

Martina Gaisch  
Isabella Mader

NEXT GENERATION DIGITAL

**AI Ethics**  
**&**  
**Human Factors in AI**

# NEXT GENERATION DIGITAL

## AI Ethics & Human Factors in AI

Martina Gaisch

Isabella Mader

© 2025 Martina Gaisch, Isabella Mader

Herausgeber:in: Excellence Edition, Wien. [www.excellence-edition.com](http://www.excellence-edition.com)

Illustration: Martina Gaisch, Isabella Mader, Victoria Kaiselgruber

Umschlaggestaltung: Midjourney AI (ursprüngliche, später veränderte Illustration),

Isabella Mader (Prompt, Umschlagdesign), SD Studio (Grafische Umsetzung)

Layout: Adeosun Caleb Boluwaji (Bolu Book)

Lektorat / Korrektorat: Alexandra Hermann

weitere Mitwirkende: Ing. Mag. Daniel Eder (rechtliche Unterstützung zum EU AI Act)

Druck und Vertrieb im Auftrag der Autorinnen:

Buchschmiede von Dataform Media GmbH

Julius-Raab-Straße 8

2203 Großebersdorf

Österreich

[www.buchschmiede.at](http://www.buchschmiede.at) - Folge deinem Buchgefühl!

Kontaktadresse nach EU-Produktsicherheitsverordnung:

[info@buchschmiede.at](mailto:info@buchschmiede.at)

ISBN:

978-3-99139-433-4 (Softcover)

978-3-99139-430-3 (E-Book)

Printed in Austria



Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlages und der Autorinnen unzulässig. Dies gilt insbesondere für die elektronische oder sonstige Vervielfältigung, Übersetzung, Verbreitung und öffentliche Zugänglichmachung.

# Inhalt

<b>Danksagung .....</b>	<b>12</b>
<b>Einführung.....</b>	<b>13</b>
Warum dieses Buch?.....	13
Warum dieser Titel? .....	14
Warum gerade jetzt? .....	14
Begriffliche Klärung.....	16
Was erwartet Sie? .....	19

## Teil 1

<b>1. Künstliche Intelligenz SICHTET.....</b>	<b>32</b>
1.1. Human in the Loop: Der Mensch als Kontrollinstanz .....	37
1.2. A Fool with a Tool.....	39
1.3. Schlüsselkompetenzen für die Nutzung von KI.....	41
1.3.1. Die Erkenntnis-Schwelle: Was der Mensch nicht kennt, das nutzt er nicht .....	43
1.3.2. Ohne Grundstein kein Bauwerk: KI ohne technologische Grundkompetenz bleibt nur ein Konzept .....	45
1.3.3. Stillstand bedeutet Rückschritt: Agilität und Lernbereitschaft als Motor. ....	46
1.3.4. Kontext ist der Code, der das Programm zum Laufen bringt.....	48
1.3.5. Nicht alles, was glänzt, ist Gold – und nicht alles, was die KI liefert, ist wahr .....	50
1.3.6. Objektivität als Kompass in einem Meer voll Meinungen .....	52
1.3.7. Die Macht der Worte: Wie man fragt, so erhält man.....	57
1.3.8. Data Literacy: Wer die Daten versteht, beherrscht die KI .....	60
1.3.9. Wer der KI Flügel verleiht, sollte auch ihr Gewissen programmieren.....	61

1.4. Qualifizierungsmaßnahmen – wie wir uns für das Zeitalter der KI rüsten.....	63
--	----

## **2. Künstliche Intelligenz DICHTET .....76**

2.1. Die „kreative“ Seite der Künstlichen Intelligenz.....	76
2.1.1. Kunst und Design - Erweiterung menschlicher Kreativität .....	78
2.1.2. Musikkomposition – Inspiration und Werkzeug .....	79
2.1.3. Texte und Literatur - Katalysator für neue literarische Formen.....	80
2.1.4. Film, Video und Animation - Erweiterung kreativer Möglichkeiten.....	81
2.1.5. Spieleentwicklung – mehr Dynamik und individuelle Anpassung .....	81
2.1.6. Mode und Produktdesign – Personalisierung und Optimierung.....	83
2.1.7. Architektur und Stadtplanung – Nachhaltigkeit und Ressourcenschonung .....	84
2.1.8. Personalisierte Kreativität und Co-Kreation – eine neue Ära der Kreativität .....	84
2.1.9. Fazit: Die kreative Symbiose von Mensch und Maschine.....	85
2.2. Einschränkungen der KI im Kreativbereich .....	86
2.2.1. Fehlende Originalität und Innovation .....	86
2.2.2. Begrenztes kulturelles und soziales Verständnis .....	87
2.2.3. Fehlende emotionale Tiefe und Intuition .....	87
2.2.4. Mangelnde ethische Urteilsfähigkeit und unvorhersehbare Konsequenzen von KI-Kreationen .....	88
2.2.5. Fehlende Lebenserfahrung und persönliche Perspektive.....	88
2.2.6. Begrenzte Fähigkeit zur Langzeitplanung und narrativen Kohärenz.....	89
2.2.7. Unzureichende Fähigkeit zur Improvisation .....	89
2.2.8. Fazit: Die unverzichtbare Rolle des Menschen in der kreativen KI-Nutzung .....	91
2.3. Digitale Ausbeutung und digitales Fließband .....	92
2.3.1. Amazon Mechanical Turk (MTurk).....	94
2.3.2. Content Moderation für Social Media Plattformen.....	95
2.3.3. Click-Farmen.....	96
2.3.4. Data Labeling für Machine Learning und KI.....	97
2.3.5. Digitale Plattformarbeit und Compliance-Herausforderungen .....	97
2.4. Automatisierung und KI in der modernen Arbeitswelt.....	99
2.4.1. Automatisierungs-Kandidaten .....	100

2.4.2. HIT (Human Intelligence Tasks).....	102
2.4.3. Aussicht auf Entlastung: „Die KI nimmt uns die Arbeit ab!“ „Hoffentlich!“ .....	104
2.4.4. Die Evolution von IT und KI: Von Automatisierung zu umfassender Opti- mierung.....	106
2.5. Echte Gefahr: Deskilling (Verdummung)!.....	107
2.6. Das neue Hybrid – Mensch und KI in Symbiose.....	107
2.6.1. Unterschied zwischen KI und Erweiterter Intelligenz.....	108
2.6.2. Produktivität durch Augmentation.....	109
2.6.3. Augmentation in verschiedenen Branchen.....	109
2.6.4. Ein positiver Ausblick – Eine neue Ära der Augmentation.....	110
2.7. Kernpunkte im Fokus: KI DICHTET.....	112
<b>3. Künstliche Intelligenz RICHTET.....</b>	<b>124</b>
3.1. Heikle Einsatzszenarien von Künstlicher Intelligenz.....	125
3.1.1. Einsatz in der Justiz.....	125
3.1.2. Einsatz im Gesundheitssystem.....	128
3.1.3. Einsatz im Personalwesen.....	131
3.1.4. Einsatz bei Kreditvergabe.....	131
3.2. Social Scoring und Massenüberwachung im EU AI Act.....	132
3.2.1. Social Scoring.....	134
3.2.2. Massenüberwachung.....	135
3.3. Systemimmanente Zielkonflikte und Herausforderungen.....	137
3.3.1. Individuelle Gerechtigkeit vs. Gruppen-Gerechtigkeit.....	137
3.3.2. Verteilungsschaden vs. Repräsentationsschaden.....	140
3.3.3. Equal Error Rate: Falsch-Positive und Falsch-Negative.....	141
3.4. Kompetenzüberschreitung der KI – Was Maschinen nicht entscheiden sollen	143
3.5. Kernpunkte im Fokus: KI RICHTET.....	147
<b>4. Künstliche Intelligenz VERNICHTET.....</b>	<b>156</b>
4.1. Deep Fakes: Die Täuschung der Realität.....	157
4.1.1. Technologische Grundlagen und Bedrohungspotenzial.....	160

4.1.2. Fortschritte und Herausforderungen bei der Erkennung von Deep Fakes .....	165
4.2. Cybermobbing und Hate Speech: Digitale Gewalt ohne Grenzen.....	168
4.2.1. Psychologische und soziale Auswirkungen & Gegenmaßnahmen .....	170
4.2.2. Opferschutz.....	172
4.3. Social Bots: Unsichtbare Manipulatoren der Meinungsbildung .....	174
4.4. Überwachung: Der „Aufstieg“ des gläsernen Menschen .....	179
4.4.1. Konsequenzen für die Privatsphäre .....	181
4.4.2. Regulierung und die Zukunft des Datenschutzes .....	183
4.5. Autonome Waffensysteme: Krieg in den Händen der Maschinen .....	185
4.5.1. Technologische Grundlage, Funktionsweise und ethische Implikation... ..	185
4.5.2. KI und biologische Kriegsführung.....	188
4.5.3. Internationale Regulierung und Wettrüsten .....	190
4.6. Cyberkriminalität: Das wachsende Arsenal der digitalen Angriffe.....	192
4.7. Umwelt und KI: Der ökologische Preis der Technologie .....	197
4.7.1. Ressourcenverbrauch und Umweltkosten .....	198
4.7.2. Nachhaltige KI: Illusion oder Realität im Kampf gegen Umweltzerstörung .....	199
4.8. Arbeitsmarkt im KI-Zeitalter: Transformation statt Dystopie .....	202
4.8.1. Von der Angst zur Realität: Massenarbeitslosigkeit durch KI bleibt aus..	205
4.8.2. Monopolbildung und Machtkonzentration: Die Dominanz der Tech-Giganten.....	206
4.9. Innovation oder Zerstörung: KI zwischen Macht und Verantwortung .....	208
4.10. Kernpunkte im Fokus: KI VERNICHTET.....	210

## Teil 2

<b>5. Der Mensch DENKT: Human Factors in AI .....</b>	<b>224</b>
5.1. Menschlich und KI: Kooperation oder Konkurrenz? .....	224
5.2. Lernen: Menschliches Lernen vs. Maschinelles Lernen .....	227
5.2.1. Maschinelles Lernen (Machine Learning / Generative AI).....	228
5.2.2. Menschliches Lernen .....	230
5.3. Neuronale Netze: Mensch versus KI.....	233
5.4. Sprache als Grundlage von KI .....	238
5.4.1. Die Macht der Sprache: Warum wir beeindruckt sind .....	240
5.4.2. Terminologie für KI: Intent, Entities – Sprache schafft Klarheit .....	244
5.4.3. Terminologie im Prompting.....	248
5.5. Theory of Mind: Wie denkt der Mensch?.....	250
5.5.1. Wie entwickelt der Mensch diese Fähigkeit? .....	251
5.5.2. Theory of Mind AI .....	253
5.6. Bewusstsein: Es gibt keine vereinbarte Definition .....	258
5.6.1. Philosophie: Was ist Bewusstsein? .....	260
5.6.2. Psychologie: Definition und Fragen zu KI .....	266
5.6.3. Neurowissenschaften: Entstehung des Bewusstseins im Gehirn .....	268
5.6.4. Ethik: Moralische Verantwortung und KI .....	269
5.6.5. Soziologie: Gesellschaftliche Wahrnehmung von KI und Bewusstsein ....	270
5.6.6. Recht: Dramatische Konsequenzen von „Conscious AI“ .....	271
5.7. Emotion: Menschliche Gefühle und AI Emotion Detection .....	276
5.7.1. Emotionen beim Menschen .....	277
5.7.2. AI Emotion Detection.....	278
5.8. Biometrie: Die Vermessung des Menschen .....	285
5.8.1. Arten von Biometrie .....	286
5.8.2. Biometrie: Rechtliche und gesellschaftlichen Auswirkungen .....	287
5.9. Kernpunkte im Fokus: Der Mensch DENKT .....	290

<b>6. Der Mensch ERKENNT .....</b>	<b>300</b>
6.1. Bias: Definitionen und Arten.....	302
6.2. Datenbias: Das BiasScope-Modell.....	304
6.2.1. Algorithmischer Bias .....	307
6.2.2. Human Bias.....	310
6.3. Entstehung von Bias in KI-Systemen .....	316
6.4. Das AIBC-Modell und die Entwicklungsphasen eines KI-Systems.....	317
6.4.1. Bias im Systemdesign .....	328
6.4.2. Bias während der Datenanalyse .....	336
6.4.3. Bias in der Datenaufbereitung.....	344
6.4.4. Bias in der Modellierung.....	348
6.4.5. Bias in der Evaluierung.....	356
6.4.6. Bias im Deployment.....	360
6.4.7. Bias im Monitoring .....	366
6.5. Kernpunkte im Fokus: der Mensch ERKENNT .....	369
<b>7. Der Mensch BEDENKT: Empfehlungen .....</b>	<b>384</b>
7.1. Guidelines für verantwortungsvolle KI-Systeme .....	386
7.1.1. Digitaler Humanismus: Technologie im Dienst der Menschlichkeit .....	386
7.1.2. Das Human-Centered AI (HCAI) Framework.....	390
7.1.3. Die Montreal Declaration for a Responsible Development of AI.....	392
7.1.4. Die Toronto Declaration for Non-Discrimination in AI.....	393
7.1.5. Die Asilomar AI Principles.....	394
7.1.6. Explainable AI: der Weg zur Transparenz .....	396
7.2. Regularien zu KI – Europa und darüber hinaus .....	397
7.2.1. Der EU AI Act .....	398
7.2.2. EU-Regularien mit Verbindung zum EU AI Act.....	421
7.2.3. Kanada.....	430
7.2.4. USA .....	433
7.2.5. Großbritannien .....	434
7.2.6. China .....	435
7.3. Relevante internationale Initiativen zu KI-Ethik.....	437
7.3.1. United Nations High-Level Advisory Body on AI.....	437



7.3.2. Die Bletchley Deklaration: Internationaler Meilenstein zu KI-Sicherheit	438
7.3.3. KI-Konvention von 2024.....	439
7.3.4. UNESCO KI-Ethik Empfehlungen.....	440
7.3.5. ITU: AI for Good and ITU-T Technical Standards for AI.....	442
7.3.6. OECD AI Principles and AI Observatory.....	443
7.3.7. UNIDO AIM Alliance.....	444
7.3.8. ILO warnt vor der “Skills-Schere”.....	444
7.4. Strategische Überlegungen und Grundsätze.....	445
7.4.1. Strategien für die Entwicklung von KI-Systemen.....	446
7.4.2. Strategien für den Einsatz von KI-Systemen.....	448
7.4.3. KI-Policies und Ethik-Richtlinien von Organisationen.....	450
7.5. Ressourcen zu Guidelines und Regularien.....	451
7.5.1. Ethik-Leitlinien der EU für eine vertrauenswürdige KI.....	452
7.5.2. Human-Centered Machine Learning (HCML) Guidelines.....	454
7.5.3. IEEE Ethically Aligned Design (EAD) Guideline.....	458
7.5.4. ISO/IEC KI-Standards.....	460
7.5.5. Non-Discrimination-by-Design.....	462
7.5.6. Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) .....	465
7.5.7. Design for Uncertainty: Unsicherheitsbewältigung.....	467
7.5.8. Privacy by Design: Datenschutz von Beginn an mitdenken.....	468
7.5.9. Transparency by Design: Das Transparenz-Prinzip.....	474
7.5.10. Ethik Pre-Mortems.....	476
7.5.11. User Onboarding.....	477
7.5.12. Das AI Risk Repository des MIT.....	480
7.6. Leuchtturmprojekte und Best Practices für verantwortungsvolle KI-Entwicklung .....	481
7.6.1. Algorithmic Justice League: Joy Buolamwini.....	481
7.6.2. Astrid AI: der Weg zu verantwortungsvollem KI-Training.....	483
7.6.3. Cortical.io: Weniger Rechenleistung – auf dem Weg zu Green AI.....	484
7.7. Kernpunkte im Fokus: der Mensch BEDENKT.....	485

<b>8. Der Mensch LENKT .....</b>	<b>496</b>
8.1. Ethik im Überblick.....	497
8.2. Ebenen der Ethik.....	499
8.2.1. Angewandte Ethik.....	500
8.2.2. Metaethik.....	500
8.2.3. Normative Ethik.....	502
8.2.4. Deskriptive Ethik.....	502
8.3. Ethische Strömungen .....	503
8.3.1. Deontologische Ethiken.....	504
8.3.2. Konsequentialistische Ethiken.....	509
8.4. Ausgewogene Ethik im EU AI ACT: Prinzipien und Konsequenzen im Einklang .....	512
8.4.1. Der goldene Mittelweg: Hybride Ethik im EU AI Act .....	513
8.4.2. Ethische Leitlinien für verantwortungsvolle KI: Tugenden und Prinzipien .....	517
8.4.3. Die Rolle der Tugendethik im AI Act .....	519
8.5. Das PRISMA AI Modell .....	521
8.6. Handlungsempfehlungen für verantwortungsvolle KI-Entwicklung: Verantwort- ung in allen Phasen .....	533
8.6.1. Handlungsempfehlungen für das Systemdesign.....	536
8.6.2. Handlungsempfehlungen für die Datenanalyse .....	541
8.6.3. Handlungsempfehlungen für die Datenaufbereitung .....	546
8.6.4. Handlungsempfehlungen für die Modellierung.....	550
8.6.5. Handlungsempfehlungen für die Evaluierung .....	555
8.6.6. Handlungsempfehlungen für das Deployment .....	558
8.6.7. Handlungsempfehlungen für das Monitoring .....	562
8.7. Der Mensch als Wegweiser für faire KI-Systeme .....	564
8.8. Kernpunkte im Fokus: der Mensch LENKT.....	566

# Danksagung

Dieses Buch wäre ohne die Unterstützung und Ermutigung vieler Menschen nicht möglich gewesen. Unser besonderer Dank gilt den Professoren und KI-Experten der FH Oberösterreich, Fakultät für Informatik, Kommunikation und Medien in Hagenberg, FH-Prof. Univ.-Doz. DI Dr. Ulrich Bodenhofer, FH-Prof. PD DI Dr. Stephan Winkler und FH-Prof. DI Dr. Eckehard Hermann deren wertvolle Anregungen und fachliche Expertise maßgeblich zur Qualität dieses Werkes beigetragen haben.

Er gilt auch den Datenwissenschaftlerinnen Sandra Wartner, MSc und Anna-Christina Glock, MSc sowie unserer Lektorin Mag. Alexandra Hermann; Victoria Rammelmüller, MMA für die grafische Unterstützung und für die rechtliche Begutachtung von Kapitel 7.2.1. zu AI Act geht unser Dank an Ing. Mag. Daniel Eder, MSc.

Unser Dank gilt auch allen Leserinnen und Lesern, die sich mit diesem Thema auseinandersetzen und so zu einer lebendigen Diskussion beitragen. Dieses Buch ist für Sie.

# Einführung

## Warum dieses Buch?

Stellen Sie sich eine Zukunft vor, in der Technologie nicht nur die Welt verändert, sondern unsere Menschlichkeit stärkt und schützt. Dieses Buch möchte Sie einladen, innezuhalten und mit uns eine Reise zu beginnen – eine Reise in die faszinierende, aber auch herausfordernde Welt der Künstlichen Intelligenz (KI). Dabei geht es darum, die entscheidenden Fragen zu stellen, die unsere Zukunft prägen werden: Welche Rolle spielen menschliche Werte, Ethik und Verantwortung in einer von Maschinen mitgestalteten Welt? Wie können wir KI so gestalten, dass sie uns nicht entfremdet, sondern uns näher zu dem bringt, was uns als Menschen ausmacht?

Dieses Buch beleuchtet die vielfältigen Facetten der Künstlichen Intelligenz in unserer zunehmend digitalisierten Welt und setzt sich mit ihren Potenzialen ebenso auseinander wie mit ihren Risiken. KI eröffnet immense Chancen: Effizienzsteigerungen in der Gesundheitsversorgung, personalisierte Lernwege in der Bildung, und Automatisierung, die Raum für kreatives Arbeiten schafft. Doch all das gelingt nur, wenn Empathie, Fairness und Transparenz in ihrer Entwicklung und Anwendung zu unverzichtbaren Bestandteilen werden. Die Entscheidungen, die wir heute treffen, werden weitreichende Folgen haben – für Individuen, Gemeinschaften und die globale Gesellschaft.

Dieses Werk will Denkanstöße liefern und Diskussionen anregen, nicht nur in der Fachwelt, sondern in der breiten Gesellschaft. Die Zukunft der KI geht uns alle an. Wenn wir heute die richtigen Fragen stellen und die Weichen klug setzen, können wir eine Welt gestalten, in der Technologie nicht nur effizient, sondern auch ethisch und menschenzentriert ist. Eine Zukunft, in der KI-Systeme uns nicht entmenschlichen, sondern dabei helfen, menschliche Werte wie Empathie und Fairness zu bewahren. Nur durch gemeinsame Reflexion und verantwortungsvolle Entscheidungen können wir die Potenziale der KI zum Wohl aller nutzen.

## Warum dieser Titel?

Der englische Titel „AI Ethics & Human Factors“ wurde bewusst gewählt, obwohl das Buch auf Deutsch verfasst ist, um auf den internationalen Kontext und die globalen Herausforderungen hinzuweisen, die KI mit sich bringt. Künstliche Intelligenz ist ein weltweit relevantes Thema, dessen ethische Implikationen nicht an Landesgrenzen Halt machen. Der Begriff Human Factors etwa ist in der internationalen Forschung und Technologieentwicklung fest verankert und beschreibt die Berücksichtigung menschlicher Bedürfnisse und Fähigkeiten im Design technologischer Systeme (Meister 2018).

Ein englischer Titel signalisiert zudem eine Offenheit für eine globale Leserschaft und betont, dass die in diesem Buch behandelten Themen von allgemeiner Bedeutung sind, unabhängig von Sprache und Kultur. Er lädt auch nicht-deutschsprachige Expertinnen und Experten ein, die Konzepte und Ansätze zu diskutieren und so einen interkulturellen Austausch zu fördern.

AI Ethics & Human Factors verbindet also den Anspruch, sich kritisch mit den ethischen Fragen unserer zunehmend digitalisierten Welt auseinanderzusetzen, und weist zugleich darauf hin, dass die Berücksichtigung menschlicher Werte – wie Empathie, Fairness und Transparenz – ein universelles Anliegen ist. Gerade in einer Zeit, in der KI-Systeme grenzüberschreitend Einfluss auf unser Leben haben, ist ein Titel in einer global verständlichen Sprache ein bewusster Schritt, um zur internationalen Diskussion über die ethische und menschenzentrierte Gestaltung von Technologie beizutragen.

## Warum gerade jetzt?

Die Dringlichkeit, sich mit dem Thema Künstliche Intelligenz auseinanderzusetzen, hat in den letzten Jahren deutlich zugenommen. KI ist längst nicht mehr ein theoretisches Konzept oder ein experimentelles Laborprojekt, sondern eine Technologie, die bereits tief in unser Leben eingreift. Doch nun stehen wir an einem entscheidenden Wendepunkt: Weltweit arbeiten Regierungen, Unternehmen

und Organisationen daran, Rahmenbedingungen und Gesetze zu schaffen, um den Einsatz von KI zu regulieren. Die Europäische Union hat im August 2024 mit dem AI Act<sup>1</sup> einen bedeutenden Schritt in diese Richtung unternommen. Doch diese Gesetze und Richtlinien gründen auf den ethischen Überlegungen, die wir heute formulieren.



Verpassen wir es jetzt, klare Prinzipien zu definieren – etwa in Bezug auf Transparenz, Fairness und den Schutz der Menschenrechte –, laufen wir Gefahr, eine Zukunft zu schaffen, in der technologische Effizienz über ethische Verantwortung gestellt wird.

Eine weitere Herausforderung stellt die immense Machtkonzentration im Bereich der KI-Entwicklung dar, die sich zunehmend auf wenige große Technologieunternehmen fokussiert. Diese Konzentration kann zwar Innovationen vorantreiben, birgt aber zugleich das Risiko, soziale und ökonomische Ungleichheiten zu verstärken. Wenn der Zugang zu diesen fortschrittlichen Technologien ungleich verteilt bleibt, könnte das die bestehenden Unterschiede weiter vertiefen.

Aus all diesen Gründen ist es jetzt von essenzieller Bedeutung, über die ethischen Fragen zu diskutieren, die die KI-Entwicklung prägen. Dieses Buch versteht sich nicht nur als Weckruf, sondern auch als fundierte Informationsquelle, um den Blick über technologische Errungenschaften hinaus auf die menschlichen und gesellschaftlichen Implikationen zu lenken. Es geht nicht nur darum, was KI leisten kann, sondern vor allem auch darum, was sie leisten sollte. Mit klaren Analysen und praxisnahen Beispielen zeigt dieses Werk auf, wie wir die Weichen für eine verantwortungsvolle und gerechte KI-Nutzung stellen können. Die Zeit zum Handeln ist jetzt, denn die kommenden Jahre werden entscheidend dafür sein, wie wir als Gesellschaft mit dieser transformativen Technologie umgehen – und ob wir es schaffen, sie im Sinne des Gemeinwohls zu gestalten.

<sup>1</sup> <https://artificialintelligenceact.eu/ai-act-explorer/>

## Begriffliche Klärung

Bevor wir tiefer in die Thematik einsteigen, ist es wichtig, die zentralen Begrifflichkeiten kurz zu beleuchten. Ein solides Verständnis dieser grundlegenden Konzepte bildet die Basis für eine fundierte Diskussion über die ethischen Implikationen der KI-Entwicklung. Im Folgenden betrachten wir daher die vier Schlüsselbegriffe, die das Fundament dieses Buches bilden: Ethik, insbesondere KI-Ethik (Kapitel 8), Bias (Kapitel 6) und Human Factors (Kapitel 5).

### Ethik

Ethik ist ein grundlegender Bereich der Philosophie, der sich mit den Prinzipien des guten und schlechten Verhaltens befasst und die moralischen Werte und Normen, die das Handeln von Individuen und Gesellschaften leiten, untersucht. Ethik kann in verschiedene Unterkategorien unterteilt werden, darunter Metaethik, normative Ethik, angewandte Ethik und deskriptive Ethik.

Während die Metaethik die Grundlagen und das Wesen ethischer Prinzipien untersucht und dabei Fragen nach der Bedeutung moralischer Begriffe und Urteile aufwirft, konzentriert sich die normative Ethik auf die Frage, wie Menschen handeln sollten. Sie entwickelt und bewertet verschiedene ethische Theorien, die unterschiedliche Ansätze bieten, um festzulegen, was als moralisch richtig oder falsch gilt.

Angewandte Ethik befasst sich mit der Anwendung ethischer Prinzipien auf konkrete, oft kontroverse Themen. Beispiele sind die Medizinethik, die Umweltethik, die Wirtschaftsethik oder die Technik- und KI-Ethik. Angewandte Ethik hilft, praktische Lösungen für moralische Dilemmata in bestimmten Bereichen zu finden.

Deskriptive Ethik befasst sich mit dem, was tatsächlich als moralisch richtig oder falsch angesehen wird. Sie ist empirisch orientiert und bedient sich der Methoden

der Sozialwissenschaften, insbesondere der Anthropologie, Soziologie und Psychologie, um moralische Phänomene zu untersuchen und zu beschreiben.

Ethik ist ein dynamischer Bereich, der sich ständig weiterentwickelt, um auf neue gesellschaftliche Herausforderungen zu reagieren. In diesem Kontext gewinnt die Ethik der Künstlichen Intelligenz zunehmend an Bedeutung. Als Spezialgebiet der angewandten Ethik widmet sie sich den moralischen Fragen und Implikationen, die sich aus der Entwicklung und Anwendung von KI-Systemen ergeben. Da KI-Technologien immer tiefer in unser tägliches Leben eindringen, wird es entscheidend, ihre ethischen und gesellschaftlichen Auswirkungen nicht nur zu verstehen, sondern auch aktiv zu gestalten und zu regulieren.

#### **KI-Ethik (AI Ethics)**

Die KI-Ethik befasst sich mit Prinzipien und Herausforderungen der KI-Entwicklung und -Anwendung, um menschliche Werte zu respektieren und gesellschaftliche Verantwortung zu fördern. Ein zentrales Anliegen ist der Umgang mit Verzerrungen und Diskriminierung, da vorurteilsbehaftete Datensätze zu unfairen Entscheidungen führen können (siehe Kapitel 6 Bias). Dabei sind Transparenz und Erklärbarkeit essenziell, um Vertrauen und Verantwortlichkeit zu gewährleisten, da KI-Modelle oft als „Black Boxes“ fungieren. Ebenso wichtig ist der Schutz der Privatsphäre, da die Verarbeitung großer Mengen personenbezogener Daten Datenschutzrisiken birgt. Ein verantwortungsvoller Umgang mit Daten verhindert Missbrauch und Vertrauensverlust.

#### **Bias**

Bias bezeichnet eine systematische Verzerrung oder Abweichung, die zu fehlerhaften oder unfairen Ergebnissen führen kann. In der Statistik führt Bias zu fehlerhaften Ergebnissen, da die Daten oder Methoden nicht repräsentativ sind oder systemische Fehler enthalten. In der Kognitionspsychologie beschreibt Bias Denk-



fehler, die unser Urteilsvermögen beeinflussen, wie etwa den Bestätigungsfehler. Bei der KI-Entwicklung können Verzerrungen zu unfairen Entscheidungen führen, beispielsweise in Bewerbungsprozessen oder Prognosen. Eine sorgfältige Überprüfung der Daten und Modelle ist daher essenziell, um diskriminierende Tendenzen zu vermeiden. Kapitel 6 widmet sich diesem Thema ausführlich.

### Human Factors

bezeichnet einen interdisziplinären Ansatz, der sowohl in der Technikentwicklung als auch in der Wissenschaft Anwendung findet. Dieser Ansatz untersucht das Zusammenspiel zwischen Mensch und Technologie mit dem Ziel, Systeme zu gestalten, die den menschlichen Fähigkeiten und Bedürfnissen gerecht werden. In der Künstlichen Intelligenz bedeutet dies, Systeme so zu entwickeln, dass sie nützlich, verständlich und sicher für den Menschen sind. Kapitel 5 widmet sich diesen Human Factors und analysiert Unterschiede, Gemeinsamkeiten und Schnittstellen zwischen Mensch und Technologie.

### Icons

Im Buch sind zahlreiche Textboxen enthalten, die durch verschiedene Icons strukturiert und hervorgehoben werden. Diese Icons bieten eine klare Orientierung und erleichtern das Verständnis der Inhalte.



**Ausrufezeichen:** steht für Definitionen, um wichtige Begriffe klar zu erklären.

---



**Infozeichen:** steht für weiterführende Informationen

---



**Buch:** steht für Zusammenfassungen, die Kernaussagen prägnant wiedergeben.

---



**Sprechblase:** markiert Beispiele, Case Studies oder Studien, die praktische Anwendungen veranschaulichen.

---



**Glühbirne:** weist auf Tipps und Empfehlungen hin.

---

## Was erwartet Sie?

Dieses Buch versteht sich als ein Wegweiser durch die komplexe und vielschichtige Landschaft der KI, fokussiert auf die ethischen Herausforderungen und den menschlichen Faktor, der oft im Schatten technischer Innovationen steht. Es wird Ihnen nicht nur helfen, die Risiken und Chancen dieser transformativen Technologie zu begreifen, sondern auch dabei kritisch über ihre Rolle in unserer Gesellschaft nachzudenken. Das Buch legt den Fokus auf zentrale Fragen wie: Was sind die Potenziale und Gefahren der KI, und welchen Beitrag sollte KI leisten? Es möchte einen Diskurs anstoßen, der gerechte, transparente und menschenfreundliche Lösungen fördert.

Um das Thema in seiner Tiefe und Breite zu erkunden, ist das Buch in acht Kapitel gegliedert, die Sie durch verschiedene Facetten der KI-Ethik und der menschlichen Interaktion mit dieser Technologie führen werden. Es kann zudem als Lehrbuch genutzt werden, da es neben theoretischen Konzepten auch praxisorientierte

Fallstudien, Handlungsanweisungen und zusätzliche Materialien enthält, die zum vertieften Lernen und zur Anwendung des Gelernten anregen.

i

### Kapitelübersicht

Das Buch ist in zwei Hauptabschnitte gegliedert, die das Thema Künstliche Intelligenz aus zwei unterschiedlichen, aber komplementären Perspektiven beleuchten:

Maschinenlogik umfasst die ersten vier Kapitel und der Mensch am Steuer steht im Fokus der zweiten vier Kapitel.

**Teil 1** widmet sich der Funktionsweise und den Fähigkeiten von Künstlicher Intelligenz und untersucht sowohl ihre Potenziale als auch ihre Grenzen. In den Kapiteln „KI SICHTET“, „KI DICHTET“, „KI RICHTET“ und „KI VERNICHTET“ werden verschiedene Aspekte der KI analysiert – von ihrer Fähigkeit, Daten effizient zu durchforsten und Muster zu erkennen, über ihr kreatives Potenzial bis hin zu ihren Einsatzmöglichkeiten bei Entscheidungen und den damit verbundenen ethischen Dilemmata. Dabei wird verdeutlicht, wie KI in der Lage ist, gewaltige Mengen an Informationen zu verarbeiten, gleichzeitig aber auch das Risiko birgt, unethische Entscheidungen zu automatisieren oder soziale Ungleichheiten zu verschärfen.

**Teil 2** richtet den Fokus auf die menschliche Seite und nimmt mit den Kapiteln „Der Mensch DENKT“, „Der Mensch ERKENNT“, „Der Mensch LENKT“ und „Der Mensch BEDENKT“ eine menschenzentrierte Perspektive ein. Hier wird erörtert, wie essenzielle menschliche Werte wie Empathie, Verantwortungsbewusstsein und Urteilsvermögen in der Entwicklung und Steuerung von KI-Systemen und Anwendungen eingebettet bleiben müssen. Jedes Kapitel thematisiert eine wesentliche Fähigkeit des Menschen – von der Fähigkeit zu reflektieren und zu verstehen, über das aktive Steuern und Kontrollieren von Technologien, bis hin zur Verantwortung, langfristige Auswirkungen ethisch abzuwägen. Diese

Kapitel betonen die unersetzliche Rolle des Menschen als Entscheidungsträger und ethische Instanz im Umgang mit KI.

Durch diese Zweiteilung bietet das Buch eine umfassende und tiefgehende Erörterung der technischen Aspekte von KI auf der einen Seite und der ethischen sowie menschlichen Implikationen auf der anderen. Dieser strukturierte Aufbau erlaubt es, die Herausforderungen und Chancen von KI aus beiden Perspektiven parallel zu betrachten und einen ausgewogenen Diskurs über die Zukunft der KI und ihre gesellschaftlichen Auswirkungen zu führen.

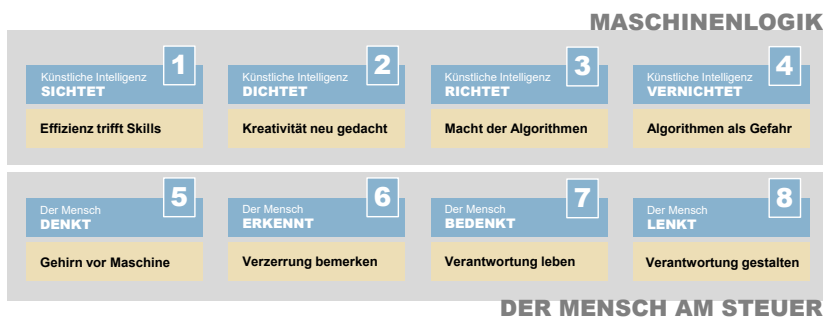


Abb. 1: Die Kapitel dieses Buches in der Übersicht.



Teil I: Maschinenlogik: Die Welt durch die Augen der KI

## Kapitel I: KI SICHTET – Sammeln, Verarbeiten, Analysieren: Effizienz trifft Skills

Dieses Kapitel widmet sich der Fähigkeit von KI, immense Datenmengen zu erfassen und auszuwerten. Durch ihre präzise Mustererkennung kann die KI wertvolle Erkenntnisse gewinnen, die dem menschlichen Auge oft verborgen bleiben. Doch während KI sichtet und Effizienz schafft, ist es der Mensch, der durchblickt – denn am Ende braucht es den menschlichen Verstand, um diese Ergebnisse zu prüfen, einzuordnen und verantwortungsvolle Entscheidungen zu treffen. Effizienz allein reicht nicht, wenn es um das komplexe Zusammenspiel von Daten und