eigenes, ein Humanum, entgegenzusetzen. Die Bewohner des 21. Jahrhunderts müssen sich ein neues (realistisches!) Selbstbild sowie ein gereiftes Verantwortungsbewusstsein und eine im Maßlosen maßvolle politische Ordnung zur Steuerung und Beherrschung ihrer Lebensverhältnisse schaffen. Oder sie werden wieder einmal scheitern.

12 Ständiges Beschäftigtsein und die Definition des eigenen Geltungswertes durch Erwerbsarbeit sind relativ junge Phänomene. Siehe dazu Eberhard Straub: Vom Nichtstun. Berlin 2009.

13 Susan Lund/James Manyika: »Five lessons from history on AI, automation, and employment«, McKinsey & Co., Nov. 2017. <https://tinyurl.com/yywentoq/>

14 Odo Marquard: »Zeitalter der Weltfremdheit? Beitrag zur Analyse der Gegenwart«, in: ders.: Apologie des Zufälligen. Stuttgart 1986, S. 15. Marquard spricht in Bezug auf seine eigene Zeit von einem durch die Verunsicherung hervorgerufenen »Klima hysterischer Angst«, das uns schwer in der Gegenwart wiederzufinden ist. Bezeichnenderweise beschränkt sich dieses Klima auf die »westliche Welt«.

15 Marquard spricht von einem »Kontinuitätssinn«. Ohne Traditionen, Gebräuche usw. könnten wir »unseren Kontinuitätsbedarf und [...] unseren Vertrauensbedarf nicht mehr decken und den Wirklichkeitswandel nicht mehr aushalten« (ebenda, S. 19).

16 Ebenda, S. 19.