

Roberto Simanowski  
Todesalgorithmus

Das Dilemma der  
künstlichen Intelligenz

© Passagen Verlag

Passagen Thema  
herausgegeben von  
Peter Engelmann

Passagen Verlag

Deutsche Erstausgabe

Dieses Buch wurde ermöglicht durch die freundliche Unterstützung der Passagen Freunde – Freundeskreis des Passagen Verlags.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de/> abrufbar.

Alle Rechte vorbehalten  
ISBN 978-3-7092-0461-0  
2., durchgesehene Auflage 2021  
© 2020 by Passagen Verlag Ges. m. b. H., Wien  
Grafisches Konzept: Gregor Eichinger  
Satz: Passagen Verlag Ges. m. b. H., Wien  
<http://www.passagen.at>  
Druck: Ferdinand Berger & Söhne GmbH, 3580 Horn

# Inhalt

Vorbemerkung	11
Würde des Menschen	15
<i>SpamAssassin</i>	15
<i>Terminator 7</i>	17
<i>Lebenswertverrechnung</i>	19
<i>Weichensteller nach 9/11</i>	22
<i>Moral Machine</i>	25
<i>Tötungsregeln</i>	28
<i>Verfassungswidrigkeit</i>	33
<i>Opferlogik</i>	37
<i>Technikkultur</i>	42
<i>Sachordnungen</i>	48
Herrschaft der Maschinen	55
<i>HALs Enkel</i>	55
<i>Anthrobscene</i>	60
<i>KI-Diktatur</i>	65
<i>Asimovs viertes Gesetz</i>	74
<i>Posthumanismus</i>	83
List der Vernunft	89
<i>Hegel im Silicon Valley</i>	89
<i>Willensfreiheit</i>	91
<i>Geschäftsführer des Weltgeistes</i>	97

<i>Digitaler Pantheismus</i>	100
<i>Geist und Technik</i>	106
<i>Sozialingenieure</i>	113
Nachbemerkung	121
Preisrede zur Vergabe des Tractatus-Preises 2020	127
Anmerkungen	141
Editorische Notiz	157

## Vorbemerkung

Dieses Buch handelt vom Todesalgorithmus, von der Klimakrise und von der Heimkehr des Menschen ins Paradies diesseits der Willensfreiheit. Der Todesalgorithmus sitzt am Steuer autonomer Fahrzeuge und hat bei einem Unfall über Leben und Tod zu entscheiden. Ein Dilemma, weil die Entscheidung im Funktionsrahmen der künstlichen Intelligenz (KI) nicht mehr spontan erfolgen kann, sondern vorprogrammiert werden muss. Es muss vorab bestimmt werden, ob das Auto dem Kind oder den Rentnern ausweichen soll. Wie auch immer die Entscheidung ausfällt, sie verstößt gegen das in Deutschland und anderen Staaten geltende Aufrechnungsverbot von Menschenleben.

Während dieses ethische Dilemma aus den Erfolgen der KI-Forschung resultiert, ist die Klimakrise Resultat des vorangegangenen technischen Fortschritts: von den auf fossilen Brennstoffen basierenden Verbrennungsmotoren und Industrieanlagen bis zum Plastikmüll und Flugverkehr. Der Hoffnung, diese Krise mit smarterer Technik, „grünen Codes“ und einer CO<sub>2</sub>-Grenzausgleichs-Steuer zu lösen, steht die Ansicht gegenüber, dass es einer grundlegenden Veränderung des menschlichen Lebensstils im Kontext einer Postwachstumsgesellschaft bedarf. Gegen die ökologischen Notwendigkeiten sprechen ökonomische und politische Bedenken, die nicht zuletzt aus der

Kondition des modernen Menschen resultieren, sich und die Gegenwart den anderen und der Zukunft vorzuziehen. Der Mensch, so scheint es, kann das eigene Überleben nicht sichern – jedenfalls nicht, solange die Mängel der demokratischen Staatsform die der *Conditio Humana* verdoppeln. Kann die künstliche Intelligenz es, wenn sie stark genug ist und, vom Menschen dazu beauftragt, das Regime übernimmt: als besorgte Nanny oder gütiger Diktator? Weder die technische noch die politische Entwicklung sind reif für ein solches Szenario. Aber die Forschung zur KI läuft auf Hochtouren und der Zweifel an der Zukunftsfähigkeit der Demokratie motiviert inzwischen selbst Politologen zu ketzerischen Gedankenexperimenten.

Die Selbstentmachtung des Menschen durch sein eigenes Geschöpf wäre die Vollendung eines Aufbruchs, der mit dem Griff nach der Erkenntnis begann. Überlässt der Mensch die Regelung seiner Angelegenheiten der KI, wie es ansatzweise schon geschieht, wenn Apps ihn von A nach B navigieren und Algorithmen die Partnerwahl übernehmen, kehrt er im Grunde an den Anfang zurück: als er noch unfähig war, Gut und Böse zu unterscheiden. Diese Rückkehr enthebt ihn nicht nur der Notwendigkeit, selbst zu entscheiden, sondern auch der Gefahr, sich dabei zu irren. Ist es die Heimkehr ins Paradies? Wäre ein solches Paradies wünschenswert? Die Antwort hängt davon ab, worin man den Sinn des Lebens sieht und wieviel Autonomie, Nudging und Social-Scoring darin enthalten sein soll.

Die imaginierte Zukunft – das ist der gedankliche Ausgangspunkt dieses Essays – beginnt heute, mit dem autonomen Fahren. Der Todesalgorithmus am Lenkrad autonomer Fahrzeuge ist die Testfahrt für eine Gesell-

schaft, in der die KI das Steuer übernimmt. In den dystopischen Szenarien der Science-Fiction beherrscht oder vernichtet die KI den Menschen. In den Phantasien der Technikenthusiasten halten sich Menschen intelligente Maschinen als Sklaven oder transfigurieren selbst zur Maschine. Die hier entwickelte Perspektive bringt beide Pole zusammen: Die KI ergreift zwar die Macht über den Menschen, aber nur, um ihn vor dem selbstverschuldeten Untergang zu bewahren.

Das Buch entwickelt dieses Szenario mit philosophischer Neugier, ohne am aktuellen Technikstand zu haften. Er ist methodisch inspiriert durch das *What if*-Prinzip des spekulativen Designs, das zur Imagination von Zukunftsszenarien einlädt und mit entsprechenden Produktideen eine gesellschaftliche Debatte über bestehende Akzeptanzgrenzen anstoßen will.<sup>1</sup> Die Schreibmotivation ist allerdings nicht die Verschiebung von Akzeptanzgrenzen, sondern das Augenmerk auf diese: die Grenzen und die Verschiebung. Die dazu vorgenommene Spekulation erfolgt in eher konservativer Form; als klassischer Essay, von dem Adorno sagte: „In Freiheit denkt er zusammen, was sich zusammenfindet in dem frei gewählten Gegenstand.“ Das Resultat ist – dafür warb jüngst ein anderer Philosoph – ein „vagabundierendes Denken“, das „nicht auf das Stimmige, sondern auf das Stimulierende“ zielt – und weniger Recht haben will als gehört werden.<sup>2</sup>





## Würde des Menschen

### *SpamAssassin*

Das Überleben der digitalen Technologien hängt von Anfang an in hohem Maße auch von ihrem Tötungsinstinkt ab. Diese Gleichung gilt schon für die E-mailkommunikation, die nie erfolgreich gewesen wäre ohne die Lösung des Spam-Problems. Der berühmte Algorithmus, der hier früh für Ordnung sorgte, heißt „SpamAssassin“. Ein martialischer Name für einen Filter, der das Rauschen im neuen Kommunikationsmedium durch die Trennung der erwünschten von der unerwünschten Post reduzieren soll. Ein passender Name, denn am Ende geht es von Anfang an immer genau darum: Was überlebt und wer nicht.

Die Müllbeseitigung ist das erste Übungsfeld der künstlichen Intelligenz. Hier lernt die Software Muster zu erkennen und Objekte zu klassifizieren. Hier trifft sie Entscheidungen unter Aufsicht und schließlich auch ohne menschliche Kontrolle: erst in einer gleichen Situation, dann in einer ähnlichen. Alle aktuellen und künftigen Phantasien über intelligente Kühlschränke, autonome Autos, autarke Roboter und andere Formen der künstlichen Intelligenz, die heute die Öffentlichkeit begeistern oder beunruhigen, stammen vom Selbstlernmechanismus des Werbemüllmörders.<sup>1</sup>

Der jüngere Bruder des SpamAssassin ist der Todesalgorithmus, der in selbstfahrenden Autos im Notfall entscheidet, ob das Fahrzeug eher in eine Gruppe von Fußgängern oder auf ein Kind oder gegen eine Hauswand fährt. Über diesen Todesalgorithmus gibt es inzwischen hitzige philosophische und auch schon juristische Debatten. Denn als sicher gilt: Das autonome Auto kommt, und wahrscheinlich noch vor den autonomen Waffen und den mechanischen Haustieren. Die Entwicklung ist politisch gewollt, da man zu Recht davon ausgeht, dass autonom fahrende Autos nicht nur die Fahrkosten und den Energieverbrauch drastisch senken werden, sondern auch die Zahl der Unfälle. Immerhin verfügt der Bordcomputer über die Fahrerfahrung *aller* Computer, verarbeitet viel mehr Information viel schneller als der Mensch, wird niemals müde, fährt nie betrunken und textet nicht am Lenkrad. Es wäre moralisch unverantwortlich, selbstfahrende Autos nicht einzuführen.

Zugleich gilt als sicher, dass es weiterhin Unfallsituationen geben wird, bei denen Todesopfer nicht vermeidbar, sondern nur wählbar sind. Das führt zum ethischen Problem, die Algorithmen der selbstfahrenden Autos mit einer spezifischen Entscheidungsmoral ausstatten zu müssen, die unvermeidlich dem Grundprinzip der deutschen Verfassung widerspricht. Die Unantastbarkeit der Menschenwürde ist in Gefahr, wenn die Technik verlangt, Abwägungen, die sich philosophisch verbieten, für den Ernstfall zu programmieren.

## *Terminator 7*

In der wahrscheinlich berühmtesten Fake-Werbung unserer Zeit fährt ein Auto durch eine Landschaft, die deutlich in der Vergangenheit liegt. Alle Menschen halten in der Arbeit inne und schauen mit Ehrfurcht auf etwas, das sich schließlich als ein Mercedes der Luxusklasse erweist. Kurz darauf stoppt dieser Mercedes vor zwei Mädchen, die auf der Straße spielen. Der Bremsvorgang geht offenbar auf einen Sensor zurück, der Objekte vor dem Fahrzeug wahrnimmt. Als das Auto weiterfährt, nun unterlegt mit spannungsgeladener Musik, kommt ein rennender Junge mit einem Drachen ins Bild und eine junge Frau, die ihm beim Wäscheaufhängen glücklich hinterherschaut. Das Zusammentreffen des Jungen mit dem Mercedes überrascht nicht, wohl aber der Zusammenprall. Hat der Sensor diesmal versagt?

Die Antwort kommt rasant in kleinen Stücken: Kurz vor dem Aufprall sieht man für eine Millisekunde das Bild von Hitler, dann ruft die Frau mit dem Wäschekorb erschrocken „Adolf?“, auf dem Ortseingangsschild steht „Braunau am Inn“, aus der Vogelperspektive formen sich die Glieder des überfahrenen Jungen auf der Straße zu einem Hakenkreuz, der daraufhin eingeblendete Werbesatz für Mercedes' Bremssystem lautet: „Erkennt Gefahren, bevor sie entstehen“.<sup>2</sup>

Dieses Video aus dem Jahr 2013 hat im Frühjahr 2016 über 5 Millionen Views; mehr als 21 000 finden es gut, knapp 2 000 finden es schlecht, die rund 2 500 Kommentare verteilen sich entsprechend. Der Hinweis, dass es sich nicht um einen autorisierten Mercedes-Werbespot handelt, sondern um die Abschlussarbeit von Studenten der Filmakademie Ludwigsburg, verringert so wenig wie