



Ralf T. Kreuzer · Marie Sirrenberg

# Künstliche Intelligenz verstehen

Grundlagen – Use-Cases –  
unternehmenseigene KI-Journey

**EBOOK INSIDE**



**Springer Gabler**

---

# Künstliche Intelligenz verstehen

---

Ralf T. Kreutzer · Marie Sirrenberg

# Künstliche Intelligenz verstehen

Grundlagen – Use-Cases –  
unternehmenseigene KI-Journey

Ralf T. Kreutzer  
Campus Schöneberg, Hochschule für  
Wirtschaft und Recht  
Berlin, Deutschland

Marie Sirrenberg  
Bad Wilsnack, Brandenburg, Deutschland

ISBN 978-3-658-25560-2      ISBN 978-3-658-25561-9 (eBook)  
<https://doi.org/10.1007/978-3-658-25561-9>

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

Springer Gabler

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Gabler ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

*Jede hinreichend fortschrittliche Technologie ist  
von Magie nicht zu unterscheiden.*

*Clarke, A. C. (2008). Biokapital. Die Versöhnung  
von Ökonomie, Natur und und Menschlichkeit.  
Berlin: Berlin Verlag.*

Arthur C. Clarke

*Wir widmen dieses Werk der Hochschule für  
Wirtschaft und Recht, Berlin,  
die allen Fleißigen optimale Lehr-, Lern- und  
Forschungsmöglichkeiten bietet.*

---

## Vorwort

Ein Begriff dominiert in zunehmendem Maße die Diskussionen zum Thema Digitalisierung: **Künstliche Intelligenz** (KI) oder auch Artificial Intelligence (AI). Chinesische Unternehmen wie *Sensetime* stellen die Künstliche Intelligenz sogar in das **Zentrum der vierten industriellen Revolution**, in der sich die meisten Wirtschaftsnationen der Welt heute befinden. *Sensetime* hat eine gute Berechtigung, dies zu tun – schließlich ist es zurzeit das wertvollste KI-Start-up der Welt.

Und das kommt nicht von ungefähr: China hat in seinem Master-Plan „Made in China 2025“ die Künstliche Intelligenz als einen von zehn industriellen Bereichen definiert, in denen China eine globale Führungsrolle übernehmen möchte. Und die dort formulierten Ziele können sich sehen lassen. Bis 2030 möchte China nicht nur ein globales **AI-Innovation-Center** werden, sondern die chinesische KI-Branche soll dann auch einen Wert von ca. 130 Mrd. EUR aufweisen – und die KI-unterstützte Industrie das Zehnfache davon. China hat erkannt, dass es sich bei der Künstlichen Intelligenz um die **Mutter aller neuen Technologien** handelt.

Von einer solchen generalstabsmäßigen Planung sind Deutschland und Europa (noch) weit entfernt. Bezüglich der möglichen Ursachen stellen sich mehrere Fragen:

- Liegt die ungenügende Beschäftigung mit der Künstlichen Intelligenz darin begründet, dass teilweise noch **keine umfassende Vorstellung** darüber vorliegt, was die Künstliche Intelligenz für Unternehmen und ganze Branchen leisten kann?
- Sind es primär die **engen rechtlichen Rahmenbedingungen**, die es hier wirkenden Unternehmen schwer machen, die für die Künstliche Intelligenz notwendigen Datensätze aufzubauen und auszuwerten?
- Oder **fehlen (bewährte) Konzepte**, um das Potenzial der Künstlichen Intelligenz im eigenen Umfeld erfolgreich auszuschöpfen?

Wir sollten uns vor Augen führen, dass wir bereits heute permanent mit KI-Anwendungen in Kontakt kommen. Wer einen digitalen persönlichen Assistenten wie *Alexa* oder *Google Home* nutzt, greift auf KI-Anwendungen zu. Wer sich beim Übersetzen durch *Google Translate* oder das deutsche Start-up *DeepL* unterstützen lässt, profitiert

von der Künstlichen Intelligenz. Wer Gesichtserkennungssysteme einsetzt, verwendet KI-Algorithmen. Wenn Radiologen Röntgenbilder und CT-Scans durch Computer auswerten lassen, kommen KI-gestützte Expertensysteme zum Einsatz. Außerdem werden zunehmend Roboter eingesetzt – und das nicht nur in der Produktion. Das autonome Fahren ist ein weiteres KI-Einsatzfeld, das sich eines Roboters als Fahrer bedient. Damit wird deutlich:

**Die Künstliche Intelligenz ist in unserem Alltag längst angekommen.**

Wir wollen durch dieses Werk einen Beitrag leisten, dass (noch) mehr Menschen das mit der Künstlichen Intelligenz einhergehende Potenzial verstehen und erkennen. Gleichzeitig wird verdeutlicht, welche **Rahmenbedingungen** für einen **verantwortlichen Umgang** mit Künstlicher Intelligenz notwendig sind. Schließlich wird eine überzeugende **KI-Journey zur unternehmerischen Erschließung des KI-Potenzials** vorgestellt. Schließlich ist eines sicher:

**Die Künstliche Intelligenz wird das Leben von Menschen und Unternehmen – eingebettet in die Möglichkeiten der Digitalisierung – noch nachhaltiger verändern, als sich das viele heute vorstellen können.**

Um hierauf – als Mensch und Unternehmen – besser vorbereitet zu sein, wurde dieses Werk verfasst. Es soll Mut machen, sich mit diesem Thema (frühzeitig) auseinanderzusetzen und nachhaltig wertschöpfende Einsatzfelder zu identifizieren und zu nutzen – bevor es andere tun. Es soll vor allem Neugierde und Interesse dafür wecken, in welchen verschiedenen Einsatzfeldern Künstliche Intelligenz ihre Wirkungen entfalten kann. Hierbei gilt:

**Die Künstliche Intelligenz wird sich sehr schnell von einer Nice-to-have-Technologie zu einer Must-have-Technologie entwickeln.** Schließlich ist die Künstliche Intelligenz keine Innovation wie viele andere, sondern eine Basis-Innovation, die in den nächsten Jahren in alle Wirtschafts- und Lebensbereiche vordringen wird.

Gut, wenn man darauf vorbereitet ist.

Berlin  
Königswinter  
Bad Wilsnack  
Mai 2019

Ralf T. Kreutzer  
Marie Sirrenberg

---

# Inhaltsverzeichnis

<b>1 Was versteht man unter Künstlicher Intelligenz und wie kann man sie nutzen?</b> .....	1
1.1 Was ist der Kern der Künstlichen Intelligenz? .....	2
1.2 Welche Ziele lassen sich mit Künstlicher Intelligenz erreichen? .....	16
1.3 Einsatzfelder der Künstlichen Intelligenz .....	26
1.3.1 Natural-Language-Processing (NLP)/Verarbeitung natürlicher Sprache .....	28
1.3.2 Natural-Image-Processing/Computer-Vision/Bildverarbeitung .....	36
1.3.3 Expert-Systems/Expertensysteme .....	41
1.3.4 Robotics/Roboter .....	44
1.4 Welche globalen wirtschaftlichen Effekte werden durch die Künstliche Intelligenz ausgelöst? .....	52
Literatur .....	68
<b>2 Grundlagen und Treiber der Künstlichen Intelligenz</b> .....	73
2.1 Moore's Law und die Effekte der Exponentialität .....	74
2.2 Digitalisierung und Dematerialisierung von Produkten, Dienstleistungen und Prozessen .....	74
2.3 Vernetzung von Produkten, Services, Prozessen, Tieren und Menschen .....	76
2.4 Big Data .....	78
2.5 Neue Technologien .....	84
2.6 Investitionen in Künstliche Intelligenz .....	100
Literatur .....	103

<b>3</b>	<b>Anwendungsfelder der Künstlichen Intelligenz – Best Practices</b> .....	107
3.1	Produktionsbereich .....	108
3.2	Dienstleistungssektor .....	125
3.2.1	Customer-Services: von einfachen Chatbots zu digitalen persönlichen Assistenten .....	126
3.2.2	Marketing und Vertrieb .....	156
3.2.3	Einzelhandel .....	180
3.2.4	Gesundheitswesen .....	186
3.2.5	Energiesektor und Smart Home .....	199
3.2.6	Mobilitäts-/Transportsektor .....	210
3.2.7	Wartungs- und Instandhaltungssektor .....	216
3.2.8	Sicherheitssektor – Social Scoring .....	221
3.2.9	Bildung und Human-Resource-Management .....	227
3.2.10	Finanzdienstleistungen .....	241
3.2.11	Kreativbranche .....	246
3.3	Militärsektor .....	257
	Literatur .....	261
<b>4</b>	<b>KI-Challenge – wie Künstliche Intelligenz im Unternehmen zu verankern ist</b> .....	271
4.1	3-Horizonte-Modell als Orientierungsrahmen .....	271
4.2	Erfassung der KI-Reife des eigenen Unternehmens .....	274
4.3	Entwicklung einer KI-Journey im eigenen Unternehmen .....	278
4.3.1	Phase 1: Umfassende Informationsbeschaffung .....	279
4.3.2	Phase 2: Systematische Vorbereitung des KI-Einsatzes .....	282
4.3.3	Phase 3: Entwicklung von KI-Anwendungen .....	285
4.3.4	Phase 4: Integration von KI-Anwendungen und KI-Ergebnissen in das Unternehmen .....	308
	Literatur .....	313
<b>5</b>	<b>Ausblick</b> .....	317
	Literatur .....	328
	<b>Stichwortverzeichnis</b> .....	331

---

## Über die Autoren



**Prof. Dr. Ralf T. Kreutzer** ist seit 2005 Professor für Marketing an der Berlin School of Economics and Law sowie Marketing und Management Consultant, Trainer und Coach. Er war 15 Jahre in verschiedenen Führungspositionen bei Bertelsmann, Volkswagen und der Deutschen Post tätig, bevor er 2005 zum Professor für Marketing berufen wurde.

Prof. Kreutzer hat durch regelmäßige Publikationen und Vorträge maßgebliche Impulse zu verschiedenen Themen rund um Marketing, Dialog-Marketing, CRM/Kundenbindungssysteme, Database-Marketing, Online-Marketing, Digitalen Darwinismus, Dematerialisierung, digitale Transformation, Change-Management, strategisches sowie internationales Marketing sowie im Bereich Künstliche Intelligenz gesetzt und eine Vielzahl von Unternehmen im In- und Ausland in diesen Themenfeldern beraten und Führungskräfte auf Middle- und Top-Management-Ebene trainiert und gecoacht. Prof. Kreutzer ist ein gefragter Keynote-Speaker auf nationalen und internationalen Konferenzen. Er moderiert auch World-Café-Formate und weitere interaktive Formen der Gruppenarbeit.

Seine jüngsten Buchveröffentlichungen sind „B2B-Online-Marketing und Social Media (2015, zusammen mit Andrea Rumler und Benjamin Wille-Baumkauff), „Dematerialisierung – Die Neuverteilung der Welt“ (2015, zusammen mit Karl-Heinz Land), „Digital Darwinism – Branding and Business Models in Jeopardy“ (2015, zusammen mit Karl-Heinz Land), „Digitaler Darwinismus – der stille Angriff auf Ihr Geschäftsmodell und Ihre Marke“ (2. Auflage, 2016, zusammen mit Karl-Heinz Land), „Kundenbeziehungsmanagement im digitalen Zeitalter“ (2016), „Digital Business Leadership – Digitale

Transformation – Geschäftsmodell-Innovation – agile Organisation – Change-Management“ (2017, zusammen mit Tim Neugebauer und Annette Pattloch), „Digitale Markenführung“ (2017, zusammen mit Karl-Heinz Land), „Praxisorientiertes Online Marketing“ (3. Auflage, 2018), „Social-Media-Marketing kompakt“ (2018), „E-Mail-Marketing kompakt“ (2018), „Führung und Organisation im digitalen Zeitalter – kompakt“ (2018), „Digital Business Leadership – Digital Transformation, Business Model Innovation, Agile Organization, Change Management“ (2018, zusammen mit Tim Neugebauer und Annette Pattloch), „Toolbox für Marketing und Management“ (2018) sowie „Online-Marketing – Studienwissen kompakt (2. Auflage, 2019),.

Kontakt:

Prof. Dr. Ralf T. Kreutzer

Professor für Marketing an der Berlin School of Economics and Law sowie Marketing und Management Consultant

Alter Heeresweg 36

53639 Königswinter

[kreutzer.r@t-online.de](mailto:kreutzer.r@t-online.de)

[www.ralf-kreutzer.de](http://www.ralf-kreutzer.de)



**Marie Sirrenberg** ist IT-Beraterin Marketing & Vertrieb beim Atlassian Premium Solution Partner STAGIL. Zuvor war sie als NPS-Spezialistin bei dem SaaS-Start-up zenloop tätig, baute dort Vertriebsstrukturen auf und beriet vor allem erfolgreiche E-Commerce-Unternehmen zum Thema Feedback-Management. Bei dem mittelständischen Industrieunternehmen WDM begleitete sie Digitalisierungsprozesse mit den Schwerpunkten Online-Kommunikation und SEO.

Während ihres Master-Studiums International Marketing Management an der HWR Berlin bei Herrn Professor Ralf T. Kreutzer fokussierte sie sich auf Digitalisierungsthemen wie Social-Media-Strategien, digitale Bildung und Künstliche Intelligenz sowie Design-Thinking. Die umfassende Bedeutung der Künstlichen Intelligenz für zukünftige

Geschäftsmodelle war Anlass für ihre Masterarbeit „The future role of artificial intelligence in the service sector – challenges, tasks, recommendations“.

Kontakt:

Marie Sirrenberg (M.A.)

Große Straße 1

19336 Bad Wilsnack

[marie.sirrenberg@web.de](mailto:marie.sirrenberg@web.de)

---

## Abkürzungsverzeichnis

AAL	Ambient Assisted Living
Abb.	Abbildung
Abschn.	Abschnitt
AGI	Artificial General Intelligence
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AIR	Artificial-Intelligence-Roman
AKI	Acute Kidney-Injury
AP	Associated Press
API	Application-Programming-Interface (Anwendungsprogrammierschnittstelle)
AR	Augmented Reality
B2B	Business-to-Business
B2C	Business-to-Consumer
BaaS	Backup as a Service
BAT	<i>Baidu, Alibaba und Tencent</i>
BCI	Brain-Computer-Interface
BEO	Bot-Engine-Optimization
BKA	Bundeskriminalamt
BMI	Brain-Maschine-Interface
bspw.	beispielsweise
bzgl.	bezüglich
bzw.	beziehungsweise
CAD	Computer-Aided Design
CDO	Chief Digital Officer
CFI	Leverhulme Centre for the Future of Intelligence
CPS	Cyber-Physical-System (cyber-physisches System)
CRM	Customer-Relationship-Management
CT	Computertomografie
CUI	Conversational User-Interface

---

DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
d. h.	das heißt
DICaaS	Data-Intensive Computing as a Service
DSGVO	Datenschutz-Grundverordnung
DWH	Data-Ware-House
EEG	Elektroenzephalografie
EMG	Elektromyografie
ERP	Enterprise-Resource-Planning
ETFs	Exchange-Traded Funds
EU	Europäische Union
FAQs	Frequently Asked Questions
FLOPS	Floating Point Operations Per Second
fMRI	funktionelle Magnetresonanztomografie
FTF	fahrerlose Transportfahrzeuge
GAFAMI	<i>Google/Alphabet, Apple, Facebook und Amazon</i>
ggf.	gegebenenfalls
GUI	Graphical User-Interface
HBP	Human-Brain-Project
HPCaaS	High Performance Computing as a Service
HR	Human Resources
HaaS	Humans as a Service
IaaS	Infrastructure as a Service
insb.	insbesondere
IoE	Internet of Everything
IoT	Internet of Things
IT	Informationstechnologie
ITS	<i>Intelligent</i> Tutoring-System
Kap.	Kapitel
KDD	Knowledge-Discovery in Databases
KI	Künstliche Intelligenz
LPWAN	Low-Power-Wide-Area-Network
MaaS	Mobility as a Service
MINT	Mathematik, Informatik, Naturwissenschaft, Technik
ML	Machine-Learning (Maschinen-Lernen)
MOOC	Massive-Open-Online-Course
MRT	Magnetresonanztomografie
MUaaS	Music-as-a-Service
MVP	Minimum Viable Product
NASA	National Aeronautics and Space Administration
NLG	Natural-Language-Generation
NLP	Natural-Language-Processing
NLU	Natural-Language-Understanding

---

NPS	Net-Promoter-Score
OECD	Organisation for Economic Co-operation and Development
PaaS	Platform as a Service
RFID	Radio-Frequency-Identification
ROI	Return on Investment
RPA	Robotic-Process-Automation
SaaS	Software as a Service
SDK	Software-Development-Kit
SEO	Search-Engine-Optimization
SLAM	Simultaneous Localization and Mapping
SPS	Speicherprogrammierbare Steuerungen
SST	Self-Service-Technologien
STS	Speech-to-Speech
STT	Speech-to-Text
TaaS	Transportation as a Service
TK	Techniker Krankenkasse
TTS	Text-to-Speech
TTT	Text-to-Text
u. a.	unter anderem
u. E.	unseres Erachtens
u. U.	unter Umständen
UX	User-Experience
vgl.	vergleiche
VMES	Virtual-Manufacturing-Execution-System
VR	Virtual Reality
XAI	Explainable Artificial-Intelligence

# Was versteht man unter Künstlicher Intelligenz und wie kann man sie nutzen?

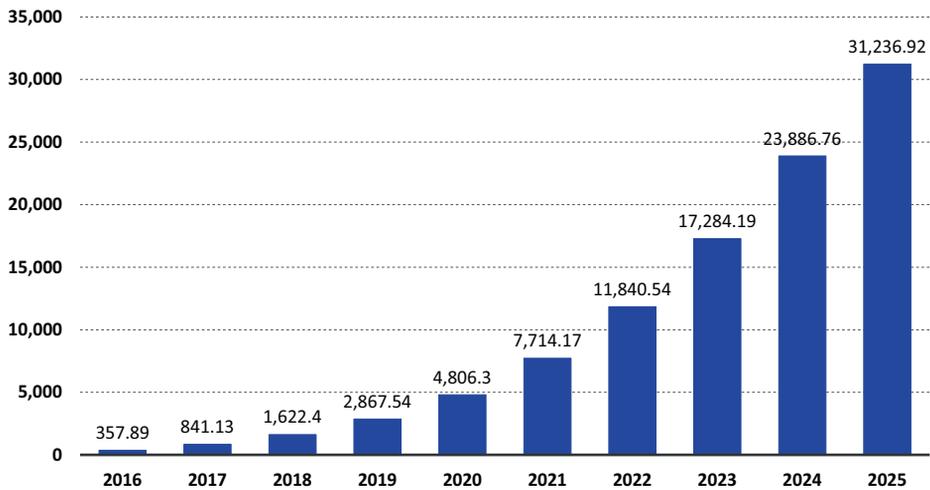
# 1

*Impossible is not a fact! It is an opinion!*

Künstliche Intelligenz ist ein relativ neues Forschungsfeld, das erst langsam aus dem Dunstkreis der Spezialisten heraustritt. Meistens begegnet sie uns in einer Art und Weise, die zunächst nicht an Künstliche Intelligenz denken lässt. Wir stellen nur fest, dass etwas einfacher und leichter von der Hand geht als früher. Denken Sie nur an **digitale persönliche Assistenten** wie *Alexa*, *Google Home* oder *Siri*, die Ihnen per Spracheingabe das gewünschte Musikstück von *Spotify* abspielen, Einkaufslisten erstellen oder sogar Einkäufe initiieren, Termine für Sie vereinbaren, Begriffe für Sie erklären oder ggf. die komplette Steuerung Ihres Smart Homes übernehmen. Wir sprechen von digitalen persönlichen Assistenten, weil diese keine physisch erlebbaren Assistenten mehr sind.

Wenn Sie sich die **Übersetzungshilfen im Internet** zunutze machen, sei es *Google Translate* oder das deutsche Start-up *DeepL*, greifen Sie ebenfalls auf KI-Anwendungen zu. Auch bei Systemen zur **Gesichtserkennung** (etwa zur Einlasskontrolle in Unternehmen) werden KI-Algorithmen genutzt. Bei der Auswertung von Krankenakten oder von Röntgenbildern und CT-Scans durch Computer kommen KI-gestützte **Expertensysteme** zum Einsatz. Ein schier unerschöpfliches Einsatzfeld der Künstlichen Intelligenz stellen **Roboter** dar. Deren intensiver Einsatz ist oft (noch) auf die Produktion und Logistik-Aufgaben beschränkt. Auch das autonome Fahren ist ein KI-Einsatzfeld, das sich eines Roboters als Fahrer bedient. Die größte Herausforderung der Künstlichen Intelligenz stellt nach wie vor die umfassende **Nachbildung des menschlichen Gehirns** dar.

Lohnt es sich für Sie als Student, als Manager, als Unternehmen oder als Land, sich intensiv mit den Entwicklungen rund um Künstliche Intelligenz zu beschäftigen? Wir meinen uneingeschränkt: „Ja“! Untermauern möchten wir diese Einschätzung durch die Zahlen in Abb. 1.1. Hier wird deutlich, welche **Umsätze mit Künstlicher Intelligenz** weltweit von 2016 bis 2025 (in Millionen US\$) erwartet werden. Diese Zahlen sprechen unserer Meinung nach für sich!



**Abb. 1.1** Prognose zum Umsatz mit Unternehmensanwendungen im Bereich Künstliche Intelligenz – weltweit von 2016 bis 2025 (in Mio. US\$). (Quelle: Statista 2018, S. 7)

Bevor wir vertiefend in die verschiedenen Einsatzfelder der Künstlichen Intelligenz einsteigen, klären wir zunächst, was unter Künstlicher Intelligenz genau zu verstehen ist und welche Ziele sich dadurch erreichen lassen. Dann werden verschiedene Einsatzfelder vertiefend betrachtet, um die Bandbreite der KI-Nutzbarkeit zu veranschaulichen. Anschließend beleuchten wir die globalen wirtschaftlichen Effekte, die durch Künstliche Intelligenz ausgelöst werden.

## 1.1 Was ist der Kern der Künstlichen Intelligenz?

Bevor wir uns der „künstlichen“ Intelligenz nähern, lohnt sich ein Blick auf die **Intelligenz** selbst. Statt den Blick zu verengen und nur einen Intelligenzquotienten zu betrachten, hilft es, sich vor Augen zu führen, was menschliche Intelligenz heute ausmacht oder ausmachen kann. Dazu gehört, dass wir Intelligenz in ihren relevanten Ausprägungen als **multiplen Intelligenzansatz** erfassen, der folgende Bereiche abdeckt (vgl. Gardner et al. 2011, S. 490–498):

- Sprachliche Intelligenz
- Musikalische Intelligenz
- Logisch-mathematische Intelligenz
- Räumliche Intelligenz
- Körperlich-kinästhetische Intelligenz
- Intrapersonale und interpersonale Intelligenz

- Naturalistische und existenzielle Intelligenz
- Kreative/schöpferische Intelligenz

Die **Vielfalt von Intelligenz** zeigt schon an dieser Stelle, warum es noch einige Zeit dauern wird, bis Maschinen in der Lage sein werden, die den Menschen angeborenen und erlernten Felder der Intelligenz in ihrer Gesamtheit abzudecken. Deshalb ist es u. E. nicht nur verfrüht, sondern auch überzogen, sich bei der KI-Diskussion auf Schreckensszenarien zu konzentrieren, bei denen KI-Maschinen die Weltherrschaft übernehmen. Ein durchaus lesenswertes Beispiel hiervon stammt vom Bestseller-Autor *Frank Schätzing* und trägt den schönen Titel „Die Tyrannei des Schmetterlings“.

- ▶ **Merk-Box** Künstliche Intelligenz (KI) bzw. Artificial Intelligence (AI) umfasst zwei Bereiche. Hierzu zählt zunächst die Erforschung, wie „**intelligentes**“ **Verhalten** Probleme löst. Basierend auf den so gewonnenen Erkenntnissen werden Systeme entwickelt, die automatisiert „**intelligente**“ **Lösungen** erzeugen (sollen). Die Herangehensweise beschränkt sich nicht darauf, Lösungen so zu erarbeiten, wie dies Menschen tun würden. Es wird vielmehr angestrebt, Ergebnisse zu finden, die außerhalb des Lösungsraums der Menschen liegen. Der Kern der Künstlichen Intelligenz ist Software!

Es gibt verschiedene Ansätze, um das Wesen der **Künstlichen Intelligenz** begrifflich zu erfassen. Die folgende sehr flexible Definition von Rich (1983) ist für die grundlegende Klärung u. E. am besten geeignet. Sie lautet: „Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.“ Diese Kennzeichnung der Künstlichen Intelligenz verdeutlicht, dass die Grenzen des Machbaren immer wieder neu definiert werden. Oder haben Sie vor zehn, 15 oder 20 Jahren damit gerechnet, dass uns selbstfahrende Autos 2019 als ein fast schon normales Phänomen erscheinen werden?

Eine etwas präzisere Definition lautet: Künstliche Intelligenz bezeichnet die Fähigkeit einer Maschine, kognitive Aufgaben auszuführen, die wir mit dem menschlichen Verstand verbinden. Dazu gehören Möglichkeiten zur Wahrnehmung sowie die Fähigkeiten zur Argumentation, zum selbstständigen Lernen und damit zum eigenständigen Finden von Problemlösungen. Es können drei **Arten von Auswertungen** – kombiniert oder isoliert – eingesetzt werden:

- **Description** (Beschreibung des „Ist“)
- **Prediction** (Vorhersage des „Wird“)
- **Prescription** (Empfehlung des „Was“)

Bei der Entwicklung der Künstlichen Intelligenz kam es zu einem interessanten **Phänomen**. Die ersten Aufgaben der Künstlichen Intelligenz waren für Menschen schwer, für KI-Systeme dagegen einfach zu bearbeiten (etwa komplexe Rechenprozesse).

Solche Aufgaben ließen sich durch formale mathematische Regeln exakt bewältigen. Sich anhand dieser Regeln durch eine große Datenfülle zu arbeiten, war eine der leichtesten Aufgaben für KI-Systeme. Dagegen ist es für Computer oft viel schwerer, solche Aufgaben zu bewältigen, die für Menschen leicht zu meistern sind, bei deren Lösung nicht nur auf mathematische Regeln zu setzen ist. Dies ist bei der Sprach- und Objekterkennung der Fall. Ein Mensch kann sehr einfach erkennen, wann es sich bei einem physischen Objekt um einen Tisch und wann um einen Stuhl handelt. Beide weisen zwar meistens vier Beine auf, aber die Funktion ist eine andere. Um dies zu lernen, muss dem KI-System häufig eine Vielzahl von Bildern gezeigt werden. Trotzdem kann dieses System vielfach – noch nicht – den eigentlichen „Sinn der Objekte“ erkennen.

Wenn ein KI-System durch eine große Zahl von Fotos die **Unterscheidung zwischen Schäferhund und Wolf** gelernt hat, kann man das System sehr leicht in die Irre führen, wenn ein Schäferhund auf einem Bild mit Schnee zu sehen ist. Dann kann es passieren, dass der Schäferhund als Wolf erkannt wird, weil viele Wolf-Fotos im Hintergrund Schnee aufweisen. Oder anders herum: Trägt ein Wolf auf einem Foto ein Halsband für die Leine, wird das KI-System mit Sicherheit einen Schäferhund vermuten, weil auf den Trainingsfotos für die KI-Algorithmen sicherlich kaum Wölfe mit Halsband vorkamen. Soviel zur (jetzigen) Intelligenz der Computer.

Den Inhalten der Künstlichen Intelligenz nähern wir uns am besten über die Abb. 1.2. Ein wichtiges Element der Künstlichen Intelligenz sind die sogenannten **neuronalen Netze**. Dieser Begriff kommt ursprünglich aus den Neuro-Wissenschaften. Dort bezeichnet ein neuronales Netz die Verbindung zwischen Neuronen, die als Teil des Nervensystems bestimmte Funktionen ausüben. Die Computer-Wissenschaften versuchen, solche neuronalen Netze nachzubilden. Ein besonderes Merkmal von ihnen ist, dass Informationen in den Netzen nicht über lineare Funktionen verarbeitet werden. Außerdem findet eine parallele Verarbeitung von Informationen statt, die durch die Verknüpfung der Neuronen und die speziellen Verarbeitungsfunktionen ermöglicht wird. Auf diese Weise können auch sehr komplexe, nicht-lineare Abhängigkeiten der Ursprungsinformationen abgebildet werden. Entscheidend ist, dass neuronale Netze diese



**Abb. 1.2** Leistungsbestandteile der Künstlichen Intelligenz

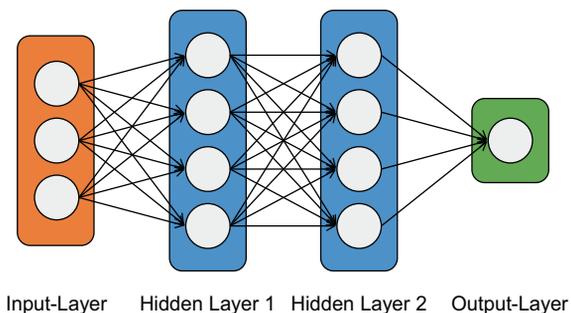
Abhängigkeiten selbstständig erlernen. Dies basiert auf Erfahrungsdaten (auch Trainingsdaten genannt), mit denen diese Systeme zu Beginn gefüttert werden (vgl. Lackes 2018).

Unter einem **neuronalen Netzwerk** ist ein System von Hard- und Software zu verstehen, dessen Aufbau sich am menschlichen Gehirn orientiert. Es stellt damit gleichsam das **Masterpiece der Künstlichen Intelligenz** dar. Ein neuronales Netzwerk verfügt in der Regel über eine große Anzahl von Prozessoren, die parallel arbeiten und in mehreren Schichten angeordnet sind (vgl. Abb. 1.3). Die erste Schicht (**Input-Layer** oder **Eingabe-Schicht**) erhält die Rohdaten. Diese Schicht kann mit den Sehnerven in der menschlichen Sehverarbeitung verglichen werden. Jede nachfolgende Schicht (hier **Hidden Layer 1** und 2) erhält den Output der vorhergehenden Schicht – und nicht mehr die Daten, die in den vorgelagerten Schichten verarbeitet wurden. Analog dazu empfangen Neuronen im menschlichen System, die weiter vom Sehnerv entfernt sind, Signale von den Neuronen, die ihnen näher liegen. Dieser natürliche Prozess wird in neuronalen Netzwerken nachempfunden. Zur Verarbeitung der Daten kann eine sehr große Zahl von Hidden-Layers eingesetzt werden – häufig nicht nur 100, 1000 oder 10.000. Das KI-System lernt von jedem Übergang zu einer anderen Schicht (idealerweise) dazu. Die letzte Schicht (**Output-Layer** oder **Ausgabe-Schicht**) erzeugt die Ausgabe der Ergebnisse des KI-Systems (vgl. Rouse 2016).

Jeder **Verarbeitungsknoten** verfügt über einen eigenen Wissensbereich. Dieser umfasst nicht nur die Regeln, mit denen er ursprünglich programmiert wurde. Dazu gehören vielmehr auch das Wissen und die Regeln, die im Zuge des sogenannten Machine-Learnings ergänzend oder korrigierend erarbeitet wurden. Das bedeutet, dass die „Maschine“ selbstständig dazu lernt und sich damit vom ursprünglichen „Wissen“ mehr oder weniger weit entfernen kann (vgl. vertiefend Schölkopf und Smola 2018).

- **Merk-Box** Die „Maschine“ emanzipiert sich im Zuge ihres Einsatzes zunehmend von den ursprünglichen Eingaben (Daten und Regeln). Im Vergleich zu den klassischen regelbasierten Systemen (hier werden Daten so verarbeitet, wie es im Vorfeld definiert wurde) soll die Künstliche Intelligenz sich eigenständig entwickeln und lernen, um auf Basis der so gewonnenen Erfahrungen noch bessere Ergebnisse zu erzielen. Die initial eingesetzten Algorithmen

**Abb. 1.3** Verschiedene Schichten bei neuronalen Netzwerken



stellen nur den Nährboden für die Entwicklung neuer Algorithmen dar. Wenn sich neue Algorithmen im Laufe des Lernprozesses als aussagekräftiger erweisen, arbeitet die „Maschine“ selbstständig mit diesen weiter. Dieser Prozess wird **Machine-Learning** genannt.

Um diesen Lernprozess zu unterstützen, sind die verschiedenen Ebenen vielfältig miteinander verbunden. Wie Abb. 1.3 verdeutlicht, sind die Eingänge jedes Knotens einer Ebene „n“ mit vielen Knoten der vorangestellten Ebene „n-1“ verbunden. Eine Ausnahme bildet der Input-Layer, der auch nur einen Knoten aufweisen kann (Abb. 1.3 zeigt hier drei Knoten). Zusätzlich sind die Ausgänge der Ebene „n“ mit den Eingängen der nachfolgenden Ebene „n+1“ verbunden. Die beschriebenen Verknüpfungen ermöglichen die stufenweise Informationsweitergabe von Layer zu Layer. Die zweite Ausnahme hinsichtlich der Anzahl der Knoten liefert die Output-Layer. Dort kann es einen (so in Abb. 1.3) oder auch mehrere Knoten geben, von denen Antworten abgelesen werden können.

Zur Beschreibung von neuronalen Netzen kann zum einen die **Tiefe des Modells** herangezogen werden. Diese wird durch die Anzahl der Schichten definiert, die zwischen Eingang und Ausgang liegen. Hier wird von den sogenannten versteckten Schichten des Modells gesprochen (deshalb Hidden Layer). Neuronale Netze können auch durch die **Breite des Modells** beschrieben werden. Dazu wird die Anzahl der versteckten Knoten des Modells bzw. die Anzahl der Ein- und Ausgänge pro Knoten berücksichtigt. Variationen des klassischen neuronalen Netzwerk-Designs erlauben verschiedene Formen der Vorwärts- und Rückwärtsausbreitung von Informationen zwischen den Ebenen.

Lassen Sie uns noch etwas tiefer das wichtige **Machine-Learning** (ML, auch **Maschinen-Lernen**) betrachten. Hierbei kommen – wie bereits angedeutet – Algorithmen zum Einsatz, die in der Lage sind, eigenständig zu lernen und sich folglich eigenständig zu verbessern. Unter einem Algorithmus ist eine programmierte Anweisung zu verstehen, die eingegebene Daten in vordefinierter Form verarbeitet und zur Ausgabe von darauf basierenden Ergebnissen führt. Beim Machine-Learning kommen ganz spezielle Algorithmen zum Einsatz – konkret sogenannte **selbst-adaptive Algorithmen**. Hierdurch können die Maschinen eigenständig lernen, ohne dass Programmierer in den laufenden Lernprozess eingreifen müssen. Es bedarf großer Datenmengen. Nur dann können die Algorithmen so trainiert werden, dass sie in der Lage sind, vordefinierte Aufgabenstellungen immer besser zu meistern – ohne dafür erneut programmiert zu werden. Dazu wird auf Erkenntnisse zugegriffen, die u. a. durch Deep Learning gewonnen werden (siehe im Folgenden). Um diese Lernprozesse zu fördern, sind folglich große, qualitativ hochwertige Datenmengen als „Trainingsmaterial“ erforderlich. Mit diesen sogenannten **Trainingsdaten** werden die neuen Algorithmen generiert. Nach der Fertigstellung werden diese fortlaufend mit weiteren **Inputdaten** überprüft, um so zu verbesserten Entscheidungsgrundlagen zu kommen. Um die Leistung des Algorithmus auf Basis bereits gesammelter Erfahrungen zu verbessern, kommen sogenannte **Feedbackdaten** zum Einsatz (vgl. Agrawal et al. 2018, S. 43).

Beim Machine-Learning und damit bei der Entwicklung immer leistungsstärkerer Algorithmen können verschiedene **Arten des Lernens** unterschieden werden (vgl. McKinsey 2018b, S. 2–6; Gentsch 2018, S. 38 f.):

- **Beaufsichtigtes Lernen – Supervised Learning**

Bei diesem Lernprozess kennt das KI-System bereits die richtigen Antworten und muss die Algorithmen „nur noch“ so anpassen, dass die Antworten möglichst präzise aus dem vorhandenen Datensatz abgeleitet werden können. Das Ziel bzw. die Aufgabe des Algorithmus ist hier folglich bereits bekannt.

Bei diesem Lernansatz müssen Menschen jedes Element der Input-Daten kennzeichnen. Zusätzlich sind die Output-Variablen zu definieren. Der Algorithmus wird auf die eingegebenen Daten trainiert, um die Verbindung zwischen den Eingangsgrößen und den Output-Variablen zu finden. Hierbei kommen u. a. die Methoden Lineare Regression, Lineare Diskriminanzanalyse sowie das Entscheidungsbaumverfahren zum Einsatz. Sobald das Training abgeschlossen ist – typischerweise, wenn der Algorithmus ausreichend genau ist – wird der Algorithmus auf neue Daten angewendet.

Die Aufgabe eines solchen KI-Systems könnte darin bestehen, die bekannten Preise für verschiedene Automodelle durch deren Merkmale (bspw. Marke, PS-Stärke, Art des Motors, Ausstattungsmerkmale) zu erklären. Hier lernt das System aus einem komplett vordefinierten Datensatz selbstständig, die relevanten Erklärungsmuster zu erkennen.

- **Nicht-überwachtes Lernen – Unsupervised Learning**

Das KI-System verfügt bei dieser Lernform nicht über vordefinierte Zielwerte und muss Ähnlichkeiten und damit Muster in den Daten eigenständig erkennen. Dem Anwender sind solche Muster im Vorfeld folglich nicht bekannt; vielmehr ist es die Aufgabe des Algorithmus, diese selbstständig zu erkennen. Die durch das System gewonnenen Erkenntnisse können folglich auch außerhalb des bisher „menschlich Vorstellbaren“ liegen.

Hierzu erhält der Algorithmus unbeschriftete Daten. In diesen soll der Algorithmus eigenständig eine Struktur erkennen. Dazu identifiziert der Algorithmus Daten-Gruppen, die ein ähnliches Verhalten oder ähnliche Merkmale aufweisen. Hierbei werden u. a. die Methoden der hierarchischen und der K-Means-Clusterung verwandt. Eine spannende Aufgabenstellung hierfür besteht darin, Menschen in den sozialen Medien zu erkennen (Mustererkennung), die besonders anfällig dafür sind, falschen Mitteilungen zu glauben, diese positiv zu kommentieren und weiterzuleiten. Hier könnte sich zeigen, dass es sich besonders um solche Personen handelt, die besonders häufig Katzenfotos liken oder meist erst zwischen 22.00 und 22.30 Uhr in den sozialen Medien aktiv sind. Solche Erkenntnisse können außerhalb dessen liegen, was ggf. vermutet wurde. Dass solche Anwendungen – aus unserer Sicht leider – bereits zum Einsatz kamen, zeigen die Analysen der US-Präsidenten-Wahl (2016) und das Brexit-Votum (2016).

- **Verstärkendes Lernen – Reinforcement-Learning**

Bei diesem Lernprozess liegt zu Beginn der Lernphase kein optimaler Lösungsweg vor. Das System muss iterativ durch einen Trial-and-Error-Prozess eigenständig Lösungswege ausprobieren, um diese anschließend zu verwerfen und/oder weiterzuentwickeln. Dieser iterative Prozess wird durch „Belohnungen“ (bei guten Lösungs-ideen) sowie „Bestrafungen“ (bei schlechten Ansätzen) vorangetrieben. Dieses Lernkonzept wird häufig eingesetzt, wenn nur wenige Trainingsdaten vorliegen oder das ideale Ergebnis nicht klar definierbar ist. Es kommt auch zum Einsatz, wenn erst aus der Interaktion mit der Umwelt etwas gelernt werden kann.

Im Zuge dieses Lernprozesses trifft der Algorithmus eine Entscheidung und handelt entsprechend. Dann enthält dieser eine Belohnung, wenn die Aktion die Maschine zu einer Annäherung an das Ziel führt. Alternativ erfährt das System eine Bestrafung, wenn es sich vom Ziel entfernt. Der Algorithmus optimiert seine Aktionen selbstständig, indem er sich laufend selbst korrigiert.

Diese Lernvariante wurde bei dem in Abschn. 1.2 beschriebenen Wettstreit zwischen dem *Go*-Weltmeister und dem Computer *AlphaGo* eingesetzt. Durch die Simulation verschiedener Partien gegen sich selbst und durch die dabei gemachten Erfahrungen „Sieg“ (Belohnung) und „Niederlage“ (Bestrafung) konnte das System seine Strategien kontinuierlich verbessern.

Eine spezielle Ausgestaltung der neuronalen Netze und eine Teilmenge des Machine-Learnings stellt das sogenannte **Deep Learning** dar (vgl. Abb. 1.2; vgl. Arel et al. 2010, S. 13; Kelly 2014, S. 6–8; McKinsey 2018b, S. 6; Domingos 2015). Deep Learning ist eine Art des maschinellen Lernens, die eine größere Bandbreite an Datenressourcen verarbeiten kann, weniger Datenvorverarbeitung durch den Menschen erfordert und oft genauere Ergebnisse liefern kann als herkömmliche maschinelle Lernansätze. Das „Deep“ bezieht sich auf die große Anzahl der Schichten des neuronalen Netzes. Hierzu werden spezielle Netzwerke aufgebaut, die sehr große Mengen an Eingabedaten aufnehmen und über mehrere Schichten verarbeiten können. Hierzu kommen besondere Optimierungsmethoden zum Einsatz, die eine noch umfangreichere innere Struktur aufweisen als klassische neuronale Netze. Hierbei werden tief liegende Muster und Korrelationen erkannt, die die vorhandenen Datenpunkte miteinander verbinden.

Um anspruchsvollere Aufgaben zu bewältigen, können Computer heute aus ihren eigenen Erfahrung lernen und neue Inputdaten in Relation zu den schon vorhandenen Daten bringen. Dazu ist es nicht mehr erforderlich, dass Menschen diese Daten zunächst formal spezifizieren. Die Maschine lernt nach und nach, komplexe Konzepte aus einfacheren Elementen zusammensetzen. Die Visualisierung dieser Zusammenhänge kann durch Diagramme erfolgen, die aus einer Vielzahl von Schichten besteht und dadurch „Tiefe“ erlangen (vgl. Abb. 1.3). Deshalb wird von „Deep Learning“ gesprochen. Ein Beispiel hierfür stellt die Handschriftenerkennung dar. Hier müssen Bildpunkte sukzessiv erkannt und mit Inhalt angereichert werden. Durch klassische Programmierung kann eine Erkennung unterschiedlichster Handschrift praktisch nicht geleistet werden. Hier bedarf es der Konzepte, die „von selbst“ lernen.

- ▶ **Merk-Box** KI-Anwendungen weisen Grundfähigkeiten der Wahrnehmung sowie des Verstehens, Lernens und Handelns auf (vgl. Bitkom und DFKI 2017, S. 29).

In diesem Kontext wird von **Neuro-Computing** (auch **Neural Computing**) gesprochen. Darunter werden Technologien subsumiert, die neuronale Netzwerke verwenden, die das menschliche Gehirn simulieren. Diese werden für bestimmte Aufgaben trainiert, bspw. für die Mustererkennung in großen Dateien. Das umfassende Ziel von KI-Anwendungen kann mit dem Begriff **Knowledge-Discovery** (auch Knowledge-Discovery in Databases, KDD) beschrieben werden. Es geht um die „Wissenserkennung in Datenbanken“. Hierzu kommen verschiedene Techniken zum Einsatz, die versuchen, bisher unbekannte fachliche Zusammenhänge – die sogenannte Kernidee – in großen Datenbeständen zu erkennen. Diese „Kernidee“ soll gültig, neu und nützlich sein und idealerweise ein bestimmtes Muster zeigen. In Abgrenzung zum Data-Mining umfasst Knowledge-Discovery nicht nur die Verarbeitung von Daten, sondern auch die Bewertung der erzielten Ergebnisse.

- ▶ **Merk-Box** Zur **Entmystifizierung der Künstlichen Intelligenz** könnte man ganz platt formulieren: Der **Kern der Künstlichen Intelligenz** besteht darin, selbstständig große Datenmengen zu verarbeiten, darin eigenständig Muster zu erkennen und auf deren Grundlage autonom Entscheidungen und/oder Vorhersagen zu treffen. KI-Anwendungen sind bei solchen Aufgabenstellungen häufig schneller und – je nach System – auch kostengünstiger.

Eine Besonderheit der neuronalen Netze stellt deren **Anpassungsfähigkeit** innerhalb eines bestimmten Einsatzfeldes dar. Dies führt dazu, dass sich diese Netze selbstständig verändern und damit weiterentwickeln können. Dabei fließen jeweils die Erkenntnisse ein, die auf Grundlage der sogenannten „Erstausbildung“ durch die Trainingsdaten sowie durch die Verarbeitung weiterer Daten gewonnen wurden. Eine große Bedeutung kommt der Gewichtung der jeweiligen Input-Streams zu. Das KI-System gewichtet selbstständig jene Dateneingaben höher, die dazu beitragen, richtige Antworten zu erhalten.

Um ein neuronales Netzwerk zu trainieren, wird es zunächst mit großen Datenmengen versorgt. Gleichzeitig ist dem Netzwerk mitzuteilen, wie der Output aussehen soll. Um ein **Netzwerk zur Identifikation von Gesichtern bekannter Schauspieler** zu trainieren, wird dem System im Zuge der Erstausbildung eine Vielzahl von Fotos von Schauspielern, Nicht-Schauspielern, Masken, Statuen, Tiergesichtern und so weiter zur Verfügung gestellt. Jedes einzelne Foto wird mit einem Text versehen, der die Inhalte des Fotos möglichst gut beschreibt. Das können zum einen die Namen der dort abgebildeten Schauspieler sein, zum anderen Hinweise darauf, dass es sich nicht um einen Schauspieler handelt, sondern um eine Maske oder ein Tier.

Durch die **Bereitstellung von beschreibenden Informationen** kann das Modell seine **internen Gewichtungen** anpassen. So lernt es, seine Arbeitsweise kontinuierlich zu verbessern. Es können bspw. die Knoten A, B und D dem Knoten BB der nächsten Schicht

mitteilen, dass es sich bei dem Eingabebild um ein Foto von *Daniel Craig* handelt. Dagegen meint Knoten C, auf dem Bild *Roger Moore* erkannt zu haben (etwa weil auf dem Foto neben dem Schauspieler ein *Aston Martin* zu sehen ist, den beide in *James-Bond*-Filmen nutzen). Bestätigt jetzt das Trainingsprogramm, dass auf dem Foto tatsächlich *Daniel Craig* abgebildet ist, wird der Knoten BB das Gewicht der Eingabe von Knoten C verringern, weil dieser eine falsche Bewertung vornahm. Gleichzeitig wird das System die Gewichte für die Knoten A, B und D erhöhen, weil deren Ergebnisse korrekt waren.

Jeder Knoten entscheidet eigenständig, welche Eingaben von der vorherigen Schicht in welcher Form an die nächste Schicht gesendet werden. Um zu diesen Entscheidungen zu kommen, verwenden neuronale Netze Regeln und Prinzipien. So können ein gradientenbasiertes Training, Fuzzy-Logik, genetische Algorithmen und Bayes'sche Methoden eingesetzt werden. Hier können u. a. **Grundregeln über die Beziehungen verschiedener Objekte** im zu modellierenden Raum erarbeitet werden. So kann etwa einem **Gesichtserkennungssystem** Folgendes mitgeteilt werden (vgl. Rouse 2016):

- Augenbrauen befinden sich über den Augen
- Schnurrbärte sind unter der Nase (deshalb auch Oberlippenbart)
- Bärte sind über und/oder neben dem Mund, auf den Wangen und am Halsansatz zu finden
- Bärte finden sich überwiegend bei Männern; allerdings gibt es auch Damenbärte
- Augen liegen auf gleicher Höhe nebeneinander
- Augen befinden sich rechts und links oberhalb der Nase
- Der Mund liegt unter der Nase
- Etc.

Solche Regeln, die dem System bei der initialen Bereitstellung von Material mitgegeben werden (sogenannte **Preloading-Rules**), können das Training beschleunigen und das Modell leistungsfähiger machen. Sie bauen auch **Annahmen über die Art des Problemraumes** auf, die sich später entweder als irrelevant und nicht hilfreich oder sogar als falsch und damit kontraproduktiv erweisen können. Deshalb ist die Entscheidung darüber, ob und welche Regeln im Vorfeld zu definieren sind, von großer Bedeutung.

Zusätzlich möchten wir Sie auf einen weiteren wichtigen Aspekt hinweisen: die **Fairness von Künstlicher Intelligenz**. Menschen, die Preloading-Rules definieren und zu Trainingszwecken Daten in die Systeme einspeisen, sind per se voreingenommen – das liegt in unserer Natur. Damit können die hier verwendeten Regeln wie auch die Daten einen Bias (eine Verzerrung) aufweisen, der sich auf spätere Bewertungen und Entscheidungsempfehlungen auswirkt (bspw. bei Kreditwürdigkeitsprüfungen) – ohne (leicht) erkannt werden zu können.

Ein **Beispiel aus der Rechtsprechung** kann diese Gefahr eindrucksvoll belegen. In den USA sollte ein KI-System Gerichtsurteile fällen. Dazu wurde es anhand von alten Gerichtsurteilen trainiert. Beim Einsatz wurde ein interessantes Phänomen festgestellt: Wenn man bei den Angeklagten die Hautfarbe von Weiß auf Schwarz verändert, ging

plötzlich das Strafmaß nach oben. Hierbei wurde deutlich, dass Vorurteile, die in den alten Urteilen lagen, vom KI-System unreflektiert auf die neuen Rechtsfälle übertragen wurden (vgl. Hochreiter 2018, S. 3).

- ▶ **Merk-Box** Eine große Fehlerquelle bei KI-Anwendungen liegt im: **Bias in – Bias out!**

Damit Sie solchen Gefahren durch Verzerrungen in den Datensätzen vorbeugen, müssen Sie zwingend auf die **Ausgewogenheit der Trainingsdaten** achten. Einen Ansatz, um dies zu erreichen, stellt der Austausch von ausgewogenen Datensätzen zwischen Unternehmen dar. So stellte *IBM* im Jahr 2018 die Datensätze von einer Million Gesichtsfotos zur Verfügung, um damit Systeme zur Gesichtserkennung zu trainieren (vgl. Rossi 2018, S. 21). Kämen zu Trainingszwecken der KI-Systeme bspw. nur europäische oder nur asiatische Fotos zum Einsatz, würden die Ergebnisse im Hinblick auf einen globalen Einsatz verfälscht. Zusätzlich sollte im verantwortlichen **KI-Programmierer-Team** eine hohe Diversity (nach Alter, Geschlecht, Nationalität etc.) vorliegen, damit weder den Trainingsdatensätzen noch den Preloading-Rules (unbewusste) Stereotype und Vorurteile der Programmierer mitgegeben werden.

Wie schnell es zu solchen Verzerrungen kommen kann, zeigt eine Studie aus dem Jahr 2017 (vgl. Lambrecht und Tucker 2017). Hier wurde erkannt, dass *Facebook*-Werbeanzeigen geschlechterdiskriminierend ausgespielt wurden. Es handelte sich um Jobanzeigen aus dem MINT-Bereich, die Frauen weniger häufig ausgespielt wurden als Männern. Diese unbewusst eingebaute Diskriminierung resultierte daraus, dass junge Frauen auf *Facebook* eine begehrte Zielgruppe sind. Folglich ist es bei ihnen teurer, eine Werbeanzeige zu platzieren. Wenn der Algorithmus also die Wahl hatte, bei gleichen Klick-Raten zwischen einem Mann und einer Frau zu entscheiden, fiel die Wahl auf die kostengünstigere Variante – in diesem Fall den Mann.

- ▶ **Merk-Box** Damit Sie mögliche Verzerrungen in Ihren Daten vermeiden, sollten Sie unterschiedliche (verlässliche) Datenquellen heranziehen. Eine hohe Diversität in Ihren Teams führt – quasi automatisch – dazu, dass sich Vorurteile oder Stereotypen in KI-Systemen vermeiden lassen, welche zu fehlerhaften Erkenntnissen führen. Ein **Data-Audit** kann hier wertvolle Unterstützung leisten, indem es systematisch die Qualität der einlaufenden Daten überprüft.

Eine spannende Ergänzung bzw. eine Alternative zum **Training von Robotern** stellt das Demonstration-based Training (demonstrationsbasierte Einarbeitung) dar. Die Programmierung von Robotern (insb. für Produktionsprozesse) ist eine komplexe, zeitaufwendige und teure Aufgabe, die ein hohes Maß an Expertenwissen erfordert. Wenn sich Aufgaben, Prozesse und/oder die Produktionsumgebung verändern, müssen die dort eingesetzten Roboter neu programmiert werden. Hier bieten sogenannte **Wandelbots** eine Lösung an. Mit **Demonstration-based Teaching** können Roboter programmiert werden,

ohne hierfür neue Programme zu schreiben. Indem Robotern vorgeführt wird, wie eine bestimmte Aufgabe auszuführen ist, lernt das Steuerungsprogramm selbstständig die dafür erforderlichen Abläufe. Hierdurch können die Task-Experten Roboter auch in dynamischen und komplexen Umgebungen unterrichten – ohne dafür über Programmierkenntnisse verfügen zu müssen. So kann der Roboter neue Aufgaben in wenigen Minuten erlernen, ohne dass hierfür Spezialistenwissen erforderlich wäre. Während des Lernvorgangs erfassen die Sensoren des Roboters sowie ggf. weitere externe Sensoren die Eigenschaften der Umgebung, die für den Lernprozess notwendig sind (vgl. Wandelbots 2019).

- ▶ **Merk-Box** Sie sollten prüfen, inwieweit es möglich ist, beim Training von Robotern die demonstrationsbasierte Einarbeitung zu nutzen.

Neben der Vermeidung von fehlerbehafteten Datensätzen stellt die **Transparenz über die Prozesse der Entscheidungsfindung bei KI-Systemen** eine große Herausforderung dar. Da die KI-Maschine selbstständig zu Ergebnissen und Entscheidungen gelangt, stellt sich für die Nutzer und insb. für die Betroffenen die Frage nach dem „Warum“. Schließlich möchte man sein Schicksal, sei es bei einer finanziellen Anlageentscheidung (Stichwort Robo-Advisor), bei der Ablehnung eines Kreditantrags oder beim autonomen Fahren, nicht einer Black Box anvertrauen. Vielmehr möchte man schon wissen, warum in welchen Situationen so oder so entschieden wurde bzw. wird.

Die Aufgabe hier lautet: **Explainable Artificial Intelligence** (XAI; auch erklärbare Künstliche Intelligenz). Hierunter versteht man den Versuch, eine **Black Box „Künstliche Intelligenz“** zu vermeiden und eine **Grey Box „Künstliche Intelligenz“** zu schaffen, die zumindest eine partielle Nachvollziehbarkeit von Ergebnissen und Entscheidungen ermöglicht. Es wird angestrebt, den Prozess und die Ergebnisse des KI-Einsatzes besser nachvollziehbar zu machen. Hier werden verschiedene Felder unterschieden:

- **Transparenz der Daten**

Da die Qualität und „Unbestechlichkeit“ der Künstlichen Intelligenz mit den verfügbaren Datengrundlagen steht und fällt, sollte es dem interessierten Anwender möglich sein, die Datengrundlagen der KI-Anwendung zu prüfen. Werden in diesen Daten Verzerrungen oder irrelevante Grundgesamtheiten erkannt, ist den Ergebnissen des KI-Systems nicht zu trauen. Es ist darauf hinzuweisen, dass es wiederum oft eines Expertenwissens bedarf, um diese Datengrundlagen kritisch zu überprüfen; für Laien ist das meist nicht möglich. Hier könnten Zertifizierungsprozesse mit entsprechenden Prüfsiegeln für die verwendeten Daten weiterhelfen.

- **Transparenz der Algorithmen**

Besonders wichtig bei KI-Anwendungen ist es, zu erkennen, aufgrund welcher Algorithmen bestimmte Ergebnisse erzielt wurden. Da die Maschine selbstständig dazulernt, ist dieser Prozess nicht leicht nachzuvollziehen. Für die Akzeptanz der

KI-Ergebnisse ist es aber unverzichtbar, zumindest die Haupteinflussfaktoren einer Entscheidung erkennen zu können (bspw. bei einer Bonitätsbewertung von Kunden oder bei der Empfehlung, welche Personen für eine neues Angebot über welchen Kanal wann angesprochen werden sollten). Mit einer solchen Transparenz steht und fällt die Bereitschaft, sich KI-Systemen anzuvertrauen. Schließlich möchte sich (heute) niemand auf Systeme und deren Entscheidungen verlassen, die nicht nachvollziehbar sind.

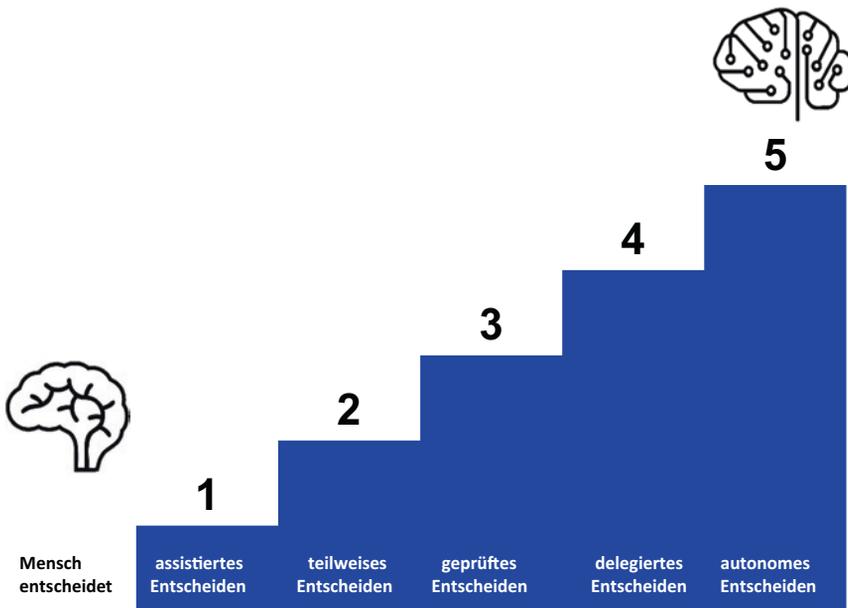
- **Transparenz der Auslieferung von Daten**

Hierbei geht es darum, die Ergebnisse für die Nutzer und/oder die Betroffenen so aufzubereiten, dass auch eine mathematisch und/oder statistisch wenig geschulte Person die gewonnenen Erkenntnisse verstehen kann – soweit diese nicht bereits automatisiert in Prozesse einfließen. Es soll nachvollzogen werden können, warum bei einer KI-gestützten Bonitätsbewertung Person A im Vergleich zu Person B keinen Kredit erhält. Bei einer KI-unterstützten Erarbeitung von Gerichtsurteilen soll deutlich werden, warum für den Angeklagten X eine Bewährungsstrafe vorgeschlagen wird und weshalb der Angeklagte Y ins Gefängnis soll.

Ohne **Explainable Artificial Intelligence** bleiben KI-Anwendungen eine Black Box, die eine kritische Analyse der zugrunde liegenden Prozesse sowie der gewonnenen Ergebnisse schwer oder unmöglich macht. Dann können Anwender nicht nachvollziehen, wie ein Ergebnis zustande kommt – und es fällt schwerer, den Ergebnissen trotzdem zu „trauen“.

- ▶ **Merk-Box** Setzen Sie von Anfang an auf **Explainable Artificial Intelligence** – selbst dann, wenn Ihre KI-Spezialisten lieber darauf verzichten möchten. Ohne eine gewisse Transparenz über Daten, Prozesse und Ergebnisse werden Sie bei den Anwendern nur schwer eine Akzeptanz für KI-Anwendungen und deren Ergebnisse gewinnen.

Technologien der Künstlichen Intelligenz können nach ihrem Automatisierungsgrad unterschieden werden. Das **Fünf-Stufen-Modell** in Abb. 1.4 visualisiert die mögliche **Arbeitsteilung zwischen menschlichem und maschinellm Handeln**. Der Grad der Automatisierung von Entscheidungen hängt von der Komplexität der Fragestellung und der Leistungsfähigkeit des jeweils eingesetzten KI-Systems ab. Welche rechtlichen, ethischen und auch wirtschaftlichen Fragen mit der jeweils gewählten Arbeitsteilung verbunden sind, machen folgende Beispiele deutlich. Beim **assistierten Entscheiden** unterstützt ein KI-System den Menschen bei seinen Entscheidungen. Dies kann ein KI-Algorithmus sein, der Kaufvorschläge bei *Amazon* unterbreitet oder bei *Google* zu einer Autovervollständigung bei unseren Suchen führt. Die Autokorrektur im Smartphone stellt ein weiteres Beispiel für das assistierte Entscheiden dar. Viele lustige und ärgerliche Beispiele machen deutlich, dass manche Nutzer dem KI-System hier ein autonomes Entscheiden erlauben.



**Abb. 1.4** Fünf-Stufen-Modell der Automation des Entscheidens. (Quelle: In Anlehnung an Bitkom und DFKI 2017, S. 62)

Beim **teilweisen Entscheiden** werden dem Nutzer durch das KI-System bereits Entscheidungen abgenommen (vgl. Abb. 1.4). Dies ist bei Suchprozessen im Internet bzw. in den sozialen Netzen der Fall. Hier werden dem Nutzer nach bestimmten (nicht transparenten) Algorithmen Informationen präsentiert bzw. vorenthalten. So entsteht die sogenannte Filter-Bubble, in der jeder in einer eigenen (Schein-)Welt lebt, die von der Realität mehr oder weniger weit entfernt sein kann (vgl. grundlegend Pariser 2017). Die heute verfügbaren KI-basierten Übersetzungsprogramme sollten auch nur für ein teilweises Entscheiden genutzt werden. Bei einer kritischen Analyse der heute erzielten Übersetzungsergebnisse können – gerade bei komplexeren Sachverhalten – immer noch viele Fehler festgestellt werden.

Beim **geprüften Entscheiden** werden Entscheidungsideen des Menschen durch ein KI-System überprüft – quasi als Anwendung einer Cross-Validierung (vgl. Abb. 1.4). Wenn das KI-System und der Mensch zum gleichen Ergebnis kommen, muss es passen. Beim **delegierten Entscheiden** werden (Teil-)Aufgaben vom Menschen ganz bewusst an ein KI-System verlagert. Dies ist vielfach bei Qualitätskontrollen in der Produktion der Fall; hier entscheiden entsprechende Systeme eigenständig, ob ein Produkt den Qualitätsanforderungen entspricht oder nicht. Beim **autonomen Entscheiden** werden ganze Aufgabenkomplexe an ein KI-System verlagert und dort ohne weitere menschliche Intervention durchgeführt. Dies ist beim Robo-Advisor der Fall, der eigenständige