Grenzen der schwachen Künstlichen Intelligenz

§1 Grenzfragen ohne Androiden

In der Forschung zur Künstlichen Intelligenz (KI) lassen sich die Ansätze der starken und der schwachen Künstlichen Intelligenz unterscheiden. Die *starke KI* will eine Intelligenz erschaffen, die mindestens alles kann, was Menschen als intelligente Leistungen vollbringen können. Einige Vertreter der starken KI wollen letztlich eine künstliche Person erschaffen. Die *schwache KI* will Maschinen mit Fertigkeiten versehen, die bei Menschen mit Intelligenz vollbracht werden. An anderer Stelle habe ich die schwache KI als Methodik (als Heuristik einer Philosophie des Geistes) verteidigt und zugleich den Umfang einer Computationalen Theorie des Geistes (CTM) bzw. der Kognition (CTC) kritisch betrachtet.[1] Insofern es m.E. starke Einwände gegen eine allgemeine CTM oder CTC gibt, gibt es die entsprechenden Einwände gegen die starke KI. Hier geht es nun, um die Problematik, welche Grenzen man darüber hinaus (selbst) bei der schwachen KI einräumen muss.

Schwache KI erhebt nicht die Ansprüche der starken KI, die nach *science fiction* und Versprechungen klingen. Der Unterschied zwischen diesen Sorten der KI kann zwar definitorisch gezogen und einsichtig gemacht werden, er gehört jedoch nicht zum Allgemeinwissen und wird auch nicht von allen Akteuren im KI-Bereich gemacht. So kann der Erfolg von Systemen der schwachen KI den Eindruck eines allgemeinen Fortschritts in der KI hervorrufen und die Akzeptanz des Narrativs der starken KI erhöhen. Dieses Narrativ ist nicht bloß überzogen – was zunächst harmlos sein könnte – sondern weltanschaulich aufgeladen. KI wird als die Lösung der gegenwärtigen gesellschaftlichen Probleme (von Armut bis Klimawandel) vorgesehen und als Überwindung allgemeiner menschlicher Beschwerden (wie Krankheit und Altern) verkündet.[2] Die damit einhergehende technizistische Sicht auf gesellschaftliche Probleme wird diesen in der Regel nicht gerecht – selbst, wenn ‚gut gemeint'[3] – sondern lässt politische und gesellschaftliche Problemlagen und Konfliktlagen aus den Augen treten. Außerdem bleiben keine Alternativen, wenn die

---

1 Vgl. „Was leistet eine Computationale Theorie der Kognition, und was nicht?" und „Schwache Künstliche Intelligenz als Heuristik".

2 Paradigmatisch: Ray Kurzweil, *The Singularity is Near*.

3 Vgl. etwa: Steven Pinker, *Enlightenment Now*.

technischen Versprechungen sich nicht einlösen lassen. KI-Versprechen setzen hier eine längere Geschichte der Technologiegläubigkeit fort.[4]

Auf der einen Seite kann so der Erfolg zunächst sinnvoller KI-Systeme zur Stärkung nicht sinnvoller und fragwürdiger Weltanschauungen (wie des Transhumanismus[5]) beitragen.

Auf der anderen Seite kann der Unglaube an die überzogenen Versprechungen der starken KI dazu beitragen, die Gefahren, die sich mit Anwendungen der schwachen KI, wenn sie umfassend in unseren Alltag einkehren, zu übersehen. Ethisch und philosophisch relevanter als Debatten um Androiden etc. sind heute Fragen nach den Grenzen und Auswirkungen gegenwärtiger KI.

Schwache KI hat Grenzen nicht nur bezüglich des Umfangs der Modellierung kognitiver Prozesse. Die Grenzen betreffen u.a. Eingrenzungen im *Verständnis* von ‚Intelligenz' und Grenzen der gelingenden Interaktion mit Systemen der KI. Die zweite Art der Grenzen und Beschränkungen schließt ethische Fragen ein.

§2 Grenzen durch Operationalisierungen

Eine Operationalisierung einer intelligenten Leistung schränkt deren Verständnis immer auch ein auf die berücksichtigten Modelle, die wiederum geplante Operationalisierungen und Grenzen dessen, was sich gerade umsetzen lässt, vor Augen haben.

Operationalisierungen schließen in der theoretischen Modellierung der betrachteten Entitäten mutmaßlich irrelevante Eigenschaften aus (etwa die Hautfarbe einer Person) und machen Vereinfachungsannahmen für handhabbare Algorithmen (etwa statistische Unabhängigkeit in den Daten).

Weitere mit einer Operationalisierung verbundene Festlegungen und Vereinfachungen erfolgen mit der Wahl der Softwarearchitektur (dem Programmier-Paradigma und der Programmiersprache sowie den verwendeten Datenstrukturen).[6]

---

4 Vgl. David Noble, *The Religion of Technology*.

5 Vgl. kritisch: Nicholas Agar, *Humanity's End*, oder auch: Julian Nida-Rümelin & Nathalie Weidenfeld, *Digitaler Humanismus*.

6 Vgl. z.B. George Luger & William Stubblefield, *AI Algorithms, Data Structures, and Idioms in Prolog, Lisp, and Java*; Joseph Bigus & Jennifer Bigus, *Constructing Intelligent Agents Using Java*.

Eine erfolgreiche Operationalisierung in einem erfolgreichen System kann eine entsprechende verkürzte Behandlungsweise eines Problems verankern – etwa, indem nur die behandelbaren Kernanwendungsfälle im Weiteren behandelt werden. Dies gilt umso mehr, wie sich solche Systeme als Prestige-Objekte einer Expertenkultur präsentieren. Der Umstand, dass sich ein Problem nicht mit dem investierten Aufwand (an Geld, Zeit und Expertenwissen) lösen lässt, drängt es an den Rand, während die Erfolgsbereiche nun festlegen, was es heißt, mit dem Ausgangsproblem umzugehen. Der scheinbare Erfolg bestimmt auch die Interpretation der Ergebnisse.

Bei den heute im Mittelpunkt des Interesses an KI stehenden Systemen des ‚Maschinellen Lernens' kommen Fragen nach der (einseitigen) Auswahl der Trainingsdaten und der Festlegung des bewertenden Feedbacks an die Lernleistungen solcher Systeme hinzu.[7]

Leistungen von KI-Systemen treten oft als definitorisch für ‚intelligent' auf, ähnlich wie Intelligenz sprichwörtlich das sein soll, was Intelligenztests messen. Dies reduziert das Verständnis von ‚Intelligenz' auf algorithmische kognitive Leistungen zum Nachteil anderer geistiger Leistungen, die auch Intelligenz erfordern (etwa kreative Leistungen oder das Konstruieren eines Artefaktes). Zu Beginn von Standardeinführungen zur KI wird ‚Intelligenz' (d.h. das *explanandum* bekannt aus dem menschlichen Fall) gerne *gleichgesetzt* mit zielgerichtetem Problemlöseverhalten im Rückgriff auf Wissen.[8]

Die Interaktion mit einem KI-System drängt den Benutzer, sich den Formaten und Beschränkungen der Problembehandlung, die das System ausmachen, anzupassen, d.h. selber sich dem System anzupassen als umgekehrt. Es werden Abhängigkeiten verankert, die sich nicht einfach zurückdrehen lassen.[9]

§3 Ethische und politische Grenzen

Scheinbar erfolgreiche KI-Systeme laden dazu ein, sich auf sie zu verlassen, auch dann, wenn den Benutzern ihre Funktionsweise nicht durchsichtig ist. Dies trägt sowohl zur Mystifizierung der Leistungen solcher Systeme bei als auch zur unkritischen Übernahme ihrer Resultate (etwa von Vorhersagen oder Einschätzungen z.B. ökonomischer

---

7 Vgl. Katharina Zweig, *Ein Algorithmus hat kein Taktgefühl*.

8 Vgl. z.B. George Luger, *Artificial Intelligence*, oder: Stuart Russell & Peter Norvig, *Künstliche Intelligenz*.

9 Vgl. schon früh: Joseph Weizenbaum, *Die Macht der Computer und die Ohnmacht der Vernunft.*

Entwicklungen). So verlassen sich Börsen und Finanzmärkte auf Systeme, welche die Börsianer selbst kaum durchschauen, obwohl darin ein Risiko zu Börseneinbrüchen liegt.

In der Programmierung gilt der sprichwörtliche Warnhinweis „Computer tun genau, was man ihnen sagt – nicht, was man meinte." Anekdoten aus der Informatik und viele Szenarien in *science fiction* Literatur oder Filmen handeln von genau solchen Fällen, wo scheinbar harmlose Befehle zu bizarren oder katastrophalen Ergebnissen führen, weil die Spezifikation der Aufgabe nicht (hinreichend) einschloss, die zu bewahrenden Rahmenbedingungen zu beachten (z.B. das selbstfahrende Auto, das möglichst schnell zum Ziel fahren soll, fährt in einem Wechsel von starker Beschleunigung und Vollbremsung bei Rücksichtslosigkeit gegen andere Verkehrsteilnehmer, oder der semi-autonome Rasenmäher, der Zusammenstöße vermeiden soll, fährt nur noch in einem kleinen Kreis – etc.). Was im Einzelfall eher ein amüsantes Versagen mit sich bringt, kann mit Ausmaß und Reichweite des KI-Systems katastrophale Konsequenzen haben (ein semi-autonomes Krankenhaussystem entscheidet, die effektivste Methode der Schmerzlinderung ist das Abschalten der Geräte auf der Intensivstation). Hier müsste eine Heuristik der Vorsicht und der Begrenzung der Wirkweite von KI-Systemen eingreifen.[10]

Ethisch relevant sind nicht allein Schäden, die KI-Systeme als Nebeneffekte mitverursachen, sondern auch gesellschaftliche Nebeneffekte, wie die gesellschaftliche Verteilung des Zugangs zu diesen Systemen.

Rechtlich geklärt werden muss im Einzelfall, wer für die Schäden einer Fehlfunktion haftbar zu machen ist. Dies betrifft sowohl automatisierte Beurteilungen (in Bildungseinrichtungen oder einem Kreditinstitut) mittels einer Software als auch mobile Systeme (von selbstfahrenden Robotern bis zu Drohnen) oder das ‚smart home' des *Internet of Things*.[11]

Im militärischen Bereich, der traditionell als Hauptförderer der KI auftritt, stellen sich Fragen nach der Rüstungsbegrenzung speziell im Bereich semi-autonomer Waffen.[12]

---

[10] Dies erinnert an die ‚Heuristik der Furcht' in Hans Jonas' Technikphilosophie (vgl. *Das Prinzip Verantwortung*), sobald man diese aus Jonas' metaphysischer Konstruktion löst. Die fehlende Berücksichtigung von Nebeneffekten findet sich mindestens schon in der Legende von König Midas.

[11] Vgl. Mark Coeckelbergh, *AI Ethics*, Kap. 8; vgl. auch: Deborah Johnson, *Computer Ethics*, Kap. 7.

[12] Vgl. Toby Walsh, *It's Alive*, S. 236-51. Solche Waffen sind natürlich nicht ‚autonom' im philosophisch traditionellen Sinn des Wortes (sich selbst Gesetze gebend und frei entscheidend), obwohl sie „autonom" genannt werden. Es wäre jedoch irreführend diesen starken Autonomiebegriff als Beurteilungsbasis anzulegen. Diese Systeme sind *semi-autonom* in der Art eines Computerspiels: innerhalb der programmierten Regeln werden einzelne Schritte aus einer Menge möglicher Schritte ausgewählt, um den Gegner (sei es der menschliche Spieler des Computerspiels oder die gegnerischen Soldaten bei einem Drohneneinsatz) zu schlagen. Diese Auswahl wird nicht mehr von einem menschlichen Operator des Systems kontrolliert.

Zugleich verstärken entsprechende Systeme und Software-Agenten gezielt den *bias*, der sich in *social media* (wie Facebook und Twitter) findet. Systeme, deren Erfolg in Klicks auf entsprechende Buttons und Links gemessen wird, lernen die Benutzer in eine entsprechende Richtung zu lenken (etwa durch Präsentation ähnlicher oder reißerischer Inhalte) und konterkarieren so die vermeintliche Meinungsvielfalt dieser Internetmedien.

KI-Systeme (z.B. der großen Werbungsvermarkter wie Google und Facebook) tragen bei zum um sich greifenden ‚Surveillance Capitalism'[13] der Speicherung und auswertenden Nutzung möglichst vieler durch Computergebrauch generierten Benutzerdaten und werfen damit Fragen ihrer Regulierung auf.

Der Ubiquität der KI-Systeme und den Versprechen beschleunigter Digitalisierung laufen die Bemühungen zur Regulierung hinterher. Wie in anderen Bereichen der Gesetzgebung folgt auf den Erlass des Gesetzes die Suche nach Gesetzeslücken von – in der Regel ökonomisch – motivierten Akteuren im Regelungsbereich. Je mehr Profit und Einfluss sich mit dem Einsatz von KI-Systemen verbinden, umso mehr Spezialisten (wie Anwaltskanzleien) werden von interessierten Akteuren darauf angesetzt, Regelungslücken zu finden, wobei sich mit der Nutzung dieser Lücken ein entsprechend hohes Risiko von politisch unerwünschten Effekten oder Nebeneffekten der KI-Systeme ergibt. Das Risiko geht weniger von superintelligenten KI-Systemen aus, welche ihre menschlichen Überwacher austricksen wollen, denn von einer Gesellschaft, die auf Profit- und Einflussmaximierung basiert und in der kurzfristig nutzenrationales Handeln zu unübersehbaren Folgen führt (wie bei den Auswirkungen der massiven Nutzung fossiler Energieträger und des Verbrennungsmotors auf das Klima oder der Anhäufung von atomaren oder Plastikmüll).

Manuel Bremer, 2021

[13] Vgl. Shoshana Zuboff, *The Age of Surveillance Capitalism*.