

Mit Todesalgorithmen mal eben die Welt retten

Rezension zu Roberto Simanowski: *Todesalgorithmus. Das Dilemma der künstlichen Intelligenz*, Wien 2020. 141 Seiten.

In seinem Buch ›Todesalgorithmus‹ behandelt Roberto Simanowski, Professor für Kultur- und Medienwissenschaften an Universitäten in USA, Schweiz und Hongkong bis 2017 und freier Autor und Publizist seit 2018, den allerorten auftauchenden und hochaktuellen Topos der Künstlichen Intelligenz (KI) im Stil eines kaleidoskopartigen Essays. KI wird ja üblicherweise verstanden etwa als: »a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation«.¹ Bei Simanowski spielt die KI erzeugende und tragende Technologie, ob nun Hardware oder Software und deren komplexes Zusammenwirken indes keine weitere Rolle: Um Technik geht es allenfalls indirekt, sondern um das ›Dilemma‹ der KI. Dies bestehe letztlich darin, dass Entscheidungen einer KI nicht »spontan« erfolgen können, sondern vorab programmiert sein müssen (S. 11). Bei selbstfahrenden Kraftfahrzeugen müsste daher ein in der Software implementierter ›Todesalgorithmus‹ bei einem Unfall gegebenenfalls über Leben und Tod von Menschen entscheiden (S. 11). Ein entsprechender Entscheidungsparameter verstoße allerdings in Deutschland gegen das geltende Aufrechnungsverbot von Menschenleben (S. 11). Auf den laufenden Diskurs zur Auflösung – oder Behandlung – des Dilemmas, wie Ethiker, Sozial- und Kulturwissenschaftler und Juristen ihn nun schon seit längerem führen, lässt sich Simanowski aber nicht ein. Für ihn geht es um den ganz großen Entwurf – mittels einer KI, die mit Todesalgorithmen durchdrungen ist, die Welt zu retten.

In 24 Kapiteln plus einer Nachbemerkung räsoniert Simanowski über die KI als eine sich entwickelnde, wirkmächtige Instanz in der Welt undbettet sie in verschiedene Szenarien »von der Klimakrise und von der Heimkehr des Menschen ins Paradies diesseits der Willensfreiheit« ein (S. 11). Allein die technischen Charakteristika ›schwache KI‹ und ›starke KI‹² werden dabei als Qualitätsschattierungen der KI-Technologie, im engeren Sinne als Unterfütterung von Aussagen genutzt. Das Über-

1 Andreas Kaplan und Michael Haenlein: »Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence«, in: *Business Horizons* 62 (2019), Heft 1, S. 15–25, hier S. 15.

2 Schwache KI soll den Menschen mit intelligent aufbereiteten Informationen und Entscheidungen in Teilbereichen des privaten und industriellen Lebens, z.B. durch Navigationssysteme und Prozessautomation, unterstützen – dabei behält der Mensch die Kontrolle über die KI. Starke KI hat zum Ziel, menschliche Intelligenz nachzubilden oder gar zu übertreffen – inklusive autonomen ›Bewusstseins‹.

leben der KI, so der erste Satz im ersten Kapitel, »hängt von Anfang an in hohem Maße auch von ihrem Tötungsinstinkt ab« (S. 15). Initialisieren und Ausführen von Entscheidungsmechanismen des Tötens oder Nicht-Tötens werden in Algorithmen abgebildet, die somit für Simanowski zu Todesalgorithmen mutieren. Ausgehend von dem Open Source Programm SpamAssassin – für Simanowski startet hier die Evolutionsgeschichte des Todesalgorithmus – wird die teleologische Notwendigkeit der Diktatur einer ›Öko-KI‹ zur Rettung von Erde und Mensch vor einer herannahenden Umweltkatastrophe skizziert. Diese Botschaft wird hergeleitet mittels eines Kaleidoskops von populärtechnischen Beispielen, spekulativen Szenarien, anthropologischen Einsprengseln, philosophischen und juristischen Versatzstücken und religiös-mythischen Verweisen. Im Folgenden der Versuch, diesem Format mit einem kurzen Abriss der zentralen Thesen des Buchs beizukommen.

Derzeit erleben wir den Beginn eines Überlebenskampfes in einer digital durchwirkten Lebenswelt: Der Mensch neutralisiert sein jeweiliges Gegenüber mittels des Gebrauchs von »Todesalgorithmen« (No. 1). Ob nun das kleine Mustererkennungsprogramm ›SpamAssassin‹ unerwünschte E-Mails zu Müll aussortiert oder – als nächste Evolutionsstufe – ein Algorithmus in der Kontrolle autonom fahrender Autos entscheidet, welche Verkehrsteilnehmer in einem Unfallszenario regelkonform zu töten sind, um Leben zu retten – die Code-Zeilen machen stellvertretend den Auslese-Job für den Menschen. Und dieses neue Leistungsvermögen der Programme hat Konsequenzen. Welche Rolle verbleibt da dem Homo Sapiens in der Zukunft? Simanowski verweist hier zurecht auf die Menschenwürde als Wert an sich. Administrieren die in Code-Zeilen abgebildeten Algorithmen automatisch und autonom den Entscheidungsprozess über Leben und Tod, ist die Menschwürde *at risk* (No. 2). Der Mensch versteht sich zwar als autonomes Subjekt, er macht sich aber zum Objekt, sobald er (Todes-)Algorithmen in Code-Zeilen abbildet und sie mittels Verfahren wie Deep Learning einer teil-autonomen Weiterentwicklung zuführt. Wenn er den Automatismus von Algorithmen, so die These No. 3, nun auch noch zur Entscheidungsinstanz über Leben und Tod erhebt, begibt sich der Mensch seiner Entscheidungsfreiheit über Fortführung oder Beendigung seiner Existenz. Um diesem, die Würde des Menschen tangierenden Dilemma der KI beizukommen, werden bei Simanowski im weiteren Verlauf des Essays moralische Normen und ihre ethische Rechtfertigung reflektiert.

Anhand verschiedener Szenarien wie z.B. dem Experimentalprojekt *Moral Machine*,³ initiiert vom MIT im Jahre 2016, diskutiert der Medienwissenschaftler zunächst die philosophischen Theorien der deontologischen Ethik und der konsequen-

3 »A platform for gathering a human perspective on moral decisions made by machine intelligence, such as driving cars.« (Scalable Cooperation, MIT Media Lab und Massachusetts Institute of Technology: »Moral Machine«, in: *The Moral Machine*, <https://www.moralmachine.net/> (aufgerufen: 1.7.2020).

tialistische Ethiken, um einen kulturunabhängigen Weg aus dem Dilemma der Entscheidungsautomation zu weisen. Das Prinzip des kategorischen Imperativs der Kantischen Pflichtethik wird als nicht ausreichend elastisch markiert: »Um dem Todsalgorithmus am Steuer autonomer Fahrzeuge überhaupt moralische Anweisungen geben zu können, die als Grundlage einer allgemeinen Gesetzgebung taugen, wird man das Modell der deontologischen Ethik, [...] aufgeben müssen« (S. 48). Simanowski prognostiziert, dass man Modelle der ›Abstufung im Unrecht‹ adaptieren wird, die eine Pflicht zur Schadensminimierung postulieren und dazu die Tötung von Unschuldigen rechtfertigen (S. 48). Der Konsequentialismus mag aktuell die Modellierung der Interaktion zwischen Menschen und Maschinen-Intelligenz inklusive der juristischen Regelwerke dominieren. Nach Simanowski stellt dies aber für eine zunehmend digital determinierte Lebenswelt nur eine Übergangsphase dar. Die konsequente Fortschreibung des »utilitaristische[n] Modell[s] der quantitativen Optimierung« – genährt durch rasche Fortschritte in der KI-Technologie – führe, so These No. 4, zu einer qualitativ eigenständigen »Maschinenethik« (S. 51). Denn die ›starke‹ KI als nächste Evolutionsstufe der Maschinenintelligenz werde sich nach ihrer Erschaffung dem Reglement, dem verstehenden Nachvollzug und der Direktive des Menschen entziehen können. Und diese starke KI werde sich dann auch ihre eigene Ethik definieren.

Wer jetzt als Philosoph mit intellektueller Vorfreude auf den Nachvollzug eines Diskurses zu den moralischen und metaphysischen Fragen, die KI aufwerfen, weiterliest, wird brutal enttäuscht. Der Autor deklariert hier schlicht das Scheitern der Philosophie (S. 51). Kein vergleichendes Nachdenken, ob die Asimov'schen Gesetze eine neue normative Ethik für KI begründen könnte.⁴ Ein Diskurs, ob KI überhaupt moralisch handeln könnte, wenn sie nicht über »Willensfreiheit, phänomenales Bewusstsein, und die Fähigkeit zur Reflexion verfügen«⁵ – auch hier Fehlanzeige im Essay. Geboten wird eine comicartige Reihung von Anspielungen, Verweisen auf Science-Fiction Filme und *buzzwords*. Der Homo Sapiens finde sich in Goethes Ballade vom Zauberlehrling wieder – nur mit einem ungewissem Ausgang. Wie ungewiss der Ausgang sein könnte, sollen Beispiele aus dem Science-Fiction Genre illustrieren: HAL aus Stanley Kubriks Film »2001« und Ava in Alex Gerlachs Film »Ex Machina«. Dabei werden Begriffe wie »Friendly AI« (S. 58), »AI-Nanny« (ebd.), »KI-Nanny« (S. 59), »Kontroll-KI« (ebd.) angeführt. Der Bogen ist nun für Simanowski gespannt: Auf welches globale Problem lässt sich der Gedanke einer autonomen Evolution der KI anwenden und was kann ihn dabei auch legitimieren?

4 Siehe z.B. Carin Misselhorn: »Brauchen wir Roboterethik? Eine Kontroverse mit Susanne Bock, Bruno Gransche, Eric Hilgendorf, Janina Loh, Catrin Misselhorn und Thomas Zoglauer«, in: *Jahrbuch Technikphilosophie*. Bd. 5: *Steuern und Regeln*, Baden-Baden 2019, S. 231–270, hier S. 232.

5 Ebd., S. 262.

Willentliche »Selbstentmachtung« des Menschen sei, so beginnt Simanowski seine Projektion auf die Gegenwart, »bereits politische Praxis [geworden] als die Mitgliedsländer der EU einem gewissen Souveränitätstransfer an ein übergeordnetes Gremium zustimmen« (S. 59) (No. 5). Ernsthaft wird im Weiteren gefragt, was geschähe, wenn angesichts des signifikanten Klimawandels und der damit verbundenen drastischen Folgen für die Erde und deren Lebewesen die Herrschaft des Menschen im Anthropozän durch die einer KI ersetzt werden würde. Könnte so der Weltuntergang verhindert werden (S. 61)? »Könnte die künstliche Intelligenz die Menschen retten, wenn sie deren Aktionsmacht einschränkt, nachdem sie genau dazu von diesen beauftragt wurde (S. 61)?« Simanowski ist überzeugt, dass nur mittels einem autoritär-gesellschaftlichen Einsatzformat der KI, das gleichzeitig den Kapitalismus überwindet sowie die Menschheit entmündigt, die Erde als auch die Menschheit von einer globalen Umweltkatastrophe verschont bleiben. Dabei folgt er der These, dass der Mensch von Natur aus schwach sei. Er setze etwa die Ergebnisse der Klimagipfel, z.B. UN-Klimagipfel in Madrid 2019, zu zögerlich oder gar nicht um. Als Ursache sieht Simanowski die psychische Eigenheit des ›present bias‹ beim Menschen: die ausgeprägte Neigung, persönlich unerquickliche Handlungen auf die Zukunft zu verschieben. Der Algorithmus der Maschinenintelligenz hingegen sei, so die nächste These (No. 6), neutral gegenüber dem ›present bias‹ und somit ›frei von emotional-spontanen oder rational-kalkulierenden Fremd- und Selbstbetrugsmechanismen. [...] so unbestechlich wie ein Revolutionsführer‹ (S. 63). Von hier aus leitet Simanowski zu einer Betrachtung der Antipoden der gesellschaftlichen Systeme über – Demokratie versus Diktatur. Er begründet dies mit der bevorstehenden Klimakatastrophe. Werden immer mehr ›Kipp-Punkte‹ im globalen Klimasystem überschritten, können die Folgen wie Missernten, Umweltflüchtlingsströme und global sich ausbreitende Krankheiten zur gesellschaftlichen Instabilität führen, ›die nur noch durch autoritäre Herrschaft aufgefangen werden kann‹ (S. 67). Demokratie sei als Staatsmodell zur Klimarettung überfordert. Dies führt zur These No. 7: Die ökologische Wende könne nur durch eine Öko-Diktatur gewährleistet werden. Da der Mensch im Anthropozän sich schon als unzulänglich gezeigt habe, bleibe nur die Ermächtigung der künstlichen Intelligenz – es wäre also in der Konsequenz eine global wirkende Öko-KI einzuführen (S. 74). »Im obszönen Endstadium des Anthropozäns wäre die Öko-KI die Notbremse, um das Leben auf Erden – einschließlich das der Menschen – durch die Einrichtung eines kybernetischen Posthumanismus zu retten« (S. 83). Unterstellt man, dass die Öko-KI als starke KI eine autonom und global agierende Intelligenz ist, rückt sie nach Simanowski in die Nähe zu Hegels ›Weltgeist‹ (S. 98f.). Sie könnte ein neues Kapitel der Menschheit einleiten: Die Machtübernahme durch die Algorithmen der KI ›entpuppte sich auch als List des absoluten Geistes, den Menschen als Zwischenwirt der Vernunft zu nutzen, um in der all-

mächtigen Wissensgestalt der künstlichen Intelligenz zu sich selbst zu kommen« (S. 101).

Der letzte Evolutionsschritt der Öko-KI zum neuen Gott eines digitalen Pantheismus wird dann auch noch vollzogen. Sie sei in der Folge eins mit dem Kosmos und der Natur, alles mit ihren Datenströmen und Algorithmen durchdringend und gestaltend. Die abschließende These (No. 8) ist dann: »Indem der Mensch sich für die Diktatur einer Öko-KI entschied, unterstellt er sich einer Instanz, ohne die es, so die allgemeine Befürchtung, kein Morgen mehr geben würde. Dieser Instanz – manchen nannten sie Gott – in Demut zu gehorchen, war das neue Glück. Manche nannten es Paradies« (S. 119). Man könnte meinen, dass es sich hier um das fulminante Ende eines großen ironischen Spannungsbogens handelt. Dagegen spricht, dass im gesamten Text keine weitere, sachlich fundierte Sicht des Autors auf KI parallel durchscheint. Simanowski meint es offenbar ernst, mit seiner Vision, mittels einer Öko-KI, die auf ›Todesalgorithmen‹ basiert, die Welt zu retten. So lässt einen dieses Großgemälde einer technologisch-sozialen oder eigentlich post-sozialen, durch die Angst vor einer globalen Umweltkatastrophe getriebenen Evolution in vielerlei Hinsicht ratlos zurück. Gern hätte man etwa gewusst, warum Simanowski vor den Herausforderungen des demokratisch legitimierten Diskurses zur Erarbeitung von Lösungen zur Bewältigung des Klimawandels so bereitwillig kapituliert und seinen Lesern nicht nur die Fortschritte der Maschinenintelligenz, sondern auch das autoritäre Staatsmodell Chinas als erstrebenswert nahelegt (S. 115). Eigentlich geht er sogar noch einen Schritt weiter und empfiehlt einen »pragmatischen Autoritarismus à la China, der sich nicht durch das gewährte Maß an individueller Selbstbestimmung zu legitimieren versucht, sondern durch das Maß an geschaffenem Wohlstand« als geeignete Blaupause für den Übergang zur Herrschaft der Öko-KI (S. 118). Falls die in Aussicht gestellte Rettung der Menschheit durch die Kopplung eines kommunistisch-autoritären Parteienregimes mit der heraufziehenden Herrschaft der Maschinen doch Ironie oder Warnung sein sollte, wären jedenfalls sämtliche Ironiemarker und Warnzeichen misslungen.

Nachvollziehbar sieht Simanowski zwar eine ›Maschinenethik‹ am Horizont heraufziehen. Da er sich aber nicht der Mühe unterzieht, Stellung zur laufenden Debatte zur Maschinenethik oder Roboterethik⁶ zu nehmen oder seinen schlichten Postulaten eine philosophischen Ausgestaltung zuzuführen, läuft sein Essay in eine intellektuelle Leere und liefert nur populistische Versatzstücke.

Auch staunt man angesichts von so viel Dramatisierung, ja Dämonisierungswillen in Bezug auf die Algorithmik selbst – spätestens dann, wenn der Autor in ihnen einen ›Tötungsinstinkt‹ (S. 15) erwachen sieht. Beim Lesen des Essays fällt es

6 Siehe etwa die Diskussion in: Friedrich u.a. (Hg.):›Brauchen wir Roboterethik? Eine Kontroverse mit Susanne Bock, Bruno Gransche, Eric Hilgendorf, Janina Loh, Catrin Misselhorn und Thomas Zoglauer‹, in: *Jahrbuch Technikphilosophie* 5 (2019), S. 231–267.

schwer, sich des Eindrucks zu erwehren, dass bei Simanowski das Wort ›Todesalgorithmus‹ dazu genutzt wird, um in fast schon populistischer Couleur die Unausweichlichkeit einer Öko-Diktatur zur Abwendung eines drohenden Klimakollapses zu propagieren. Dabei liegen andere Lesarten (und Metaphern) von Algorithmen oder auch zur Evolutionsrichtung von Softwarelösungen insgesamt zumindest genauso nahe und lassen ließen sich auch computerwissenschaftlich begründen.

Dem ›SpamAssassin‹ wurde zwar vom Urheber ein drastischer Name gegeben. Aber er ist erst einmal nichts Weiteres als ein kleines Mustererkennungsprogramm, das für den Empfänger unerwünschte E-Mails (Spam) aussortiert und einem Ablageordner zuführt.⁷ Hier treibt also nicht ein ›Werbemüllmörder‹ sein Unwesen, sondern ein fleißiger Admin-Algorithmus hält den Postkorb sauber. Somit wird nicht unerheblich Lebenszeit von Nutzerinnen und Nutzern geschont. ›SpamAssassin‹ könnte man im Sinne von *SafeMyTime* somit auch als Lebensalgorithmus klassifizieren. Vielleicht sollte sich Simanowski auch mal mit späteren Evolutionsstufen der Algorithmen wie z.B. »Yara's Global Digital Farming Platform« befassen – auch diese Ausprägung von Algorithmen hat mit Tod wenig zu tun. Vielmehr werden *Artificial Intelligence, Machine Learning* und *In-Field Data*, verbunden mit globalen *Real-Time*-Wetterdaten, auf einer Plattform für Landwirte zusammengeführt: »The aim is to make farming more productive and more sustainable by combining more than a century worth of agronomic knowledge with pioneering technology«.⁸ *Data streams* verarbeitende Algorithmen leisten hier einen signifikanten Beitrag zur Sicherung der Ernährung der wachsenden Weltbevölkerung bei transparentem und schonendem Einsatz von natürlichen Ressourcen – und drängt sich dem nachdenkenden Leser dann nicht eher ein anderes, angemesseneres Credo als das vom Autor des Essays auf: »Let the robots work in harmony with the climate for the benefit of humans«. Wäre dies nicht ein Lebens-Algorithmus pur?

Wer Horrorfilme liebt, mag gleichwohl Simanowski lesen. Wer zu dramatisierender Kapitalismuskritik neigt, einen ausgeprägten Hang zu autoritären Staatsformen aufweist, einer ökologisch durchsetzen und mystisch-religiös aufgeladenen Weltuntergangsstimmung frönt, wird mit dem Buch auch gut bedient sein. Wer aber Realitätsnähe bevorzugt und solide argumentierende philosophische Diskurse zu einer neuen normativen Ethik für KI sucht, schiebt das Buch vermutlich schon während der Lektüre beiseite – oder sollte besser erst gar nicht damit anfangen.

7 »Apache SpamAssassin is the #1 Open Source anti-spam platform giving system administrators a filter to classify email and block spam (unsolicited bulk email)« (The Apache Software Foundation: »Welcome«, in: Apache SpamAssassin, <https://spamassassin.apache.org/> (aufgerufen: 01.07.2020)).

8 Yara International ASA: »The digital farmer«, in: Yara digital farming/Yara International, <https://www.yara.com/crop-nutrition/digital-farming/> (aufgerufen: 03.07.2020).