

Inhaltsverzeichnis

Vorwort.	V
I. Zum Einstieg	1
1. Anwendungsbereich des AI Acts	1
a) AI Systems.	1
(1) Einführung.	1
(2) A Deeper Dive: Qualitative und quantitative Aspekte von AI Systems	5
(3) Kein Anfang und kein Ende? AI-Infrastruktur und AI Systems.	15
(4) Lösungsansätze aus dem Bereich AI Safety, Computa- tion und UML	22
b) Risikobasierter Ansatz	27
c) Persönlicher Anwendungsbereich	29
(1) Provider/Anbieter	30
(2) Product manufacturer/Hersteller	32
(3) Deployer/Betreiber	32
(4) Importer und Distributor/Einführer und Händler	32
d) Räumlicher Anwendungsbereich.	33
e) AI Literacy (AI-Kompetenz)	36
(1) Operationalisierung von AI Literacy	38
i. Programmziele	38
ii. Schulungsbedarf, einschließlich Zielgruppen und Fachkenntnisniveaus.	39
iii. Schulungsinhalte.	40
iv. Schulungsmethoden	41
v. Schulungshäufigkeit	41
vi. Evaluierung	42
(2) Fazit	42
2. Gegenstand dieses Handbuchs	43
II. Design	45
1. Die Idee.	45
2. Legal Requirements Engineering	46
3. Zweckbestimmung	47
a) Usability Engineering	52
b) Human Factors Engineering	53
4. Initiale Risikokategorisierung und -bewertung	57
III. Entwicklung	61
1. Anforderungen an ein AI System	61
a) Qualitätsmanagementsystem	61
(1) Arten des Qualitätsmanagements	62

(2) Struktur und Bestandteile	63
(3) Software- und AI-System-Qualitätsmodelle	66
(4) Operationalisierung der rechtlichen Anforderungen von Art. 17	69
i. Erste Linie	71
ii. Zweite Linie	71
iii. Dritte Linie	73
b) Risikomanagement	75
(1) Risikoidentifizierung	76
(2) Bewertung des Risikos	78
i. Faktoren für die Abschätzung der Eintrittswahr- scheinlichkeit	79
ii. Faktoren zur Bewertung der Auswirkungen	80
(3) Risikominderung	85
(4) Risikoakzeptanz	86
(5) Risikotoleranz	87
(6) Etablierung von Controlling-, Überwachungs- und Incident-Response-Prozessen	87
i. Controlling	87
ii. Monitoring	87
iii. Incident-Response-Prozesse	88
c) Modell-Anforderung: Menschliche Aufsicht	90
d) Genauigkeit, Robustheit und Cybersicherheit von AI Sys- tems (Art. 15 AI Act)	98
(1) Überblick	98
i. Programmsatz für den AI System Life Cycle (Art. 15.1)	98
ii. Benchmarking und Normung (Art. 15.2)	104
iii. Dokumentationspflichten (Art. 15.3)	105
iv. Resilienz (Art. 15.4)	106
v. Cybersicherheit (Art. 15.5)	108
(2) Einzelerläuterungen	116
i. Bedeutung von harmonisierten Normen und Stan- dards im AI Act	116
ii. Beständige Funktion („consistent performance“) während des gesamten AI System Life Cycles: Universal Lessons in Machine Learning	120
iii. Was sind Feedback Loops und wie sind sie zu behandeln?	124
iv. „Engineering Safety in Machine Learning“	129
(a) Standardisierung: Gewonnene Erfahrung	129
(b) AI Engineering Best Practices	132

(c) Praxisbeispiel: Sicherheit und Robustheit von AI Systems im Automotive-Bereich.	135
(d) Inhärent sicheres Design im Machine Learning. .	137
(e) Ein Framework für das Testing von AI Systems .	140
(f) AI Red Teaming	141
(g) NIST-Testplattform „Dioptra“	143
2. Überblick über Data Governance und Data Management	
(Art. 10).	144
a) Machine Learning und Trainingsdaten in a nutshell	144
b) Verpflichtende Qualitätskriterien für Training, Validierung und Testing von AI Systems (Art. 10.1).	145
c) Data Governance und Data Management (Art. 10.2).	151
d) Standardisierung von Data (Quality) Management.	151
e) Definitionen und Metriken	153
(1) Trainingsdaten („training data“), Art. 3.29	153
(2) Validierungsdaten („validation data“), Art. 3.30	154
(3) Validierungsdatensatz („validation data set“), Art. 3.31 ..	154
(4) Testdaten („testing data“), Art. 3.32.	155
f) Die einzelnen Verfahren.	157
(1) Relevant design choices (Art. 10.2.a)	157
(2) Data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection (Art. 10.2.b)	159
i. Datenquellen in der Machine-Learning-Praxis.	159
ii. Datasheets for Datasets.	160
iii. Datenschutzrecht.	165
(3) Relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation (Art. 10.2.c)	173
i. Annotation und Labelling.	173
ii. Data Cleaning	178
iii. Updating	179
iv. Enrichment	179
v. Aggregation	179
(4) Formulation of assumptions, in particular with respect to the information that the data are supposed to measure and represent (Art. 10.2.d)	180
(5) Assessment of the availability, quantity and suitability of the data sets that are needed (Art. 10.2.e)	181
(6) Examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination	

prohibited under Union law, especially where data outputs influence inputs for future operations (Art. 10.2.f) sowie Appropriate measures to detect, prevent and mitigate possible biases identified according to point (Art. 10.2.g)	181
(7) Identification of relevant data gaps or shortcomings that prevent compliance with this Regulation, and how those gaps and shortcomings can be addressed (Art. 10.2.h) . . .	182
g) Der Kampf gegen Bias und Diskriminierung (Art. 10.3, 10.4 und 10.5)	183
(1) Erwägungsgründe und Ethics Guidelines for Trustworthy AI der High-Level Expert Group on Artificial Intelligence (HLEG AI, 2019)	183
(2) Die Fundamental Right Agency (FRA)	187
(3) Forschung und Wissenschaft: It's not just the data, stupid!	188
(4) Internationale Standardisierung	189
(5) Hinreichend relevant, repräsentativ und so weit wie möglich fehlerfrei und vollständig in Hinblick auf die Zweckbestimmung	191
i. Die Zweckbestimmung des Systems	191
(a) Relevanz („relevance“)	192
(b) Repräsentativität („representative“)	192
(c) Fehlerfreiheit („error-free“)	193
(d) Vollständigkeit („complete“)	194
(6) Statistische Merkmale in Datensätzen, einzeln oder kombiniert	195
i. Statistische Merkmale (Statistical Characteristics)	195
ii. Kombination von Datensätzen (Combination of Datasets)	196
(7) Geografisch, kontextuell, verhaltensbezogen oder funktional typische Datensätze	197
i. Geografische Rahmenbedingungen	197
ii. Kontextuelle Rahmenbedingungen	198
iii. Verhaltensbezogene Rahmenbedingungen	199
iv. Funktionale Rahmenbedingungen	200
(8) Verarbeitung sensibler Daten zur Analyse und Mitigation von Verzerrungen (Bias)	200
(9) AI, Bias und europäisches Antidiskriminierungsrecht: ein Überblick	202
i. Rechtsnormen	203
ii. Anwendungsbereiche und geschützte Merkmale	204

iii. Direkte und indirekte Benachteiligung.	204
iv. Rechtfertigung	205
v. Positive Maßnahmen.	206
vi. Beweislastumkehr.	206
vii. Rechtsfolgen	207
viii. Anspruchsgegner	207
3. Testing und Compliance	208
a) Sandboxes	209
(1) Sandboxes im AI Act.	210
(2) Einrichtung und Betrieb von AI-Sandboxes	211
(3) Ein guter Grund für die Teilnahme an AI-Sandboxes. ...	212
(4) Sandbox-Plan.	215
(5) Exit-Berichte	215
(6) Konsequenzen	215
(7) Verarbeitung von Daten innerhalb der Sandbox	216
b) Testen unter realen Bedingungen außerhalb von AI-Regulatory-Sandboxes.	217
c) Konformitätsbewertung.	217
(1) Was?	218
(2) Wann?	220
(3) Wer und Wie?	221
(4) Keine Regel ohne Ausnahme.	224
(5) Warum?	225
d) Harmonisierte Normen, gemeinsame Spezifikationen und Konformitätsvermutung.	228
(1) Harmonisierte Standards	228
(2) Vermutung der Konformität	228
(3) Gemeinsame Spezifikationen	231
(4) Codes of Conduct and Guidelines	233
e) Inverkehrbringen	234
4. Technische Dokumentation.	236
IV. Deployment	239
1. Provider.	239
a) Das Offensichtliche	239
b) Dokumentation (Art. 18) und automatisch erzeugte Protokolle (Art. 19)	241
c) Risikomanagement	242
d) Menschliche Aufsicht	243
e) Transparenz und Bereitstellung von Informationen für Deployer und/oder Endnutzer	245
f) Post-Market Monitoring (Art. 72); Korrekturmaßnahmen und Informationspflicht (Art. 20); Meldung schwerwiegen-	

der Vorfälle (Art. 73) und Zusammenarbeit mit den Behörden	248
(1) Zielsetzung und Zweck	249
(2) Wichtige Merkmale	249
(3) Post-Market-Monitoring-Plan	251
(4) Abschlussbericht	256
(5) Zusammenspiel mit anderen Systemen und Prozessen	257
2. Deployer	259
a) Das Offensichtliche: Due Diligence, Verwendung nach Gebrauchsanweisung (Logs, menschliche Aufsicht), Transparenz, Überwachung und Informationspflicht, Berichterstattung	259
(1) Due Diligence: Die rechtliche Tiefenprüfung	259
(2) Verwendung gemäß der Betriebsanleitung (Protokolle, menschliche Aufsicht)	262
(3) Transparenz	263
(4) Monitoring und Informationspflichten	264
b) Data Governance	265
c) Grundrechte-Folgenabschätzung (FRIA)	267
(1) Wer?	269
(2) Was?	270
(3) Warum?	272
i. Vornahme von Anpassungen und Durchführung zusätzlicher Maßnahmen	273
ii. Bewertung der Akzeptanz von Restrisiken	273
(4) Endergebnis	274
d) Beratung des Betriebsrats	276
e) Datenschutz-Folgenabschätzung	277
f) Art. 25 Verpflichtungen entlang der KI-Wertschöpfungskette	278
V. Besonderheiten	283
1. Urheberrecht, TDM und Generative AI im AI Act	283
a) General Purpose AI im AI Act	283
(1) Anbieter von Modellen und Anbieter von Systemen	283
(2) Pflichten des Providers eines AI-Modells	287
(3) Verpflichtungen der Provider eines General Purpose AI Systems	291
(4) Urheberrecht und General Purpose AI Models	292
b) Training und Trainingsdaten	292
(1) Generative AI benötigt viele Daten	292
(2) Memorization und Overfitting	295
(3) Text- und Data-Mining	296

i.	Richtlinie 2019/790 zum Urheberrecht im digitalen Binnenmarkt (CDSM-Richtlinie)	296
ii.	TDM und Transformer-Modelle	298
iii.	Drei-Stufen-Test	301
iv.	Zusammenfassung	303
c)	Nutzung, Inputs und Outputs	304
(1)	Nutzung und Inputs	304
(2)	Outputs	306
2.	AI in Finanzdienstleistungen	307
a)	Zwei Einschränkungen	308
b)	Eine zusätzliche (wesentliche) Belastung	309
c)	Wo es doch einfacher wird.	309
3.	Biometrische und Emotionserkennungssysteme	311
a)	Biometrische Identifizierung	313
b)	Biometrische Verifizierung	317
c)	Biometrische Kategorisierung	317
d)	Emotionserkennung	320
	Literaturverzeichnis	325