Eduard Kaeser

Philosophische Probleme der Digitalisierung

Vortrag Zürich, Verein Risiko und Sicherheit, 30. Januar 2020

Ich möchte in meinem kleinen Referat nicht so sehr das Digitale der neuen Technologien anleuchten, als vielmehr einen Grundzug der Technik in den Fokus rücken, der nicht erst mit den neuesten Systemen auftaucht. Das heisst, mich interessiert eigentlich die anthropologische Perspektive, also die Frage: Was macht die Technik aus uns? Nicht erst die neuen digitalen Systeme werfen diese Frage auf, sie verleihen ihr nur eine gewisse zeitbedingte Dringlichkeit.

Quasi-autonome Artefakte

Ich beginne am besten mit einer Kurzdefinition: Technik ist das Delegieren von menschlichen Fähigkeiten an Geräte, Maschinen, künstliche Systeme. Das können manuelle oder intellektuelle Fähigkeiten sein. Heute stehen natürlich die smarten künstlichen Systeme im Fokus. Die jüngsten Generationen, die lernenden Maschinen mit den Algorithmen der neuronalen Netze, entwickeln dabei erstaunliche Fähigkeiten. Sie erkennen visuelle oder verbale Muster, sind fähig zum Klassifizieren von Gegenständen, sie entwickeln Spielstrategien, die dem Menschen überlegen sind, sie fungieren als Entscheidungshilfen, wickeln den Zahlungsverkehr ab, führen im Netz ein eigenes Leben als Bots, die mit uns als scheinbar ebenbürtige Gesprächspartner in Kontakt treten, kurz, sie haben sich einen Status als *quasi-autonome Artefakte* zugelegt. Dabei sehen nicht wenige Kommentatoren diesen Status als Übergangsstadium zu einer höher entwickelten maschinellen Superintelligenz an, die sich womöglich von uns Menschen abkoppeln wird. Entsprechende Utopien und Dystopien sind in den Medien Dauerbrenner, im Stil von „Beherrschen schon bald einmal smarte Staubsauger, Rasenmäher, Googleautos und Drohnen die Welt?“

Nichtintendierte Effekte

Wir kennen die klassischen Tücken der Technik. Das Gerät funktioniert nicht so, wie es funktionieren sollte, es widersetzt sich seiner Indienstnahme oder geht eigene Wege. Es zeigt Züge der Autonomie. Ingenieure rechnen routinemässig mit solchen nichtintendierten Effekten. Wie es der britische Kybernetiker Stafford Beer ausdrückte: „Der Zweck eines Systems ist das, was es tut“. Das gilt auch für einfache Werkzeuge wie Hammer oder für eine Maschine wie den Automotor. Beide können für andere Zwecke eingesetzt werden als für das Einschlagen von Nägeln oder das Bewegen eines Fahrzeugs; zum Beispiel den Hammer für Weitwurf oder den Motor als Antrieb einer Waschmaschine. Wir neigen dazu, solche Verwendungen als zweckentfremdet zu bezeich-

nen. Aber warum eigentlich? Wenn ich den Hammer zum Weitwurf gebrauche, dann ist sein Zweck eben das Geworfenwerden.

Bei quasi-autonomen Artefakten erhält diese Nichtintendiertheit nun eine neue Bedeutung. Natürlich meinen wir mit den Zwecken und Intentionen menschliche Zwecke und Intentionen. Was aber, wenn die Technologie so weit entwickelt wäre, dass sich in dem, was wir als nichtintendiert, das heisst als Unfall oder Zufall, betrachten, intendierte Wirkungen der Maschinen äusserten?

Das ist zunächst einmal ein ziemlich verquerer, wenn nicht unheimlicher Gedanke. Er steht quer zu einer Linie des Denkens, die Intentionalität allein dem Lebenden zuspricht. Tiere mögen Intentionen haben. Aber Pflanzen? Oder gar Maschinen? Wir Menschen neigen dazu, in alles Sinn, Bedeutung, Zweckmässigkeit hineinzulesen. Das Spiel des Windes in den Blättern oder eine bestimmte Wolkenkonstellation „sagen“ uns etwas. Das Flussrauschen trägt uns eine Botschaft zu. Eine zufällige Anordnung von Dingen interpretieren wir als Zeichenfolge. Nicht funktionierender Technik unterstellen wir eine „Tücke“. Und dementsprechend begegnen wir ihr auch. Mit der zunehmenden Komplexität der technischen Systeme, die unseren Alltag und Haushalt durchsetzen, wächst unser Gefühl, diese Systeme nicht mehr im Griff zu haben, und wir begegnen diesem Kontrollverlust mit der typisch menschlichen Reaktion der Intentionalisierung und Personalisierung - eigentlich die moderne Form des alten Abwehrzaubers. Unser Umgang mit der Technik ist auf weiten Strecken zeitgemässer Animismus.

Unbegreifliche Maschinen

Das Problem nimmt allgemeinere Züge an. Wir treten ein ins Zeitalter der unbegreiflichen Maschinen. Man kann in diesem Zusammenhang ein „Murphy-Gesetz“ für die Entwicklung von Geräten formulieren: Jede Technologie (samt Software) wächst ihrem Designer irgendeinmal über den Kopf. Erste Anzeichen beobachten wir bereits bei den neuronalen Netzen. Natürlich kennen die Designer den Aufbau eines solchen Netzes. Die Prinzipien seiner Architektur sind sogar erstaunlich einfach. Die Basiselemente, die „Neuronen“, arbeiten nach bestimmten wohlformulierten Regeln, aber das Ganze ist nicht durch ein Master-Programm gesteuert – ebensowenig wie unser Hirn. Das Netz wird trainiert, sich selber zu regulieren und zu korrigieren. Das geschieht dadurch, dass man ihm einen Input eingibt und beobachtet, ob ein gewünschtes Resultat als Output erzielt wird. Das Problem, das dabei entsteht, lässt sich kurz zusammenfassen: Input geht hinein, Output kommt heraus, aber wir verstehen nicht oder nur fragmentarisch, was dazwischen passiert.

Ein neuronales Netz lernt statistisch, es sortiert die Daten nach dem Kriterium der Häufigkeit, es arbeitet nicht mit Kategorien und Begriffen, die uns Menschen vertraut sind. Der „gelernte“ Algorithmus identifiziert ein Muster aus Fell, Schnurrhaaren, Schwanz, spitzen Ohren schliesslich vielleicht als „Katze“, aber selbstverständlich hat er keinen Begriff von Katze. Wenn wir ihm als Trainingsdaten Bilder vorsetzen, in denen Katzen häufig auf Kissen liegen, dann könnte er daraus auch lernen, nicht Katzen zu erkennen, sondern Kissen mit gestreiftem und mit kariertem Muster zu unterscheiden. Immer wieder kommt es vor, dass sich neuronale Netze nach erfolgreicher Trainingsphase als völlig unbrauchbar für neue Anwendungen erweisen. Die Bilderkennung von Google identifizierte zum Beispiel dunkelhäutige Menschen als „Gorillas“, die Software von Flickr sah im Eingangstor des Konzentrationslagers Dachau ein „Klettergerüst“.

Natürlich wird man mir sofort sagen, es handle sich hier um Kinderkrankheiten. Trotzdem: Die Entwicklung künstlich intelligenter Systeme hat heute schon ein Stadium erreicht, in dem ihre innere Entscheidungsstruktur nicht mehr völlig transparent ist, so dass solche Systeme den Eindruck erwecken, aus eigenem „Antrieb“ zu handeln. Sie können auch aus dem Ruder laufen und Kalamitäten verursachen – einige Programmierer dramatisieren das Szenario der nichtintendierten Folgen sogar schon zu einer „Software-Apokalypse“ hoch.

Man könnte die Situation so zusammenfassen: Die Designer der künstlich intelligenten Systeme stehen oft vor ihren Kreationen wie der sprichwörtliche Zauberlehrling vor dem Besen. Kritik kommt aus den Kreisen der KI selbst. Zum Beispiel erhob im Mai 2018 ein Forscher bei Google den Vorwurf, KI funktioniere wie mittelalterliche Alchemie. Man wisse eigentlich nicht, was man tue, sondern drehe hier und da an ein paar Schrauben, bis ein Algorithmus das gewünschte Ergebnis erziele. Viele Forscher auf diesem Gebiet würden im Dunkeln tappen und wüssten eigentlich nicht, was sie täten. Nicht nur einzelne Algorithmen funktionierten wie eine “Black Box”, bei der man nur Ein- und Ausgabe kennen und nicht verstehe, was im Inneren geschehe. Ganze Teile der KI-Forschergemeinde würden inzwischen genauso anmuten.

Mensch-Computer-Symbiose

Die Entwicklung der künstlichen neuronalen Netze steht erst am Anfang, und wir können schlicht nicht wissen, welche Generationen von KI-Systemen sich daraus noch entwickeln werden. Aber wir können bereits aus gegenwärtigen Erfahrungen und Beobachtungen ein paar Extrapolationen über das Zusammenleben von Menschen und Computern anstellen.

Man halte sich nur einmal vor Augen, wie unsere alltägliche Kommunikation von unseren kleinen smarten Begleitern geprägt ist, die wir ständig mit uns herumtragen. Das Internet ist eine

Mensch-Computer-Symbiose im wahrsten Sinn des Wortes. Wir beobachten in dieser Symbiose eine heimliche und scheinbar unaufhaltsame Gewichtsverschiebung. Viele unserer künstlich intelligenten Helfer – Rasenmäher, Staubsauger, Geschirrwaschmaschine - funktionieren deshalb gut, weil sie in einem eindeutig definierten und kontrollierten Setting operieren. Desgleichen die Roboter in den Laboratorien, Büros oder Fabrikhallen. An solchen adaptiven Automaten sind primär Industrie, Ökonomie und Militär interessiert, und es kann nicht erstaunen, dass sich hier mächtige Interessenkonglomerate bilden. Ihre implizite Strategie erinnert an die Lösung des gordischen Knotens: Statt die „intelligenten" Artefakte der Umwelt anzupassen, passt man die Umwelt den Artefakten an. In der Robotik spricht man von der „Enveloppe" des Roboters, also vom Raum, innerhalb dessen Grenzen die Maschine zuverlässig funktioniert.

Das ernsthafte Problem liegt darin, dass die dichte Vernetzung all dieser Apps eine Enveloppe unserer Lebenswelt bildet, aus der wir uns kaum noch lösen können. Und in dieser Enveloppe tendieren wir Menschen dazu, nun selber zur App zu werden. Nicht wir setzen die Algorithmen ein, die Algorithmen setzen uns ein. Vor gut zehn Jahren kamen die Software-Entwickler von Amazon auf die zündende Idee, Aufgaben, für deren Lösung die Algorithmen zuviel Zeit brauchten, durch Outsourcing an Menschen zu delegieren: Gesichter erkennen, übersetzen, Spam verbreiten, kleine Programme schreiben etc. Amazons „mechanische Türke" ist eine Plattform, wo Auftraggeber Arbeiten, sogenannte „human intelligence tasks („HITs"), in Auftrag geben. Der Mensch als Appendix der Algorithmen.

Was heisst eigentlich Intelligenz?

Nun war ständig von Intelligenz die Rede. Die Frage drängt sich auf, was genau wir eigentlich meinen, wenn wir sagen, künstliche Systeme seien intelligent.

Nun, solche Systeme simulieren Fähigkeiten, die beim Menschen Intelligenz voraussetzen. In diesem Sinn führt der Taschenrechner Operationen durch, die beim Menschen ein gewisses Mass an Intelligenz benötigen. Ist der Taschenrechner also intelligent? Allgemeiner gefragt: Ist Simulation von Intelligenz identisch mit Intelligenz? Wo liegt der Unterschied? Das entpuppt sich als äusserst vertracktes Problem. Es wird seit Alan Turings berühmtem Artikel 1950 intensiv diskutiert, und man kann füglich sagen, dass eine eindeutige Antwort aussteht. Es handelt sich ja auch um ein philosophisches Problem, und von solchen Problemen ist weniger eine Lösung zu erwarten als bestenfalls ein Perspektivenwechsel.

Was bedeutet das? Es bedeutet, dass wir die Frage, ob ein Artefakt intelligent sei, nicht so sehr als eine ingenieurale Frage nach Konstruktion und Komplexität auffassen sollten. Betrachten wir

erneut den Taschenrechner. Er führt geistlos und routiniert Operationen durch, ohne zu denken, sagen wir. Und wie wäre es, wenn er immer raffinierter arbeiten würde? Gibt es eine Schwelle, wo wir sagen würden: Jetzt führt die Maschine nicht mehr geistlos Operationen durch, jetzt „rechnet“ sie wirklich?

Das war ein Lieblingsthema des Philosophen Ludwig Wittgenstein. Er hat das einmal so beschrieben („Blaues Buch“): „Ist es möglich, dass eine Maschine denken kann? (..) Wir drücken mit dieser Frage nicht eigentlich die Schwierigkeit aus, die darin besteht, dass wir noch keine Maschine kennen, die das tun kann. Die Frage ist nicht analog zu der, die jemand vor hundert Jahren hätte stellen können: ‚Kann eine Maschine Gas verflüssigen?’ Der schwache Punkt ist vielmehr der, dass der Satz ‚Eine Maschine denkt (..)’ irgendwie unsinnig erscheint. Es ist also ob wir gefragt hätten: ‚Hat die Zahl 3 eine Farbe?’“

Ich möchte das noch einmal ganz klar herausstreichen. Was Wittgenstein hier anspricht, ist etwas anderes als ein naturwissenschaftliches oder ingenieurales Problem. Wir haben es mit einem soziokulturellen Problem zu tun: mit unserer Angleichung an die Maschinen. Wir haben zwar bis heute keine Antwort auf die Frage, ob Maschinen denken können, aber wir gewöhnen uns allmählich an die Sprache der Computeringenieure und Softwaredesigner, die ja nicht selten von den Funktionen der Computer so sprechen, als würden sie denken. Die Metaphorik hat sich sozusagen in die Umgangssprache eingeschlichen, und dadurch wird der Eindruck erweckt, der Computer sei buchstäblich zur Intelligenz erwacht. Die Populärliteratur ist voll von solchen undurchdachten Metaphern. Sie spricht andauernd von den Computern, die immer mehr menschliche Tasks übernehmen, so wie umgekehrt der menschliche Körper als „biologische Maschine“, neuerdings als „biologischer Algorithmus“ betrachtet wird.

Moralische Maschinen

Hier handelt es sich einstweilen um forschungsinterne Probleme. Sie nehmen eine bedenklichere Gestalt an, wenn man sie in einem weiteren sozialen Horizont betrachtet. Wie gesagt ist unser gesellschaftlichliches Leben immer dichter von künstlich intelligenten Systemen durchsetzt, an die wir menschliche Aufgaben delegieren. Kann man Maschinen auch moralisches Verhalten lehren? Robotiker, Informatiker, Mathematiker, Psychologen, Soziologen und vermehrt auch Philosophen tüfteln unentwegt an Automaten herum, die unter bestimmten Umständen ethisch relevante Entschlüsse fällen müssen. Bevorzugte Studienobjekte sind Robotersoldaten, die anstelle von Menschen auf moralisch heikles Terrain geschickt werden können. Z.B. künstliche Scharfschützen, angesetzt auf Terroristen. Sie wären nicht nur schneller, stärker und verlässlicher als ihre humanen Partner, sie erwiesen sich auch als immun gegenüber Panik, Stress, Rachegelüsten

und anderen emotionalen Kollateralstörungen.

Betrachten wir ein ziviles Beispiel. Jüngstes Kind aus der Roboter-Aufzucht ist Myon, Prototyp eines Roboters, den das Team um den Berliner Neuro-Robotiker Manfred Hild zu sozialem Umgang erziehen will. Computertechnisch bedeutet dies einen Wandel von der zentral programmgesteuerten zur dezentral selbstorganisierenden, lernenden Maschine. Myon sind nur einfache Grundregeln einprogrammiert, die ihm ein adaptives Verhalten ermöglichen sollen. Er ist dadurch viel flexibler als ein fest programmierter Roboter, er gewinnt an humanoider Individualität, ja er hat eine „Biografie". Wird er zu einem Quasi-Menschen, den wir allmählich als „unseresgleichen" akzeptieren müssen?

Manfred Hild zeigt sich zuversichtlich: „Ich glaube nicht, dass wir Angst zu haben brauchen, wenn wir es schaffen, (Robotern) unsere Werte zu vermitteln." Aber was sind „unsere" Werte? Jene der Robotiker? Ihrer Investoren, der globalen Digitalunternehmen, des „Westens"? Zu denken geben sollte die professionelle Sorglosigkeit der Robotererzieher. Als ob es bloss eine Frage der Technik (und der Zeit) wäre, den Artefakten Werte einzupflanzen und womöglich moralische Konflikte aus der Welt zu schaffen. Zuerst bauen, und dann schauen. Bevor wir die Roboter erziehen, sollten wir also kritisch fragen, wie und zu welcher Ideologie ihre Designer erzogen worden sind.

Der Robo sapiens übernimmt die Macht

Die Medien sprechen gerne vom etwas gruseligen Szenario einer Machtübernahme des Robo sapiens. Aber dieser Machtwille ist nichts als der kaschierte Machtwille des Homo sapiens. Wie sagte Putin: Wer in KI in Führung geht, wird die Welt beherrschen. Die superintelligente Maschine enthüllt zudem eine tiefe Widersprüchlichkeit in unserem Verhältnis zu den Artefakten. Wir wollen smarte Maschinen, die uns übertreffen, übermenschlich werden - und gleichzeitig wollen wir servile Maschinen, die in ihrem Status untermenschlich bleiben.

Die grösste Gefahr der KI liegt nicht bei der KI, sondern darin, dass der Mensch sozusagen von selbst vor ihr kapituliert. Was uns wirklich zu denken geben sollte, ist der Ohnmachtswille des Homo sapiens, sprich: die willfährige Bereitschaft, seine kognitiven, aber auch moralischen Kompetenzen an die Maschine abzutreten. Man muss sich das klar vergegenwärtigen: Wir machen uns selbst immer abhängiger von Algorithmen, die wir mit der Verantwortung für das Wohlergehen von Millionen von Menschen betrauen. Ihr Code ist die heimliche Macht. Aber diese Macht hat keinen „Willen". Und ihre Fehler sind unsichtbar. Oder vielmehr: Das wohl Be-

drohlichste an dieser Macht ist, dass sie den Begriff des Fehlers eigentlich gar nicht kennt. Das Programm irrt sich nicht. Es läuft einfach willenlos – auch über den Weltuntergang hinaus.

Haben wir die Schwelle schon überschritten?

Ich knüpfe, kurz zusammenfassend und abschliessend, an den Anfang an, die Kurzdefinition der Technik als Delegieren von menschlichen Fähigkeiten an Artefakte. Der technische Fortschritt reduziert fortwährend das menschliche Element in unseren Tätigkeiten, bis zu einem nahezu verschwindenden Grad. Das „Agens" verschiebt sich somit vom Menschen zur Maschine.

Und damit manifestiert sich ein tief ironischer Zug des Fortschritts. Er maschinisiert humane Aktivitäten und humanisiert maschinelle Aktivitäten. All die Geräte, welche die Laboratorien verlassen und sich in unser Alltagsleben mischen, werden „belebt" und „beseelt". Wir verlieren buchstäblich unsere Seele an sie. Das ist die neue Form von Animismus, auf technisch avanciertem Niveau. Dieser Animismus bleibt solange harmlos, als wir uns seiner bewusst bleiben und uns nicht von ihm unser Selbstverständnis diktieren lassen. Aber wahrscheinlich haben wir die Schwelle schon überschritten.