

Reinhard Messerschmidt

Risiken und Nebenwirkungen des
aktuellen Hypes um künstliche
Intelligenz



CAIS Report

Fellowship
Mai bis Juni 2017

Risiken und Nebenwirkungen des aktuellen Hypes um künstliche Intelligenz

„Früher hat man dem Computer ein Problem übergeben, wenn man es verstanden hatte, heute ist es andersrum.“ (Joseph Weizenbaum, zitiert nach Kees, 2015)

Sind menschliche und künstliche Intelligenz vergleichbar?

Die Geschichte der künstlichen Intelligenz (KI) ist eine Geschichte voller Missverständnisse. Diese haben schon vor Jahrzehnten zu überzogenen Erwartungen geführt und erleben im gegenwärtigen Hype eine Renaissance. Damals wie heute gibt es viel Verwirrung um den Begriff *Intelligenz*. Zunächst ist dieser, wie der Physiker und Technikphilosoph Klaus Kornwachs (2009, S. 37) treffend bemerkt, „zu verwaschen [...], weil es keine befriedigende Theorie der Intelligenz gibt“.

Bleibt der Begriff also schon beim Menschen grundsätzlich unscharf, wie zum Beispiel Kontroversen um Intelligenztests gezeigt haben, führt er auf Computer angewandt in der Wissenschaft und Öffentlichkeit häufig zu vermenschlichenden Fehlinterpretationen, aber auch umgekehrt zu einer technischen Reduktion des Menschlichen („Mensch-Maschine“). So schreibt der Informatiker und Philosoph Stefan Ullrich (2017, S. 188):

Der Analogieschluss von der Denkweise des Menschen auf die Funktionsweise der Maschine („Elektronengehirn“) ist für Experten wie Laien gleichermaßen verführerisch, da man nichts von Computern

verstehen muss, um ihn dennoch erklären zu können; freilich schreibt man ihm dann menschliche Eigenschaften zu, die durch diese Zuschreibung wiederum entwertet werden. [...] Die Maschine folgt den ihr einprogrammierten Gesetzen in den Grenzen des physikalisch Machbaren. Wenn sie eingesetzt wird, um zentrale Bereiche des menschlichen Zusammenlebens zu kontrollieren, zu steuern, zu beobachten oder zu messen, dann muss stets mitgedacht werden, dass sie nur auf den kontrollierbaren, steuerbaren, beobachtbaren und den messbaren Bereich Zugriff hat. Der moderne Mensch hat daher seine Umwelt und sein Zusammenleben weitestgehend maschinenlesbar, also berechenbar gestaltet.

Heutige und auf absehbare Zeit auch zukünftige Computer arbeiten im Kern nach dem Prinzip, welches nach dem Mathematiker Alan Turing als „Turing-Maschine“ bezeichnet wird (Turing & Copeland, 2004). Vereinfacht ausgedrückt verarbeiten sie Nullen und Einsen nach Regeln, ohne dabei jemals wie Menschen wissen zu können, was eine Null oder Eins bedeutet, wie bereits Ada Lovelace (1842, S. 44) wusste, besser bekannt als ‘the world’s first programmer’:

The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but

it has no power of anticipating any analytical relations or truths. Its province is to assist us in making available what we are already acquainted with.

Im strengen Sinne *rechnen* Computer also gar nicht, denn logische Halbleiterschaltungen in Siliziumchips folgen nur physikalischen Gesetzen, zweifellos mit in den letzten Jahrzehnten stark vervielfachter Geschwindigkeit und Speicherkapazität. Das ist die Verschärfung der Aussage, dass Computerprogramme gar nichts entscheiden können, sondern lediglich Dinge ausrechnen (Królikowski et al., 2017). Computer können im strengen Sinn nicht einmal ausrechnen, geschweige denn entscheiden, denn sie verstehen nicht, was sie tun.

Sind Computer intelligent oder gar kreativ?

Das mit den Fachbegriffen *Syntax* und *Semantik* im Sinne von *Zeichen* und *Bezeichnetes* ausgedrückte Verhältnis können Computer allenfalls in Programmen im Hinblick auf Input und Output simulieren, dies jedoch besser als jemals zuvor. Zum Vergleich: Ab einem gewissen Alter verstehen wir Menschen (im eben erwähnten wörtlichen Sinn), was $1+1=2$ oder auch wenig später das abstraktere Schema der Addition bedeuten. Es ist allerdings zu bezweifeln, dass dies eine bewusstseinslose (Turing-)Maschine jemals über das Befolgen der programmierten Regeln hinaus auch kann. Denn von Bits in den Schaltkreisen über Assemblercode bis hin zu welcher auch immer benutzten Programmiersprache tut sie nichts weiter, als stupide den hier noch relativ einfachen Regeln des Programms zu folgen, was dazu führt, bei Input $1+1$ als Output 2 zu erzeugen. Im Prinzip ist das auch bei höherer Programmkomplexität bis hin zu KI nicht anders – mathematische Regeln befolgen, um ein vorher definiertes Problem regelgeleitet zu lösen (z.B. bei Mustererkennung etc.).

Das Gleiche gilt für Chatbots: Dass Weizenbaums *Eliza* kein Verständnis vom Konzept *Vater* hatte, wenn ‚sie‘ „How is your father“ fragte, ist offensichtlich, da sie auch hier nur programmierten Input-/Output-Regeln folgt. Nicht anders, wenn auch komplexer, war es im Fall von Microsofts ‚antisemitischer‘ Tay, die, gefüttert mit den falschen Trainingsdaten, alle Juden töten wollte, aber auf Ebene des Sinnverstehens weder wusste, was *Jude* noch *töten* bedeutet. Auch hier führt

wieder nur Input zu Output, mehr nicht. Selbst wenn wir jetzt das komplexeste Modell mit den hochqualitativsten Trainingsdaten der Welt annehmen, welches uns die nahezu perfekte Illusion vermitteln könnte, dass sich eine KI mit uns unterhält, bleibt es eine Turing-Maschine, die ihren einprogrammierten Regeln folgt.

Siri, Alexa & Co sind prinzipiell kein Stück ‚intelligenter‘ als ein Taschenrechner. Dies wohl gemerkt im Sinne von Sinnverstehen, welches ohne Bewusstsein als Träger einfach kein sinnvolles Konzept ist. Das algorithmische Durchrechnen von Möglichkeiten bei fortlaufendem numerischem Abgleich ist weder kreativ noch intelligent. Auch eine Zielfunktion hat mit zielgerichtetem menschlichem Handeln so gut wie gar nichts zu tun. Allerdings wies Alan Turing im Zusammenhang mit seinem berühmten Test (im Original *imitation game*) selbst darauf hin, dass es auf die Zuschreibung durch die Gesellschaft ankommt. Wenn Menschen ein Ding als intelligent wahrnehmen, so spielt es keine Rolle, wie es im Inneren um das Ding bestellt ist.

Verstellt der KI-Hype den Blick auf Probleme bei Big Data und Trainingsdaten?

Selbst wenn wir Computern im Rahmen einer Minimaldefinition Intelligenz zubilligen, gilt auch heute die nach Weizenbaum (1978, S. 300) „wichtigste Grundeinsicht [...], dass wir zur Zeit keine Möglichkeit kennen, Computer auch klug zu machen, und wir deshalb im Augenblick Computern keine Aufgaben übertragen sollten, deren Lösung Klugheit erfordert“. Gegenwärtig wird in einem großen Teil des Digitalisierungsdiskurses (insbesondere bzgl. Big Data und KI) das Modell der Realität mit der Realität des Modells verwechselt.

Die kurzfristige Überschätzung von KI könnte zur paradoxen Situation führen, dass der ‚Big-Data-Elefant‘ seit Jahren mitten im Raum steht und weiterhin wächst und gedeiht. Trotz enormer Skandale von NSA, BND bis Facebook erkennt nur eine Minderheit die davon ausgehende Bedrohung ihrer Souveränität und Mündigkeit (vgl. Ullrich, 2014), weil die Mehrheit stattdessen über eine kommende Superintelligenz spekuliert (oder einfach Desinteresse an beiden Themen hat).

Welche Probleme können durch die Informatik gelöst werden und welche nicht?

Digitale Allwissenheits- und Allmachtsphantasien wurden und werden oft aber nicht nur aus dem Silicon Valley heraus propagiert. Dabei wäre zu fragen, ob an eine vermeintlich ‚göttliche‘ Informatik nicht Erwartungen herangetragen werden, die sie kaum erfüllen kann. So hat der Informatiker Rainer Rehak auf dem Kongress des Chaos Computer Club 2017 in seinem Vortrag betont, dass das Fach „formale (mathematisch modellierte) Probleme ganz vorzüglich“ löse, nun aber auch noch alle anderen lösen soll.

Diese schiefe Debatte versteckt und fördert auch die Macht dahinterstehender Konzerne und ihrer Profitinteressen. Künstliche Intelligenz ist eben auch ein effektives Verkaufsargument, sowohl in der Wirtschaft als auch auf dem Markt der Ideen, wenn es darum geht, Forschungsgelder einzuwerben.

Der Philosoph Luciano Floridi hält im Interview (2017, S. 94) Spekulationen über so genannte Superintelligenz (vgl. Bostrom, 2014) nicht nur für überflüssig, sondern sogar unverantwortlich. Elon Musks Äußerung, KI sei die größte Bedrohung für die Zukunft der Menschheit, erscheint im Kontext von Klimawandel, umkämpften Ressourcen und der Tatsache, dass auch heute noch ein großer Teil der Menschheit keinen Zugang zu sauberem Wasser oder genügend Nahrungsmitteln besitzt, in der Tat als zynisch, oder in Floridis Worten als „unmoralisch“ (vgl. Floridi, 2017). Ob die, wie er es ausdrückt, „verherrlichten Rechenschieber“ halten, was in der aktuellen Debatte versprochen wird, darf in vielfacher Hinsicht getrost bezweifelt werden.

Das bedeutet natürlich nicht, dass es nichts gäbe, worüber wir uns sorgen müssten. Aber die Sorgen wären aus Perspektive kritischer Stimmen aus Philosophie und Informatik durchaus an anderen Stellen nötig, wie z.B. unserem Umgang mit dem Daten- bzw. Überwachungskapitalismus (vgl. Morozov, 2015 & Zuboff, 2015) oder mit planetaren Ressourcen und der globalen gesellschaftlichen Ungleichheit (vgl. WBGU, 2018).

Literaturverzeichnis

Bostrom, Nick (2014). *Superintelligenz. Szenarien einer kommenden Revolution*. Frankfurt a.M.: Suhrkamp.

Floridi, Luciano (2017). Elon Musks Warnung vor künstlicher Intelligenz ist unmoralisch. *Die Welt/Bilanz*, 24.08.2017. Verfügbar unter <https://www.welt.de/wirtschaft/bilanz/article167943544/Elon-Musks-Warnung-vor-kuenstlicher-Intelligenz-ist-unmoralisch.html> [01.04.2020]

Kees, Benjamin J. (2015). *Algorithmisches Panopticon. Identifikation gesellschaftlicher Probleme automatisierter Videoüberwachung*. Münster: Monsenstein u. Vannerdat.

Kornwachs, Klaus (2009). Von der Macht der Computer und der Ohnmacht der Vernunft – Weizenbaum revisited. *Forum der Forschung*, Nr. 22, 33–40. Cottbus: BTU.

Królikowski, Agata, Loebel, Jens-Martin, & Ullrich, Stefan (2017). Ausrechnen statt Entscheiden. 30 Jahre IT-Innovation. *CSR und Digitalisierung*. 317–328 Berlin: Springer Gabler.

Lovelace, Ada (1842). Notes upon the Sketch of the Analytical Engine. In: Babbage, Henry P. (Hg.): *Babbage's Calculating Engines. Being a Collection of Papers Relating to Them; Their History, and Construction*, 21–50. New York: Cambridge University Press.

Morozov, Evgeny (2015). Socialize the Data Centres! *New Left Review*, 91, 45–66.

Turing, Alan Mathison & Copeland, Brian Jack (2004). *The Essential Turing: Seminal Writing in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*. Oxford: Clarendon Press.

Ullrich, Stefan (2014). Informationelle Mü(n)digkeit. *Datenschutz und Datensicherheit* 10, 696–700. Berlin: Springer.

Ullrich, Stefan (2017). *Informationstechnische Grundlagen, Werkzeuge und Praktiken des öffentlichen Vernunftgebrauchs. Die technē der Publizität*. Verfügbar unter: <https://edoc.hu-berlin.de/bitstream/handle/18452/18436/ullrich.pdf> [01.04.2020].

WBGU – Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen (2018). *Digitalisierung: Worüber wir jetzt reden müssen*. Verfügbar unter: <https://www.wbgu.de/de/publikationen/publikation/digitalisierung-worueber-wir-jetzt-reden-muessen> [01.04.2020].

Weizenbaum, Joseph (1978). *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt a.M.: Suhrkamp.

Zuboff, Shoshana (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, 30, 1, 75–89.

Abbildungsverzeichnis

Foto Titelseite: CAIS, Matthias Begenat

Kontakt

Dr. Reinhard Messerschmidt
Wissenschaftlicher Beirat der Bundesregierung
Globale Umweltveränderungen (WBGU)

Dieser Text verkörpert die Auffassung des Autors und ist keine WBGU-Publikation.