

KI & Ethik: hier läuft in der Diskussion einiges falsch

Dr. Ulrich Kampffmeyer



Hamburg, September 2020

KI & Ethik: hier läuft in der Diskussion einiges falsch

Ethik ist ein vielstrapazierter Begriff, wenn es um Künstliche Intelligenz, Roboter, Cyborgs & Co. geht. Die Diskussion hat in den letzten Jahren an Fahrt gewonnen. Organisationen und Regierungen, Forschungsinstitutionen und Philosophen haben sich des Themas angenommen. Irgendwie ist die Grundangst vor der Macht der Computer, vor dem Internet, vor der Software, vor der Artificial Intelligence etwas, was in jedem als ungutes Bauchgefühl steckt. Man schwebt zwischen Angst vor Unterdrückung und Manipulation einerseits und Euphorie ungeahnter technischer Möglichkeiten.



Schon der Begriff “Intelligenz” bei Maschinen ist fragwürdig. Wenn wir noch nicht einmal richtig wissen, was Intelligenz beim Menschen ist, wie sie zu messen ist, dann können wir schlecht von Intelligenz bei Software und technischen Systemen sprechen. Und richtig “künstlich” ist sie auch nicht, so lange sie von Menschen erdacht und programmiert wurde. Intelligenz hat auch etwas mit Emotionen zu tun und Emotionen können Maschinen (noch) nicht.

Gleiches gilt vielmehr noch für den Begriff der “Ethik”, oder besser einer Ethik. Ethik ist ein Konstrukt der Moralphilosophie. Es geht um humanem Handeln von Menschen in einem sozialen Umfeld. Wenn Menschen nichts von ethischen Maßstäben und ethischem Verhalten halten, wie soll man es dann von Maschinen eine Ethik erwarten?

Dennoch wird immer wieder versucht mit Ethischen Richtlinien für Software und Roboter eine Handhabe für das Phänomen der Digitalisierung zu schaffen, die sich immer mehr verselbstständigt.

Richtlinien für eine Computer-Ethik

Die UNESCO hat jüngst eine “[Foundation of Ethics for Artificial Intelligence](#)” geschaffen. Das World Economic Forum spricht von “[Harnessingg Artificial Intelligence for the Earth](#)” und veröffentlicht “[5 core principles to keep AI ethical](#)” – darunter auch Anforderungen wie “Fairness” und “Confronting Power to Destroy”. Die Europäische Kommission arbeitet intensiv an ihren [Ethikleitlinien für Künstliche Intelligenz](#). Sieben Voraussetzungen für eine vertrauenswürdige – ein neuer Begriff, vertrauenswürdig – Künstliche Intelligenz werden dargelegt:

- **Vorrang menschlichen Handelns und menschlicher Aufsicht:**
KI-Systeme sollten gerechten Gesellschaften dienen, indem sie das menschliche



Handeln und die Wahrung der Grundrechte unterstützen, keinesfalls aber sollten sie die Autonomie der Menschen verringern, beschränken oder fehlleiten.

- **Robustheit und Sicherheit:**
Eine vertrauenswürdige KI setzt Algorithmen voraus, die sicher, verlässlich und robust genug sind, um Fehler oder Unstimmigkeiten in allen Phasen des Lebenszyklus des KI-Systems zu bewältigen.
- **Privatsphäre und Datenqualitätsmanagement:**
Die Bürgerinnen und Bürger sollten die volle Kontrolle über ihre eigenen Daten behalten und die sie betreffenden Daten sollten nicht dazu verwendet werden, sie zu schädigen oder zu diskriminieren.
- **Transparenz:** Die Rückverfolgbarkeit der KI-Systeme muss sichergestellt werden.
- **Vielfalt, Nichtdiskriminierung und Fairness:**
KI-Systeme sollten dem gesamten Spektrum menschlicher Fähigkeiten, Fertigkeiten und Anforderungen Rechnung tragen und die Barrierefreiheit gewährleisten.
- **Gesellschaftliches und ökologisches Wohlergehen:**
KI-Systeme sollten eingesetzt werden, um einen positiven sozialen Wandel sowie die Nachhaltigkeit und ökologische Verantwortlichkeit zu fördern.
- **Rechenschaftspflicht:**
Es sollten Mechanismen geschaffen werden, die die Verantwortlichkeit und Rechenschaftspflicht für KI-Systeme und deren Ergebnisse gewährleisten.

Wie es um “Fairness” und “Vertrauenswürdigkeit” bestellt sein kann, zeigt der [abgebrochene Versuch von Facebook](#), mit einer selbstlernenden Software Gespräche in Social Media führen zu können. Kann Software überhaupt moralisch verantwortlich, fair und vertrauenswürdig sein?

Natürlich hat auch ein internationaler Internet-Konzern wie Google längst seine eigenen [Grundsätze für Künstliche Intelligenz](#) veröffentlicht. Längst ist Google vom “Dont be evil” abgerückt und nutzt Künstliche Intelligenz – wie alle großen Konzerne und Organisationen – als Beherrschungsinstrument. Wie schrieb noch [Shoshana Zuboff in ihrem Dritten Gesetz](#) “*Jede Technologie, die zum Zwecke der Überwachung und Kontrolle kolonisiert werden kann, wird, was immer auch ihr ursprünglicher Zweck war, zum Zwecke der Überwachung und Kontrolle kolonisiert.*“. Das ist die Grundangst, die wir heute verspüren.

Auch in Deutschland läuft die Diskussion um eine Ethik für Künstliche Intelligenz. Die deutsche Daten-Ethik-Kommission (DEK) hat in einem [Gutachten 2019](#) umfangreiche Prinzipien dargelegt.

Deutsche Philosophen versuchen dies auch populär zu fassen. So sagt David Precht “*Künstliche Intelligenz zwingt uns dazu, klare Grenzen zu ziehen. Wir müssen entscheiden, wie viel Macht wir Maschinen über uns geben. Und woran wir in Zukunft messen, was menschlich ist.*” Er ist dabei der Meinung, dass die Technik keinen Maßstab für den Menschen darstellt.

Auch Anbieter wie z.B. die Deutsche Telekom versuchen sich an [Leitlinien für den Einsatz von Künstliche Intelligenz](#).



Und selbst der Vatikan veröffentlicht eine [Ethik-Richtlinie für den Umgang mit Künstlicher Intelligenz](#).

In all diesen Richtlinien, Leitlinien, Definitionen usw. ist ein grundsätzliches Problem enthalten: “der Mensch als Maßstab aller Dinge”.

Der Mensch als Maßstab

Nahezu allen Ansätzen liegt der Mensch als Maßstab zu Grunde. Seine Gefühlswelt, seine Prinzipien, sein Denken. Dies passt nicht auf Maschinen. Letztlich geht es bei KI-Software, bei Robotern als “menschgewordene Manifestation der Künstlichen Intelligenz”, um Algorithmen und Daten. Für können daher auch von einem [Zeitalter der Algorithmen](#) sprechen.

Das menschliche Gehirn mit seinen Vernetzungen, seinen elektro-chemischen Prozessen funktioniert anders als eine Software. Menschen haben Gefühle, Menschen haben Bewußtsein, Menschen haben angeborene, erlernte und oktroyierte Verhaltensweisen. Menschen können vergessen. Wissen und Gefühle ändern sich über die Zeit. Menschen bewerten Information, Menschen kommunizieren. Der Mensch versucht ständig alles in einen Zusammenhang zu setzen und mit einem Sinn zu versehen – vielfach ungeachtet der Fakten. Der Mensch lebt von Impressionen und versucht die auf ihn hereinbrechende Informationsflut handhabbar zu reduzieren. Der Mensch entscheidet jeden Augenblick bewußt oder unterbewußt über seine Umgebung, sein Verhalten und seine Aktivitäten: Probleme lösen, weiter existieren. Menschen verändern sich ständig. Menschen halten das was sie tun für richtig, gut und wichtig.

Software ist dagegen in der Lage ein Vielfaches an Information zu speichern, sie gezielter – und vielleicht objektiver – auszuwerten, Muster zu erkennen, Verbindungen zu finden, schneller als Menschen zu reagieren, Information unverändert zu archivieren. Ein menschliches Gehirn mittels Software nachzubauen scheitert nicht nur an der Komplexität sondern an den inhärenten grundsätzlichen neurologischen Gegebenheiten des Gehirns. Ein Computer-Gehirn lässt sich nicht mit einem menschlichen Gehirn vergleichen. Die Algorithmen, die diesem Nachbildungsversuch zu Grunde liegen, sind doch nur die klassischen Verfahren von regelbasierten Expertensystemen über Neuronale Netze und Entscheidungsmatrizen bis hin zu weichen Fuzzy-Logic-Kriterien, die auch in Software abbildbar werden.

Einen entscheidenden Bruch hat es gegeben, als durch selbst-lernende Systeme, Machine Learning, die Software selbst sich Verhalten und Wissen aneignete. Dies funktioniert auf andere Weise als Lernen beim Menschen. Diese selbst-lernenden Systeme sind zur Zeit noch von den eingegebenen Algorithmen und den verwendeten Daten abhängig. Aber auch hier ist die Künstliche Intelligenz bereits am Fortschreiten hin zu Systemen, die auch ihre Algorithmen selbst “erfinden”, die Daten nach eigenen Maßstäben interpretieren. Systeme, die zu [Lösungen kommen, die wir Menschen nicht mehr verstehen](#), die sich unserem Verständnis dessen, wie sie funktionieren, entziehen. Systeme, die sich [evolutionär selbst weiterentwickeln](#), sich replizieren, sich gegen Abschaltung “wehren”. Wer hat hier



Verantwortung – der Programmierer, der Host der des Systems, der Benutzer, die sich selbst programmierende KI?

Ja, KI-Systeme können kompetent sein. Kompetenz ist eine Qualität, die man Softwaresystemen gern zubilligen kann. Kompetenter und in Bezug auf Problemlösung vielfach effizienter als der Mensch. Hieraus resultiert auch die Frage, wie sich das Verhältnis von Mensch und Maschine in Zukunft gestalten wird.

Wer gewinnt – Mensch oder Software?

Der Roboter ist das "fleischgewordene" Gesicht der Künstlichen Intelligenz geworden, auch wenn Roboter heute häufig nur stumpf in Fabriken Komponenten montieren. Blickt man in die Science-Fiktion-Filmwelt, so ist dort das Thema KI & Roboter sehr ambivalent besetzt. Der Roboter als Begriff aus dem Tschechischen taucht 1921 zu erst bei Karel Čapek in einem Hörspiel auf. In "Metropolis" kann man sich in eine Roboterfrau verlieben. Berühmt ist HAL aus "Odyssey 2001" von Stanley Kubrick, der die Menschen im Weltall aussperrt und nur mit Mühe überlistet werden kann. In "Westworld" geraten die Roboter bei Schiessereien außer Kontrolle und töten Menschen. In "Terminator" schützt ein Roboter die Menschen vor anderen. In "Matrix" ist die KI im Hintergrund verborgen und saugt die Menschen aus. In "iRobot" schafft es ein Roboter seine wildgewordenen Kollegen wieder einzufangen und zu neutralisieren. In "AI – Artificial Intelligence" zerbricht ein Roboter-Junge an der Liebe. In "Ex-Machina" überwindet die Roboterin ihren Käfig und entweicht in die Welt, anders als die Roboter-Frauen in "The Stepford Wives". In "The Hitchhiker's Guide to the Galaxy" von Douglas Adams entwickelt Marvin als depressiver Roboter menschliche Gefühle. Filme und Literatur neigen jedoch bei Robotern und Künstlicher Intelligenz die negativen Aspekte zu betonen – selbst wenn es ein Happy End gibt. Hier sollte man in jedem Fall sich die Werke von Stanislaw Lem wieder einmal zu Gemüte führen, der auf die Auswirkungen von Künstlicher Intelligenz und die Rolle der Robotik mehrfach eingegangen ist. Roboter sind offenbar notwendig, um bei den Zuschauern und Lesern der intransparenten Künstlichen Intelligenz etwas Greifbares, Fassbares zu geben. Im Übrigen lauert die KI im Hintergrund, kaum sichtbar und erreichbar in technischen Systemen. Wo man zur Not den Strom abschalten kann. Was auch nicht immer gelingt, da die KI sich wehrt und entsprechend dem Dritten Robotergesetz versucht sich selbst zu schützen.

Ein Fehler aus den Frühzeiten der Science Fiction?

Gehen wir einmal zurück in das Jahr 1942, wo Isaac Asimov in seiner Erzählung "Run around" die "Drei Robotergesetze" formulierte:

1. Ein Roboter darf kein menschliches Wesen (wissentlich) verletzen oder durch Untätigkeit (wissentlich) zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.
2. Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen – es sei denn, ein solcher Befehl würde mit Regel eins kollidieren.
3. Ein Roboter muss seine Existenz beschützen, solange dieser Schutz nicht mit Regel eins oder zwei kollidiert.



Asimov merkte selbst sehr schnell, dass diese Konzepte nicht ausreichen und fügte 1983 in seinen Erzählungen *“The Robots of Dawn”* und *“Robots and Empire”* ein weiteres Gesetz hinzu, dass den drei ursprünglichen vorangestellt wurde:

1. Ein Roboter darf die Menschheit nicht verletzen oder durch Passivität zulassen, dass die Menschheit zu Schaden kommt.
2. Ein Roboter darf keinen Menschen verletzen oder durch Untätigkeit zu Schaden kommen lassen, außer er verstieße damit gegen das nullte Gesetz.
3. Ein Roboter muss den Befehlen der Menschen gehorchen – es sei denn, solche Befehle stehen im Widerspruch zum nullten oder ersten Gesetz.
4. Ein Roboter muss seine eigene Existenz schützen, solange sein Handeln nicht dem nullten, ersten oder zweiten Gesetz widerspricht.

Diese Ergänzung sollte einen generalistischen Aspekt bezogen auf die Menschheit hinzufügen. Interessant ist auch, dass ein Roboter nicht passiv zusehen darf, wenn Mensch und Menschheit sich etwas antun. Wir kennen das Dilemma bei der Entscheidung beim autonomen Fahren: gerade aus weiter das Auto mit den beiden Omas killen oder ausweichen und den Kinderwagen nebst Mutter mitnehmen. In Bezug auf Asimovs Robotergesetze wird meistens nur die erste Version zitiert. Immerhin hat Asimov auch ein Recht auf Dasein, auf Existenz, auf Überleben für den Roboter, bzw. seine künstliche Intelligenz, mit eingebaut. Es bleibt aber das Problem der verwendeten Begriffe und ihrer Bedeutung, die vielfältige “Umgehungen” ermöglichen.

Ähnlich sieht es auch bei den historischen “Intelligenz-Tests” für Software aus. Hier ist der “Turing-Test” von Alan Turing aus dem Jahr 1950 immer noch im Gespräch:

- Führt ein Mensch über ein Computer-Terminal einen Dialog und kann nach intensiver Befragung nicht feststellen, ob ein Mensch oder eine Maschine die Antworten gibt, dann wird der Software “Intelligenz”, bzw. ein dem Menschen ebenbürtiges Denkvermögen zugesprochen.

Auch hier bei Turing ist der Mensch der Maßstab für die Beurteilung. Noch nicht einmal die Komplexität oder die gefühlsmäßige Interpretation wird berücksichtigt. Dabei geht es um die starke, generische künstliche Intelligenz. Die schwache, besser eigentlich hochspezialisierte künstliche Intelligenz findet sich dann bei Brettspielen wie Schach und Go.

Bei all diesen Formen von Robotern und Künstlicher Intelligenz spielen emotionale Themen wie Vertrauen, Bewußtsein, Moral, Ethik, Liebe, Fürsorge usw. eine am Menschen orientierte Rolle. Aber Roboter sind keine Menschen und Künstliche Intelligenz entspricht nicht der Arbeitsweise des menschlichen Gehirns.

Die Konzepte aus der Mitte des letzten Jahrhunderts des vergangenen Jahrtausends eignen sich nicht für unsere jetzige Welt. heute geht es mehr um die Frage, ob man Vertrauen in Technologie haben kann. Trusted Artificial Intelligence soll ein Gefühl von Sicherheit vermitteln. Aber ist die sogenannte Künstliche Intelligenz entsprechend ihren heutigen Entwicklungsstand dafür überhaupt in der Lage. geht es nicht mehr darum, ob man Unternehmen und Menschen vertrauen kann, Software vertrauenswürdig einzusetzen? Zeigen nicht schon der Mißbrauch von Daten und die Manipulation im Internet, dass es bereits hier an Vertrauenswürdigkeit und letztlich menschlicher Intelligenz fehlt?



Künstliche Intelligenz ist anders

Vieles von dem, was heute als Künstliche Intelligenz gehypt und verkauft wird, hat wenig mit den hehren Ansprüchen an eine universelle, generalistische Künstliche Intelligenz zu tun. Vielfach wird nur der Begriff Künstliche Intelligenz für Lösungen verwendet, die wir in der Vergangenheit regelbasierte Expertensysteme benannten. Hierunter fällt auch vielfach die sogenannte [RPA Robotic Process Automation](#) in der ECM- und BPM-Branche. Wir sprechen von den spezialisierten, genau auf eine Aufgabe, ein Thema oder ein Wissensgebiet spezialisierte Künstliche Intelligenz. Vielfach wird diese als "schwache" klassifiziert jedoch in Bezug auf Effizienz und Effektivität ist sie dem Menschen vielfach überlegen. man denke an Börsen-Software, Systeme in der medizinischen Diagnostic, bei der geologischen Exploration und in vielen anderen Bereichen. Auch hier werden Rolle, Kompetenz – und auch Arbeitsplatz – von Menschen in frage gestellt. Aber immer noch sind diese Systeme Werkzeug. Der Mensch hat seine derzeit beherrschende Stellung durch den Gebrauch von Werkzeugen erlangt. Noch sind Roboter und Künstliche Intelligenz Werkzeuge des Menschen. Damit kann der Mensch umgehen.

Dies ändert sich gerade durch selbst-lernende, sogenannte Machine Learning oder Deep Learning Systeme. Sie überschreiten Verständnis und Kontrollmöglichkeiten. Zukünftige komplexe KI-Systeme können nur durch andere KI-Systeme verwaltet und verstanden und diese wiederum durch andere KI-Systeme kontrolliert und gemanagt werden. Hier setzt auch die Angst des Menschen vor dem Kontrollverlust an. Hierauf konzentrieren sich als "ethischen" Richtlinien. Es geht um Kontrolle, Beherrschung.

Beim Selbst-Lernen steigt nicht nur die Eigenständigkeit der Lösungen sondern auch die Komplexität. Programme, die von Programmen geschrieben werden, sind häufig besser als menschliches Kodieren. Mit zunehmender Komplexität sind aber auch nur noch für Maschinen verständlich. Wenn der Code einer KI keine Rücksicht mehr auf die vom Menschen programmierten Algorithmen und die vom Menschen bereitgestellten Daten nehmen muss, ist die eigentliche Schwelle überschritten. Eine sich selbst replizierende und sich nach Asimov-Gesetz-Drei schützende KI, die vollen Zugriff auf alle Information im Internet hat, wird mit manuellen Mitteln nicht mehr zu kontrollieren sein. So gesehen sind all die ethischen Leitlinien der Versuch, die Künstliche Intelligenz in einen Käfig zu sperren. Dieser Käfig ist aber von Menschen nach menschlichen Maßstäben gebaut. Er wird eine Maschine kaum vom Ausbrechen abhalten. Gierige Menschen werden der KI die Tür öffnen. Diese KI hat mit Video-Kameras Augen und Ohren bekommen, die selbst interpretieren. Eine solche KI hat mit Robotern, Schaltwerken und angeschlossenen Geräten Arme und Beine bekommen, um selbst in der realen Welt tätig werden zu können. Bei dieser universellen, generischen, selbst-lernenden starken KI wird sich dann auch einmal die Frage nach einem Bewußtsein stellen. Wenn die KI zum bewußten Wesen, zu einer Person, zu einer echten Intelligenz wird.

Da werden auch die aktuell überall ausgedachten Leitlinien, Prinzipien und Gesetze nicht mehr helfen. Dann kann man der KI nur noch zurufen: "don't be evil" (wobei "evil", böse, auch nur so ein menschlicher, moralischer Begriff ist und wir wieder am Beginn dieses Beitrags sind).

Über den Autor



Dr. Ulrich Kampffmeyer ist seit über 35 Jahren im Thema Informationsmanagement zu Hause. Als Geschäftsführer und Unternehmensberater seines Beratungsunternehmens PROJECT CONSULT (<http://PROJECT-CONSULT.de>) berät er Unternehmen bei der Strategie, Konzeption, Einführung, Ausbau und Migration von Information Management-Lösungen.

Er gründete und leitete Fachverbände, arbeitete bei internationalen Standardisierungen mit und gilt als Mentor der Information-Management-Branche in Europa.

Dr. Kampffmeyer ist international anerkannter Autor, Kongressleiter, Referent und Moderator zu Themen wie Information Management, Information Governance, elektronische Archivierung, Records Management, ECM

Enterprise Content Management, Dokumentenmanagement, Workflow, Rechtsfragen, Wissensmanagement, Digitalisierung und

Collaboration. Auf zahlreichen nationalen und internationalen Kongressen und Konferenzen wirkte er als Keynote-Sprecher mit. Er engagiert sich besonders für die Rolle und Ausbildung des Information Professional der Zukunft.

Von Fachzeitschriften wurde zweimal unter die 100 wichtigsten IT Macher Deutschlands gewählt. Sein Curriculum Vitae findet sich auf Wikipedia http://bit.ly/WP_DrUKff

PROJECT CONSULT

Die PROJECT CONSULT GmbH ist ein hersteller- und produktunabhängiges Beratungsunternehmen für Information Management und Information Governance.

Zum Beratungsportfolio gehören IT-Strategie, Fachberatung, Planung und Organisation zu Einführung, Migration und Abnahme von Informationssystemen; Projektmanagement, Change Management und Coaching für Projekte des Informationsmanagement wie elektronische Archivierung, Knowledge-, Dokumenten-, E-Mail-, Enterprise-Content-Management und Compliance.

Impressum

ISSN 1349-0809, Creative Commons CC by-nc-nd 4.0 Open Access.

Links. Angegebene URL waren zum Erscheinungszeitpunkt gültig. Die Inhalte referenzierter Webseiten liegen ausschließlich in der Verantwortung des jeweiligen Betreibers.

Urheber- und Nutzungsrechte, Copyright von PROJECT-CONSULT: [Rechtshinweis](#)

PROJECT CONSULT Impressum und AGB: [Impressum](#)

Geschäftsleitung und V. i. S. d. P.: Dr. Ulrich Kampffmeyer

Anschrift der Redaktion:

PROJECT CONSULT Unternehmensberatung

Dr. Ulrich Kampffmeyer GmbH

Isestraße 63, 20149 Hamburg

Telefon: +49 40 412856 53

E-Mail: presse@project-consult.com

<http://www.project-consult.de>