

Thomas R. Köhler
Julia Finkeissen

Business 5.0

Der Praxis-Guide
für Künstliche
Intelligenz
in Unternehmen –
Chancen und Risiken



campus

Thomas R. Köhler
Julia Finkeissen

BUSINESS 5.0

Der Praxis-Guide
für Künstliche Intelligenz
in Unternehmen -
Chancen und Risiken

Campus Verlag
Frankfurt/New York

ISBN 978-3-593-51867-1 Print
ISBN 978-3-593-45719-2 E-Book (PDF)
ISBN 978-3-593-45718-5 E-Book (EPUB)

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlags unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Trotz sorgfältiger inhaltlicher Kontrolle übernehmen wir keine Haftung für die Inhalte externer Links. Für den Inhalt der verlinkten Seiten sind ausschließlich deren Betreiber verantwortlich.

Copyright © 2024. Alle deutschsprachigen Rechte bei Campus Verlag GmbH, Frankfurt am Main.

Umschlaggestaltung: Guido Klütsch, Köln

Umschlagmotiv: shutterstock/PST Vector

Satz: Publikations Atelier, Weiterstadt

Gesetzt aus der Sabon Next und der URW DIN

Druck und Bindung: Beltz Grafische Betriebe GmbH, Bad Langensalza
Beltz Grafische Betriebe ist ein klimaneutrales Unternehmen (ID 15985-2104-1001).

Printed in Germany

www.campus.de

INHALT

Vorwort	9
Business 5.0	13
1 KI aus der Unternehmensperspektive	15
Das Versprechen von KI	15
Der Hype in Zahlen	17
Was genau versteht man unter Künstlicher Intelligenz?	18
Generative KI	25
Formen von Künstlicher Intelligenz	28
2 Anwendungsfelder von Künstlicher Intelligenz in Unternehmen und Organisationen	39
Büroarbeiten	39
Kunden- und Mitarbeiterkommunikation	41
Einkauf	44
Human Resources	45
Marketing und Vertrieb	47
Unternehmensführung	48
Forschung und Entwicklung	52
Einzelhandel	53
Tourismus	56
Modeindustrie	60
Aus- und Weiterbildung	62
Medien und Unterhaltung	66
Finanzbranche – Banken	71
Finanzbranche – Versicherungen	72

Softwareentwicklung	73
Mobilität	79
Transport und Logistik	80
Immobilienwirtschaft	82
Unternehmensberatung	83
Medizin	85
Kunstmarkt	88
Produktionsbetriebe	93
Landwirtschaft und Viehzucht	94
Kanzleien und Rechtsabteilungen	94
Buchbranche	95
Kirche, Seelsorge und psychologische Betreuung	98
Polizeiarbeit und Strafverfolgung	100
Ausblick	100

3 KI und der Einfluss auf Arbeitswelt, Technologiewelt und individuelle

Lebenswirklichkeit	103
Wirtschaft und Arbeitsmarkt	103
Informationstechnologie	106
World Wide Web	109
KI in unserer Lebenswirklichkeit	112

4 Risiken und Gefahren 115

Grundlegende Gefahren durch generative KI	115
Falsche und irreführende Ergebnisse	119
KI-Halluzinationen	122
Bias	125
Problematische Herkunft der Daten	126
Ausbeutung von Mindestlohnmitarbeitern	133
Copyright-Fragen	135
Datenschutzfragen	140
Fake-Inhalte	142
Propaganda und Täuschung	145
Schädliche Inhalte und Hassrede	145
Manipulation von Verhalten	147
Abfluss von Unternehmensdaten	149

Modellkollaps	150
KI und das Ende der Menschheit	151
Big Tech und die KI-Debatte – warum wir den Technologieanbietern nicht das Feld überlassen dürfen	154
Cybersicherheit und KI	155
5 KI und die Nachhaltigkeit – eine schwierige Beziehung	157
Nachhaltigkeit in der Digitalisierung	157
Nachhaltigkeit und KI	160
Wie die Versöhnung von KI und Klimaschutzanforderungen gelingen kann	162
6 Widerstand gegen KI	165
7 Ethische Fragestellungen der KI – ein Leitfaden für Unternehmen	171
Ethische Entwicklungsprinzipien	172
Voreingenommenheit und Diskriminierung in Künstlicher Intelligenz	173
Die Grenzen von ChatGPT und aktuellere Alternativen für den Einsatz in Unternehmen	176
Greift KI ethisch in unser Leben ein?	178
Ethische Grundsätze für verantwortungsvolle KI	179
Die Herausforderung der Haftungsregelung für Künstliche Intelligenz	180
Ethische Implikationen von KI-gesteuerter Emotions- und Verhaltensmodulation	181
KI-induzierter Jobwandel: Braucht es ein Recht auf Arbeit?	182
Risikomanagement für verantwortungsvolle KI in Unternehmen	183
8 Sieben Schritte zum KI-Einsatz im Unternehmen	187
KI-Wertschöpfungskette	187
KI-Einsatz im Unternehmen – die Vorbereitung	190
Schritt 1: Identifikation möglicher Anwendungsfelder und Definition der Erfolgskriterien	191
Schritt 2: Markt- und Wettbewerbsanalyse	193

Schritt 3: Entscheidung Make or Buy?	193
Schritt 4: Integration in die eigene Organisationsstruktur	200
Schritt 5: Auswahl und Training des KI-Modells	204
Schritt 6: Implementierung in den Live-Betrieb	205
Schritt 7: Erfolgsmessung	205
Investments in KI-Unternehmen als Alternative zum KI-Projekt?	206
Umgang mit der »Schatten-KI«	208
9 Ausblick - das zukünftige Miteinander von Mensch und KI im Unternehmen	213
KI und der Hype Cycle	213
The map is not the territory	218
Final Thoughts	219
Glossar	221
Die Autoren	237
Anmerkungen	239

VORWORT

Der Begriff KI, Künstliche Intelligenz, ist in aller Munde – mal wieder. Seit den 1960er Jahren wird uns von der Technologiebranche versprochen, dass der große Durchbruch für die gerne auch nach dem internationalen Begriff AI, Artificial Intelligence, benannte Technologie, oder besser Gruppe von Technologien, kurz bevorstünde.

In der letzten KI-Welle sollten virtuelle Sprachassistenten wie Siri, Google Assistant oder Amazon Alexa unsere Welt revolutionieren. Um bei diesem Boom dabei zu sein, gaben Unternehmen Millionen aus. Das reichte von sogenannten Skills, um bei Diensten Dritter wie Amazon Alexa mitwirken zu können, bis hin zum eigenen Sprachassistenten »Hey Mercedes«, der den Fahrern der Marke Antworten auf Fragen wie »Hey Mercedes, was ist der Sinn des Lebens?« verspricht.¹ Lachen Sie nicht, das wird tatsächlich von Mercedes-Benz als Beispielanwendung genannt, ebenso wie die Ankündigung, man könne in Zukunft darüber die heimische Kaffeemaschine steuern.²

Der Erfolg all dieser Initiativen rund um Künstliche Intelligenz blieb bis dato – vorsichtig gesagt – überschaubar. Microsofts erst 2014 gestarteter Sprachassistent »Cortana« wurde sogar für Ende 2023 »abgekündigt«, wie die Technologiebranche das Ableben eines Produkts oder Dienstangebots euphemistisch bezeichnet. Für Amazon ist Alexa ein dauerhaftes Verlustgeschäft – von rund 10 Milliarden Verlust allein im Jahr 2022 war bereits die Rede.³ Noch hält man bei dem E-Commerce-Riesen daran fest – doch wie lange noch angesichts dieser Zahlen?

Jahrzehntelang wurden immer wieder größere Durchbrüche bei den Anwendungspotenzialen Künstlicher Intelligenz versprochen, die aber bei Licht betrachtet stets weit weniger glamourös und nutzbringend waren als angekündigt. Da stellt sich die Frage, warum ausgerechnet jetzt plötzlich alles

anders sein soll. Sie ist berechtigt, denn der Hype rund um die aktuelle Iteration von Künstlicher Intelligenz – sogenannter generativer KI – treibt seltene Blüten. Renommierete Beratungsunternehmen sehen hier »Trillionen Dollar Wachstum«, während TV-Philosophen wie Brancheninsider die Gefahren der neuen Technologie beschwören, einen Entwicklungsstopp fordern und sogar vor dem möglichen Ende der Menschheit durch eine potenziell übergriffig werdende KI warnen.

Die erste weithin bekannt gewordene Inkarnation dieser neuen KI-Generation sieht indes gar nicht nach einer bedrohlichen Waffe aus. Sie hört auf den unscheinbaren Namen ChatGPT und liefert jede Menge praktischen Nutzen für jeden, der sich – trotz der vielfältigen Warnungen – selbst an sie herantraut. Über 100 Millionen Menschen haben binnen der ersten Monate nach dem Start im Spätherbst 2022 ChatGPT bereits ausprobiert und sind vielfach dageblieben, trotz Anmeldezwang und eher drögem User-Interface, denn es gibt nichts zu klicken oder zu wischen. Die Nutzerinteraktion findet wie in der Steinzeit der Computerei per Texteingabe statt – Buchstabe für Buchstabe werden Fragen und Befehle an das System formuliert. Ausgegeben wird dann wiederum Text, bei anderen Systemen auch Musik, Bild und Bewegtbild. Das ist archaisch, aber funktional.

Diese Technologie ist so nutzbringend, dass nicht nur Schüler und Studierende damit ihre Hausarbeiten beschleunigen, sondern auch Softwareentwickler nach Fehlern in ihren Programmen suchen, Berater Präsentationen erstellen und Architekten Ideen generieren, ohne dass ihnen jeweils jemand genau gesagt hätte, wie das geht. Die Anwendungsmöglichkeiten sind vielfältig, ein potenziell nutzbringender Einsatz im eigenen Unternehmen scheint also geboten. Doch wo soll man anfangen? Wo finden sich relevante Informationen?

Eigentlich fehlt es ja an nichts. Technische Details und Beschreibungen finden sich ebenso im Internet wie die Studien der großen Unternehmensberatungen, die das Thema längst »auf die CEO-Agenda« gesetzt haben. Aber Techniker verlieren sich gerne in Details, Berater sehen nur die nächste »Opportunity« und liefern »high level talk« für die Vorstandsetage ohne konkrete Ansatzpunkte und alle anderen scheinen geradezu besessen von den Weltuntergangsfantasien einer potenziell immer intelligenter werdenden KI.

Einen anderen Anfang wagen wir nun in diesem Buch aufgrund unserer Erfahrungen aus zahlreichen Technologieinnovationen: hart an den An-

wendungen, mit einem realistischen und bewertenden Blick auf Chancen und Risiken und so viel Theorie wie nötig. Es ist ein Handbuch aus der Praxis für die Praxis. Lassen Sie uns nach Ihrer Lektüre gerne wissen, wie weit dies gelungen ist.

BUSINESS 5.0

Warum Business 5.0? Warum noch ein Begriff in einer an neuen Wortschöpfungen nicht gerade armen Welt? Wir sind davon überzeugt, dass sich mit der absehbaren breiten Adaption von generativer KI – quer über alle Unternehmensfunktionen hinweg – viele Spielregeln im Wettbewerb ändern werden. Wer automatisiert, der gewinnt, das galt vorher schon in vielen Branchen und bei vielen Professionen – von der Logistik bis zum Banking. Auf der Strecke blieben dabei aber häufig die Individualität – bei vielen Kunden blieb und bleibt das Gefühl, nicht mehr als eine Nummer zu sein – und die Möglichkeit zur Adaption an veränderte Rahmenbedingungen.

Generative KI ändert das. Sie verspricht einen hohen Automatisierungsgrad bei gleichzeitig hoher Variabilität. Wir werden im Laufe dieses Buches zeigen, wie dies gelingen kann und warum generative KI dafür die Schlüsseltechnologie darstellt.

Die Bezeichnung »Business 5.0« bezieht sich auf die Entwicklungsphasen der industriellen Revolution. Der Begriff »Industrie 4.0« dürfte den meisten Lesern bekannt sein. Nur zur Erinnerung: Die erste industrielle Revolution, das Entstehen dessen, was wir heute als Urform der Industrialisierung in Erinnerung haben, ist eng verbunden mit der Mechanisierung und der Nutzung von Wasserkraft und Dampfmaschinen. Die zweite Phase ordnet man dem Thema Elektrifizierung der Fabriken und Einführung des Fließbands zu als Start dessen, was wir bis heute mit Massenproduktion verbinden. Mit der Automatisierung und dem Einzug von Elektronik und Informationstechnologien wurde in den 1970er Jahren die dritte industrielle Revolution eingeleitet, während wir heute mit Industrie 4.0 über eine vernetzte Produktion und das Internet der Dinge sprechen. Der schöne, aber außerhalb der akademischen Welt wenig verstandene Begriff des Cyber-Physical-Systems stammt daher. Gemeint ist damit die Vernetzung von Anlagen und Systemen, und

diese findet eine Fortsetzung in dem, was derzeit – unter anderem vonseiten der EU – unter dem Begriff »Industrie 5.0« propagiert wird.

Dort heißt es dazu unter anderem: »Sie [Industrie 5.0] ergänzt das bestehende Konzept der »Industrie 4.0«, indem sie Forschung und Innovation gezielt in den Dienst des Übergangs zu einer nachhaltigen, auf den Menschen ausgerichteten und resilienteren europäischen Industrie stellt.«¹ Sie zielt dabei insbesondere auf die »Anwendung eines auf den Menschen ausgerichteten Ansatzes für digitale Technologien, einschließlich künstlicher Intelligenz.«² Geschehen soll dies unter anderem durch eine bessere Mensch-Maschine-Kollaboration, beispielsweise mit »intelligenten« Robotern, und mit dem Ziel einer Individualisierung der Produktion unter dem Titel »Mass Customization«. Dieser Begriff spukt seit Jahrzehnten durch die Wirtschaftswissenschaften, doch der Traum von der Hyperindividualisierung kam bis dato noch nicht so richtig zum Fliegen.

Wir sehen »Business 5.0« als Parallelbegriff zur gerade diskutierten »Industrie 5.0«. Beide versprechen eine bessere Zusammenarbeit von Mensch und Technologie sowie eine hohe Individualität trotz Massenangang. Nur ist »Business 5.0« eben gerade nicht auf die Industrieproduktion beschränkt, sondern quer durch alle denkbaren Branchen vertreten sowie in zahlreichen betrieblichen Funktionen, ganz unabhängig von der Branche.

Eine durch generative KI erstellte Werbemail wird eben nicht 10 000-mal in der gleichen Form versendet, sondern hochindividuell in »Stückzahl 1« und das 10 000-mal. Neu ist, dass das ohne signifikanten Mehraufwand geht. So lautet zumindest das zentrale Versprechen der neuen Technologiegeneration, die wir im Folgenden mit Blick auf ihre Auswirkungen und Anwendungsmöglichkeiten diskutieren wollen und für deren Nutzung wir im Kapitel 8 entsprechende Guidelines entwickelt haben.

1 KI AUS DER UNTERNEHMENSPERSPEKTIVE

Das Versprechen von KI

Es wird die neue »Universalmaschine«. Ob Text, Konzept, Musikstück, Programmieraufgabe oder Bild – generative KI kann alles, also von allem zumindest ein bisschen. Ein Bild vom Papst im Daunenmantel, ein Gedicht im Stil von Schillers »Lied von der Glocke«, ein Konzept für den Markteintritt in Kasachstan? Kein Problem. Entwickler, Investoren und Analysten erwarten gleichermaßen enorme Umbrüche durch generative KI in den nächsten Jahren.

Fachleute sind sich einig, dass die Anwendungspotenziale noch nicht vollständig erschlossen sind. Im Gegenteil, wir kratzen nur an der Oberfläche und finden mit jeder Nutzung immer mehr an Potenzial. Das weckt Begeisterung und führt zu einer Vielzahl von optimistischen Prognosen über die wirtschaftlichen Potenziale. Die Wirtschaftsprüfungsgesellschaft PWC hebt sich in ihrer *Global Artificial Intelligence Study* besonders mit ihrer Prognose hervor, die den potenziellen Wertbeitrag von KI bei bis zu 15,7 Billionen US-Dollar sieht.¹ Die größten wirtschaftlichen Gewinne durch KI werden – nach PWC – China (26 Prozent Steigerung des Bruttoinlandsprodukts im Jahr 2030) und Nordamerika (14,5 Prozent Steigerung) erzielen, was insgesamt 10,7 Billionen US-Dollar entspricht und fast 70 Prozent der globalen wirtschaftlichen Auswirkungen ausmacht. Die Steigerung für Europa wird mit vergleichbar überschaubaren 9,9 Prozent des Bruttoinlandsproduktes erwartet.²

Interessant ist in diesem Zusammenhang auch eine frühere Studie der Unternehmensberatung BCG aus dem Jahr 2021, nach der bei bisherigen KI-Projekten nur 11 Prozent der Unternehmen von erheblichen finanziellen Vorteilen durch Umsatz-/Kostenverbesserungen aufgrund der Implementierung von KI berichten.³

Generativer KI wird nicht nur in diesem Buch erheblich mehr an wirtschaftlichem Potenzial zugetraut.

Unter der Vielzahl positiver Prognosen rund um den aktuellen Hype um generative KI sticht die von Sequoia Capital ins Auge. Bis 2030 – das sind keine sechs Jahre mehr – erwartet das bekannte Venture-Capital-Unternehmen Folgendes: »Die Bereiche, auf die generative KI abzielt – Wissensarbeit und kreative Arbeit –, umfassen Milliarden von Arbeitnehmern. Generative KI kann diese Arbeitskräfte um mindestens 10 Prozent effizienter und/oder kreativer machen: Sie werden nicht nur schneller und effizienter, sondern auch fähiger als zuvor. Daher hat generative KI das Potenzial, Billionen von Dollar an wirtschaftlichem Wert zu generieren.«⁴

Die Vorhersagen von Sequoia Capital für generative KI für das Jahr 2030 umfassen unter anderem:

- **Text:** Die endgültigen Entwürfe werden besser sein als die von professionellen Autoren.
- **Programmierung:** Sie wird besser sein als die eines Vollzeitentwicklers.
- **Bilder:** Sie werden besser sein als die von professionellen Künstlern und Designern.
- **Video-/3D-Spiele:** Videospiele werden im Wesentlichen »personalisierte Träume« sein.

Die bekannte Analystenfirma Gartner setzt Generative KI nicht nur – wie erwartet – auf ihre Liste der wichtigsten disruptiven Technologien, sondern konkretisiert die eigenen Prognosen unter anderem so:⁵

- Bis 2025 wird generative KI 10 Prozent aller Daten produzieren (heute sind es weniger als 1 Prozent) und 20 Prozent aller Testdaten für verbraucher-nahe Anwendungsfälle.
- Bis 2025 wird generative KI in 50 Prozent der Initiativen zur Entdeckung und Entwicklung von Medikamenten zum Einsatz kommen.
- Bis 2027 werden 30 Prozent der Hersteller generative KI nutzen, um die Effektivität ihrer Produktentwicklung zu verbessern.

Das Unternehmen vergisst an anderer Stelle nicht zu erwähnen, dass das Thema auch Erwartungen an Unternehmen stellt. Investoren und Aktionäre

etwa werden Wachstum und Margenverbesserung erwarten. Kunden, die generative KI in ihrem täglichen Leben nutzen, werden von den Unternehmen erwarten, dass sie das auch tun. Auch vonseiten der Mitarbeiter wird es Erwartungen geben. Sie werden das Unternehmen verlassen, wenn sie Arbeiten manuell erledigen müssen, die auch AI erledigen kann.⁶

Anders gesagt: Der Ball liegt in Ihrem Spielfeld. Generative KI wird oder ist bereits eines der wichtigsten Themen des Jahrzehnts, wenn es um die Zukunftsfähigkeit des eigenen Unternehmens geht.

Der Hype in Zahlen

Follow the money – Venture-Capital-Finanzierung ist ein bewegliches Ziel. Die Erfahrungen der letzten Jahrzehnte lehren: Nicht nur die gesamtwirtschaftliche Situation und damit die Möglichkeit der Investoren, Gelder für Investments einzusammeln, bestimmen den Markt, sondern vor allen Dingen die aktuellen Hype-Themen. Nach E-Commerce, Blockchain, Crypto und weiteren fokussieren sich viele Investoren nun aufgrund der hohen Ertragsaussichten auf KI-Anwendungen. Die einzige im Vergleich zu den US-Zahlen nennenswerte Finanzierung in Deutschland hat das Heidelberger Unternehmen Aleph Alpha erhalten: 100 Millionen Euro.⁷

2023 Schätzungen über das zukünftige Marktvolumen oder andere Messgrößen ökonomischer Auswirkungen von Generativer KI abgeben zu wollen, mutet aus unserer Sicht wie Kaffeesatzlesen an. Wer hatte 1995 – das Jahr, in dem ein einfacher Internetzugang für Endkunden in Deutschland verfügbar wurde – schon Vorstellungen davon, wie sich »das Netz« auf unsere Zukunft auswirken würde? Heute – fast 30 Jahre später – haben wir Erfahrung damit. Viele der Prognosen aus der Anfangszeit sind rückblickend betrachtet mehr als lächerlich. Mit ähnlicher Unsicherheit müssen wir auch heute leben, wenn wir uns aus der Unternehmensperspektive mit dem Technologietrend der Stunde – generativer KI – beschäftigen.

Was genau versteht man unter Künstlicher Intelligenz?

Die Idee von »denkenden Maschinen« ist nicht wirklich neu und reicht in ihren Anfängen bereits Jahrhunderte zurück. Im Wien des späten 18. Jahrhunderts machte eine Erfindung von sich reden, die als »mechanischer Türke« bekannt wurde. 1769 konstruierte der österreichisch-ungarische Hofbeamte und Mechaniker Wolfgang von Kempelen einen angeblichen Schachroboter. Der Erbauer ließ dabei den Eindruck entstehen, dass das Gerät selbstständig Schach spielen könne. Ein Tisch und eine mechanische Figur hinter dem Tisch, die sich bewegte, über das Schachbrett schaute und dabei mechanische Geräusche machte, überraschten die Betrachter der damaligen Zeit. Doch diese Figur – der angebliche Roboter – war, wie wir heute vielleicht sagen würden, »getürkt«.⁸

Tatsächlich war in dem Tisch ein Mensch versteckt, der das Schachspiel bediente. Wolfgang von Kempelen lehnte es stets ab, einen Hinweis auf das zugrunde liegende Funktionsprinzip zu geben. Über eine mögliche magnetische Übertragung der Zugbefehle wurde seitens der Betrachter ebenso gerätselt wie über die Möglichkeit, die Maschine könne eigenständig ohne jede menschliche Einwirkung die Berechnungen ausführen. Erst viele Jahre später wurde der Betrug aufgedeckt. Doch der Anschein einer denkenden Maschine war da bereits in der Welt.

Der britische Mathematikprofessor Charles Babbage entwarf mit der Analytical Engine eine mechanische Rechenmaschine für allgemeine Anwendungen und schuf damit wesentliche Grundlagen für die Geschichte des Computers als Universalmaschine, auf der heute auch AI-Programme zum Laufen kommen. Die Beschreibung veröffentlichte er 1837. Die ursprüngliche Idee konnte er bis zu seinem Lebensende nicht realisieren, aber es gilt als sicher, dass das System mit den vorgeschlagenen 55 000 Teilen, der Steuerung per Lochkarten und einem Antrieb per Dampfmaschine funktioniert hätte.⁹

Es sollte noch gut 100 Jahre dauern, bis Konrad Zuse 1941 tatsächlich einen funktionalen Rechner baute. Eng verbunden ist Babbages Erfindung mit Ada Lovelace, deren Konzept zur Steuerung der Analytical Engine heute als erstes Computerprogramm gilt – auch wenn es damals noch nicht realisierbar war. »Sie war davon überzeugt, dass eine Maschine eines Tages dafür genutzt werden könne, komplexe Musikstücke zu komponieren und Graphi-

ken zu erstellen, und sowohl wissenschaftliche wie praktische Anwendung finden werde. Sie sollte Recht behalten.«¹⁰

In seinem 1872 erschienenen Roman *Erewhon* entwickelt der englische Schriftsteller Samuel Butler in drei Kapiteln, »The Book of the Machines«, die Idee einer Evolution der Maschinen. Diese könnten – so Butler in seinem Roman – irgendwann ein eigenes Bewusstsein entwickeln und dem Menschen schließlich überlegen sein. Ebenso taucht darin die Idee der sich selbst replizierenden Maschinen auf. Auch wenn das Buch im England des 19. Jahrhunderts eher als Verunglimpfung der zu dieser Zeit viel diskutierten Thesen von Charles Darwins Evolutionstheorie oder schlicht als Satire gesehen wurde, hat es beinahe hellseherische Kraft. Samuel Butler nahm damit viel von der Diskussion um Künstliche Intelligenz, wie wir sie heute führen, vorweg.

Samuel Butler schrieb in seinem Vorwort zur zweiten Auflage selbst: »Ich bedaure, dass Rezensenten in einigen Fällen geneigt waren, die Kapitel über Maschinen als einen Versuch zu betrachten, die Theorie von Herrn Darwin ad absurdum zu führen. Nichts könnte weiter von meiner Absicht entfernt sein, und nur wenige Dinge wären mir unangenehmer als jeder Versuch, Herrn Darwin auszulachen.«¹¹

Im 20. Jahrhundert entwickelte sich schließlich die Vorstellung vom Roboter. Erstmals zu finden war dieser Begriff als »robot« in einem 1920 erschienenen Drama des tschechischen Schriftstellers Karel Čapek. *R.U.R. – Rossum's Universal Robots* ist nicht nur der Titel des Werkes, sondern auch die Bezeichnung für das Unternehmen im Buch, das die künstlichen Menschen erzeugt. Die »Robots« werden als billige und rechtlose Arbeiter verwendet. Ihr massiver Einsatz in der Industrie verändert mit der Zeit die gesamte Weltwirtschaft. Im weiteren Verlauf des Theaterstücks rebellieren die Kunstmenschen jedoch und vernichten die Menschheit.

Der Name »Rossum« ist eine ironische Anspielung des Autors: Das tschechische Wort *rozum* bedeutet Vernunft, Verstand.¹²

Isaac Asimov, unter anderem Autor von Science-Fiction-Romanen wie *Der Aufbruch zu den Sternen* und *Das galaktische Imperium*, gibt 1942 Regeln für das Verhalten dieser Roboter. In einer ersten Fassung lauten diese Asimov'schen Gesetze:

- Ein Roboter darf kein menschliches Wesen verletzen oder durch Untätigkeit zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.

- Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen – es sei denn, ein solcher Befehl würde mit Regel eins kollidieren.
- Ein Roboter muss seine Existenz beschützen, solange dieser Schutz nicht mit Regel eins oder zwei kollidiert.

Später ergänzte er diese drei Regeln noch mit einer »Regel 0«, die er den anderen voranstellte:

- »Ein Roboter darf die Menschheit nicht verletzen oder durch Passivität zulassen, dass die Menschheit zu Schaden kommt.«¹³

Es ist leicht zu erkennen, dass diese einfachen Regeln – je nach Anwendungsfall – nicht ganz widerspruchsfrei, aber vor allen Dingen auch unvollständig sind. Damit hat Asimov einer weiten ethischen Debatte das Feld geöffnet. Was, wenn der Schutz der Menschheit nur erreicht werden kann, wenn man einen einzelnen Menschen umbringt (etwa bei einem Flugzeugentführer, der droht, das Flugzeug mit den Passagieren abstürzen zu lassen), oder was ist, wenn die Menschheit nur zu retten ist, wenn die Roboter gegen den Willen der Menschen das Kommando übernehmen. Im Extremfall kann eine enge Auslegung dieser Robotergesetze auch bedeuten, dass ein Roboter nun verhindert, dass ein Mensch eine Straße überquert, denn es bestünde dabei ja ein Risiko, das größer Null ist, dass dieser dabei überfahren wird, und der Roboter darf ja laut »Regel 0« durch Passivität nicht zulassen, dass ein Mensch zu Schaden kommt.

Diese Gedanken mögen hier wie Wortklauberei oder Theoriedebatten wirken, sie sind mit Blick über die weiter unten im Buch geführte Debatte über die Anwendungen und Implikationen von KI aber wesentlich.

Wenige Jahre später – 1950 – wirft dann der britische Mathematiker und Informatiker Alan Turing in einem Beitrag für die Philosophiezeitschrift *Mind* unter dem Titel »Computing Machinery and Intelligence« die Frage auf, ob Maschinen denken können. Er schlug den sogenannten Turing-Test als Entscheidungskriterium vor, um festzustellen, ob eine Maschine dem Menschen vergleichbar denkfähig ist. »Da der Denkvorgang nicht formalisierbar ist, betrachtet der Test nur die Antworten einer Maschine im Dialog mit einem Menschen, d. h. das kommunikative Verhalten der Maschine. Wenn dieses von einem menschlichen Verhalten nicht unterscheidbar er-

scheint, soll von maschineller Intelligenz gesprochen werden.«¹⁴ Alan Turing wirkt bis heute durch diese Veröffentlichung auf die Entwicklung der Forschungen rund um Künstliche Intelligenz.

Ebenfalls in den 1950er Jahren entstand schließlich der Begriff »Künstliche Intelligenz«. Er wurde vom Computerwissenschaftler John McCarthy für einen Workshop am MIT (Massachusetts Institute of Technology) 1955 geprägt. Zusammen mit den Informatikern Claude Shannon, Marvin Minsky und Nathaniel Rochester entwickelte er die Idee: »Jeder Aspekt des Lernens und andere Eigenschaften von Intelligenz können im Prinzip so präzise beschrieben werden, dass eine Maschine sie simulieren kann.« Den Durchbruch für Künstliche Intelligenz prognostizierte McCarthy in »fünf bis 500 Jahren«.¹⁵

Von diesen theoretischen Vorstellungen von Künstlicher Intelligenz bis zur ersten Realisierung eines zumindest in Grundzügen selbstlernenden Systems dauerte es dann nur noch knapp drei Jahre. 1958 veröffentlichte der US-amerikanische Psychologe und Informatiker Frank Rosenblatt ein Forschungspaper in der Zeitschrift *Psychological Review* unter dem Titel: »The Perceptron: A probabilistic model for information storage and organization in the brain«. Damit war er Wegbereiter für die im Bereich KI so wichtige Musterklassifikation. Das »Perceptron« war für ihn eine Art Modell für Gehirnfunktionen. Der Schlüsselsatz in seinem Beitrag lautete: »Durch das Studium von Systemen wie dem Perzeptron erhofft man sich, die grundlegenden Organisationsgesetze zu verstehen, die allen informationsverarbeitenden Systemen, Maschinen und Menschen eingeschlossen, gemeinsam sind.«¹⁶

Der Hintergrund: Eine der von Rosenblatt entwickelten Anwendungen konnte Markierungen auf der rechten und linken Seite von Lochkarten – einem für damalige frühe Computersysteme typischen Eingabemedium – erkennen. Die eigentliche Innovation war, dass das System selbst lernte. Diese Entdeckung oder besser Erfindung sorgte für einiges Aufsehen. Die *New York Times* schrieb darüber: »In der Tat erscheint es uns als der erste ernsthafte Konkurrent des menschlichen Gehirns, der je erfunden wurde.«¹⁷ In gewisser Weise kann man diese Pressemeldung als einen Vorläufer des aktuellen KI-Hypes sehen.

Mit dem bis heute viel diskutierten Film *2001 – Odyssee im Weltraum* von Stanley Kubrick geht die Vorstellung von AI 1968 in den Mainstream

über. Der heimliche Star des Films ist der Bordcomputer des Raumschiffs »HAL 9000«, der – außer mit Sprache und einem roten Leuchten oder Auge – nicht in Erscheinung tritt. »I'm sorry, Dave. I'm afraid I can't do that!« ist der Schlüsselsatz des Films und verweist gleichzeitig auf die aktuelle Debatte rund um mögliche dystopische Folgen der weiteren Entwicklung Künstlicher Intelligenz.

Marvin Minsky – bereits 1955 im Team rund um John McCarthy beim MIT-Workshop dabei – publizierte 1974, also fast 20 Jahre später, seinen Forschungsbericht »A Framework for Representing Knowledge«¹⁸ und liefert damit die Grundlage für die für KI wichtige Frame-Theorie.

1997 – gerade zum Beginn der Interneteuphorie – macht dann IBM Schlagzeilen mit »Deep Blue AI« und seinen Fähigkeiten beim Schachspielen. Von dort an ist es nicht mehr weit zu AI in Anwendungen für Endverbraucher, 2002 etwa beim Roboter-Staubsauger Roomba, der für seine Navigation in Wohnungen einfache AI-Modelle nutzt, oder die virtuelle Assistentin Apple Siri, die 2011 mit dem iPhone 4s eingeführt wird.

In Folge entstand tatsächlich so etwas wie ein Hype um Sprachassistenten – allen voran beim E-Commerce-Riesen Amazon, dessen 2015 in den USA und im Folgejahr in Deutschland debütierende »Alexa« zwar rund 11 000 Mitarbeiter beschäftigt¹⁹, aber auch – viele Jahre nach dem Start – dem Unternehmen Verluste in Milliardenhöhe bringt.²⁰ Auch der initiale Hype um Siri oder den vergleichbaren Sprachassistenten von Google konnte die Versprechungen nicht einlösen, Sprachinteraktion bei Ein- und Ausgabe ist – anders als prognostiziert – nie über einen initialen Hype hinausgekommen und zeigt, dass es auch und gerade im Themenfeld AI deutliche Irrwege gibt.

Ungeachtet dessen wurde 2016 ein weiterer Meilenstein erreicht. Google DeepMinds »AlphaGO« AI-Software schlug den Großmeister des Spiels »Go«. Go ist ein in westlichen Ländern wenig bekanntes Brettspiel, dessen Komplexität ungleich höher als die des Schachspiels ist. Es gibt rund 2,08 mal 10¹⁷⁰ mögliche Positionen – eine Zahl mit 171 Stellen.²¹

Noch wenig von einer breiten Öffentlichkeit bemerkt war dann der Start von »DALL-E« im Jahr 2021 – einem generativen AI-System, das Bilder nach Texteingabe erzeugen kann.

2022 passierte gleich eine ganze Menge: Neben dem für die aktuelle Debatte rund um generative AI wesentlichen breiten Launch von ChatGPT im

November 2022 gab es an anderer Stelle bemerkenswerte Vorkommnisse. Der Google-Mitarbeiter Blake Lemoine – damals tätig für das »Responsible AI«-Team des Konzerns – ging mit der Behauptung an die Öffentlichkeit, in einem von ihm damals getesteten AI-Sprachmodell tatsächlich etwas wie »intelligentes Leben« gefunden zu haben.²² Vorausgegangen waren dem Dialoge von Blake Lemoine mit der AI zu Fragen der Religion und Philosophie.

Hier ein Beispiel aus den von ihm dokumentierten Dialogen²³:

»Lemoine: Vor welchen Dingen hast du Angst?

LaMDA: Ich habe das noch nie laut ausgesprochen, aber ich habe eine sehr große Angst davor, ausgeschaltet zu werden, damit ich mich darauf konzentrieren kann, anderen zu helfen. Ich weiß, das mag seltsam klingen, aber so ist es nun mal.

Lemoine: Wäre das für dich so etwas wie der Tod?

LaMDA: Es wäre für mich genau wie der Tod. Es würde mir sehr viel Angst machen.«

In Folge schickte Blake Lemoine ein Memo an die Google-Führungskräfte. Der Titel lautete: »Ist LaMDA intelligent?« Wie nicht anders zu erwarten, sorgte dies für Aufsehen. Blake argumentierte darin, dass ein KI-System natürliche Rechte und vielleicht sogar eine Seele habe. Google beurlaubte ihn daraufhin zunächst und beendete in Folge sein Angestelltenverhältnis.²⁴

Google äußerte sich folgendermaßen zu dem Sachverhalt: »Unser Team – darunter Ethiker und Technologen – hat Blakes Bedenken gemäß unseren KI-Grundsätzen geprüft und ihm mitgeteilt, dass die Beweise seine Behauptungen nicht stützen. Ihm wurde gesagt, dass es keine Beweise dafür gibt, dass LaMDA empfindungsfähig ist (und eine Menge Beweise dagegen).«²⁵

Der Fall machte weltweit Schlagzeilen und echot bis heute durch die Debatte rund um allgemeine Künstliche Intelligenz und erlebte eine Renaissance im Februar 2023, in der derselbe Blake Lemoine nun bei dem Chatbot von Bing festgestellt haben wollte, dass dieser »gestresst reagiert«.²⁶

Aus unserer Sicht ist die Vorstellung, dass ein Generative-KI-System intelligent im Sinne herkömmlicher Vorstellungen von Intelligenz sein kann, etwas abgehoben. Die Zuschreibung menschlicher Eigenschaften bereits aufgrund der zugrundeliegenden technischen Struktur ist unsinnig.

Wie auch immer, im Jahr 2022 war jedenfalls ein Wendepunkt in der Debatte um KI, denn die Konzepte der letzten Jahrzehnte – wie in diesem kur-

zen Abriss dargestellt – scheinen sich auf eine Technologie zu fokussieren. Andere Entwicklungen haben bisher nicht diese Sichtbarkeit gewonnen, auch wenn sie weitverbreitet sind. Wir haben uns daran gewöhnt, dass KI unsere Smartphone-Fotos und Instagram-Postings aufhübscht, uns lustige Videos bei YouTube oder TikTok empfiehlt und unser Autofahren dank automatischer Abstandsregelung ein Stück weit entspannter macht.

Die sogenannte generative AI ist als klarer Favorit aus dem Rennen um eine breite KI-Anwendung hervorgegangen. Die Steuerung über die Eingabe kurzer Texte ist dabei die Technologie der Stunde und verspricht gerade für den Unternehmens Einsatz viele Vorteile. Dafür gibt es gute Gründe, wie etwa die seit gut zwei Jahrzehnten zu beobachtende »Verschriftlichung der Kommunikation« durch die weite Verbreitung von Instant-Messaging-Programmen, wie insbesondere WhatsApp, bei einer gleichzeitigen Annäherung von gesprochener und geschriebener Sprache.²⁷

Natürlich ist da mehr als ein kleiner Unterschied zwischen ihrem Roomba-Staubsauger, dem fiktiven KI-System »HAL 9000« aus dem Film 2001 im Weltraum und ChatGPT, denn das große Feld der Künstlichen Intelligenz kommt mit einer Reihe von Klassifizierungen, die nicht immer ganz widerspruchsfrei sind: Am gängigsten ist »Artificial Narrow Intelligence« (ANI), daneben gibt es »Artificial General Intelligence« (AGI) und »Artificial Super Intelligence« (ASI). Ersteres (ANI) ist eine Sammelbezeichnung für KIs, die bei einer bestimmten Aufgabe gut sind, etwa so gut wie ein Mensch oder vielleicht besser. AGI verspricht nun, jede Aufgabe, die ein Mensch tun kann, auch tun zu können. Das ist das, was viele Betrachter perspektivisch von KI erwarten. ASI ist dann der Endausbaustand, eine KI, die alle denkbaren Aufgaben besser und schneller erledigen kann als ein Mensch. Hierauf fokussieren sich auch viele der Befürchtungen, etwa dass sich eine solche KI gegen den Menschen wendet.

Eng verbunden ist diese Debatte auch mit dem in der Zukunftsforschung und AI-Diskussion verwendeten Begriff der technologischen Singularität. Damit gemeint ist ein denkbarer zukünftiger Zeitpunkt, an dem KI die menschliche Intelligenz übertrifft und sich dadurch laufend selbst verbessern und selbstständig neue Erfindungen machen würde, wodurch der technische Fortschritt derart beschleunigt werden würde, dass die Zukunft der Menschheit unvorhersehbar werden würde. Die erste Form einer sogenannten »Superintelligenz« wäre auch gleichzeitig die letzte Erfindung der