

KÜNSTLICHE IGNORANZ

NICHT-WISSEN-KÖNNEN/WOLLEN/SOLLEN/DÜRFEN

Ein Vortrag/Essay von Anil K. Jain

Englische Übersetzung: Artificial Ignorance – What AI Cannot, Would Not Want to, Should Not, and Must Not Know

Die Künstliche Intelligenz (KI) ist die Rettung der Menschheit. Und sie ist ihr Untergang. Beide Thesen finden ihre leidenschaftlichen Verfechter. Der KI-Diskurs (jenseits der Ingenieurwissenschaften) oszilliert zwischen Utopie und Dystopie, und kennt scheinbar keine Zwischenstufen. Er ist verliebt in das Extrem. Werden uns also die intelligenten Maschinen in das Reich der Freiheit (und nicht nur von A nach B) befördern, weil sie uns vom ökonomischen Zwang zu entfremdeter Arbeit erlösen und ein Schlaraffenland des unbegrenzten Konsums und der ständig für alle verfügbaren Dienstleistungen ermöglichen)? Oder werden die intelligenten Maschinen uns bald nicht nur vollends überflüssig machen, sondern uns – ihre Herren (und »Damen«) – gar entmachten und versklaven? Der Misanthrop hofft letzteres, der Optimist befürchtet ersteres. Nur eines scheint sehr unwahrscheinlich: dass die KI unser Leben nicht radikal verändern wird. So haben denn auch brave Wissenschaftler im Jahr 2015 ein »Digital-Manifest« verfasst, in dem sie nicht etwa zur Revolution gegen die Macht KI-Maschinen aufrufen, sondern angesichts der zu erwartenden Veränderungen warnend den Zeigefinger heben und eine »digitale Aufklärung« einfordern. Damit meinen sie nicht die Reduzierung des aufgeklärten Denkens auf 1 und 0, auf ja und nein, auf gut oder schlecht. Sie wollen der »digitalen Revolution« Regeln auferlegen, so dass »wir alle von den Früchten der digitalen Revolution profitieren: Wirtschaft, Staat und Bürger gleichermaßen« (Helbig et al. 2015). Dazu soll, unter anderem, die informationelle Selbstbestimmung unterstützt und Transparenz verbessert werden (vgl. ebd.).

Leider hat ihnen wohl niemand mitgeteilt, dass Revolutionen selten geregelt ablaufen. Und dass im Kontext der Digitalisierung, insbesondere im Bereich der Künstlichen Intelligenz, die Dimension des Nicht-Wissens bzw. der »Ignoranz« von entscheidender Bedeutung ist – so paradox sich das vielleicht auch anhören mag. Die Bedeutung der Ignoranz im Feld der KI (die der »natürliche Feind« der »Transparenz« ist) werde ich deshalb im folgenden nicht nur im Hinblick auf das Nicht-Wissen-Können, sondern auch im Hinblick auf das Nicht-Wissen-Wollen, -Sollen und -Dürfen beleuchten. Zunächst müssen wir uns allerdings fragen: Ist Künstliche Intelligenz tatsächlich intelligent bzw. kann sie überhaupt jemals intelligent sein? Ja, gibt es so etwas wie Intelligenz überhaupt?

1. DIE KONSTRUKTE DER KI UND IHRE NOTWENDIGE »BESCHRÄNKTHEIT«

Schon im Begriff der »Künstlichen Intelligenz« wird wie selbstverständlich unterstellt, dass es erstens Intelligenz gibt und dass diese zweitens künstlich »modelliert« werden kann. Das Modell der Intelligenz, auf das dabei überwiegend zurückgegriffen wird bzw. das als Maßstab der Bewertung der erreichten »künstlichen« Intelligenzleistung dient, ist zumeist die menschliche Intelligenz. Selbstkritisch betrachtet muss man allerdings zur Einsicht gelangen, dass wir Menschen eigentlich doch recht »beschränkte« Geschöpfe sind. Schon die Dialektik, also sozusagen die Synthese zwischen -1 und 1, überfordert nicht nur die Mathematik (nein, die aufhebende »Lösung« lautet nicht 0!) und die Maschinen, die auf ihr beruhen, sondern auch die meisten humanen »Geister«. Wie intelligent kann also eine Maschine sein, die sich ein so bescheidenes Vorbild genommen hat?

Die erste Beschränkung der KI, die ihre »Beschränktheit« erzeugt, liegt also im gewählten Modell und Maßstab: dem Menschen. Dieser wird dabei gleichzeitig systematisch überschätzt und unterschätzt. Die Überschätzung des Menschen betrifft dabei nicht nur seine Tauglichkeit als Modell der (künstlichen) Intelligenz, sondern auch, was seine Bereitschaft und Kompetenz zur Interaktion mit intelligenten Maschinen angeht. Hier spielen nicht nur diffuse (Berührungs-)Ängste eine Rolle, sondern Menschen *können* sich manchmal einfach nicht so verhalten, wie die mehr oder weniger intelligenten Maschinen es (von ihnen) erwarten. Ein Beispiel, das wohl viele schon einmal erlitten haben, ist der Dialog mit (zunehmend »intelligenten«) Sprachcomputern einer Hotline. Selbst wenn die Spracherkennung funktioniert, bedeutet das noch lange nicht, dass man die gewünschte Hilfestellung erhält – wenn man nämlich ein Anliegen formuliert, das den Erwartungshorizont der Maschine übersteigt. Hier hilft dann nur noch weiter, sich mit einem »echten« Supportmitarbeiter verbinden zu lassen – sofern diese Möglichkeit überhaupt vorgesehen ist. Allerdings, auch das zeigt die »Praxis«, hat man natürlich keine Gewähr, dass das menschliche Gegenüber ein »offenes Ohr« hat und die eigenen Erwartungen erfüllt. Dies gilt insbesondere dann, wenn man auf Personen trifft, die nicht gewillt sind, von ihren Routinen und Formalismen abzuweichen. Unflexibilität und mangelndes Einfühlungsvermögen sind also nicht nur Probleme der künstlichen Intelligenz, sondern allgemein von regelfixierten Vorgehensweisen. Die algorithmische Programmierung von KI-Systemen tendiert dazu, derartiges »bürokratisches« Verhalten zu reproduzieren, das schon so manchen zur Verzweiflung getrieben hat.

Um wirkliche Akzeptanz zu finden, müssten KI-Systeme also lernen, eine Flexibilität und »Einfühlsamkeit« an den Tag zu legen, zu der auch viele Menschen nicht fähig (oder willens oder »ermächtigt«) sind. Für die künstliche Intelligenz aber ist es schon aufgrund ihrer technischen-algorithmischen Grundlagen notwendigerweise schwierig, die Regelfixierung zu überwinden. Diese Problematik wird noch durch einen »Blaming-Effekt« verschärft. So wie etwa Rassisten das Fehlverhalten

eines Individuums auf die gesamte stigmatisierte Gruppe übertragen (siehe etwa den Diskurs um »Flüchtlingskriminalität«), ist es wahrscheinlich, dass Fehlfunktionen oder singuläre Fehlentscheidungen von KI-Systemen der gesamten Technik angelastet wird. Auch wenn etwa die Gesamtzahl der Toten im Straßenverkehr durch KI-Steuerung signifikant reduziert werden könnte: schon ein spektakulärer Unfall mit Todesfolge durch ein autonom fahrendes Fahrzeug würde wohl die gesellschaftliche Akzeptanz dieser Technologie insgesamt gefährden. Wir lassen uns lieber von betrunkenen Mitmenschen überfahren als von fehlgeleiteten Mobilitätssystemen. Die künstliche Intelligenz müsste also, um von uns akzeptiert zu werden, nahezu fehlerfrei arbeiten und trotzdem maximal situativ flexibel sein: ein offensichtlicher Optimierungskonflikt.

Dagegen werden eigene Fehler von uns allzu zu gerne ignoriert. Seien wir ehrlich: Was wir von den Maschinen erwarten, erfüllen wir selbst in keiner Weise. Wir sind sehr häufig inkompetent, unberechenbar und, ja, »dumm«. In der Organisationswissenschaft trägt dem die »Stupidity-Based Theory of Organizations« (Alveson/Spicer 2012) Rechnung. In der (anwendungsorientierten) Informatik immerhin ist der DAU – der Dümme Anzunehmende User – schon lange als Bezugsgröße bestimmend. Für die KI aber ist der DAU eine schwierigere Herausforderung als etwa im Design graphischer Benutzeroberflächen. Denn so wie die (scheinbar) Dummen für die (scheinbar) Klugen eine Störung darstellen und jene zwingen, die Welt »dümmer« einzurichten, als es aus ihrer Perspektive wünschenswert (und notwendig) wäre, behindert die menschliche Dummheit systematisch das Entwicklungspotential der künstlichen Intelligenz – indem sie die KI-Systeme zwingt, sich der Dummheit der Anwender anzupassen, sie zu imitieren, um von ihnen angenommen zu werden. Denn wer immer alles besser weiß und jede Schachpartie gewinnt, der wird eben nicht geliebt. Künstliche Intelligenz wird sich also auch künstlich dümmer stellen müssen, als sie es ohnehin zwangsläufig ist (weil sie ja allgemein an einem dummen Modell, dem Menschen, orientiert ist und zudem von dummen, d.h. »menschlichen« Entwicklern programmiert wird, die es nicht besser wissen/können).

Auch damit aber noch nicht genug: KI ist – in der aktuellen Praxis – nicht nur zur Dummheit verdammt und trotzdem zwangsläufig »ignorant« gegenüber bestimmten Bedürfnissen der Benutzer. Sie ist vor allem ignorant gegenüber dem, was notwendig wäre, um tatsächlich utopische Potentiale zu entfalten. Denn Technik – und ihre Entwicklung – ist niemals neutral. Die gesellschaftlichen Verhältnisse (ihre Ungerechtigkeiten und Schief lagen) »verhärten«, manifestieren sich in ihr. Insofern war es naiv von Marx anzunehmen, dass die kapitalistische Wirtschaft jene Produktivkräfte entwickeln würde, die zur Befriedigung der Bedürfnisse im kommunistischen »Reich der Freiheit« benötigt werden.

Ich fordere deshalb sozusagen eine »kopernikanische Wende« in der Betrachtung der KI. Frei nach Kant formuliert: Man versuche daher einmal, ob wir nicht in den Aufgaben der Künstlichen Intelligenz damit besser fortkommen, dass wir annehmen, die Ignoranz (und nicht die Intelligenz) ist ihr bestimmendes Element.

2. DIE »NATÜRLICHEN« GRENZEN DES WISSENS UND DAS NICHT-WISSEN-KÖNNEN

Im Jahr 1979 publiziert der französische Philosoph Jean-François Lyotard einen Text, den er im Auftrag des »Conseil des universités du Québec« verfasst hat. Dieser Text (der im Original den Titel »La condition postmoderne« trägt) wurde zu einer der programmatischen Schrift des Postmoderne-Diskurses. Thema des eher knappen Bandes ist das Wissen. Dieses sieht Lyotard immer mehr zu einer umkämpften Ressource werden, und so bemerkt er: »Es ist denkbar, dass die Nationalstaaten in Zukunft ebenso um die Beherrschung von Informationen kämpfen werden, wie sie um die Beherrschung der Territorien und dann um die Verfügung und Ausbeutung der Rohstoffe und billigen Arbeitskräfte einander bekämpft haben« (S. 26). Alles, was sich nicht in verwertbare Information übersetzen lässt, bleibt dabei allerdings auf der Strecke (vgl. ebd.: S. 23). Gleichzeitig wird das Wissen notgedrungen immer mehr zu einer Frage des Regierens (vgl. ebd.: S. 35) – und damit zum Gegenstand widerstreitender Positionen. Aber auch gleichsam von innen heraus delegitimiert die fortschreitende Wissenschaft den (ehemals gültigen) Absolutheitsanspruch des Wissens (vgl. ebd.: S. 112ff.). Das nährt vor allem das Misstrauen an den großen Metaerzählungen (etwa von der Emanzipation des Subjekts). Diese Delegitimierung des Wissens und der Zweifel an den Metaerzählungen sei entsprechend kennzeichnend für die »postmoderne« Gegenwart, so Lyotard (vgl. auch ebd.: S. 13ff.).

Allerdings: Man darf gestrost bezweifeln, dass es jemals tatsächlich sicheres, unanzweifelbares Wissen gegeben hat. Wie anders könnte man sich sonst die verzweifeln, »zwanghaften« Versuche der Theologen und – später – der Positivisten und Rationalisten erklären, Gewissheit herzustellen. Descartes' Streben nach absoluter Gewissheit etwa ist, wie er selbst im Vorwort zu seinen »Meditationes« darlegt, aus der allgemeinen Verunsicherung geboren. Sein Versuch darum: zum Scheitern verdammt. Historisch gab es immer Kämpfe um Deutungshoheit. Zwischen Katholiken und Protestanten, Christen und Muslimen, Empiristen und Idealisten, Materialisten und Hermeneutikern etc. Keine/r konnten sie abschließend für sich entscheiden.

Es gab und gibt diese Kämpfe schon alleine deshalb, weil es offenbar »natürliche« Grenzen des Wissens gibt. Immer wieder gerne werden die Untersuchungen von Werner Heisenberg »Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik« hier zum Beleg herangezogen. Man spricht in diesem Zusammenhang auch von der »Heisenbergschen Unschärferelation«: Ort und Impuls eines Teilchens können nicht beliebig genau (gleichzeitig) bestimmt werden. Die eigentliche (epistemologische) Konsequenz, die wir aus diesen quantenphysikalischen Überlegungen ziehen sollten, ist jedoch nicht die »Fokussierung« auf die Unschärfe, sondern dass es, wenn wir Heisenbergs Argumentation folgen, »objektive« Grenzen des Wissens gibt: Dinge, die man nicht wissen kann und über die

es deshalb niemals vollständige Gewissheit geben wird. Aber auch wenn wir Heisenbergs Argumentation nicht folgen (z.B. weil wir es mangels Wissen über die quantentheoretischen Grundlagen gar nicht können), so müssen wir trotzdem von bestimmten »natürlichen« Grenzen des Wissens ausgehen: jenen Grenzen des Wissens, die in unserem Wahrnehmungsapparat und unserem tatsächlichen Verständnisvermögen begründet liegen.

Ohne Zweifel gibt es nämlich Grenzen unserer Wahrnehmung – und damit auch dessen, was wir »sinnvoll« wissen können. Wir können zwar versuchen, diese Grenzen, z.B. durch technische Mittel, auszuweiten. Trotzdem können wir den »Sinn« von etwa Mikrowellenstrahlung, die wir nicht mit unseren Augen wahrnehmen können, nur erfahren, wenn wir sie auf anderen Wegen wahrnehmen können: zum Beispiel wenn sie Hitze erzeugt, die wir »spüren« können. Ist es uns nicht möglich, eine solche »sinnliche« Verknüpfung zu machen, bleibt das Wissen um die Existenz dieser Form der elektromagnetischen Strahlung abstrakt und »leer« – Phänomenologen (und Praktiker) wissen das.

Man könnte nun versuchen, die Grenzen (der Wahrnehmung) immer weiter auszuweiten und mit Sinn zu verknüpfen – um dadurch »klüger« zu werden. Tatsächlich sind die Grenzen des Wissens und der Wahrnehmung entscheidend für Intelligenz. Allerdings weniger, indem sie begrenzend wirken, sondern vielmehr, indem sie die eigentliche Grundlage unserer Intelligenz ausmachen. Denn im Bereich der Intelligenz gilt paradoxerweise: weniger (Wissen) ist mehr. Das scheint zunächst im Widerspruch mit unserer Alltagserfahrung zu stehen, wo wir dazu tendieren, Wissen und Intelligenz im wesentlichen in eins zu setzen. Wissen ist sozusagen die alltagspraktische »Objektivierung« von Intelligenz. Dieser »Objektivierung« liegt die Annahme zugrunde, dass man nur durch Intelligenz zu Wissen gelangen kann. Schüler wissen es aber besser: Fleiß ist hier der wesentlich vielversprechendere Ansatz. Und auch ein Buch ist nicht »klug«, nur weil viele Informationen in ihm stehen. Es kommt vielmehr auf die hergestellten, sinnhaften Zusammenhänge an.

Was aber bedeutet es, sinnvolle Zusammenhänge herzustellen? – Es bedeutet vor allem Selektion und Reduktion, also nicht die beliebige und gleichgewichtige Verknüpfung von allem mit allem, sondern die Konzentration auf bestimmte, relevante Verbindungen. Und so gilt selbst in der tradierten Kunst des Denkens, der klassischen Logik, nicht etwa die möglichst vollständige Ableitung für besonders intelligent und »elegant«, sondern die möglichst einfache logische Gedankenverbindung – zumindest wenn man der Argumentation Dignagas folgt, der im 5. Jahrhundert für eine Reduktion des bis dahin in Indien favorisierten fünfstufigen Syllogismus plädierte. Dieser fünfstufige Syllogismus folgte der Form:

1. These: Auf dem Berg gibt es Feuer.
2. Begründung: Weil auf dem Berg Rauch zu sehen ist.
3. Beispiel: Wo Rauch ist, ist auch Feuer wie in der Küche, anders als im See.
4. Anwendung: Auf diesem Hügel ist Rauch.
5. Schlusssatz: Auf diesem Hügel ist Feuer.

Dignaga schlug nun vor, dieses Muster auf drei Schritte zu verkürzen:

1. These: Auf dem Berg gibt es Feuer.
2. Begründung: Weil auf dem Berg Rauch zu sehen ist.
3. Beispiel: Wo Rauch ist, ist auch Feuer, wie in der Küche; anders als im See.

Damit folgt der von Dignaga vorgeschlagene Syllogismus einer ähnlichen Form wie der Syllogismus in der »Logik« des Aristoteles. Allerdings gibt es durchaus Gründe dafür, dass hier (wie dort) der logische Zirkel nicht vollständig geschlossen ist und Teile der Argumentationskette (nämlich die eigentlich tatsächlich notwendige »Rückführung«) in eigenständiger Denkleistung zu ergänzen sind. Intelligente Personen, so die offenbar zugrunde liegende Annahme, sind in der Lage, selbstständig die Gedankenkette zu vervollständigen. Wenn uns also jemand alles haarklein, bis ins letzte Detail, erklärt, gehen wir möglicherweise ganz zu Recht davon aus, dass er/sie uns für einigermaßen geistig beschränkt hält.

Im Bereich der Wahrnehmung ist die Reduktion sogar von noch viel zentralerer Bedeutung als innerhalb der Logik. Ein fünfstufiger Syllogismus mag umständlich erscheinen. Bei der Wahrnehmung droht jedoch der Kollaps des System, wenn unzureichend »gefiltert« wird. So ist unserer Wahrnehmungsapparat nicht nur von vorne herein limitiert, etwa indem nur ein schmales Band der elektromagnetischen Strahlung für uns überhaupt »sichtbar« ist. Auch in den folgenden Schritten ist die Filterung vielleicht wichtigste Aufgabe. Nur ein Bruchteil der ankommenden Information wird aktiv verarbeitet. Personen die Defizite im Bereich der Wahrnehmungsfilerung aufweisen, wie etwa im Krankheitsbild des Autismus, haben erhebliche Probleme, sich in der Umwelt zurecht zu finden. Es bleibt nur ein unstrukturiertes Rauschen, so wie wenn wir zu dicht an ein pointilistisches Gemälde oder einen Megascreen herantreten und nichts »sinnvolles« mehr erkennen können. Denn der Sinn ist auch Ergebnis der Filterung. Das Alles ist nichts (»Gescheites«). Etwas ist immer etwas Bestimmtes.

Auch an diesem Beispiel sehen wir also, dass es nicht nur »sinnvoll« sein kann, sondern notwendig ist, die Frage der (künstlichen) Intelligenz von der anderen Seite her zu betrachten: der – unvermeidlichen – Ignoranz (gegenüber bestimmten Gegenständen). Man könnte auch formulieren: das (richtige) Maß an Ignoranz bestimmt das Maß der Intelligenz. Diese beruht folglich ihrem »Wesen« nach auf der »Kunst der Unterscheidung« in Wichtiges und weniger Wichtiges. Alle Unterschiede aber sind – gemäß dieser Aussage – letztlich konstruiert: hervorgebracht, eine *Leistung* unseres Wahrnehmungs- und Denkapparates. Folgen wir dieser Intelligenzauffassung, so ist Intelligenz notwendig subjektiv, denn was wichtig oder unwichtig und was sinnvoll mit etwas anderem verknüpfbar ist, kommt immer auf den jeweiligen Standpunkt und die jeweilige Situation an. Objektiv betrachtet (d.h. losgelöst vom subjektiven Standpunkt) gibt es entsprechend keine Intelligenz, sondern nur die Willkür der Unterscheidung. Unter anderem deshalb ist Intelligenz nicht objektiv messbar. Das Bonmont »Intelligenz ist das, was der Intelligenztest

misst«, drückt dieses subjektivistische Dilemma der Intelligenz treffend aus. Und deshalb halten wir uns auch alle ganz zurecht für die Klügsten: Weil wir es – aus unserer subjektiven Perspektive – sind. Nietzsches Schrift »Weil ich so klug bin« ist darum vielleicht auch weniger Zeichen von sich abzeichnendem Größenwahn gewesen, sondern vielleicht der weisen Einsicht in die radikal subjektive Natur der Intelligenz geschuldet. Es stellt sich allerdings bei der künstlichen Intelligenz die Frage: Wer oder was ist hier das Subjekt? (Sprich: für wen »macht« sie Sinn?) Unabhängig von dieser bestenfalls politisch, soziologisch und psychologisch (aber nicht philosophisch) beantwortbaren Frage bleibt nur festzuhalten: Vielleicht sollte man eher von künstlicher Ignoranz sprechen (sofern bzw. gerade, wenn es sich um tatsächliche intelligente Maschinen gemäß dem oben dargelegten Verständnis von Intelligenz handelt).

3. NICHT-WISSEN-WOLLEN

Das Nicht-Wissen-Wollen macht sozusagen der Kernbereich der Ignoranz aus, denn diese bedeutet nicht nur einen Zustand der Unwissenheit, sondern impliziert dem allgemeinen Verständnis nach darüber hinaus auch ein Nicht-Wissen-Wollen. Ignoranz im eigentlichen Sinn kann also von »reiner« Unwissenheit dadurch abgegrenzt werden, dass etwas zwar gewusst werden könnte, aber nicht gewusst werden will. Ignoranz ist die Entscheidung zum Nichtwissen.

Im Bereich der KI-Maschinen könnte man die Simulation des humanen Nicht-Wissen-Könnens (etwa durch Filtertechnologien) auch als Nicht-Wissen-Wollen interpretieren. Schließlich verzichtet man so bewusst auf bestimmte, eigentlich »erreichbare« Informationen – aufgrund der »intelligenteren« Ergebnisse und aus informationsverarbeitungsökonomischen Gründen. Obwohl Ignoranz wenig populär ist (wer würde sich wohl selbst gerne als »ignorant« bezeichnen?), macht ein Nicht-Wissen-Wollen auch für Menschen durchaus Sinn. Ja, Ignoranz ist mitunter sogar überlebensnotwendig – und manchmal ist es eben schon recht »dumm«, etwas wissen zu wollen.

Das gilt nicht nur dann, wenn etwas gar nicht gewusst kann, aber man trotzdem viel Energie aufwendet, um es herauszufinden. Viel häufiger ist der Fall, dass Wissen belastend ist – und wir etwas deshalb nicht wissen wollen. Dies kann bewusst oder unbewusst geschehen. Das bewusste Nicht-Wissen-Wollen könnte man als »Ablenkung« bezeichnen. Wir versuchen, unseren Fokus auf andere Dinge zu lenken, etwas zu übersehen. Das unbewusste Nicht-Wissen-Wollen wird gemeinhin als »Verdrängung« bezeichnet. Verdrängung spielt gemäß der Psychoanalyse eine zentrale Rolle in der psychischen Ökonomie. Allerdings wird sie – auch von der Psychoanalyse – überwiegend als problematisch angesehen. Verdrängung widerspricht dem aufklärerischen Geist, der alles Wissen ans Tageslicht zerrren will. Und die Psychoanalyse stellt sozusagen einen Höhepunkt in der Bewegung der Aufklärung dar, verspricht sie doch dem Subjekt – mittels der therapeutischen Reflexion – Selbstperfektionierung und Befreiung auch von inneren Zwängen. Sie ist somit Ausdruck des aufklärerischen Begehrens nach Wissen und Kontrolle auch über das eigene Ich. Diese Kontrolle muss jedoch hart erarbeitet werden und erfordert es, die – durch das Über-Ich – unterdrückten und verdrängten Regungen des »Es« hervorzukehren, um so »Heilung« (durch Wissen) zu erfahren. Allerdings besteht immer auch die Gefahr, dass die der Verdrängung zugrunde liegenden Traumata etc. sich so (neue) Bahn brechen. Wir haben es also mit einer Dialektik der Verdrängung (und des Nicht-Wissen-Wollens) zu tun, die ich (verallgemeinernd) als »Dialektik von Reflexion und Deflexion« bezeichnen möchte (vgl. Jain 2000).

Der reflexive Impuls, der darin besteht, alles – auch sich selbst – permanent zu thematisieren und in Frage zu stellen, ist nämlich seiner »Natur« nach exzessiv. Er kann nicht Halt machen. Das gilt erst recht für reflexive Lernprozesse – und

das sogar auch im eher »formalistischen« Verständnis des reflexiven (d.h. hier: rekursiven) Lernens innerhalb der KI. So wird etwa auf der Website des »Artificial Intelligence Lab« der »University of Michigan« gewarnt: »the agent [in a reflexive learning system] does not consider the possible costs of learning a particular piece of knowledge. These costs hinge on the usefulness of knowledge: reflexive systems learn everything, even knowledge that does not promise to enhance the agent's behavior.«

Reflexivität tendiert also zum Exzess. Sie ist nicht nur (anderen) lästig, sondern (für einen selbst) belastend, indem sie (er)fordert, immer weiter voran zu schreiten. Deshalb ruft sie Gegenkräfte auf den Plan, die Gegenkräfte der Deflexivität, der Ablenkung, der Verdrängung, des Nicht-Wissen-Wollens. Deflexivität ist aber nicht nur ein Schutz vor den Exzessen der Reflexivität, sondern gewissermaßen dialektischer Ausdruck der Kräfte der Beharrung, die gegen die von den reflexiven Impulsen geforderte Veränderung ausgerichtet sind. Deflexivität ist Innovationsignorant beziehungsweise: sie fürchtet das Unbekannte, die Veränderung. Andererseits muss jede Innovation, um sich in der Praxis zu behaupten, greifbar und wirklich werden. Und das bedeutet zugleich: irgendwann ist Schluss mit Reflexion, muss man aufhören, zu hinterfragen. Es gilt, zur Tat zu schreiten und alles auszublenden, was einen vom Weg abbringen könnte. Störrisch und beharrlich. Wider das (scheinbar) bessere Wissen der anderen und ihre Kritik. Das mag vielleicht nicht immer klug sein, aber es ist notwendig.

Und diese Form der (deflexiven) Ignoranz, dieses abwehrende Nicht-Wissen-Wollen ist deshalb auch für KI-Systeme eine wesentliche Fähigkeit zur Entscheidungsfindung und zur Erreichung von Handlungsfähigkeit. Es muss eine Stoppregel geben, die reflexive Exzesse, den infiniten Regress, verhindert. Solche Regeln rein formalistisch und »zahlenmäßig« zu definieren ist allerdings schwierig. Wenn man etwa die Anzahl der Durchläufe durch einen Algorithmus begrenzt, wer sagt einem, dass nicht genau beim nächsten Mal (zu dem es aber eben nicht kommt) das ideale Ergebnis herausgekommen wäre? Menschen sind hier durch »Bauchgefühle« und »angeborenen Starrsinn« geleitet. Sie scheitern damit häufig. Aber sie haben damit manchmal (selten sogar durchschlagenden oder »Epoche machenden«) Erfolg. Auch hier ist wiederum entscheidend, das Richtige nicht wissen zu wollen. Was das ist, lässt sich meist nur retrospektiv bestimmen. KI-Systeme müssten also, um diese Fähigkeit des Rechten-Nicht-Wissen-Wollens zu simulieren, in die Zukunft blicken können – oder einen eigenen »Starrsinn« entwickeln. Starrsinnige, Nicht-Wissen-Wollende, unsere Kritik deflexiv abwehrende Maschinen werden wir aber wohl kaum ertragen wollen und können – ein Dilemma für die künstliche Ignoranz. Sie ist so (manche werden ergänzen wollen: ganz wie wir auch) zum ewigem Lernen verdammt und wird nie etwas (abschließend) wissen. Denn tatsächlich ist das Nicht-(Mehr-)Wissen-Wollen ja der eigentliche »Garant« des Wissens – zumindest im Hinblick auf das Handeln. Ein relevanter Teil der »eigentlichen« Intelligenz liegt,

so könnte man (zwischen-)resümieren, in der richtigen Form und dem richtigen Zeitpunkt der nicht-wissen-wollenden Ignoranz.

4. NICHT-WISSEN-SOLLEN

Das Nicht-Wissen hat neben der epistemologischen (Nicht-Wissen-Können) und der »pragmatischen« (Nicht-Wissen-Wollen) aber auch eine ethisch-normative Dimension: Bestimmte Dinge soll man nicht wissen.

Rechtlicher Ausdruck für dieses Nicht-Wissen-Sollens sind z.B. die Regelungen des Datenschutzes. Hier wird davon ausgegangen, dass Individuen eine Privatsphäre besitzen und die Herrschaft über ihre persönlichen Informationen behalten sollten. Auch KI-Systeme müssen diesen Normen des Datenschutzes genügen. Dazu dürfen sie entweder gar nicht in der Lage sein, bestimmte Informationen zu erhalten. Oder sie müssen diese Informationen »vergessen« bzw. so vor Zugriffen abschirmen, dass ein »Missbrauch« unmöglich ist. Diesem Modell des Datenschutzes liegt offenbar die Annahme zugrunde, dass bestimmtes Wissen den Subjekten (des Wissens) »gehört«. Es handelt sich um »privates« Wissen. Doch gerade dieses »private« Wissen weckt Begehrlichkeiten – nicht nur von staatlichen Überwachungsorganen. Es ist für Unternehmen eine hoch interessante Ressource zur Mehrwertgenerierung, da durch solches Wissen zielgerichtete, individualisierte Werbung an die Nutzer von Diensten »ausgespielt« werden kann. Das Geschäftsmodell ganzer Segmente der IT-Branche beruht nahezu ausschließlich auf dieser Wissensware, die die Privatsphäre bedroht. Und je intelligenter die Systeme der Vernetzung der Information, desto größer die Gefahr.

Die zunehmende »Durchsichtigkeit« des Einzelnen im Zuge der Digitalisierung im Verbund mit Techniken der »algorithmischen Intelligenz« (wie man die KI wohl treffender bezeichnen würde) wird also einerseits als eine Bedrohung empfunden. Andererseits gilt »Transparenz« als notwendige Bedingung der Akzeptanz neuer Technologien. Das gilt auch für die algorithmische beziehungsweise künstliche Intelligenz. Wir wollen wissen und verstehen wie diese Technologien funktionieren – und vor allem was mit dem generierten Wissen geschieht. Hieran schließt sich aber zwingend die Frage an: Wem gehört das Wissen? Dem Einzelnen? Der Gesellschaft? Den Firmen, die die Technologien zur Generierung des Wissens entwickeln und anwenden? (Oder gar den KI-Systemen?)

Das machen wir offenbar sehr stark davon abhängig, was der konkrete Inhalt des Wissens ist. Handelt es sich um Wissen/Informationen über konkrete Personen, so tendieren wir dazu, das »Eigentum« an diesem Wissen den jeweiligen Personen zuzusprechen. Nicht so stark an Personen gebundenes Wissen soll dagegen der Allgemeinheit zur Verfügung stehen. Dem stehen allerdings die ökonomischen Verwertungsinteressen entgegen. Das Patentrecht ist ein Versuch, hier einen Ausgleich zu schaffen. Es offeriert den Patentinhabern die Möglichkeit, Wissen zeitlich begrenzt exklusiv ökonomisch zu verwerten. Dafür müssen sie es aber veröffentlichen – und so der Allgemeinheit überantworten. Das Gegenmodell dazu ist die Geheimhaltung des Wissens. Auch hier gelten offenbar unterschiedliche ethische Normen für Individuen und »Korporationen«. Dem einzelnen billigen

wir zu, dass er/sie sein Wissen für sich behält (vor allem wenn es »persönliches« Wissen ist). Die Geheimhaltung von (nützlichem) Wissen durch private Gruppenakteure wie Firmen empfinden dagegen viele als unmoralisch.

Im Bereich der künstlichen Intelligenz erzeugt diese widersprüchliche Auffassung zum (Nicht-)Wissen-Sollen eine Reihe von Problemen: Soll man etwa künstliche Intelligenz wie ein Individuum betrachten und ihr deshalb zugestehen, bestimmtes Wissen zu verheimlichen und ausschließlich für sich zu benutzen? Oder ist die künstliche Intelligenz nur »Instrument« ihrer Eigner (und besitzt kein eigenes »Recht«)? Dann würde man tendenziell bejahen, dass das von der KI generierte Wissen der Allgemeinheit zur Verfügung stehen muss (zumindest nach Ablauf von bestimmten »Schutzzeiten«).

Dieses »Eigentumsproblem« des KI-generierten Wissens wird noch dadurch verschärft, dass bestimmte Technologien innerhalb der KI (wie etwa neuronale Netze) streng genommen gar kein veröffentlichbares Wissen bereitstellen, sondern nur »Anwendungen«. Die konkreten Handlungsregeln, die der Anwendung zugrunde liegen, können – ähnlich wie häufig bei humanen Alltagsexperten – gar nicht expliziert werden. Oder um ein anderes Beispiel in diesem Problemfeld zu nennen: Wer ist der (Tantiemen-berechtigte) Urheber der Komposition einer KI-gestützten Kompositionssoftware, in die der humane »Komponist« nur ein paar Parameter eingegeben hat? »Gehört« das Ergebnis nicht eher der KI (oder ihrem »Schöpfer«) als dem Anwender? Spätestens wenn KI-Systeme (sei es intendiert oder als »Nebenfolge«) eine Art (Selbst-)Bewusstsein entwickeln, wird man künstlicher Intelligenz zudem bestimmte »Subjektrechte« und damit private »Wissenshoheit« nicht verweigern können.

Das aber sind letztlich »technische« bzw. juristische Probleme. Selbst wenn man sie lösen könnte, blieben die mit ihnen verbundenen ethisch-moralischen Probleme des Nicht-Wissen-Sollens bestehen. Diese sind im Bereich der KI so wenig lösbar wie im Allgemeinen. Ja, sie treten verschärft zu Tage, denn sie lassen sich wohl kaum algorithmisch bewältigen. Denn zumindest ich tendiere hier dazu, den »Traditionalisten« innerhalb der Ethik weitgehend zuzustimmen, die sich auf Aristoteles berufen. Dieser legte dar, dass ein ethisches Gut immer im Einklang mit den Regeln der Gesellschaft stehen muss. Ethische Tugenden sind folglich ein Ergebnis der Gewöhnung (vgl. NE, Buch 2-1). Außer, dass man Mäßigung und das Mittlere (anstatt dem Extremen) anstreben soll, gibt es gemäß Aristoteles nämlich keine allgemeinen und »vernünftigen« Regeln im Bereich der Ethik, sondern letztere gehören zum Bereich der »Verstandestugenden«. Dies führt auch dazu, dass er Gerechtigkeit und die Achtung der Gesetze gleichsetzt (vgl. ebd., Buch 5-2). Das aber bedeutet letztlich nichts anders als, wie der »Sophist« Thrasymachos, zu sagen, dass Normen letztlich ein bloßer Ausfluss von Macht sind (vgl. Platon: *Politeia*, Buch 1). Das aufgeklärte Denken konnte freilich weder einen solchen »Realismus« noch den bloßen Verweis auf die »Sitte der Väter« als Grundlage der »Moral« akzeptieren. Und so versuchte Kant mit seinem »Kategorischen

Imperativ« eine allgemeine und auf rationale Einsichtsfähigkeit gegründete Formel für das »richtige« Handeln zu formulieren: »Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde« (KdpV). Würde man nun allerdings versuchen, diesen ethischen Grundsatz Kants algorithmisch in KI-Systemen zu verankern, würde man schnell feststellen: Eine solche Regel ist »unpraktikabel«, weil zu abstrakt. Ohne Referenzen auf bestimmte (konventionelle, allgemein geteilte) Werte ist es nämlich gar nicht möglich zu bestimmen, was Grundlage eines solchen allgemeinen Gesetzes sein könnte. Das Sollen der Ethik ist nicht rational-theoretisch, sondern nur empirisch-praktisch bestimmbar. Jeder Versuch, Ethik algorithmisch zu implementieren, muss an der (räumlichen und zeitlichen) Kontingenz des ethisch-moralischen Empfindens scheitern. Die künstliche Intelligenz könnte bestenfalls ethische Normen (aus der Beobachtung menschlichen Verhaltens und mit entsprechendem Training) »erlernen«. Wie zum Beispiel von KI-Systemen mit »schützenswertem« Wissen, das nicht gewusst werden soll, umgegangen wird, ist so aber nicht eindeutig vorzubestimmen – außer es werden explizite Anweisungen gegeben. Diese aber können kaum für alle Anwendungsfälle erschöpfend gegeben werden. (Regel-)Konflikte und die Verletzung von (Daten-)Schutzrechten sind damit sozusagen »vorprogrammiert«.

5. NICHT-WISSEN-DÜRFEN

Das Dilemma der Regel-Konflikte gilt verschärft, wenn es um das Nicht-(Wissen-)Dürfen beziehungsweise überhaupt den Bereich verbindlicher gesetzlicher Regelungen und ihre Implementierung in KI-Systeme geht. Der Bereich des Datenschutzes (und seine Beziehung zum Nicht-Wissen-Sollen) wurde bereits oben angesprochen. Allerdings gilt für KI-Systeme wie für humane Akteure: in unserer modernen Welt, die hochgradig verrechtlicht ist, müssen rechtliche Normen »zwingend« berücksichtigt werden – wenn man sich nicht außerhalb des Gesetzes stellen will. Tendenziell sind darum alle Handlungen (und Nicht-Handlungen) von KI-Systemen durch rechtliche Regelungen betroffen. Und das führt zwangsläufig zu Konflikten – nicht nur ggf. mit erlernten »moralischen« Regeln (denn das Gesetz steht bekanntlich nicht immer in Einklang mit der Moral), sondern indem sich die rechtlichen Regeln gegenseitig widersprechen (können) und auch in ihrer Auslegung uneindeutig sind. Das Recht wurde schließlich nicht von Logikern erdacht und auch für Menschen, ja, selbst für (höchste) Richter, ist die Abwägung zwischen verschiedenen juristischen Normen mitunter eine nur schwer zu lösende Aufgabe (an der sie scheitern). Jede Rechtsnorm bedarf der Auslegung – und diese ist, genauso wie das Recht selbst, Veränderungen und dem Wandel der Zeit (und damit der Gesellschaft und Kultur) unterworfen.

Man kann also nicht einfach KI-Systeme mit Gesetzen »füttern« und ihnen »vorschreiben«, sich an diese zu halten. Sie müssen im Stande sein, diese – immer wieder neu – zu interpretieren und bei den zwangsläufig sich ergebenden Normenkonflikten abzuwägen. Eine eindeutige Hierarchie der gesetzlichen Regelungen und Normen gibt es dabei nicht. Und selbst die Anwendung der »Roboter Gesetze«, die Isaac Asimov in seinem Science-Fiction-Roman »I, Robot« erdacht hat, um die Menschen vor Schaden durch autonom agierende Maschinen zu bewahren, ist nicht »trivial«. Denn das erste und (oberste) Gesetz lautet zwar: »Ein Roboter darf kein menschliches Wesen verletzen oder durch Untätigkeit zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.« Doch wie soll die künstliche Intelligenz entscheiden, wenn sowohl durch die eine wie durch die andere Handlung oder wenn sowohl durch Handeln wie Nichthandeln jemand zu Schaden kommen würde? Nach (Opfer-)Zahlen? Oder soll es (auch) andere Bewertungskriterien, wie die gesellschaftliche Relevanz/Leistungsfähigkeit der Person(en), geben? Man sieht, selbst eine so scheinbar eindeutige (erste) Regel ist interpretationsbedürftig und eine zwingende Auslegung gibt es nicht.

Hinzu kommt für die künstliche Intelligenz aber noch ein ganz anderes Problem. Menschen, die sich immer an alle Gesetze und Regeln halten, betrachten wir eher als dumm (oder zumindest doch als »unsouverän«). Denn für aufgeklärte Menschen gehört der souveräne (selbstgesetzgebende) Umgang mit rechtlichen Vorschriften sozusagen zu den Kernkompetenzen der Intelligenz. Intelligente Menschen halten sich nicht starr an Regeln, sondern nutzen den Interpretations-

spielraum maximal aus oder überschreiten auch schon mal rechtliche Grenzen – wenn die Regel im spezifischen Fall absolet ist, wenn es ihnen nutzt, wenn es anderen nutzt (und, sofern sie »moralisch« handeln, wenn der Schaden nicht vorhanden oder gering ist). Maschinen würden wir solche Souveränität freilich ungern zugestehen. Doch damit verdammen wir sie zwangsläufig zur »Dummheit« der starren Regelverfolgung.

Dabei hätten KI-Systeme selbst im Fall einer Regelverletzung keinerlei rechtliche Konsequenzen zu befürchten. Sie würden »ungestraft« bleiben, denn sie sind (noch) keine rechtsfähigen Subjekte. Die (juristische) Verantwortung tragen – je nach Blickwinkel und Auslegung – ihre Eigner, ihre Anwender, ihre Konstrukteure, ihre »Kontrolleure«. Zur rechtlichen Schuldfähigkeit gehört nämlich der »freie Wille« – obwohl man mit guten, mittlerweile auch neurobiologisch untermauerten Gründen bezweifeln kann, dass der menschliche Wille tatsächlich so frei ist, wie es die Juristen unterstellen (vgl. z.B. Sprenger/Gevorkian: Hingespinst Willensfreiheit). Im Fall von KI-Systemen müsste ein freier Wille allerdings in jedem Fall ausgeschlossen werden, denn die algorithmische Intelligenz kann nur wollen, was ihr der Algorithmus vorgibt. Das aber kann, trotz aller Probleme der Regel-Auslegung, nur sein: halte die Gesetze ein. Alles andere wäre selbstverständlich illegal. Und deshalb *darf* ein KI-System noch nicht einmal potentiell wissen, wie man Gesetze umgeht. Diese künstliche Ignoranz gegenüber den Gesetzen muss die künstliche Intelligenz (juristisch) zwingend verwirklichen.