

L'homme machine?

Der neue, alte Streit um das Menschenbild und die Spaltung der Fakultäten

Ulrich Nortmann

(2007)

1. Laptop oder Lyrik

Am Projekt künstlicher Intelligenz, kurz: KI, scheiden sich die Geister. Unter anderem deshalb, weil man das KI-Projekt so verstehen kann, dass es am traditionellen Menschenbild rüttelt. Das traditionelle Bild: Menschen sind, bewahre, *keine* maschinellen Systeme; denn diese sind vergleichsweise dumm, wir sind es nicht. Sollte es dagegen, so wie es die KI will, eines Tages gelingen, Maschinen richtig intelligent zu machen, sie z. B. natürliche Sprachen tadellos verstehen zu lassen und sie damit dem menschlichen Fähigkeitsniveau anzunähern, so würde die Auffassung, dass Menschen keine (organischen) Maschinen seien, ein ganzes Stück weniger für sich haben als vorher.

Die Grenzlinie zieht sich an mancher Universität quer durch ein und dieselbe Fakultät. Es gibt einerseits in Philosophischen Fakultäten des Landes die traditionell arbeitenden Philologen (und -innen), ausgestattet – leider auch nicht mehr ganz flächendeckend – mit einem ausgeprägten Sinn für die Feinheiten der menschlichen Sprachverwendung sei es in alltäglichen, sei es in künstlerisch ambitionierten Zusammenhängen. Das ist auch gut so. Technische Neuentwicklungen bedeuten diesen Kollegen oft wenig. Es gibt andererseits ausgesprochen technologisch orientierte Sprachforscher wie die Vertreter des aufstrebenden Faches Computerlinguistik. Computerlinguisten wollen Rechner Sprachverarbeitung und -erzeugung lehren. Viele haben dabei vermutlich das Ideal „semantischer Maschinen“ im Blick, die Sprache wirklich verstehen und nicht nur Sprachverständnis mehr oder weniger gelungen simulieren können, durch einigermaßen geschicktes Hin- und Herschieben von Zeichen oder Lauten.

Die philologische Partei reagiert verständlicherweise befremdet, wenn sie auf der einen Seite ein derzeit (noch?) vergleichsweise geringes Niveau maschineller Sprachverarbeitung und -synthese registriert: Jeder sieht es, die sprachliche Qualität Google-übersetzter Texte ist ein Witz. Und wenn sie andererseits einen gleichwohl

ungebrochenen technologischen Optimismus bei den „Technikern“ ausmacht: Merken die nicht, wo sie eigentlich stehen? Den Bastlern wiederum erscheinen die „Traditionalisten“ altmodisch oder sogar ein bisschen naiv in ihrer Unterschätzung der Möglichkeiten von geeignet konstruierten und programmierten Maschinen. Man geht einander, was fachliche Dinge betrifft, im Allgemeinen aus dem Wege. Manchmal allerdings kommen die Angehörigen beider Parteien in Gremien unvermeidlicherweise zusammen. Dann geht es unter Umständen giftig zu, wie wenn, sagen wir, in der Berufungskommission eine Wahl zwischen eher lyrischen oder eher maschinenorientierten fachlichen Profilen ansteht.

2. Künstliche Intelligenz

Bei den KI-Skeptikern ist zuweilen die Auffassung anzutreffen, Computer seien nicht wesentlich mehr als unkreative Datenspeicher. Datenspeicher mit einer, zugegeben, sehr hohen Aufnahmekapazität und einer gleichfalls sehr hohen Zugriffsgeschwindigkeit. Klar ist demgegenüber jedoch, bei aller Sympathie für die Sache der Traditionalisten: Computer lassen sich so programmieren, dass sie nicht Daten lediglich speichern, sondern aus den gespeicherten Daten auch etwas zu machen wissen, ebenso wie Gehirne es tun. Dieses „etwas“ kann auch bei Computern ziemlich weit gehen.

Man sollte sich von dieser Tatsache nicht durch den Umstand ablenken lassen, dass Schreibsoftware mit ihren teils infantilen Merkmalen, mit Worttrennungen, die gelegentlich den Eindruck völligen Unverständnisses nahelegen, und mit den oft störenden Mechanismen zur vorauseilenden Erfüllung vermuteter Benutzerabsichten eher verdrießt: Wäre doch *Word* nur nicht so elend dumm gestrickt, dass es immer wieder besser als ich selbst zu wissen „glaubt“, welchen Text ich als nächstes erzeugen will. Ja, dieses Zeug, mag es sich inzwischen auch beinahe unentbehrlich gemacht haben, es verdrießt wahrlich eher, als dass es den Eindruck eines auch nur ansatzweise intelligenten Gegenübers vermitteln könnte. Für eine Einschätzung der Erfolgsaussichten des KI-Projekts ist jedoch die grundsätzliche Frage zu diskutieren, wie weit maschinelle Intelligenz überhaupt reichen kann. Von kontingenten Fehlentwicklungen bei der einen oder anderen Software hat man dabei durchaus abzusehen, wenn's auch zuweilen schwerfällt, den Ärger unter der Decke zu halten.

Gegenüber dem Schimpf des letzten Absatzes ist beispielsweise Folgendes festzuhalten. Computer können so programmiert werden, dass sie dazu in der Lage sind, aus gespeicherten Mengen empirischer Daten Hypothesen über gesetzesartige Zusammenhänge zwischen Größen oder sonstigen Gegebenheiten zu gewinnen: Hypothesen von immerhin achtbarem Theoretizitätsgrad. Man kann Rechner zu Theorem-Beweisern machen, die sich sogar darauf „verstehen“ (verstehen?), auf der Basis vorausgegangener, von ihnen abgespeicherter Erfahrungen mit fehlgeschlag-

nen Beweisversuchen ihre Strategien in modifizierter Form an neuen Beweisaufgaben zu erproben. Daher kann zunächst einmal alles in allem das Fazit nur lauten: Es gibt künstliche Intelligenz, also Intelligenz, die an von Menschen gebauten Maschinen realisiert ist.

Man darf allerdings die Frage stellen, ob es technologiepolitisch richtig ist, die weitere Entwicklung dieses Bereichs mit einem sehr großen Mitteleinsatz zu fördern. Wir sehen dramatischen Problemen entgegen, die sich aus der Ressourcenverknappung infolge Übervölkerung des Planeten ergeben. Angesichts dieser Probleme werden manche der im angewandten KI-Bereich entwickelten Verfahren, vom interaktiven Fußgänger-Navigationssystem bis hin zu der per Funkchip sich für den Einkauf empfehlenden Tomate (nachdem die mit ihr zu belegende Pizza bereits im Einkaufswagen elektronisch geortet wurde), den nächsten Human-Generationen im Rückblick wahrscheinlich als abgedrehte Super-Duper-Verfeinerungs-Technologien von äußerst begrenzter Bedeutung erscheinen. Auch tragen viele Menschen ihre etwaigen Anliegen immer noch lieber anderen Menschen vor, als eine Plastikkarte in einen Schlitz zu stecken. Nicht wenige sind wohl auch bereit dazu, sich die von ihnen bevorzugte Umgangsweise von Mensch zu Mensch etwas kosten zu lassen. Das schafft dringend erwünschte Human-Beschäftigung. Daher wäre es ein Fehler, eine immer stärkere Durchdringung des Bereichs alltäglicher Verrichtungen mit KI-Anwendungen zu forcieren. Mit technologischen Kanonen auf unproblematischen Kleinkram schießen zu wollen ist keine gute Sache, solange es wesentlich Dringenderes zu tun gibt. Doch dies: Wie wollen wir leben? ist partiell eine andere Diskussion als die Selbstbild-Diskussion: Wer sind wir?

3. Kreativität bei Maschinen?

Die für unser Selbstbild relevante Frage ist, wie weit denn das gehen kann, was eben mit einigen Beispielen angedeutet wurde. Sind von geeignet programmierten Computern große theoretische Würfe zu erwarten wie die Erzeugung von fruchtbaren Gesetzeshypthesen, die von den empirischen Daten *sehr* weit entfernt liegen? Ist mit kreativen Überschreitungen der durch bestimmte Axiomensysteme abgesteckten Beweismöglichkeiten zu rechnen, wie sie für menschliche Theorem-Beweiser jedenfalls vorstellbar sind? Vorstellbar etwa im Anschluss an Kurt Gödels gefeierten Satz und Beweis in einer für Uneingeweihte etwas rätselhaft titulierten Arbeit, die vor rund 80 Jahren, 1931, erschienen ist: „Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I“. Der Titel lässt immerhin ein wenig von der Lehre anklingen, die man am Ende aus dem Opus ziehen kann: Probleme, die sich „formal“ nicht klären lassen, nämlich nicht mit den argumentativen Mitteln irgendeines zur Hand genommenen Axiomensystems, das jenen bestimmten Bedingungen genügt, die es zu einem System vom *Principia mathematica*-

Typ machen, sie können mit klug von Fall zu Fall neu ausgedachten Argumentationsstrategien womöglich doch entschieden werden.

Viele sind zweifellos geneigt, auf Fragen wie die eben gestellten spontan mit „nein“ zu antworten. Doch die Angelegenheit ist, näher besehen, nach heutigem Erkenntnisstand als offen zu betrachten, ob das nun gefällt oder nicht. Man möchte sogar einen Schritt weiter gehen und suggestiv fragen: Sind wir Menschen selbst denn etwas anderes als maschinelle, aus Materie gebildete Systeme – und warum sollten dann, falls dem so ist, Leistungen der fraglichen Qualität nicht auch eines Tages für von Menschen gebaute maschinelle Systeme in den Bereich des Möglichen rücken? Gewiss, es bleiben Unterschiede. Die Materiesorte ist, solange Computer nicht aus *wet ware* bestehen, hier eine andere als dort, und es gibt für uns keinen menschlichen und wohl auch sonst keinen personalen Konstrukteur.

Allerdings darf der KI-Freund auf diese Weise nicht im Ernst argumentieren wollen: nicht auf der Grundlage der Prämisse, Menschen seien Maschinen. Es steht ja zur Debatte, ob „seine“ Maschinen mit Menschen intellektuell gleichziehen können, und das läuft in gewisser Weise gerade auf die Frage hinaus, ob Menschen hinsichtlich des Intellekts nichts anderes als Maschinen sind. Aus analogen Gründen sollte aber auch der Kontrahent nicht meinen, dass das Gehirn schon deshalb, weil es Sitz von Bewusstsein und Denken ist, weit mehr als eine organische Maschine sei. Denn dies ist wiederum das, was zur Debatte steht. Es ist erst noch zu klären – und viele arbeiten weltweit daran –, worin Bewusstsein eigentlich besteht und ob eine Maschine, sei sie organisch oder nicht, der Sitz von Bewusstsein und von Denkaktivität sein kann, bis hin zum Erbringen von Leistungen der höchsten Originalitätsstufe.

4. Der Mensch eine Maschine?

Bei der Maschinen-Thematik ist überdies Gelassenheit angebracht. So wie ein zeitgenössischer Philosoph, der in Bielefeld lehrende Ansgar Beckermann, sie an den Tag legt, der einmal geäußert hat: Wenn es Maschinen mit derartig schönen Eigenschaften gebe, wie wir Menschen sie besitzen, dann habe er nicht das geringste dagegen, eine Maschine zu sein. Mancher kann vielleicht nicht anders, als mit dem Wort „Maschine“ die Vorstellung von einfachen, ja primitiven Mechanismen zu verbinden. Er soll sich in diesem Fall für die Zwecke der Diskussion um maschinelle und menschliche Intelligenz das allzu negativ besetzte Wort ersetzt denken durch „physikalisches System“ oder durch irgend etwas anderes in dieser Art, vergleichsweise neutral. Julien de la Mettrie, ein philosophierender Arzt, wählte indessen in der Absicht zu provozieren bewusst das andere Wort, nämlich für den Titel seines 1747 veröffentlichten Buches „L’homme machine“. Sicher ist nur, dass Menschen, sollten sie denn rein physikalische Systeme sein, einen sehr besonderen, äußerst

bemerkenswerten „Maschinen“-Typus repräsentieren würden. Das ist offenbar auch der Tenor der Äußerung Beckermanns.

Doch wie steht es mit einem weiteren, beispielsweise von Ethikern unterschiedlicher fachlicher Provenienz in die Diskussion gebrachten und ohne Zweifel sehr wichtigen Punkt: Untergraben Ergebnisse der Lebenswissenschaften jede gesellschaftlich wünschenswerte Ausformung eines übergreifenden Ethos? Der Zusammenhang mit unserem Thema besteht darin, dass wir sozusagen als Komplement zur KI-Bewegung, wo diese mit Maschinen den Menschen approximieren will, eine in der umgekehrten Richtung operierende Neuro-Bewegung erleben, die dazu antritt, den Menschen gewissermaßen als biologische Maschine vor Augen zu führen. Nehmen wir an, diese in Deutschland mit den Namen Gerhard Roths und Wolf Singers verbundene Bewegung bringt es irgendwann dahin, dass der Mensch, wie der sogenannte „Naturalismus“ in der Philosophie es will, komplett ins Naturgeschehen integriert erscheint. Dieses Geschehen wird, nach dem Stand der Theorie-Bildung in den zuständigen naturwissenschaftlichen Disziplinen, als etwas begriffen werden, das, von einer Prise quantenmechanischen Zufalls abgesehen, unter deterministischen Gesetzen steht. (Selbst wenn viele solcher Gesetze uns aufgrund ihrer Komplexität nicht zugänglich sein sollten.) Werden wir uns am Ende einer derartigen, denkbaren Entwicklung eine Gesellschaft eingehandelt haben, in der es äußerst unfriedlich zugeht? Nämlich deshalb, weil nach erfolgreicher Demontage von Willensfreiheit und Verantwortung jeder unter Hinweis auf die physischen Voraussetzungen seines Verhaltens sich selbst Übelstes zu tun erlauben und dafür Nachsicht verlangen dürfte?

5. *Homo sapiens* und *Sphex ichneumoneus*

Man kann eine solche Sorge vieler Theologen, Philosophen und Kulturwissenschaftler nachvollziehen. Und doch verhält es sich auch hiermit nicht so, wie es zunächst den Anschein hat. Es sei hypothetisch gesetzt, sagen wir ruhig einmal: als „der schlimmste Fall“, dass alles menschliche Handeln und Entscheiden in der skizzierten Weise als Teil eines rein physischen Geschehens durch eine Mischung aus physikalischer Determination und gewissen Zufallsanteilen gesteuert ist. In welchem Sinne dabei immer noch Platz für Autonomie bleibt, so paradox dieser Anspruch zunächst klingen mag, hat der amerikanische Philosoph Daniel Dennett auf für mich überzeugende Weise erklärt: in seinem schon 1984 veröffentlichten Buch *Elbow Room. The Varieties of Free Will Worth Wanting*. (Deutsch: *Ellenbogenfreiheit. Die erstrebenswerten Formen freien Willens*, 1986 und neu aufgelegt 2005.)

Für das Vorliegen von Freiheit und Autonomie bei der Ausführung einer Handlung bzw. der Genese einer entsprechenden Entscheidung, so die Grundidee, kommt es gerade nicht darauf an, dass der Zufallsanteil groß ist und infolgedessen

alles auch anders laufen könnte (oder in der Rückschau: anders hätte laufen können). Was mir durch Zufall widerfahrt (als „meine“ Entscheidung), stammt nicht wirklich von mir, oder? Es kommt im Gegenteil darauf an, dass es anstelle eines Anderskönnens eine Determination des Handelnden durch die „richtigen“ Faktoren gibt. Nämlich durch ein Zusammenwirken von bei der Handlungsentscheidung gegebenen äußeren Umständen mit relativ dauerhaften, den Handelnden prägenden Dispositionen, Charaktereigenschaften, Wünschen und Überzeugungen, bzw. mit deren neuronalen Korrelaten; wobei solche Dispositionen, Eigenschaften etc. in der vorausgegangenen Entwicklung des handelnden Individuums durch die Mühle kritischer Reflexionen gegangen sein müssen mit dem Ergebnis, dass der Handelnde sich ihrer bewusst wurde und zu ihnen „ja“ sagen konnte. (Das macht die Qualität aus, die hier mit der „Richtigkeit“ von Faktoren gemeint ist.) Andernfalls ergäbe sich das Bild eines Akteurs, der durch Antriebe bis hin zur Dimension von neurotischen Zwängen oder Süchten, die er eigentlich verwirft und nicht als Momente seiner Persönlichkeit bejahen will, zu bestimmten Handlungen bewegt wird und insofern unfrei ist.

Bei Dennett ist es im übrigen noch wichtig, dass die maßgeblichen Reflexionen, wenn von Autonomie die Rede sein können soll, eine gewisse Variationsbreite und ein gewisses Mindest-Komplexitätsniveau realisieren müssen; ein Niveau, von dem angenommen wird, dass *homo sapiens* es aufgrund seiner eigentümlichen stammesgeschichtlichen Entwicklung im Allgemeinen zu erreichen in der Lage ist. Durch eine solche Koppelung von Freiheit an ein bestimmtes kognitives Niveau wird sichergestellt, dass von ihr nur da die Rede ist, wo die Bewertungsprozeduren, die Denk- und Reaktionsweisen des in den Blick genommenen Individuums nicht allzu einfachen Schemata verhaftet bleiben. Etwa solchen Schemata, wie sie bei Dennetts bekannt gewordenem Negativ-Beispiel vorgefunden werden: bei der aus menschlicher Sicht ganz unfrei „handelnden“, weil in ihrem Verhalten durch äußerst einfache und starre Muster der Reaktion auf Schlüsselreize determinierten Grabwespe *sphex ichneumoneus*.

Freiheit und Autonomie werden so freilich der Tendenz nach zu einer graduierbaren Angelegenheit: Der Mensch einerseits und die Grabwespe andererseits besetzen Außenpositionen eines Spektrums, das vielfältige Zwischenformen zulässt. Der einzelne Mensch braucht nicht einmal an jedem Tag seines Lebens im gleichen Umfang frei zu sein. Und es kommt, wenn man die Sache konsequent zu Ende zu denken versucht, Glück ins Spiel. Ein Quantum Glück – im stärksten Sinne des *vollkommen zufälligen, glücklichen Widerfahrnisses* – gehört wahrscheinlich dazu, einen Charakter auszubilden, der seinen Träger zur Autonomie qualifiziert und ihn weit genug sich entfernen lässt von einem allzu einfach konfigurierten Reiz-Reaktions-Automaten oder von einem zu automatisiert einer bestimmten Tradition verhafteten bloßen Produkt des Einflusses anderer.

Ergebnisse von Zufallsprozessen, zunächst im Hirn auf einer mit Quantentheorie beschreibbaren Mikroebene der Atome, Elektronen ... sich einstellend, sodann gegebenenfalls eingespeist als Anfangszustände in die zeitlichen Entwicklungen größer dimensionierter Systeme vom Typus des „deterministischen Chaos“ (charakteristische Eigenschaft solcher Systeme: äußerst sensitiv gegenüber kleinsten Änderungen bei den Anfangsbedingungen!), sie könnten dafür sorgen, dass in Gehirnen etwas passiert, was wir dann auf einer anderen als der neurobiologischen Beschreibungsebene etwa so ausdrücken würden: Plötzlich erhielt er (oder sie) einen Anstoß dazu, ein eingefahrener Denk-Gleis zu verlassen, eine Konstellation einmal aus einer veränderten Perspektive zu betrachten etc. etc. – und bald, der Rest dann vielleicht wieder in determinierten Bahnen, zeigte er oder sie sich als irgendwie veränderter, komplexerer, stärker reflektierter ... Charakter. Wenn es ungefähr so laufen sollte, dann wäre die Konsequenz in der Tat: Für den Charakter, den du hast (zu deiner Zufriedenheit, so nehmen wir 'mal an) und der für das Prädikat „frei, gelegentlich“ gut ist, spielen auch Glücksfälle eine Rolle, denk' nicht zu viel an „Verdienst“.

6. Dominiert Dennett am Ende Descartes?

Auf diese Weise betrachtet, mit *à la* Dennett vorzusehenden Fällen der Determination, jawohl, aber wenigstens einer Determination in der „richtigen“ Weise, erscheint selbst der „schlimmste Fall“ vielleicht nicht mehr gar so schlimm. Aber wie dem auch sei, sogar unter den Bedingungen des schlimmsten Falles wäre die Sinnhaftigkeit unserer bislang geübten – und in ihrer Genese dann natürlich im Wesentlichen determinierten – Praxis von Verantwortungszuschreibung, moralischem Tadeln und juristischem Verurteilen nicht notwendigerweise gemindert. Wir müssten uns nur dazu bequemen, uns in voller Konsequenz auf eine im Übrigen nicht grundstürzend neue Sichtweise einzulassen. Dahingehend nämlich, dass wir es bei Tadel (moralisch) und Verurteilung (juristisch) mit Arten der kausalen Einwirkung auf Menschen zu tun hätten, welche geeignet sind, gewisse bei den Adressaten der Aktion vorliegende innere Parameter in einer bestimmten Richtung zu verändern; so zu verändern, dass in Zukunft mit sozial erwünschtem Verhalten zu rechnen wäre. Aufzugeben wäre nur eine besondere Art von Freiheitsmetaphysik. Sie scheint für das in gesellschaftlicher Hinsicht so wichtige Funktionieren moralischer und juristischer Praxis nicht nötig zu sein.

Eher „therapeutisch“ zu nennende Konzeptionen der Bestrafung von Rechtsverletzern stehen schon immer diesem Gedanken nahe. Sie brechen mit der Tradition, die Legitimität von Bestrafungen von Personen für bestimmte Handlungen überwiegend daran festzumachen, dass diese Personen auch anders hätten handeln können. Es geht ihnen nicht um den Blick zurück, sondern mit dem Blick nach vorn um das unter Umständen hoch angesetzte Ziel: nicht nur für ein zukünftiges An-

ders-Können bei der für „schuldig“ befundenen Person zu sorgen, nein besser noch für ein zukünftiges Anders-Müssen. Dabei kann es auch um durchaus banal erscheinende Effekte gehen, etwa: Wir versuchen, das Überzeugungssystem der Person nachhaltig zu beeinflussen („ich weiß jetzt, was mir blüht, wenn ich noch einmal ...“). Die jüngeren Fortschritte in der Hirnforschung und das nicht von allen, aber von einigen Beteiligten gezeichnete Bild zusammengeschrumpfter Spielräume für punktuelles menschliches Anders-Können bieten aus dieser Perspektive viel weniger Anlass zur Aufregung, als gemeinhin angenommen wird.

Die Tradition, das ist zum Beispiel auch René Descartes mit seinem Anliegen einer scharfen Grenzziehung zwischen determinierten Maschinen und, aus seiner Sicht, beseelten und mit Freiheit begabten Menschen. *Non laudantur automata*, schreibt Descartes 1644 in seinen *Principia philosophiae*, den „Prinzipien der Philosophie“: Wir loben die Automaten nicht dafür, dass sie alle Bewegungen, zu denen sie eingerichtet sind, akkurat ausführen, denn sie führen sie mit Notwendigkeit aus. Es gibt kein Anderskönnen. Die Wahrheit dürfte eher sein, dass wir die Automaten deshalb nicht loben und tadeln (abgesehen von gelegentlichen emotionalen Ausbrüchen vor begriffsstutzig oder unberechenbar erscheinenden Personal-Computern), weil wir wissen: Diese Systeme sind, bis auf weiteres, in der Regel nicht konstruktiv so eingerichtet, dass wir, indem wir sie lobten oder tadelten, ein verbessertes zukünftiges Verhalten bewirken könnten. Lob und Tadel zu verteilen wäre hier sinnlos, weil die Konstruktion nicht danach ist. Das ist bei Menschen bekanntlich anders.