

Klein | Dennerlein | Ritschl (Hg.)

# Health Care und Künstliche Intelligenz

Ethische Aspekte verstehen –  
Entwicklungen gestalten



## **Eine Arbeitsgemeinschaft der Verlage**

Brill | Schöningh – Fink · Paderborn  
Brill | Vandenhoeck & Ruprecht · Göttingen – Böhlau · Wien · Köln  
Verlag Barbara Budrich · Opladen · Toronto  
facultas · Wien  
Haupt Verlag · Bern  
Verlag Julius Klinkhardt · Bad Heilbrunn  
Mohr Siebeck · Tübingen  
Narr Francke Attempto Verlag – expert verlag · Tübingen  
Psychiatrie Verlag · Köln  
Ernst Reinhardt Verlag · München  
transcript Verlag · Bielefeld  
Verlag Eugen Ulmer · Stuttgart  
UVK Verlag · München  
Waxmann · Münster · New York  
wbv Publikation · Bielefeld  
Wochenschau Verlag · Frankfurt am Main

**PD Dr. Andreas Klein** lehrt Ethik, Dogmatik und Philosophie an der Evangelisch-Theologischen Fakultät der Universität Wien und Ethik im Gesundheitswesen an mehreren Hochschulen und Ausbildungsstätten.

**Prof. Dr. Sebastian Dennerlein** lehrt und forscht zur Regulation und Förderung von Lernen am Arbeitsplatz an der University Twente.

**Prof. Dr. Helmut Ritschl** leitet das Institut für Radiologietechnologie an der FH JOANNEUM in Graz und beschäftigt sich mit der digitalen Transformation im Gesundheitswesen und deren Implikationen.

Andreas Klein / Sebastian Dennerlein /  
Helmut Ritschl (Hg.)

# **Health Care und Künstliche Intelligenz**

Ethische Aspekte verstehen - Entwicklungen gestalten

Narr Francke Attempto Verlag · Tübingen

Umschlagabbildung: DALL-E 3 über ChatGPT4; Prompting: Christof Wolf-Brenner, Robert Gutounig & Sebastian Dennerlein

Bibliografische Information der Deutschen Nationalbibliothek

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

DOI: <https://doi.org/10.36198/9783838562575>

© 2024 · Narr Francke Attempto Verlag GmbH + Co. KG  
Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Alle Informationen in diesem Buch wurden mit großer Sorgfalt erstellt. Fehler können dennoch nicht völlig ausgeschlossen werden. Weder Verlag noch Autor:innen oder Herausgeber:innen übernehmen deshalb eine Gewährleistung für die Korrektheit des Inhaltes und haften nicht für fehlerhafte Angaben und deren Folgen. Diese Publikation enthält gegebenenfalls Links zu externen Inhalten Dritter, auf die weder Verlag noch Autor:innen oder Herausgeber:innen Einfluss haben. Für die Inhalte der verlinkten Seiten sind stets die jeweiligen Anbieter oder Betreibenden der Seiten verantwortlich.

Internet: [www.narr.de](http://www.narr.de)  
eMail: [info@narr.de](mailto:info@narr.de)

Einbandgestaltung: siegel konzeption | gestaltung  
CPI books GmbH, Leck

utb-Nr. 6257

ISBN 978-3-8252-6257-0 (Print)

ISBN 978-3-8385-6257-5 (ePDF)

ISBN 978-3-8463-6257-0 (ePub)



# Inhalt

Vorwort . . . . .	17
Einleitung   Andreas Klein, Sebastian Dennerlein und Helmut Ritschl .	19
1.    Hinführung . . . . .	19
2.    Begriffliche Annäherungen . . . . .	22
2.1  Künstliche Intelligenz (KI) . . . . .	22
2.2  Maschinelles Lernen – Machine Learning (ML) . . . . .	28
2.3  (Künstliche) Neuronale Netze (KNN) . . . . .	31
2.4  Deep Learning (DL) . . . . .	34
3.    Veranschaulichung einiger ethischer Herausforderungen und Lösungsansätze . . . . .	35
4.    Zu den Beiträgen dieses Buches . . . . .	39
Literatur . . . . .	44
Abschnitt 1: Grundlagen zu KI und erste ethische Überlegungen . . . . .	49
Data Science und Künstliche Intelligenz   Wolfgang Granigg und Klaus Lichtenegger . . . . .	51
1.    Von Big Data zur Künstlichen Intelligenz . . . . .	52
2.    Was ist Künstliche Intelligenz? . . . . .	53
3.    Eine kurze Geschichte der KI . . . . .	54
3.1  Die Anfänge: Rechnen und Codes knacken . . . . .	54
3.2  Logik und Symbole . . . . .	55
3.3  Maschinelles Lernen . . . . .	56
3.4  Statistisches Lernen . . . . .	56
3.5  Von der Natur das Lernen lernen . . . . .	57
3.6  Die Deep-Learning-Revolution . . . . .	58
3.7  Technische Infrastruktur . . . . .	59
3.8  What else? . . . . .	60
4.    Wie lernen Computerprogramme? . . . . .	61
4.1  Supervised Learning . . . . .	61

4.2	Unsupervised Learning . . . . .	63
4.3	Reinforcement Learning . . . . .	63
5.	Einige Herausforderungen im ML . . . . .	64
5.1	Underfitting und Overfitting . . . . .	64
5.2	Ausdruckskraft vs. Erklärbarkeit . . . . .	66
	Literatur . . . . .	66
	Wahrscheinlichkeit und Statistik – manchmal gegen unsere Intuition   Klaus Lichtenegger, Raphaelae Raab und Wolfgang Granigg . . . . .	69
1.	Zugänge zur Statistik . . . . .	70
2.	Verzerrungen in den Daten . . . . .	72
3.	Die Crux mit dem Mittelwert . . . . .	73
4.	Regression zur Mitte . . . . .	75
5.	Das Simpson-Paradoxon . . . . .	76
6.	Fallstricke bei der Datenvisualisierung . . . . .	77
7.	Wahrscheinlichkeit wider die Intuition . . . . .	78
8.	Bedingte Wahrscheinlichkeiten . . . . .	80
9.	Der Satz von Bayes . . . . .	81
	Literatur . . . . .	84
	Die Hintergründe von KI im Gesundheitswesen verstehen lernen   Marco Tilli, Michael Melcher, Debora Stickler und Raphaelae Raab . . . . .	85
	Vom Problem zum Machine Learning . . . . .	86
	Beispiel 1: Modellieren von BIP und Kindersterblichkeit . . . . .	86
	Was ist eine Lineare Regression? . . . . .	87
	Beispiel 2: Gruppieren von Brustkrebs-Merkmalen . . . . .	89
	Beispiel 3: Klassifikation COVID-19 vs. Grippe aufgrund der Symptome . . . . .	90
	Beispiel 4: Bildverarbeitung und -klassifikation . . . . .	93
	Aufbau eines CNNs . . . . .	95
	Arten von CNNs . . . . .	96
	Beispiel 5: Befunde verstehen und schreiben . . . . .	97
	Explainable AI im Medizinwesen . . . . .	98
	Was ist xAI eigentlich? . . . . .	98
	Wo xAI angewandt wird . . . . .	99
	Literatur . . . . .	100

Ethische Perspektiven eines verantwortungsbewussten Umgangs mit Künstlicher Intelligenz   Andreas Klein . . . . .	103
1.    Hinführung . . . . .	103
2.    Wie intelligent ist Künstliche Intelligenz – oder kann sie werden? . . . . .	104
3.    Ethik und die Frage nach dem guten Handeln . . . . .	108
4.    Ethik für KI . . . . .	109
4.1  Ethik-Leitlinien für eine vertrauenswürdige KI . . . . .	110
4.2  Ethische Verpflichtungen . . . . .	112
4.3  Grundrechte und KI . . . . .	113
4.4  Vier ethische Grundsätze (Prinzipien): . . . . .	115
5.    Abschließende Würdigung und Ausblick . . . . .	117
Literatur . . . . .	121

Abschnitt 2: Anwendungsbeispiele von KI-Anwendungen in unterschiedlichen Domänen des Gesundheitswesens . . . . .	127
--	-----

Artificial Intelligence und Machine Learning in der medizinischen Bilddatenverarbeitung   Wolfgang Birkfellner . . . . .	129
1.    Einführung . . . . .	129
2.    Welche Daten werden verwendet? . . . . .	131
3.    Eine unverbindliche Anleitung für Experimente . . . . .	133
4.    Anwendungsmöglichkeiten . . . . .	134
5.    Stärken, Schwächen und Bedrohungen . . . . .	136
6.    Auswirkungen auf das Berufsbild . . . . .	137
7.    Ausblick und Herausforderungen . . . . .	139
Literatur . . . . .	140

ChatGPT als Arzt?   Lars Mehnen, Stefanie Gruarin, Mina Vasileva und Bernhard Knapp . . . . .	143
1.    Zusammenfassung . . . . .	143
2.    Einführung . . . . .	143
3.    Beschreibung der Untersuchungsmethode . . . . .	144
3.1  Schritt 1: Ursprung der klinischen Fallvignetten . . . . .	144
3.2  Schritt 2: Verwendung von ChatGPT . . . . .	144
3.3  Schritt 3: Bewertung der richtigen Antworten . . . . .	145
3.4  Schritt 4: Darstellung der diagnostischen Genauigkeit . . . . .	145



4.	Ergebnisse des Experiments . . . . .	145
4.1	Diagnostische Genauigkeit von ChatGPT bei häufigen Erkrankungen . . . . .	145
4.2	Diagnostische Genauigkeit von ChatGPT bei seltenen Erkrankungen . . . . .	146
5.	Diskussion der Ergebnisse aus dem Experiment . . . . .	147
5.1	ChatGPT (Version 3.5 und 4) erreicht bemerkenswerte Genauigkeiten . . . . .	147
5.2	Lernt ChatGPT nur auswendig? . . . . .	147
5.3	ChatGPT kann / soll keinen menschlichen Arzt ersetzen .	149
	Literatur . . . . .	150

### Aktuelle Anwendungsszenarien und -beispiele von KI-Systemen in Diagnostik und Therapie | Bianca Buchgraber-Schnalzer und Bernhard

Neumayer . . . . .	153	
1.	Einleitung . . . . .	153
2.	Medizinische Bildgebung . . . . .	155
2.1	Bildrekonstruktion . . . . .	155
2.2	Bildanalyse . . . . .	157
3.	Kardiologische Erkrankungen . . . . .	159
4.	Mentale Gesundheit bzw. psychische Erkrankungen . . . . .	160
5.	Physiotherapeutische KI-Support-Tools . . . . .	162
6.	Kognitive Beeinträchtigungen und Demenz . . . . .	163
7.	Dermatologie und chronisches Wundmanagement . . . . .	165
8.	Diskussion . . . . .	167
	Literatur . . . . .	168

### Federated Learning | Hannes Hilberger, Helmut Ahammer und Markus

Bödenler . . . . .	175	
1.	Einleitung . . . . .	175
2.	Technische Grundlagen . . . . .	177
3.	Herausforderungen mit Federated Learning . . . . .	183
4.	Aktuelle Anwendungen von Federated Learning im Gesundheitsbereich . . . . .	188
5.	Zusammenfassung und Ausblick . . . . .	190
	Literatur . . . . .	192

Medizinprodukte mit KI in der klinischen Praxis   Martin Baumgartner, Aaron Lauschensky, Hannes Perko, Tobias Allgeier, Stefan Beyer und Klaus Donsa . . . . .	197
1.    Einleitung . . . . .	197
1.1    Hintergrund und Bedeutung von KI in der klinischen Praxis	197
1.2    Zielsetzung und Struktur des Kapitels . . . . .	198
2.    Beispiel 1: Regelbasierter Algorithmus beim telemedizinischen Monitoring von Patienten mit Herzinsuffizienz . . . . .	199
2.1    Beschreibung des Medizinprodukts . . . . .	199
2.2    Funktionsweise der KI-Anwendung . . . . .	200
2.3    Bewertung aus medizinischer, regulatorischer, technischer und ethischer Sicht . . . . .	202
3.    Beispiel 2: Deep-Learning-basierte KI-Anwendung zur EEG-Analyse . . . . .	204
3.1    Beschreibung des Medizinprodukts . . . . .	204
3.2    Funktionsweise der KI-Anwendung . . . . .	206
3.3    Bewertung aus medizinischer, regulatorischer, technischer und ethischer Sicht . . . . .	208
4.    Fazit . . . . .	210
4.1    Was bedeutet das „Prädikat“ „Medizinprodukt mit KI“ für den Aufwand der Entwicklung und auch später im Routineeinsatz? . . . . .	210
4.2    Welche Vorteile bietet der Einsatz von KI in Medizinprodukten im Vergleich zu herkömmlichen Produkten ohne KI? . . . . .	211
4.3    Welche Herausforderungen und Risiken sind mit der Integration von KI in Medizinprodukten verbunden? . . . .	212
4.4    Welche Fähigkeiten und Schulungen sind erforderlich, um Lösungen, die KI einsetzen, in Medizinprodukten anzuwenden? . . . . .	212
Literatur . . . . .	213

Moderner Datenschutz und vertrauenswürdige KI   Lea Demelius, Michael Jantscher und Andreas Trügler . . . . .		217
1.	Künstliche Intelligenz im Gesundheitsbereich . . . . .	217
1.1	Vertrauenswürdige KI . . . . .	218
1.2	Datenschutz und Privatsphäre . . . . .	219
2.	Technische Datenschutz-Maßnahmen für KI-Anwendungen . . . . .	221
2.1	Homomorphe Verschlüsselung . . . . .	221
2.2	Differential Privacy . . . . .	223
2.3	Entwicklungen im Bereich Maschinelles Lernen . . . . .	225
3.	Anwendungen und Beispiele . . . . .	227
3.1	KI-Analyse von Patient:innenakten . . . . .	227
3.2	Mobilität und Ausbreitung von Infektionskrankheiten . . . . .	229
4.	Zusammenfassung . . . . .	229
	Literatur . . . . .	231
Ethische Aspekte von KI in der präklinischen Krebsforschung   Claire Jean-Quartier und Fleur Jeanquartier . . . . .		235
1.	Einleitung . . . . .	235
2.	Beispiele für ethische Aspekte von KI in der präklinischen Krebsforschung . . . . .	237
2.1	Ersatz von Tierversuchen durch in silico-Ansätze . . . . .	237
2.2	Transparenz von KI und Verständlichkeit von Modellen . . . . .	238
2.3	Nachhaltige KI und moralische Entscheidungsprinzipien . . . . .	241
2.4	Offene Forschung im Sinne der Zugänglichkeit zum Nutzen der Gesellschaft . . . . .	242
	Literatur . . . . .	244
Digitalisierung in der Pharmaindustrie   Sarah Stryeck und Johannes Khinast . . . . .		249
1.	Einführung in die Digitalisierung in der pharmazeutischen Industrie . . . . .	249
2.	Digitalisierung in der Wirkstoffentdeckung und -entwicklung . . . . .	254
3.	Digitalisierung in der Produktion . . . . .	255
4.	Herausforderungen bei der Digitalisierung der PI . . . . .	259

5.	Chancen durch KI-gestützte Verfahren in der PI . . . . .	261
5.1	Effizientere Versorgung mit Arzneimitteln (aus Europa) . .	261
5.2	Qualität und Transparenz . . . . .	262
5.3	Technologiesouveränität und Nachhaltigkeit . . . . .	262
5.4	Bessere Patient:innenversorgung . . . . .	263
	Literatur . . . . .	264
Abschnitt 3: Ethische und rechtliche Aspekte von KI-Anwendungen im Gesundheitswesen . . . . .		267
Kompetenzen ethischer Reflexionen   Andreas Klein . . . . .		269
1.	EU Ethik-Leitlinien und KI-Anforderungen . . . . .	269
1.1	Verwirklichung einer vertrauenswürdigen KI: Anforderungen an KI-Systeme . . . . .	270
1.2	Die Bewertungsliste (ALTAI) . . . . .	276
2.	Folgerungen aus den Ethik-Leitlinien für die Praxis . . . . .	277
2.1	Ethikkodizes . . . . .	278
2.2	Ethikkommissionen . . . . .	280
2.3	Der AI Act . . . . .	283
3.	Das MEESTAR-Modell . . . . .	289
	Literatur . . . . .	292
	Anhang . . . . .	294
Methoden und Tools zur ethischen Reflexion in der agilen Entwicklung von Künstlicher Intelligenz   Sebastian Dennerlein, Christof Wolf-Brenner, Robert Gutounig, Stefan Schweiger und Viktoria Pammer-Schindler . . . . .		301
1.	Keine ethisch verantwortungsvolle KI ohne Reflexion . . .	301
2.	Zum Verständnis von ethischer Reflexion und relevanten Charakteristiken . . . . .	302
3.	Zur Verortung ethischer Reflexion im Entwicklungsprozess . . . . .	304
4.	Von ethischen Prinzipien zu deren Berücksichtigung in der Praxis . . . . .	306
5.	Darstellung und Illustration von sieben Methoden und Tools zur ethischen Reflexion . . . . .	307
5.1	Methoden und Tools 1: Data Skills Framework . . . . .	309

5.2	Methoden und Tools 2: Data Ethics Maturity Model . . . . .	311
5.3	Methoden und Tools 3: Assessment List for Trustworthy AI (ALTAI) . . . . .	313
5.4	Methoden und Tools 4: MEESTAR – Modell zur Ethischen Evaluierung Soziotechnischer Arrangements . . . . .	315
5.5	Methoden und Tools 5: DEDA – Data Ethics Decision Aid	317
5.6	Methoden und Tools 6: Ethics in Tech Practice – A Toolkit	318
5.7	Methoden und Tools 7: Artificial Intelligence Incident Database (AIID) . . . . .	320
6.	Diskussion offener Herausforderungen in der ethisch reflektierten Gestaltung von KI . . . . .	320
7.	Reflexionsfragen . . . . .	323
	Literatur . . . . .	324
Künstliche Intelligenz in der Medizin   Matthias Wendland . . . . .		329
1.	Einleitung . . . . .	329
2.	Anwendungsgebiete der KI in der Medizin . . . . .	330
3.	Spezifische Risiken der KI in der Medizin . . . . .	331
3.1	Fehlerhafte Diagnostik und Therapieentscheidungen . . . . .	331
3.2	Verzerrungen (Biases) . . . . .	332
3.3	Datenschutz und Datenmissbrauch . . . . .	333
Regulatorische Rahmenbedingungen für KI-basierte Medizinprodukte   Sabrina Linzer, Christoph Matoschitz und Klaus Donsa . . . . .		341
1.	Einleitung . . . . .	341
1.1	Hintergrund und Bedeutung der regulatorischen Rahmenbedingungen . . . . .	341
1.2	Zielsetzung des Kapitels . . . . .	342
2.	Regulatorische Anforderungen für Medizinprodukte mit KI	342
2.1	Medizinprodukteverordnung . . . . .	342
2.2	Klassifizierung von Medizinprodukten . . . . .	345
2.3	Konformitätsbewertungsverfahren und CE-Kennzeichnung . . . . .	345
2.4	Anforderungen an die Technische Dokumentation . . . . .	349
2.5	Entwicklung von KI-basierter Software als Medizinprodukt	349

3.	Wertvolle Orientierungshilfen bei der Entwicklung und beim Einsatz in der klinischen Praxis . . . . .	350
3.1	Praktische Umsetzung: Normen, Spezifikationen und Leitfäden . . . . .	350
3.2	Verantwortung von Herstellern und Anwendern von Medizinprodukten . . . . .	355
3.3	Haftungsfragen bei Fehlern oder Schäden durch KI-Anwendungen . . . . .	356
3.4	Datenschutz und Datensicherheit . . . . .	358
4.	Fazit . . . . .	359
	Literatur . . . . .	360
	Abschnitt 4: Konsequenzen von KI für die Gesundheitsversorgung.	
	Transformation der Handlungsfelder in Gesundheitsberufen . . . . .	363
	Einbettung von KI und Ethik in Curricula der Gesundheitsberufe am Beispiel eines cMOOCs   Helmut Ritschl, Waltraud Jelinek-Krickl, Rupert Beinbauer, Julia Tomanek, Bianca Buchgraber-Schnalzer und Marco Tilli . . . . .	365
1.	Einbettung neuer inhaltlicher Entwicklungen im beruflichen Handlungsfeld der Gesundheitsberufe . . . . .	365
2.	Beschreibung der neuen Modulkonstruktion: Didaktik, Kompetenz, Kompetenzlevels, Lernziele . . . . .	368
3.	Strategie zur Identifikation von Themen und Inhalten zu KI-Anwendungen in einer konkreten Lehrveranstaltung .	372
4.	Diskussion der Tiefe und der Methode der Wissensvermittlung – didaktische Reduktion . . . . .	375
5.	Mögliche Erfolgsfaktoren zur Einbettung der neuen Lehrinhalte zum Thema KI und Ethik . . . . .	377
6.	Muster eines cMOOCs zur Einführung in die KI für Gesundheitsberufe am Beispiel des Handlungsfeldes Radiologietechnologie . . . . .	378
	Literatur . . . . .	383
	Veränderung des Berufsbildes für Fachärzt:innen der Radiologie   Erich Sorantin, Ariane Hemmelmayr und Michael Georg Grasser . . . . .	387
1.	Hinführung und erste Überlegungen . . . . .	387

2.	Der Workflow in der Radiologie als Ausgangspunkt möglicher Transformationen . . . . .	389
2.1	Veränderung des Arbeitsfeldes „Clincial Decision Support“	389
2.2	Veränderung des Arbeitsfeldes in der Bildakquisition und Rekonstruktion . . . . .	390
3.	Veränderung des radiologischen Befund-Arbeitsplatzes und der Befunderstellung . . . . .	393
4.	Der blinde Fleck – Cybersicherheit und Datenschutz in der Radiologie – ein neues Handlungsfeld rückt immer näher	393
4.1	Datenschutzrechtliche Sicherheitsaspekte . . . . .	396
5.	Zusammenfassung . . . . .	399
	Literatur . . . . .	399

Künstliche Intelligenz und die Veränderung der Handlungsfelder von nicht-ärztlichen Gesundheitsberufen   Helmut Ritschl, Andreas Jocham, Wolfgang Staubmann, Dalibor Jeremic, Eva Mircic, Felix Mühlensiepen und Lucia Ransmayr . . . . .	403
Ad (I): Exemplarische Entwicklungen der Gesundheits- und Krankenpflege durch KI-Anwendungen . . . . .	405
Ad (II): Exemplarische Entwicklungen der Diätologie durch KI-Anwendungen . . . . .	407
Ad (III): Exemplarische Entwicklungen in der Radiologietechnologie durch KI-Anwendungen . . . . .	410
Ad (IV): Exemplarische Entwicklungen in der biomedizinischen Analytik durch KI-Anwendungen . . . . .	412
Ad (V): Exemplarische Entwicklungen der Logopädie durch KI-Anwendungen . . . . .	414
Ad (VI): Exemplarische Entwicklungen der Physiotherapie durch KI-Anwendungen . . . . .	416
Fazit und Schlussfolgerung aus den Betrachtungen der nicht-ärztlichen Gesundheitsberufe . . . . .	417
Literatur . . . . .	419

Was wollen wir von dem, was wir technisch können, realisieren?   Christof Wolf-Brenner, Nina Wolf-Brenner und Martin Semmelrock . .	425
1. Einführung . . . . .	425
1.1 Eine typische Aufnahme . . . . .	426
1.2 Herausforderungen im Aufnahmeprozess . . . . .	429
2. Eine Vision für KI im Aufnahmeprozess . . . . .	431
2.1 Self-Service Triage . . . . .	431
2.2 Schätzung des täglichen Zustroms und Abstroms von Patient:innen . . . . .	432
2.3 Empfehlungen zur Auswahl der Laboruntersuchungen . . .	433
3. Ethische Herausforderungen und Erwägungen . . . . .	433
Literatur . . . . .	436
 KI zur Optimierung von Patient:innen-Flüssen im Gesundheitswesen   Daniel Pözl, Robert Darkow, Susann May, Gernot Reishofer und Helmut Ritschl . . . . .	439
1. Hintergrund / Ausgangssituation . . . . .	439
2. Gesundheitskommunikation mittels KI-basierten Chatbots und NLP-Übersetzer zur Unterstützung der Patient:innen-Flüsse . . . . .	441
3. Gesundheitsvorsorge/Gesundheitsbeobachtung mittels AI gestütztem SMART Health Monitoring . . . . .	445
4. Autonome KI-gesteuerte Drohnen zur Unterstützung in Medikamentenzulieferung, Notfallmedizin, Katastrophenmanagement sowie Search and Rescue . . . . .	449
5. Fazit für die Unterstützung von Patient:innen-Flüssen durch KI Anwendungen . . . . .	450
Literatur . . . . .	450
 Verzeichnis der Autorinnen und Autoren . . . . .	455





## Vorwort

Das vorliegende Buch geht zu guten Teilen auf eine Webinarreihe mit dem Titel „AI in Healthcare & Ethics“ vom Herbst 2021 zurück. Diese Vortragsreihe verfolgte die Absicht, Expert:innen, Studierende und die interessierte Öffentlichkeit zusammen zu bringen, um durch den Austausch von Wissen, Sorgen und Ideen einen konstruktiven Dialog und gemeinsamen Weg zu fördern. Sie bestand aus insgesamt sieben Sessions zu jeweils zwei Vorträgen durch ausgewiesene Expert:innen in ihrem Tätigkeitsfeld mit anschließender Diskussion. Die Expert:innen präsentierten grundlegende Konzepte zu Künstlicher Intelligenz (KI) sowie aktuelle Anwendungsbeispiele im Kontext des Gesundheitswesens und diskutierten diese mit den Teilnehmer:innen. Dabei wurden die zahlreichen Chancen und Weiterführungen durch KI-basierte Systeme beispielsweise im Rahmen von Diagnostik und Therapie aufgezeigt, aber auch die aktuellen Begrenzungen und grundsätzlichen Problematiken. Vor dem Hintergrund der gut besuchten Webinarreihe, die für alle Beteiligten eine erhebliche Bereicherung darstellten, legte sich den Herausgebern die Überlegung nahe, die Inhalte und Einsichten einer breiteren Öffentlichkeit in Form eines Sammelbandes zur Verfügung zu stellen.

Der vorliegende Band greift die reichhaltigen Aspekte aus Präsentation und Diskussion auf und führt sie durch Integration neuer Erkenntnisse und Themen vertieft weiter. Die Entwicklung und Anwendung der KI schreitet derart rasant voran, dass beinahe jeder einzelne Beitrag fortwährend mit neuen Studien ergänzt werden müsste. Insofern bedarf es auch eines gewissen Mutes zur Lücke und zur Fragmentarizität – wie überhaupt jedes Leben stets nur fragmentarisch bleiben kann („vita brevis“).

Das **Konzept des Buches** zielt zum einen darauf ab, einen **gut verständlichen Überblick** über die zentralen Themenstellungen zu bieten und damit auch als **Lehrbuch** zu dienen. Zum anderen möchten die einzelnen Beiträge auch **in die Tiefe** gehen und aus den verschiedenen Forschungsgebieten auch ‚Insidern‘ neue Erkenntnisse erlauben. Diese ambitionierte Doppelstrategie, nämlich spezifische Fachkompetenzen im KI-Bereich mit nachvollziehbaren Darstellungen zu verbinden, ermöglicht den Bezug zu vielfältigen Handlungsfeldern, so dass unterschiedliche Ausbildungs- und Berufsgruppen davon profitieren können. Hier sind etwa zu nennen die Berufe rund um die Herstellung und Konstruktion von KI-Systemen für kon-

krete Anwendungsszenarien, sodann die zahlreichen Gesundheitsberufe im ärztlichen und nicht-ärztlichen Bereich und in der medizinischen Forschung samt ihren Standesvertretungen, weiterhin auch Ethiker:innen, die hierzu ein valides Reflexionsverhältnis erarbeiten wollen oder sollen, vielfältige Ausbildungsstätten und Hochschulen und schließlich End-User (z. B. Patient:innen, pflegende Angehörige usw.) oder auch Entscheidungsträger in Gesellschaft und Politik. Insofern adressiert das Buch eine breite Leserschaft.

In allen jenen Handlungsfeldern geht es letztlich um einen **verantwortungsvollen Umgang** mit diesen neuen und teilweise auch disruptiven Entwicklungen, die schon jetzt deutlich in den gesellschaftlichen Subsystemen als Herausforderungen wahrgenommen werden. Quer durch sämtliche Beiträge wird deutlich, dass Verantwortung hier nicht einfach verschoben und delegiert werden kann bzw. darf, sondern auf jeder Ebene rational und proaktiv zu übernehmen ist. Eine einseitige (aber durchaus klassische) Verantwortungszuschreibung entweder an Teams von Entwickler:innen und Betreiber:innen oder umgekehrt an Anwender:in und User ist je für sich unplausibel und unreflektiert. Verantwortung besteht stets für die eigenen Handlungen.

Die Herausgeber und Autor:innen dieses Buches beabsichtigen, die Themenstellungen rund um KI im Gesundheitswesen für eine große Leser:innengruppe zugänglich zu machen. Es sollte für jede Gruppe möglich sein, sich ein eigenes, begründetes Urteil zu bilden. Denn auch hier geht es um einen fortwährenden, (vorläufig) unabschließbaren kritischen Diskurs. Erst die aktive Teilnahme am Diskurs schafft Bewusstsein, und erlaubt uns die Zukunft rund um das Thema konstruktiv mitzugestalten und Verantwortung zu übernehmen.

Dank gilt daher zunächst allen Teilhabenden der seinerzeitigen Webinarreihe sowie den beitragenden Autor:innen. Wir danken aber insbesondere auch dem Verlag Narr Francke Attempto, der von Beginn an seine Bereitschaft zu dieser Publikation erklärt hat. Seitens des Verlages ist es vor allem Herr Stefan Selbmann, der sämtliche Vorarbeiten und Prozesse umsichtig und geduldig begleitet und betreut hat. Darüber hinaus gilt auch der Verlagsgemeinschaft UTB großer Dank für die Übernahme des Buches in diese Reihe. So möge das vorliegende Sammelwerk nun seine Leser:innen und Anwendung in Ausbildung und Praxis finden.

Andreas Klein, Sebastian Dennerlein und Helmut Ritschl  
Wien, Twente und Graz im Winter 2023

# Einleitung

Andreas Klein, Sebastian Dennerlein und Helmut Ritschl

## 1. Hinführung

Künstliche Intelligenz (KI) hat spätestens seit der öffentlichen Verfügbarkeit von ChatGPT im November 2022 die breite Bevölkerung erreicht und dort erhebliche Wirkungen erzielt. Seitdem überschlägt sich die mediale Berichterstattung im Tagesrhythmus mit Neuerungen, Veränderungen, euphorischen oder kritischen Stellungnahmen hierzu. Ergänzt wurde das Spektrum an ‚Conversational Agents‘ nach ChatGPT durch die neue Bing-Suche im Edge-Browser, Copilot-Funktionen in Windows 11 oder Bard von Google.<sup>1</sup> Weitere (Gratis-) Anbieter:innen stehen bereits in den Startlöchern. Was generative KI-Systeme dieser Art können, wird jedem rasch begreiflich, der sie ausprobiert: Sie liefern erstaunliche Ergebnisse zu unterschiedlichsten, in normaler Sprache formulierten Suchanfragen – und die Antworten sind zudem in überzeugend guter Sprachqualität formuliert. Schrittweise werden weitere Modi (Funktionen) wie etwa Bild- bzw. Grafikgenerierung (z. B. Dall-E in der Bing-Suche) integriert und der Öffentlichkeit zur Verfügung gestellt.

Was KI überhaupt ist und dementsprechend KI-Systeme charakterisiert, wo ihre Probleme liegen, welche Potenziale und existenziell-gesellschaftlichen Herausforderungen sie in sich bergen, ist weit weniger bekannt. Vielfach (bzw. Vielen) reicht es schon, dass sie (irgendwie) funktionieren und einen Output generieren, der sowohl individuell als auch öffentlich oder unternehmerisch äußerst nützlich zu sein scheint.

Längst ist **KI** auch **im Gesundheitswesen** angekommen, selbst wenn sie dort aufgrund der heiklen Problemkonstellation äußerst zögerlich integriert wird. Die Vorteile und Chancen, die auch im vorliegenden Sammelband in den unterschiedlichsten Domänen dargestellt werden, liegen auf der

---

1 Erwähnt sei an dieser Stelle, dass es freilich auch Open-Source-Projekte gibt, die hier nicht übergangen werden sollten, allerdings in der Öffentlichkeit kaum eine relevante Rolle spielen. Beispielsweise wäre auf die Open Source Bildgenerations-Plattform hinzuweisen: <https://stablediffusion.fr/>.

Hand: KI-Systeme eröffnen Verbesserungen in so gut wie allen Bereichen des Gesundheitswesens. In Zukunft wird kein Weg daran vorbeiführen und unzählige Forschungsprogramme sind bereits angelaufen, wie schon ein flüchtiger Blick in entsprechende Literaturdatenbanken oder nationale wie internationale Förderprogramme zeigt.

Die Zögerlichkeit einer standardisierten Integration von KI-Systemen hat aber auch berechtigte Gründe. Der Gesundheitsbereich ist ein äußerst prekäres gesellschaftliches System, so dass hier mit außerordentlicher Vorsicht vorgegangen werden muss. Gemäß AI Act, also dem im Dezember 2023 beschlossenen EU-weiten regulatorischen Instrument<sup>2</sup>, handelt es sich hier um einen technologischen **Hochrisikobereich**, der adäquater Rahmenbedingungen bedarf. Fehler an irgendeiner Stelle können gravierende, inakzeptable Konsequenzen nach sich ziehen. Darüber hinaus sind zahlreiche ethische Fragestellungen zu bearbeiten und klären, nicht zuletzt die, wie sich die gesamte Interaktion im Gesundheitswesen durch die Integration KI-gestützter Systeme transformieren könnte und in welcher Weise dies als wünschenswert oder kritikbedürftig betrachtet wird. Sämtliche Schnittstellen im Gesundheitswesen wären hiervon betroffen, bis hinein in die Weiterentwicklung von Gesundheitsberufen und ihre Kompetenzen. Eine umsichtige und umfassende Abwägung von Chancen und Risiken im Sinne der Patient:innensicherheit ist also unentbehrlich.

Diesen vielschichtigen Bezügen und Perspektiven ist der vorliegende Band verpflichtet. Die **Anlage des Buches** ist dabei so gestaltet, dass in vier Abschnitten thematisch zusammengehörende Themenkomplexe dargestellt werden. Im **ersten Teil** werden schwerpunktmäßig Grundlagen der bzw. zu KI erläutert. Hierzu zählt neben den wichtigen theoretischen und technischen Konzepten und deren Beziehungen zum Gesundheitswesen auch eine ethische Grundorientierung mit fundamentalen Beurteilungsoptionen. Im **zweiten Abschnitt** werden Anwendungsbeispiele ethisch verantwortungsvoller KI-Systeme im Gesundheitswesen erörtert. Im **dritten Teil** werden ethische und rechtliche Aspekte von KI beschrieben und diskutiert. Der **vierte Bereich** des Buches widmet sich den Konsequenzen von KI im Blick auf die Gesundheitsversorgung bzw. auf die Handlungsfelder einzelner Gesundheitsberufe.

---

2 Siehe hierzu auch aktuell: <https://www.europarl.europa.eu/news/en/press-room/2023/1206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>.

Sämtliche Beiträge stellen sich als Querschnittsthematik unterschiedlichen **ethischen Fragen**, die sich unmittelbar aus konkreten KI-Anwendungen im jeweiligen Kontext ergeben. Hierin spiegelt sich eine Grundüberzeugung des gesamten Buches und aller Autor:innen wider: (Neue) Technologien sind nicht wertneutral, sondern beruhen auf Zielsetzungen, moralischen Zwecksetzungen und Wertungen, da Hersteller:innen ihrerseits in sozialen Kontexten mit inhärenten moralischen Ansprüchen und Zielsetzungen verwurzelt sind und Technologien zudem unmittelbar ethische Fragen auslösen. Die sog. **Neutralitätsthese**, wonach Technik ganz allgemein zweckfrei und insofern wertneutral ist und konkrete Zwecke allererst bei der Verwendung ins Spiel kommen, erscheint mittlerweile selbst als naiv (vgl. Fenner 2022: 270 ff). Schon für die Herstellung oder Planung technologischer Systeme sind bestimmte Zwecke und Realisierungsvorstellungen leitend<sup>3</sup>, damit aber auch die Inkaufnahme eventueller Risiken oder Nachteile. Insofern sind solche Konstrukte nicht wertneutral, sondern umgekehrt enorm **werthaltig** – und deshalb bedürfen sie auch von Anfang an der ethisch-kritischen Reflexion. Moralische **Verantwortung** auf allen Ebenen kann demzufolge als gemeinsame Querschnittsmaterie aller Beiträge dieses Buches betrachtet werden (vgl. auch Nida-Rümelin 2011; Bayertz 1991, 1995; Wittwer 2002).

Im Rahmen dieser **Einleitung** sollen bereits ein paar **Grundlagen und Begriffe** vorgestellt und reflektiert werden, um die folgenden Sektionen vorzubereiten und Leser:innen schrittweise an die Thematik heranzuführen. Dies betrifft zunächst einmal den Begriff der „Künstlichen Intelligenz“ selbst. Hieran schließen sich weitere Bereiche an wie beispielsweise Machine Learning, (Künstliche) Neuronale Netze, Deep Learning, aber auch verschiedene etablierte Lernmethoden für KI-Systeme. In einschlägigen Diskursen werden an dieser Stelle regelmäßig unterschiedliche Verständnisse vorausgesetzt, so dass bereits hier mögliche Missverständnisse vorprogrammiert sind (auf spezifisch philosophische Erörterungen darüber, was überhaupt unter einem „Begriff“ zu verstehen ist, wird hier verzichtet; vgl. Seiffert 1975: 15 ff). Eine historische Darstellung der Theorie- und Entwicklungsgeschichte von KI wird

---

3 Dies hat auf seine Weise v. a. Martin Heidegger (1967) im Kontext von „Zeug“ und „Werkzeug“ (§ 15) schon deutlich hervorgehoben (68): „*Ein Zeug ‚ist‘ strenggenommen nie. Zum Sein von Zeug gehört je immer ein Zeugganzes, darin es dieses Zeug sein kann, das es ist. Zeug ist wesenhaft ‚etwas, um zu ...‘. Die verschiedenen Weisen des ‚Um-zu‘ wie Dienlichkeit, Beiträglichkeit, Verwendbarkeit, Handlichkeit konstituieren eine Zeuganzheit. In der Struktur ‚Um-zu‘ liegt eine Verweisung von etwas auf etwas.“ Das „herzustellende Werk“ ist „das *Wozu*“ des Werkzeugs (70).*

hier aus Platzgründen ausgespart (vgl. Heinrichs / Heinrichs / Rütter 2022: 1ff; Russell / Norvig 2012: 39ff; Eberl 2018).

Die Einleitung beschließen wird ein **Kurzüberblick über die einzelnen Beiträge**.

## 2. Begriffliche Annäherungen

### 2.1 Künstliche Intelligenz (KI)

Betritt man den Themenbereich „Künstliche Intelligenz“ (engl. Artificial Intelligence), fällt zunächst eine unüberschaubare Anzahl von Definitionen, Beschreibungen und Charakterisierungen auf. Hinzu kommen unterschiedliche (technische oder mathematische) Konzepte, die ebenfalls unter KI subsumiert werden. Dieser Umstand führt in Debatten fortwährend zu Unklarheiten und wechselseitigen Missverständnissen. Dementsprechend fallen auch die jeweiligen Einschätzungen recht different aus: Handelt es sich hier überhaupt um eine Form von Intelligenz oder doch nur um verständnislose Algorithmen? Sind künftige KI-Systeme eine massive Bedrohung für die Menschheit und den Planeten Erde (z. B. ChaosGPT<sup>4</sup>) – wogegen dann entsprechende Maßnahmen zu ergreifen wären (so etwa Elon Musk oder die Rede von einem Fukushima der KI<sup>5</sup>) – oder doch nur ein Mailüftchen im großen Pool technologischer Entwicklungen?

Zumindest hat die breite Bevölkerung durch die omnipräsente Zugänglichkeit zu ChatGPT, GPT-3.5 bzw. 4, Bard, usw. mittlerweile einen Eindruck davon gewonnen, was sich mit diesen großen (und teilweise schon **multimodalen**) **Sprachmodellen (LLM: Large Language Model)** erreichen lässt. Durch die sprachliche Zugänglichkeit (Conversational Agent/AI) werden diese generativen Systeme auch schon ausgiebig genutzt. Kürzlich hat Google mit Gemini bereits die nächste, offenbar deutlich stärkere (multimodale) Generation in Umlauf gebracht. Das bedeutet allerdings nicht, dass man

---

4 Vgl. dazu etwa: <https://www.youtube.com/watch?v=g7YJIpkk7KM>, oder: <https://the-decoder.de/chaosgpt-ist-der-erste-konkrete-versuch-mit-ki-die-menschheit-zu-vernichten/>.

5 Siehe etwa das Interview mit Thomas Metzinger: <https://www.spektrum.de/news/interview-die-unterschaetzten-risiken-der-kuenstlichen-intelligenz/1377620> (zuletzt abgerufen: 17.12.2023).

auch verstanden hat, was diese Systeme sind, was sie ausmacht, wie sie funktionieren und worin ihre Problemfelder liegen.

Eine zentrale Schwierigkeit mit dem Begriff „Künstliche Intelligenz“ steckt in einem angemessenen Verständnis von „**Intelligenz**“. Diese Problematik betrifft jedoch nicht nur künstliche Systeme (vgl. Müller 2024), sondern auch biologische. Je nachdem, was in den Intelligenzbegriff integriert wird, differiert die Zuschreibung, welche Entitäten Intelligenz haben (können), zu welchen Zeiten oder Gelegenheiten (z. B. nur im Wachzustand) oder in welchen Graden. Je höher die Anforderungen an den Intelligenzbegriff angesetzt werden, umso weniger Entitäten fallen letztlich hierunter – und manche Entitäten, wie etwa der Mensch, auch nur zu bestimmten Gelegenheiten. Anstatt sich mit dem Intelligenzbegriff bei künstlichen Systemen herumzuschlagen, wäre es auch möglich, direkt z. B. von **Maschinellern Lernen (Machine Learning: ML<sup>6</sup>)** zu sprechen. Einem begrifflichen Wechsel steht jedoch entgegen, dass sich einerseits der Begriff KI längst gesellschaftlich und wissenschaftlich etabliert hat und sich andererseits auch bei einem Sprachwechsel analoge Probleme ergeben dürften.

Hinzu kommt der (erschwerende) Umstand, dass mit KI unterschiedliche Bereiche adressiert werden können: ein wissenschaftliches Forschungsgebiet, eine technologische Methode, konkrete Applikationen (wie etwa im Gesundheitswesen) oder philosophische Erörterungen zu grundlegenden Begriffen und Verständnissen.

Häufig werden mit „Intelligenz“ ganz bestimmte grundlegende **Fähigkeiten** oder **Kompetenzen** verbunden (vgl. Legg / Hutter 2007<sup>7</sup>; König u. a. 2022), wie etwa (logisches) Denken, Planen, Problemlösung, Wahrnehmung, Lernen, Informationsverarbeitung, Verarbeitung von (natürlicher) Sprache und evtl. Metakognition (Denken über das Denken oder über Lernprozesse<sup>8</sup>). Diese Fähigkeiten müssen (biologische oder künstliche) Systeme nicht zu allen Zeiten in gleicher Weise ausüben. Darüber hinaus kann mit dieser Zuschreibung Bezug zu bestimmten **Interaktionen mit der Umwelt** oder **Akteursfähigkeiten** genommen werden: z. B. räumliche, soziale, körperlich-kinästhetische, musikalische, sprachliche, logisch-mathematische oder inter- und intrapersonelle Intelligenz. Umstritten ist jedoch (vgl. Heinrichs / Heinrichs / Rüter 2022: 16 ff), inwieweit sämtliche dieser Fähigkei-

---

6 Siehe dazu vor allem die ersten drei Buchbeiträge.

7 Siehe dazu auch den ersten Buchbeitrag von Klein.

8 Siehe dazu auch den Buchbeitrag von Dennerlein u.a.



ten – und in welcher Reihenfolge und Gruppierung – realisiert sein müssen oder etwa nur grundlegende (z. B. Lernen oder Anpassungsfähigkeit).

Mit dem Gesagten wird nun ein weiterer wichtiger Aspekt in der Debatte um KI sichtbar, nämlich eine explizite oder implizite Ausrichtung an menschlicher Intelligenz, also eine sog. **anthropozentrische Verkürzung** (vgl. König u. a. 2022: 23; Müller 2024). Dies kann dann jedoch zu einer einseitigen, vereinfachten Sicht auf den Diskussions- und Untersuchungsgegenstand führen. In der Ethik würde an dieser Stelle der bekannte Vorwurf des **Speziesismus** ansetzen, wonach Intelligenz primär oder ausschließlich dem Menschen vorbehalten wäre (Artegoismus). Es ist jedoch inzwischen weitgehend Konsens, auch Tieren eine bestimmte Form von Intelligenz – wenn auch in unterschiedlichen Graden – zumindest nicht abzusprechen; von Phänomenen wie Schwarmintelligenz oder kollektiver Intelligenz ganz zu schweigen. Orientiert man sich beim Intelligenzbegriff an bestimmten Fähigkeiten, erweitert sich per se auch die Gruppe der zu inkludierenden Akteure; samt künstlichen Systemen.

Eine weitere wichtige **Unterscheidung** für eine KI-Charakterisierung betrifft jene **zwischen Verhalten und Denken**. Bereits die klassische Arbeit zu KI von Stuart Russell und Peter Norvig (2012: 22 ff) konstruierte ein Koordinatensystem nach den Unterscheidungen Verhalten/Denken und (faktisch) menschlich/ideal. Hieraus ergeben sich folgende Orientierungsmöglichkeiten für eine KI-Charakterisierung: (faktisches) menschliches Denken oder menschliches Handeln, rational-ideales Denken oder rational-ideales Handeln. Russell und Norvig entscheiden sich für die Orientierung an **rationalem Handeln**, weil „eine perfekte Rationalität ein guter Ausgangspunkt für die Analyse“ sei (Russell / Norvig 2012: 26) und zudem die Schwächen des faktischen menschlichen – und häufig unintelligenten – Denkens und Verhaltens vermieden wird. Darüber hinaus würde diese Orientierung den Fortschritt eher befördern. Weiters bestehe der Vorteil gegenüber einer Orientierung am rein rationalen Denken darin, nicht auf logische oder mathematische Operationen festgelegt zu sein, sondern auf die Tauglichkeit in der Praxis fokussieren zu können. Nachteilig erweist sich an diesem Konzept allerdings die einseitige Privilegierung des Menschen und der Idealität, wodurch sowohl der Abstand zum Tier als auch zum alltäglichen Handeln offensichtlich wird.<sup>9</sup>

---

9 Nebenbei sei erwähnt, dass auch der Begriff der Rationalität ebenso strittig ist wie jener der Intelligenz. Vgl. hierzu etwa Hahn 2017.

Deutlich konträr zum Gesagten fassen Russell und Norvig ihr Verständnis von KI wie folgt zusammen (2012: 14):

„Wir definieren die KI als die Lehre von Agenten, die Wahrnehmungen aus der Umgebung erhalten und Aktionen ausführen. Jeder dieser Agenten implementiert eine Funktion, die Wahrnehmungsfolgen auf Aktionen abbildet, und wir beschreiben unterschiedliche Möglichkeiten, diese Funktionen darzustellen, wie zum Beispiel reaktive Agenten, Echtzeitplaner, neuronale Netze und entscheidungstheoretische Systeme.“

Von idealer Rationalität im Handeln ist hier keine Rede (mehr).

Um herauszufinden, ob bestimmte (biologische oder künstliche) Systeme Intelligenz aufweisen, wurden bislang häufig gewisse (hypothetische) Tests herangezogen, wie etwa der **Turing-Test**, das **Chinese-Room-Argument** oder auch **Intelligenztests**. Allen Turing (1950) schlug – in behavioristischer Manier – vor, dass man ab dem Zeitpunkt Intelligenz unterstellen müsse, ab dem ein menschlicher Kommunikator nicht mehr adäquat entscheiden kann, ob er mit einem Menschen oder einer Maschine kommuniziert (vgl. Oppy / Dowe 2021). Der Fokus liegt hier ganz offensichtlich wieder auf der menschlichen Interpretation bzw. Bewertung eines bestimmten Outputs. Unter diesen Gesichtspunkten wären Menschen wohl unter gewissen Umständen bereit, Sprachassistenten oder ChatGPT Intelligenz zuzusprechen<sup>10</sup>, während dies bei Tieren nur eingeschränkt möglich wäre. Ähnlich dürfte es sich bei Intelligenztests verhalten, da diese bereits von neueren KI-Systemen bewältigt wurden.

Das **Chinese-Room-Argument** von John Searle (1980) widerspricht vehement dieser Schlussfolgerung und generell einer solchen Herangehensweise. Sein Gegenbeispiel lautet: In einem geschlossenen Raum befindet sich ein Mensch, der kein Chinesisch versteht. Dieser erhält Fragen in chinesischer Schrift (die er natürlich nicht versteht) und gibt sogar sinnvolle Antworten ebenfalls in chinesischer Schrift. Dies ist möglich durch eine entsprechende Anleitung in seiner Muttersprache. Die Personen außerhalb des Raumes schließen aus den Antworten, dass der Mensch Chinesisch beherrscht, obwohl er das faktisch nicht tut. Searle schließt hieraus, dass es sich hier nicht um Intelligenz handelt, da die Befolgung einer **Syntax** (wie in Computern) noch nicht zur **Semantik** befähigt, denn hierfür brauche es

---

10 Mittlerweile wurden zahlreiche Abwandlungen des Turing-Tests entwickelt, um ihn differenzierter zu gestalten. Vgl. dazu etwa Legg / Hutter 2007: 426f.

**Intentionalität** bzw. **Verstehen**. Folglich können Computer per se auch nicht intelligent sein.

Searles Gedankenexperiment und vor allem seine Schlussfolgerungen stoßen allerdings auf zahlreiche Kritik (Bringsjord / Govindarajulu 2022; Beckermann 2008: 313 ff; Pauen 2001: 149 ff). Recht unklar ist bei Searle, was er hier mit „Verstehen“ meint und welchen Mehrwert dieses Verstehen (praktisch) erbringt (Legg / Hutter 2007: 438 f). Offenbar geht es um (das Erfassen von) Bedeutung und Referenz. Aber selbst in der **Hermeneutik**, also der Lehre vom Verstehen und seinen Bedingungen, stellt in einer Konversation ein adäquater Outcome, also eine plausible Antwort, den Gesprächspartner weitgehend zufrieden; wir können nicht in die Köpfe der anderen schauen (vgl. etwa den Klassiker: Gadamer 1990). Jedenfalls erscheint es voreilig, z. B. aktuellen generativen KI-Systemen semantische Fähigkeiten abzusprechen. Für **Funktionalisten** (vgl. Beckermann 2008: 141 ff; Pauen 2001: 128 ff; Esfeld 2005: 85 ff), gegen die sich Searles Kritik richtet, besteht hier kaum ein Problem, da diese sich mit den jeweils funktionalen Rollen zufriedengeben, also primär am Verhalten von Systemen interessiert sind.

Unklar ist bei Searle darüber hinaus, warum Computer keine Intentionalität (vgl. Müller 2024) oder künftig sogar Bewusstsein (so etwa Chalmers) haben können. Thomas Metzinger (2003: 171) verschärft die Kritik an Searle um einen weiteren Aspekt:

„Wir sollten ein System spätestens dann als bewusstes Objekt behandeln, wenn es uns gegenüber auf überzeugende Weise demonstriert, dass die philosophische Frage nach dem Bewusstsein für es selbst ein Problem geworden ist, zum Beispiel wenn es eine eigene Theorie des Bewusstseins vertritt, d. h. wenn es mit eigenen Argumenten in die Diskussion um künstliches Bewusstsein einzugreifen beginnt.“<sup>11</sup>

---

11 Ergänzend sei hier erwähnt, dass gemeinhin zwischen Bewusstsein und phänomenalem Bewusstsein unterschieden wird. Bewusstsein meint hier zunächst, dass ein System auf einer Basisebene kognitive Zustände hat (z. B. die Wahrnehmung von etwas) und sich auf einer Metaebene dieser Basiszustände bewusst ist. Bewusstsein steht hier im Gegensatz zu „unbewusst“, „unterbewusst“ oder „nicht bewusst“. Demgegenüber ist phänomenales (qualitatives) Bewusstsein ein solches, wie es sich auf der Basisebene anfühlt, diese Zustände zu haben. Hier besteht das sog. „harte Problem“ des Bewusstseins (Chalmers 1996), also ob es möglich ist, dieses phänomenale Bewusstsein auch zu erklären. Um ein klassisches Beispiel (Nagel 1974) zu verwenden: Können wir jemals erklären oder verstehen, wie es ist, eine Fledermaus zu sein (what is it like to be a bat).

Hinzuweisen ist noch auf die Unterscheidung zwischen „starker“ und „schwacher“ KI, die ebenfalls auf John Searle zurückgeht. **Starke KI** würde (für Searle) demnach selbst „Geist“ sein und insofern Verstehen und kognitive bzw. mentale Zustände aufweisen.<sup>12</sup> Dies ist freilich ein ambitioniertes Ziel, von dem unklar ist, wie weit sich aktuelle Systeme dieser Vision bereits annähern. Man spricht hier gegenwärtig auch von einer allgemeinen oder generellen künstlichen Intelligenz (**AGI: Artificial General Intelligence**), bei der – wie der Name andeutet – von allgemeineren rationalen Standards unter verschiedenen Herausforderungen ausgegangen wird, also ein „umfassender Problemlöser“ (Heinrichs / Heinrichs / Rütter 2022: 18). Dem entspricht auch eine bekannte Definition von Intelligenz von Legg und Hutter (2007: 402): „Intelligence measures an agent’s ability to achieve goals in a wide range of environments.“ Dies wäre unter Bedingungen **multimodaler Funktionen**, also der Verbindung mehrerer Fähigkeiten in einer KI, durchaus erwägenswert. Bei den neuesten Generationen generativer KI-Systeme wird über die Zuschreibung von AGI jedenfalls laut nachgedacht (Bubeck u. a. 2023; Lake / Baroni 2023).

Derzeitige KI-Systeme bewegen sich vornehmlich auf dem Niveau einer **schwachen** (weak) **KI**, so dass diese Systeme nur eine besondere Leistungsfähigkeit für die Erreichung ganz konkreter Ziele darstellen. Gegenüber einer starken KI, so Searle, würden sie „Geist“ lediglich *simulieren* oder uns *so erscheinen, als ob* sie Geist wären, jedoch nicht selbst Geist sein. Wie der „Geist“ jedoch evolutionär in das biologische System „Mensch“ gelangt, wird nicht beantwortet. Im **menschlichen Gehirn** jedenfalls laufen zunächst auch nur binäre physische Prozesse ab, indem Neuronen entweder ein Signal zum Feuern aussenden – oder eben nicht.

Als letzte, alles überragende Option steht die Vision einer sog. **Superintelligenz** im Raum, die sich durch die völlige Überlegenheit gegenüber menschlichen Fähigkeiten in allen Bereichen auszeichnen würde (Bostrom 2014; Russell 2020: 143 ff; Heinrichs / Heinrichs / Rütter 2022: 169 ff). Ob es wünschenswert ist, eine echte starke KI oder gar eine Superintelligenz überhaupt zu entwickeln, ist nicht zuletzt eine Frage der Ethik.

---

12 In jüngeren Arbeiten fordert Searle für „Geist“ und somit für eine starke KI auch qualitative Zustände, also ein phänomenales Bewusstsein. Es ist jedoch fraglich, ob diese Forderung gerechtfertigt ist, wie weit qualitative Zustände von KI-Systemen realisiert werden könnten und ob dies überhaupt – aus ethischen Gründen – wünschenswert wäre.

Eine andere (spannende) Frage ist, ob es möglich sein könnte, dass KI-Systeme ihrerseits höher entwickelte „Nachkommen“ entwickeln könnten – was in der Evolutionsgeschichte auf biologischem Weg offenbar gelungen ist. In diesem Zusammenhang dürfte es zumindest zu denken geben, dass manche KI-Systeme bereits sog. **emergente Systemeigenschaften** hervorgebracht haben sollen, also solche Fähigkeiten, die im gesamten Trainings- und Lernprozess nirgends intendiert oder erwartet waren.<sup>13</sup>

Zusammenfassend kann hier auf die sehr brauchbare und recht umfassende Definition von KI durch die „unabhängige hochrangige Experten-Gruppe für Künstliche Intelligenz“ (HEG-KI) der Europäischen Kommission (2019) in ihren Ethik-Leitlinien verwiesen werden (Nr. 143):

„Künstliche-Intelligenz-(KI)-Systeme sind vom Menschen entwickelte Software- (und möglicherweise auch Hardware-) Systeme, die in Bezug auf ein komplexes Ziel auf physischer oder digitaler Ebene agieren, indem sie ihre Umgebung durch Datenerfassung wahrnehmen, die gesammelten strukturierten oder unstrukturierten Daten interpretieren, Schlussfolgerungen daraus ziehen oder die aus diesen Daten abgeleiteten Informationen verarbeiten und über die geeignete(n) Maßnahme(n) zur Erreichung des vorgegebenen Ziels entscheiden. KI-Systeme können entweder symbolische Regeln verwenden oder ein numerisches Modell erlernen, und sie können auch ihr Verhalten anpassen, indem sie analysieren, wie die Umgebung von ihren vorherigen Aktionen beeinflusst wird.“

Diese Definition beinhaltet eine ganze Reihe von Aspekten und Fähigkeiten, wobei hier offen bleiben kann, inwieweit dies bereits auf aktuelle KI-Systeme zutrifft.

In den folgenden Abschnitten sollen weitere Begriffe und Konzepte überblicksartig erläutert werden, die für die gesamte Diskussion von zentraler Bedeutung sind.

## 2.2 Maschinelles Lernen - Machine Learning (ML)

ML kann als **Oberbegriff** für unterschiedliche Arten von **künstlich generiertem Wissen aus Erfahrung** betrachtet werden. Dabei geht es um die Entwicklung lernfähiger Systeme und Algorithmen, die nicht – wie etwa ein Taschenrechner – auf festgelegte Reaktionsweisen beschränkt sind

---

13 Siehe hierzu einige Hinweise im ersten der beiden Beiträge von Andreas Klein in diesem Band.

(vgl. insgesamt zu diesen Abschnitten Heinrichs / Heinrichs / Rüter 2022). Das **Lernen** wird durch **Beispiele** unter Zuhilfenahme von Methoden der Statistik und der Informationstheorie initiiert.<sup>14</sup> Am Ende soll ein **verallgemeinerbares** Ergebnis erzielt werden. Der jeweilige Algorithmus entwirft aufgrund der zur Verfügung gestellten **Trainingsdaten** ein **statistisches Modell**, wobei eine Testung gegen die Trainingsdaten vorgenommen wird – und der Algorithmus schrittweise verbessert wird. Das System versucht **Muster** und **Gesetzmäßigkeiten** in den Trainingsdaten zu erkennen und diese für weitere Verwendungen – z. B. Prognosen – auszuwerten. Von besonderer Bedeutung ist hier das **Bewerten unbekannter Daten**, so dass ML auch (in begrenzter Weise) für unbekannte Szenarien einsatzfähig ist.

Eine technische Voraussetzung für die Leistungsfähigkeit von ML besteht in der erheblichen Weiterentwicklung von Computersystemen bzw. der Hardware-Basis. Die Einsatzgebiete von ML sind umfangreich und reichen von Übersetzungen, Textproduktionen, Spracherkennung, Spielen, visuellen Erkennungen, Diagnoseverfahren bis hin zu Sicherheitskonzepten und autonomem Fahren.

Hinsichtlich der Lernmethoden werden primär drei unterschiedliche Ansätze verfolgt, die jeweils ihre Vor- und Nachteile haben: 1) **Überwachtes Lernen** (Supervised Learning), 2) **Unüberwachtes Lernen** (Unsupervised Learning) und 3) **Bestärkendes Lernen** (Reinforcement Learning).

Beim **überwachten Lernen** wird dem System eine **umfangreiche**, bereits **klassifizierte Datenmenge** zur Verfügung gestellt, woraus das System ein entsprechendes **Modell** entwickelt. Insofern wird hier mit Vorgaben gearbeitet, z. B. mit Typen (Klassifizierungen), wobei die Eingabe- und Ausgabewerte bekannt sind. Metaphorisch ausgedrückt: Es wird dem System gesagt, was es überhaupt sehen soll, z. B. Katzen. Anhand dieses Modells kann das System daraufhin auch unbekannte Daten (Informationen) entsprechend analysieren und bewerten. Überwachtes Lernen wird häufig eingesetzt, z. B. bei Handschriftenerkennung, Vorhersagen eines Verkaufspreises oder künftigen Nachfragen nach einem Produkt. Ein **Nachteil** dieses Lerntyps sind einerseits die benötigten großen Datenmengen und andererseits die vorgegebenen Klassifizierungen. Als Unterkategorien können hier noch das **teilüberwachte Lernen**, das **aktive Lernen** und das **selbstständige Lernen** genannt werden.

---

14 Siehe hierzu vor allem die drei ersten Beiträge in diesem Sammelband.