

HANSER



Leseprobe

zu

Handbuch Data Science und KI

von Stefan Papp, Wolfgang Weidinger, Katherine Munro, Bernhard Ortner, Annalisa Cadonna, Georg Langs, Roxane Licandro, Mario Meir-Huber, Danko Nikolić, Zoltan Toth, Barbora Vesela, Rania Wazir, Günther Zauner

Print-ISBN: 978-3-446-46947-1

E-Book-ISBN: 978-3-446-47245-7

E-Pub-ISBN: 978-3-446-47410-9

Weitere Informationen und Bestellungen unter

<https://www.hanser-kundencenter.de/fachbuch/artikel/9783446469471>

sowie im Buchhandel

© Carl Hanser Verlag, München

Inhalt

Geleitwort	XV
Vorwort	XIX
1 Einführung	1
1.1 Was sind Data Science, Machine Learning und Künstliche Intelligenz?	2
1.2 Datenstrategie	9
1.3 Von der Strategie zu den Anwendungsfällen	11
1.3.1 Datenteams	11
1.3.2 Daten und Plattformen	18
1.3.3 Modellierung und Analyse	19
1.4 Implementierung von Anwendungsfällen	19
1.4.1 Iterative Erkundung von Anwendungsfällen	20
1.4.2 End-to-End-Datenverarbeitung	23
1.4.3 Datenprodukte	23
1.5 Beispiele für reale Anwendungsfälle	24
1.5.1 Digitalisierung der Wertschöpfungskette	24
1.5.2 Marketing-Segment-Analyse	25
1.5.3 360°-Sicht auf den Kunden	25
1.5.4 Anwendungsfälle für NGOs und Nachhaltigkeit	26
1.6 Ergebnisse liefern	27
1.7 Kurz und bündig	30
2 Infrastruktur	31
2.1 Einführung	31
2.2 Hardware	33
2.2.1 Verteilte Systeme	36
2.2.2 Hardware für KI-Anwendungen	39
2.3 Linux Essentials für Datenexperten	41
2.4 Terraform	59
2.5 Cloud	63
2.5.1 Basisdienste	66
2.5.2 Cloud-native Lösungen	70
2.6 Kurz und bündig	73

3	Datenarchitektur	75
3.1	Übersicht	75
3.1.1	Maslowsche Bedürfnishierarchie für Daten	76
3.1.2	Anforderungen an die Datenarchitektur	77
3.1.3	Die Struktur einer typischen Datenarchitektur	78
3.1.4	ETL (Extrahieren, Transformieren, Laden)	78
3.1.5	ELT (Extrahieren, Laden, Transformieren)	79
3.1.6	ETLT	80
3.2	Datenerfassung und -integration	80
3.2.1	Datenquellen	81
3.2.2	Traditionelle Dateiformate	82
3.2.3	Moderne Dateiformate	84
3.2.4	Zusammenfassung	86
3.3	Data Warehouses, Data Lakes und Lakehouses	87
3.3.1	Data Warehouses	87
3.3.2	Data Lakes und das Lakehouse	91
3.3.3	Zusammenfassung: Vergleich zwischen Data Warehouses und Lakehouses	93
3.4	Datenverarbeitung und -umwandlung	94
3.4.1	Big Data und Apache Spark	94
3.4.2	Databricks	101
3.5	Workflow-Orchestrierung	103
3.6	Ein Datenarchitektur-Use-Case	105
3.7	Kurz und bündig	110
4	Data Engineering	112
4.1	Integration von Daten	113
4.1.1	Daten-Pipelines	113
4.1.2	Entwerfen von Data Pipelines	120
4.1.3	CI/CD	123
4.1.4	Programmiersprachen	124
4.1.5	Kafka als Referenz-ETL-Werkzeug	127
4.1.6	Entwurfsmuster	132
4.1.7	Automatisierung der Stufen	133
4.1.8	Sechs Bausteine der Data Pipeline	134
4.2	Verwaltung analytischer Modelle	139
4.2.1	Modelllieferung	140
4.2.2	Modell-Update	141
4.2.3	Modell- oder Parameter-Update	142
4.2.4	Modellskalierung	143
4.3	Feedback in die operationalen Prozesse	143
4.4	Kurz und bündig	144

5	Datenmanagement	145
5.1	Datenmanagement	147
5.1.1	Datenkatalog	149
5.1.2	Data Discovery	151
5.1.3	Datenqualität	154
5.1.4	Verwaltung von Stammdaten	156
5.1.5	Gemeinsame Nutzung von Daten	157
5.2	Informationssicherheit	158
5.2.1	Datenklassifizierung	159
5.2.2	Schutz der Privatsphäre	161
5.2.3	Verschlüsselung	163
5.2.4	Secrets Management	165
5.2.5	Defense in Depth	166
5.3	Kurz und bündig	167
6	Mathematik	168
6.1	Lineare Algebra	169
6.1.1	Vektoren und Matrizen	169
6.1.2	Operationen zwischen Vektoren und Matrizen	173
6.1.3	Lineare Transformationen	175
6.1.4	Eigenwerte, Eigenvektoren und Eigendekomposition	176
6.1.5	Andere Matrixzerlegungen	178
6.2	Kalkulus und Optimierung	180
6.2.1	Ableitung	180
6.2.2	Gradient und Hessian	182
6.2.3	Gradientenabstieg	184
6.2.4	Eingeschränkte Optimierung	186
6.3	Wahrscheinlichkeitsrechnung	187
6.3.1	Diskrete und kontinuierliche Zufallsvariablen	188
6.3.2	Erwartungswert, Varianz und Kovarianz	192
6.3.3	Unabhängigkeit, bedingte Verteilungen und Bayes-Theorem	193
6.4	Kurz und bündig	195
7	Statistik – Grundlagen	196
7.1	Daten	197
7.2	Einfache lineare Regression	198
7.3	Multiple lineare Regression	206
7.4	Logistische Regression	209
7.5	Wie gut ist unser Modell?	217
7.6	Kurz und bündig	218

8	Maschinelles Lernen	220
8.1	Einführung	220
8.2	Grundlegendes: Feature Spaces	222
8.3	Klassifizierungsmodelle	226
8.3.1	K-Nearest-Neighbor-Klassifikator	226
8.3.2	Support Vector Machine	227
8.3.3	Entscheidungsbaum	228
8.4	Ensemble-Methoden	230
8.4.1	Bias und Varianz	230
8.4.2	Bagging: Random Forests	232
8.4.3	Boosten: AdaBoost	235
8.5	Künstliche neuronale Netze und das Perceptron	236
8.6	Lernen ohne Label – Struktur finden	239
8.6.1	Clustering	239
8.6.2	Lernen von Mannigfaltigkeiten	240
8.6.3	Generative Modelle	241
8.7	Reinforcement Learning	242
8.8	Übergreifende Konzepte	245
8.9	In die Tiefe gehen – Deep Learning	246
8.9.1	Convolutional Neural Networks	246
8.9.2	Training von Convolutional Neural Networks	248
8.9.3	Recurrent Neural Networks	250
8.9.4	Long Short-Term Memory	251
8.9.5	Autoencoder und U-Netze	253
8.9.6	Adversarial-Trainingsansätze	254
8.9.7	Generative Adversarial Networks	255
8.9.8	Cycle GANs und Style GANs	257
8.9.9	Andere Architekturen und Lernstrategien	258
8.10	Validierungsstrategien für maschinelle Lerntechniken	259
8.11	Schlussfolgerung	260
8.12	Kurz und bündig	261
9	Großartige künstliche Intelligenz erschaffen	262
9.1	Wie KI mit Data Science und maschinellem Lernen zusammenhängt	262
9.2	Eine kurze Geschichte der KI	266
9.3	Fünf Empfehlungen für die Entwicklung einer KI-Lösung	268
9.3.1	Empfehlung Nr. 1: Seien Sie pragmatisch	268
9.3.2	Empfehlung Nr. 2: Erleichtern Sie Maschinen das Lernen – schaffen Sie induktive Verzerrungen	271
9.3.3	Empfehlung Nr. 3: Analysen durchführen	277
9.3.4	Empfehlung Nr. 4: Hüten Sie sich vor der Skalierungsfalle	279
9.3.5	Empfehlung Nr. 5: Hüten Sie sich vor der Verallgemeinerungsfalle (so etwas wie ein kostenloses Mittagessen gibt es nicht)	289

9.4	Intelligenz auf menschlicher Ebene	294
9.5	Kurz und bündig	297
10	Natural Language Processing (NLP)	299
10.1	Was ist NLP, und warum ist es so wertvoll?	299
10.2	NLP-Datenaufbereitungstechniken	301
10.2.1	Die NLP-Pipeline	301
10.2.2	Konvertierung des Eingabeformats für maschinelles Lernen	308
10.3	NLP-Aufgaben und -Methoden	310
10.3.1	Regelbasiert (symbolisch) NLP	311
10.3.2	Ansätze des statistischen maschinellen Lernens	314
10.3.3	Neuronales NLP	323
10.3.4	Transferlernen	329
10.4	Auf dem neuesten Stand: Aktuelle Forschungsschwerpunkte für NLP	342
10.5	Kurz und bündig	345
11	Computer Vision	348
11.1	Was ist Computer Vision?	348
11.2	Ein Bild sagt mehr als tausend Worte	350
11.2.1	Das menschliche Auge	350
11.2.2	Das Bildaufnahmeprinzip	352
11.2.3	Digitale Dateiformate	357
11.2.4	Bildkomprimierung	359
11.3	Ich sehe was, was du nicht siehst	360
11.3.1	Computergestützte Fotografie und Bildmanipulation	363
11.4	Computer-Vision-Anwendungen und zukünftige Richtungen	366
11.4.1	Image-Retrieval-Systeme	367
11.4.2	Objekterkennung, Klassifizierung und Verfolgung	369
11.4.3	Medizinische Computer Vision	371
11.5	Menschen sehen lassen	375
11.6	Kurz und bündig	377
12	Modellierung und Simulation – Erstellen Sie Ihre eigenen Modelle	379
12.1	Einführung	380
12.2	Allgemeine Aspekte	381
12.3	Modellierung zur Beantwortung von Fragen	382
12.4	Reproduzierbarkeit und Lebenszyklus des Modells	384
12.4.1	Der Lebenszyklus einer Modellierungs- und Simulationsfrage	386
12.4.2	Parameter- und Output-Definition	387
12.4.3	Dokumentation	390
12.4.4	Verifizierung und Validierung	391

12.5	Methoden	395
12.5.1	Gewöhnliche Differentialgleichungen (ODEs)	396
12.5.2	Systemdynamik (SD)	397
12.5.3	Diskrete Ereignissimulation	400
12.5.4	Agentenbasierte Modellierung	403
12.6	Beispiele für Modellierung und Simulation	406
12.6.1	Dynamische Modellierung von Eisenbahnnetzen zur optimalen Wegfindung mit agentenbasierten Methoden und Reinforcement Learning	406
12.6.2	Strategien zur agentenbasierten Covid-Modellierung	409
12.6.3	Deep-Reinforcement-Learning-Ansatz für eine optimale Nachschubpolitik in einer VMI-Umgebung	414
12.7	Zusammenfassung und Lessons Learned	417
12.8	Kurz und bündig	418
13	Visualisierung von Daten	422
13.1	Geschichte	423
13.2	Welche Tools Sie verwenden sollten	429
13.3	Arten von Datenvisualisierungen	431
13.3.1	Streudiagramm	432
13.3.2	Liniendiagramm	432
13.3.3	Säulen- und Balkendiagramme	433
13.3.4	Histogramm	434
13.3.5	Tortendiagramm	435
13.3.6	Box Plot	436
13.3.7	Heat Map	436
13.3.8	Baumdiagramm	437
13.3.9	Andere Arten von Visualisierungen	438
13.4	Wählen Sie die richtige Datenvisualisierung	438
13.5	Tipps und Tricks	441
13.6	Präsentation der Datenvisualisierung	446
13.7	Kurz und bündig	447
14	Datengetriebene Unternehmen	449
14.1	Die drei Ebenen eines datengesteuerten Unternehmens	450
14.2	Kultur	450
14.2.1	Unternehmensstrategie für Daten	451
14.2.2	Die Analyse des aktuellen Stands	453
14.2.3	Unternehmenskultur und Organisation einer erfolgreichen Datenorganisation	455
14.2.4	Kernproblem: der Fachkräftemangel	463
14.3	Technologie	465
14.3.1	Die Auswirkungen von Open Source	465
14.3.2	Cloud	466

14.3.3	Auswahl des Anbieters	466
14.3.4	Data Lake aus der Unternehmensperspektive	467
14.3.5	Die Rolle der IT	468
14.3.6	Data Science Labs	468
14.3.7	Revolution in der Architektur: das Data Mesh	469
14.4	Business	470
14.4.1	Daten kaufen und teilen	470
14.4.2	Implementierung des analytischen Anwendungsfalls	472
14.4.3	Self-Service Analytics	472
14.5	Kurz und bündig	473
15	Rechtliche Grundlagen	474
15.1	Einführung	474
15.2	Rechtliche Datenkategorien	475
15.3	Datenschutzgrundverordnung	476
15.3.1	Grundsätze der Datenschutzgrundverordnung	477
15.3.2	Einwilligungserklärung	478
15.3.3	Risikofolgeabschätzung	480
15.3.4	Anonymisierung und Pseudo-Anonymisierung	481
15.3.5	Arten der Anonymisierung	481
15.3.6	Rechtmäßigkeit, Transparenz und Verarbeitung	484
15.3.7	Recht auf Datenlöschung und Korrektur	485
15.3.8	Privacy by Design	486
15.3.9	Privacy by Default	486
15.4	ePrivacy-Verordnung	487
15.5	Datenschutzbeauftragter	487
15.5.1	Internationaler Datenexport in Drittländern	488
15.6	Sicherheitsmaßnahmen	488
15.6.1	Datensicherheit	489
15.7	Datenschutz in Kalifornien im Vergleich zur DSGVO	489
15.7.1	Territoriale Gültigkeit	490
15.7.2	Opt-in versus Opt-out	490
15.7.3	Recht auf Datenexport	491
15.7.4	Das Recht, nicht diskriminiert zu werden	491
15.8	Kurz und bündig	492
15.9	Weiterführende Literatur	493
16	AI in verschiedenen Branchen	494
16.1	Automobilindustrie	498
16.1.1	Vision	499
16.1.2	Daten	499
16.1.3	Anwendungsfälle	500
16.1.4	Herausforderungen	501

16.2	Luftfahrt	502
16.2.1	Vision	503
16.2.2	Daten	504
16.2.3	Anwendungsfälle	504
16.2.4	Herausforderungen	505
16.3	Energie	506
16.3.1	Vision	506
16.3.2	Daten	507
16.3.3	Anwendungsfälle	507
16.3.4	Herausforderungen	508
16.4	Finanzen	509
16.4.1	Vision	509
16.4.2	Daten	509
16.4.3	Anwendungsfälle	510
16.4.4	Herausforderungen	512
16.5	Gesundheit	512
16.5.1	Vision	513
16.5.2	Daten	514
16.5.3	Anwendungsfälle	514
16.5.4	Herausforderungen	515
16.6	Regierung	515
16.6.1	Vision	515
16.6.2	Daten	516
16.6.3	Anwendungsfälle	516
16.6.4	Herausforderungen	520
16.7	Kunst	520
16.7.1	Vision	521
16.7.2	Daten	521
16.7.3	Anwendungsfälle	522
16.7.4	Herausforderungen	522
16.8	Produktion	523
16.8.1	Vision	523
16.8.2	Daten	523
16.8.3	Anwendungsfälle	524
16.8.4	Herausforderungen	525
16.9	Öl und Gas	525
16.9.1	Vision	526
16.9.2	Daten	526
16.9.3	Anwendungsfälle	527
16.9.4	Herausforderungen	528
16.10	Sicherheit am Arbeitsplatz	529
16.10.1	Vision	529
16.10.2	Daten	530
16.10.3	Anwendungsfälle	530
16.10.4	Herausforderungen	531

16.11 Einzelhandel	532
16.11.1 Vision	532
16.11.2 Daten	533
16.11.3 Anwendungsfälle	533
16.11.4 Herausforderungen	534
16.12 Anbieter von Telekommunikation	534
16.12.1 Vision	535
16.12.2 Daten	535
16.12.3 Anwendungsfälle	535
16.12.4 Herausforderungen	537
16.13 Transport	538
16.13.1 Vision	538
16.13.2 Daten	539
16.13.3 Anwendungsfälle	539
16.13.4 Herausforderungen	539
16.14 Lehre und Ausbildung	540
16.14.1 Vision	540
16.14.2 Daten	541
16.14.3 Anwendungsfälle	542
16.14.4 Herausforderungen	542
16.15 Die digitale Gesellschaft	543
16.16 Kurz und bündig	545
17 Mindset und Community	546
17.1 Data Driven Mindset	546
17.2 Data-Science-Kultur	549
17.2.1 Start-up oder Beratungsunternehmen?	549
17.2.2 Labs statt Konzernpolitik	550
17.2.3 Keiretsu statt Einzelkämpfertum	551
17.2.4 Agile Softwareentwicklung	552
17.2.5 Firmen- und Arbeitskultur	553
17.3 Antipatterns	556
17.3.1 Abwertung von Fachwissen	556
17.3.2 Die IT wird es schon richten	557
17.3.3 Widerstand gegen Veränderungen	558
17.3.4 Besserwisser-Mentalität	558
17.3.5 Schwarzmalerei	559
17.3.6 Pfennigfuchseriei	560
17.3.7 Angstkultur	560
17.3.8 Kontrolle über die Ressourcen	561
17.3.9 Blindes Vertrauen in die Ressourcen	561
17.3.10 Das Schweizer Taschenmesser	562
17.3.11 Over-Engineering	563
17.4 Kurz und bündig	564

- 18 Vertrauenswürdige KI 565**
 - 18.1 Rechtlicher und Soft-Law-Rahmen 566
 - 18.1.1 Normen 568
 - 18.1.2 Verordnungen 569
 - 18.2 KI-Stakeholder 571
 - 18.3 Fairness in der KI 572
 - 18.3.1 Bias 573
 - 18.3.2 Fairness-Metriken 576
 - 18.3.3 Unerwünschten Bias in KI-Systemen reduzieren 580
 - 18.4 Transparenz von KI-Systemen 581
 - 18.4.1 Dokumentieren der Daten 582
 - 18.4.2 Dokumentieren des Modells 584
 - 18.4.3 Explainability (Erklärbarkeit) 585
 - 18.5 Schlussfolgerung 587
 - 18.6 Kurz und bündig 587
- 19 Die Autor:innen 588**
- Index 593**

Geleitwort

*„Mathematical science shows what is. It is the language of unseen relations between things.
But to use and apply that language, we must be able to fully appreciate, to feel, to seize the
unseen, the unconscious.“*

Ada Lovelace

So wie die Computerkompetenz vor einer Generation eine neue Reihe von grundlegenden Fähigkeiten darstellte, die es zu erwerben galt, so stellt die Kompetenz im Bereich der künstlichen Intelligenz (KI) für unsere heutigen Generationen und darüber hinaus das Gleiche dar. In den letzten zwei Jahrzehnten hat sich die Datenwissenschaft zur mathematischen Architektur und zu der entsprechenden Sprache entwickelt, mit der wir Systeme aufbauen und in der wir mit ihnen interagieren, die unsere Sinne und Entscheidungsfähigkeiten erweitern. Es reicht also nicht mehr aus, Befehle per Mausklick an Computer zu senden, sondern es ist von entscheidender Bedeutung, dass wir in der Lage sind, KI-gestützte Empfehlungen von Computern zu interpretieren und mit ihnen zu interagieren. Derzeit verarbeiten Maschinen, d. h. Computer, die mit Sensoren (im weitesten Sinne) gekoppelt sind, ein immer breiteres Spektrum an Daten, darunter Texte, Bilder, Videos, Audio-daten, Netzwerkdiagramme und eine Vielzahl von Informationen aus dem Internet, der Privatwirtschaft und dem öffentlichen Sektor. In Anbetracht der Datenvielfalt nähern sich die Autoren dieses Buches der Datenwissenschaft als einem grundlegenden Schlüsselthema für die Gesellschaft. Sie tun dies mit großer Einsicht, aus mehreren wichtigen Blickwinkeln und in einem unterhaltsamen Stil, der bei Anfängern und Experten gleichermaßen Anklang findet.

Aus Daten Nutzen zu ziehen ist wohl das verbindende Ziel des Wissensarbeiters des 21. Jahrhunderts. Sogar Berufsbereiche, die als klassisch datenfern galten, wie Verkauf und Kunst, haben jetzt datengetriebene Teilbereiche wie Marketingautomatisierung und Computergestaltung. Zum Nutzen der Leser bringen die Autoren Erfahrungen aus erster Hand und sorgfältige Recherchen ein, um überzeugend darzulegen, welche Rolle wir alle spielen müssen, wenn wir versuchen, Daten für bessere Ergebnisse zu nutzen. In der Tat ist die Bandbreite, die in diesem Werk vermittelt wird, beeindruckend. Sie reicht von Überlegungen zur Hardwareleistung (z. B. CPU, Netzwerk, Speicher, I/O, GPU) bis hin zu den verschiedenen Rollen von Teammitgliedern beim Aufbau von Maschinen, die Muster in Daten finden können. Darüber hinaus gehen die Autoren auf die Möglichkeiten ein, mit denen Maschinen heute sehen und lesen können, nämlich Computer Vision und Natural Language Processing, was tiefgreifende Auswirkungen auf fast alle Industriebereiche hat.

Bei der Lektüre dieses Buches möchte ich Sie ermutigen, neugierig zu sein und sich eine Reihe von Fragen zu stellen, wie Ihr beruflicher Werdegang und die Gesellschaft, wie Sie sie sehen, gegenwärtig von immer fortschrittlicheren Maschinen beeinflusst werden: von den Möglichkeiten, die Ihr Smartphone bietet, bis hin zu der Art und Weise, wie Arbeits-

plätze auf dem Markt durch Automatisierungstools umgestaltet werden. Hier sind einige Fragen, die Ihnen den Einstieg erleichtern sollen:

- Wie verschiebt sich das Verhältnis der Aufgaben, mit denen Sie Ihre Zeit verbringen, mit dem Aufkommen von immer fortschrittlicheren Maschinen in Ihrem Arbeitsbereich?
- Was bedeutet es, wenn Maschinen über Wahrnehmungsfähigkeiten verfügen, die denen des Menschen entsprechen, also sehen, hören, riechen, schmecken, tasten und mehr?
- Wie gehen wir als Gesellschaft mit der Voreingenommenheit und dem Vertrauen in Daten um?
- Wie können wir den Bau und die Nutzung von Maschinen, die lernen, inklusiver gestalten?
- Welche eindeutig menschlichen Fähigkeiten können Sie hervorheben, um Organisationen, die Ihnen am Herzen liegen, zu mehr Wettbewerbsfähigkeit und Nachhaltigkeit zu verhelfen?

Ich habe mich davor gehütet, den Begriff „denkende Maschinen“ oder „künstliche allgemeine Intelligenz“ zu verwenden, um eine Übertreibung zu vermeiden. Worauf ich Ihre Aufmerksamkeit lenken möchte, ist die breite Anwendbarkeit dessen, was wir in der Forschung rund um Maschinen mit Lernfähigkeiten sehen. Aus meiner Zeit in den Labors der Universitäten Columbia und Cornell, dem Princeton Plasma Physics Laboratory, der American University of Armenia und dem von der NASA unterstützten TRISH (Translational Research Institute for Space Health), das mit TrialX zusammenarbeitet, ist mir klar, dass Maschinen in einem enorm breiten Spektrum von Bereichen Muster in Daten finden und Menschen sowohl in normalen als auch in missionskritischen Kontexten alarmieren können. Die Auswirkungen auf die menschliche Erfahrung sind also vielschichtig, und Data Scientists spielen eine wichtige Rolle bei der Entwicklung von Systemen, bei denen die menschliche Interaktion mit dem maschinellen Output eine positive Summe ergibt. Ich kann nicht genug betonen, dass ein Nullsummen-Ansatz bei der Automatisierung suboptimal ist. Unternehmer neigen jedoch dazu, einen Weg zur maximalen Summe zu finden.

Gemeinsam mit Kollegen und durch meine Arbeit beim BAI Accelerator und Covenant Venture Capital unterstütze ich Start-ups bei einer Art Tandem-Lernen: wie ein schnell wachsendes Unternehmen eine Branche umgestalten kann, indem es Marktlücken aufspürt, und wie die Erfindung eines Unternehmens lernen und neue Fähigkeiten für Kunden bereitstellen kann. In dem leistungsstarken Technologiebereich der Computer Vision, der eine tragende Säule der Datenwissenschaft ist, stechen beispielsweise drei Unternehmen hervor, die in drei sehr unterschiedlichen Industriebereichen bahnbrechend sind: Embodied, Scylla und cognaize in den Bereichen Gesundheitswesen, Sicherheit bzw. Finanzen.

- Das Vorzeigeprodukt von Embodied, Moxie, ist ein Roboter, der das emotionale Wohlbefinden und die soziale Entwicklung von Kindern unterstützt. Um dies zu erreichen, muss Moxie die Familienmitglieder auf überzeugende Weise sehen und mit ihnen kommunizieren. Er muss den emotionalen Zustand der Menschen, mit denen er interagiert, sowohl visuell als auch durch andere Hinweise verstehen, um einen sinnvollen Dialog führen zu können. Die Gesundheitsdienstleister haben also ein neues robotisches Teammitglied, mit dem sie zusammenarbeiten können. Embodied war auf der Titelseite des TIME Magazine zu sehen.

- Scylla ermöglicht es dem Sicherheitsteam eines Unternehmens, die Sicherheit proaktiv zu verbessern. Mit Echtzeit-Erkennungsfunktionen müssen Kameranetzwerke nicht mehr passiv sein, sondern können proaktiv eingesetzt werden. Die Anwendungsmöglichkeiten sind vielfältig und reichen von der Erkennung von Ausrutschern und Stürzen in Krankenhäusern und Stadien, um die Gesundheitsergebnisse zu verbessern, bis hin zur Alarmierung von Eindringlingen in Produktionsstätten und Bürogebäuden, um die Mitarbeiter besser zu schützen. Scylla wurde bereits in Forbes vorgestellt.
- cognaize unterstützt Finanzinstitute und Versicherungsunternehmen bei der Verarbeitung einer enormen Menge unstrukturierter Daten zur Risikobestimmung. Eine wichtige Erkenntnis ist, Dokumente nicht nur als Text zu betrachten, sondern auch visuelle Informationen zu berücksichtigen: Stil, Tabellen, Struktur. Darüber hinaus verfügt cognaize über einen „Human-in-the-Loop“, bei dem Kollegen und das System insgesamt kontinuierlich lernen. cognaize wurde auf dem NASDAQ-Bildschirm am Times Square vorgestellt.

In den drei oben genannten Beispielen für aufstrebende Unicorn-Start-ups arbeiten Data Scientists eng mit Ingenieuren, Analysten, Designern, Inhaltentwicklern, Fachleuten und Kunden zusammen, um Maschinen zu entwickeln, die lernen und auf nuancierte Weise mit Menschen interagieren. Das Ergebnis ist ein Wandel in der Art der Arbeit: Menschen werden auf die wichtigsten Dokumente oder Momente aufmerksam gemacht, und aus der menschlichen Erfahrung wird gelernt, um die Qualität zu verbessern. Dies steht stellvertretend für einen neuen Wandel, der KI-Kenntnisse voraussetzt: Arbeitsplätze in fast allen Bereichen der Wirtschaft werden Aspekte aufweisen, die eine maschinelle Interaktion erfordern: Menschen, die Korrekturen vornehmen, neue Fähigkeiten erlernen, auf Warnungen reagieren und diese interpretieren und eine schnellere Reaktionszeit haben, um anderen Menschen zu helfen, indem sie Maschinen zur Unterstützung einsetzen. In den kommenden Jahren bin ich gespannt auf die Rolle der Datenwissenschaft in der Schnittstellenforschung, auf neue Algorithmen und darauf, wie Menschen ihre Arbeit um ein Vielfaches verstärken können.

Als ich vor fast einem Jahrzehnt die erste Ausgabe von *The Field Guide to Data Science* mit verfasst habe, war es bemerkenswert, wie sehr sich die Disziplin weiterentwickelt hat, sowohl in Bezug auf das, was technisch erreicht wurde, als auch in Bezug auf das, was noch zu erreichen ist. Das Handbuch Data Science bringt die Disziplin in diesen beiden Dimensionen voran und trägt die Fackel weiter.

Lesen Sie weiter.

Herbst 2021

Armen R. Kherlopian, Ph.D.

Vorwort

“The job of the data scientist is to ask the right questions.”

Hillary Mason

Als ich das Vorwort der ersten Ausgabe las, wurde ich das Gefühl nicht los, dass einige Trends im Wesentlichen gleich geblieben sind, während andere ganz plötzlich auftauchten und die Gesellschaft und Unternehmen wie eine Lawine überrollten.

Wenn wir mit den Veränderungen beginnen, die die Gesellschaft tiefgreifend getroffen haben, ist die Pandemie natürlich eine davon. Abgesehen von den unzähligen Folgen, die sie für unser Leben hatte und immer noch hat, möchte ich mich auf die Facetten konzentrieren, die mit dem Thema dieses Buches zu tun haben: Data Science und KI.

Vereinfacht gesagt, hatte dies zur Folge, dass ganze Gesellschaften und unsere gesamte Lebensweise im Handumdrehen datengesteuert wurden. Kennzahlen wie die siebentägige Inzidenzrate oder Prognosen auf der Grundlage von Pandemiesimulationen lenkten unseren Alltag und setzten zeitweise sogar Grundrechte wie das Recht, die Wohnung zu verlassen, außer Kraft. Dies führte zu Diskussionen und Fragen, die jedem Data Scientist mit etwas Erfahrung vertraut und ihm im Laufe seines Berufslebens immer wieder begegnet sind, beispielsweise:

- Können wir diesen Modellen und ihren Vorhersagen vertrauen?
- Ist der gewählte KPI wirklich der richtige für diesen Zweck?
- Sind die zugrundeliegenden Daten quantitativ und qualitativ gut genug?

All diese Fragen sind berechtigt und werden, wie schon vor zwei Jahren, von einem anderen Trend befeuert: der Digitalisierung. Der Motor dafür sind Daten. Darüber hinaus verfolgen Data Scientists immer noch das gleiche Ziel:

Mithilfe von Daten verständliche Antworten auf Fragen zu geben.

Trotz aller Trends bleibt dieses Ziel gleich und wird immer eine der zentralen Säulen von Data Science sein.

Aber das ist nicht der einzige Trend, der entweder gleich geblieben oder noch stärker geworden ist. Das wichtigste anhaltende Phänomen ist der immer noch massive Hype, der durch Begriffe wie „Künstliche Intelligenz“ und „Data Science“ ausgelöst wird. Obwohl diese Bereiche unglaublich wertvoll und leistungsfähig sind, wecken Diskussionen darüber leider oft falsche Versprechungen und verzerrte Erwartungen, die wiederum zu Enttäuschungen führen. Einige Unternehmen haben in der Vergangenheit bereits große ehrgeizige Initiativen gestartet, die zu enttäuschenden Ergebnissen geführt haben, weil die Erwartungen zu hoch und die Zeitvorgaben zu kurz waren. So ist beispielsweise das vollautonome Fahren ein besonders schwieriges Problem, das es zu lösen gilt.

Dennoch bleibt künstliche Intelligenz für viele Unternehmen die Hoffnung schlechthin. Investoren sehen sie als eine universell einsetzbare Technologie, die fast überall angewendet werden kann. Die Situation ist vergleichbar mit der Entwicklung in den Neunzigerjahren, als alles, was mit dem „Internet“ zu tun hatte, einen Aufschwung erlebte. Plötzlich brauchte jedes Unternehmen eine Webseite, und es wurden erhebliche Investitionen in die Ausbildung von Webprogrammierern getätigt. Ähnlich verhält es sich heute mit allem, was mit KI zu tun hat. Auch hier sind die Investitionen enorm, und es gibt eine riesige Anzahl von Kursen zu diesem Thema. Letztendlich hat die Entwicklung des „Internets“ zu einem riesigen Ökosystem von Unternehmen und Anwendungen geführt, die das Leben von Milliarden von Menschen auf tiefgreifende Weise beeinflussen, und es scheint, dass KI einen ähnlichen Weg einschlägt.

Dies erklärt zumindest teilweise einen weiteren auffälligen Trend: die weitere Spezialisierung von Data Science Rollen mit Bezeichnungen wie „Data Translator“ oder „ML Engineer“. Dies ist eine natürliche Entwicklung, da es ein Zeichen dafür ist, dass das Feld reifer wird, aber es birgt auch die Gefahr, dass die Verantwortlichkeiten im Bereich Data Science über schlecht koordinierte Organisationen verstreut sind und somit nicht ihr volles Potenzial erreichen. In den Kapiteln 14 und 17 wird dies noch ausführlicher behandelt.

Schließlich entwickelt sich die „vertrauenswürdige KI“ als eine weitere, äußerst wichtige Bewegung innerhalb von Data Science. Dabei handelt es sich um einen Forschungsbereich, der darauf abzielt, einige bisher unerfüllte Anforderungen wie Erklärbarkeit oder Fairness zu erfüllen. Aus diesem Grund wird es als eines der neuen Kapitel in dieses Buch aufgenommen (Kapitel 18).

Angeichts all dieser Trends in Data Science ist einer der Gründe für die Gründung der Vienna Data Science Group (VDSG) in den letzten zwei Jahren noch wichtiger geworden: einen neutralen Ort zu schaffen, an dem ein internationaler und interdisziplinärer Wissensaustausch zwischen allen beteiligten Experten stattfinden kann. Wir engagieren uns nach wie vor sehr stark für die Entwicklung des gesamten Data-Science-Ökosystems (Ausbildung, Zertifizierung, Standardisierung, gesellschaftliche Wirkungsforschung etc.) in Europa und darüber hinaus.

Ein Produkt des Austauschs in unserer Gemeinschaft ist die 2. Auflage dieses Buches, das erheblich erweitert wurde, um Themen wie KI (Kapitel 9), maschinelles Lernen (Kapitel 8), NLP (Kapitel 10), Computer Vision (Kapitel 11) oder Modellbildung und Simulation (Kapitel 12) eingehender zu behandeln. Um unser Ziel zu verfolgen, die Gesellschaft über Data Science und ihre Auswirkungen aufzuklären, wurde in Kapitel 12 ein sehr relevanter und aktueller Anwendungsfall aufgenommen: ein agentenbasiertes Covid-19-Modell, das einen Eindruck zu den möglichen Auswirkungen bestimmter Maßnahmen und deren Kombination auf die Ausbreitung der Krankheit liefern soll.

Um unseren Lesern eine solide Grundlage zu bieten, wurde eine Einführung in die zugrunde liegende Mathematik (Kapitel 6) und Statistik (Kapitel 7), die in Data Science verwendet werden, aufgenommen und mit einem Abschnitt über Visualisierung (Kapitel 13) abgeschlossen.

Trotz der vielen neuen Inhalte ist das Ziel dieses Buches dasselbe geblieben und sogar noch wichtiger geworden: ein realistisches Bild von Data Science zu vermitteln.

Denn trotz aller Trends bleibt auch Data Science dasselbe: eine interdisziplinäre Wissenschaft, die eine sehr heterogene Schar von Spezialisten versammelt, die sich aus drei großen Strömungen zusammensetzt:

- Informatik/IT
- Mathematik/Statistik
- Fachwissen in der Branche, in der Data Science angewendet wird.

Die Wissenschaft zielt darauf ab, neues Wissen zu generieren, und dieses wird nach wie vor genutzt, um

- bestehende Geschäftsprozesse in einem Unternehmen zu verbessern (Kapitel 16) und
- völlig neue Geschäftsmodelle zu ermöglichen.

Data Science ist auf dem Vormarsch, und ihre direkten und indirekten Auswirkungen auf die Gesellschaft nehmen rasant zu, wie die Pandemie zeigt. In einigen Bereichen ist eine gewisse Ernüchterung eingetreten, was jedoch als gesunde Entwicklung gesehen werden kann, um dem Hype entgegenzuwirken. Die Rollen innerhalb von Data Science Teams werden immer differenzierter, und immer mehr Unternehmen setzen Data Science Projekte in die Produktion um.

Data Science ist erwachsen geworden und tritt gerade in eine neue Ära ein.

Frühjahr 2022

Wolfgang Weidinger



Die Kapitel in diesem Buch wurden von mehreren Autor:innen geschrieben, von denen manche aufgrund der besseren Lesbarkeit auf eine gendergerechte Sprache verzichtet haben. Selbstverständlich sprechen wir aber alle Personen gleichermaßen an.

Großartige künstliche Intelligenz erschaffen

Danko Nikolić

“We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon im Jahr 1955



Fragen, die in diesem Kapitel beantwortet werden:

- Was ist KI, und wie unterscheidet sie sich von der einfachen Erstellung von Modellen für maschinelles Lernen?
- Was braucht es, um ein großartiges KI-Produkt zu entwickeln?
- Was sind die häufigsten Fallen bei der Konzeption und Entwicklung einer KI, und wie können Sie diese Fallen vermeiden?

■ 9.1 Wie KI mit Data Science und maschinellem Lernen zusammenhängt

Sie fragen sich vielleicht, welche Funktion ein Kapitel über künstliche Intelligenz (KI) in einem Buch über Data Science hat? Oft wird KI nur als ein schicker Name für maschinelle Lernmodelle verstanden, Modelle, die Data Scientists ohnehin im Rahmen ihrer Arbeit erstellen. Wenn das der Fall wäre, wäre KI einfach ein Teil der Data Science, und es gäbe keinen Grund, ein separates Kapitel über KI zu schreiben, da sich der Rest des Buches um diese Technologie dreht. Nun, das ist nicht ganz richtig. Es stimmt zwar, dass einer der wichtigsten – und vielleicht auch interessantesten – Teile der KI in den maschinellen Lernmodellen liegt, aber KI besteht aus viel mehr als nur maschinellem Lernen. Bei der Entwicklung eines KI-Produkts sind einige kritische Überlegungen anzustellen, die Sie normalerweise nicht in einem typischen Buch über Data Science oder sogar in anderen Kapiteln dieses Buchs finden. Wenn Sie in einem dieser Bereiche einen Fehler machen, kann Ihr Endprodukt enttäuschen. Sie können beispielsweise in eine Situation geraten, in der zu Beginn des Erstellungsprozesses alles in Ordnung zu sein scheint, das Endprodukt jedoch nicht überzeugt und die Bedürfnisse und Erwartungen der Endnutzer nicht erfüllt.

Sehen wir uns zunächst an, welche Art von Maschinen wir heute als Beispiele für KI betrachten. Was uns sofort in den Sinn kommen mag, ist vielleicht ein Roboter. Aber nicht irgendein Roboter. Die meisten Roboter sind nicht sehr intelligent. Roboter bestehen aus mechanischen Komponenten wie Armen und Aktuatoren. Und dann gibt es noch Batterien und Sensoren. Aber das allein reicht nicht aus, um einen Roboter als KI zu bezeichnen. Es gibt viele Roboter, die zwar sehr nützlich für uns sind, aber einfach nur dumm sind. Beispiele dafür sind Staubsaugerroboter in Privathaushalten und Industrieroboter in Fabrikhallen. Ausschlaggebend dafür, ob ein Roboter das Prädikat „künstlich intelligent“ erhält oder nicht, ist, was er mit seiner gesamten Hardware selbstständig tun kann. Nur ein intelligenter Roboter, der über Fähigkeiten verfügt, die weit über die reine Programmierung von Bewegungen hinausgehen, wird die Ehre haben, als KI bezeichnet zu werden. Wir suchen hier nach einem Roboter, der eine Vielzahl unterschiedlicher Verhaltensweisen zeigt, sich in einer komplexen Umgebung zurechtfindet oder Aufgaben in einer Vielzahl neuartiger Situationen bewältigen kann. Stellen Sie sich zum Beispiel einen anthropomorphen Roboter vor, der in der Lage ist, einen Tisch voller schmutzigem Geschirr abzuräumen, dieses Geschirr dann manuell abzuwaschen und es schließlich abzutrocknen und in den Schrank zu stellen – und das alles, ohne etwas kaputt zu machen! Roboter mit derartigen Fähigkeiten gibt es noch nicht.

Um mit der Entwicklung eines solchen Roboters zu beginnen, könnte es bald klar sein, dass es nicht ausreicht, Deep-Learning-Modelle zu trainieren. Man kann sich dafür entscheiden, sich in hohem Maße auf Deep Learning zu verlassen, und doch wird der Roboter viel mehr brauchen als das, was Deep Learning bieten kann. Um die erforderliche Intelligenz des Roboters zu fördern, müssen wir Technologien entwickeln und einsetzen, die weit über die Möglichkeiten des maschinellen Lernens hinausgehen – und auch weit über das hinausgehen, was Data Science abdeckt. Dennoch werden Data Scientists eine entscheidende Rolle bei der Entwicklung solcher Roboter spielen. Das ist der Grund, warum Sie dieses Kapitel lesen.

Eine Art von Robotern hat die Aufmerksamkeit der Industrie auf sich gezogen, und es wurde auch viel investiert: unsere Autos. Es wurde viel Geld in die Entwicklung von Autos gesteckt, die in der Lage sind, selbstständig zu fahren und somit zu intelligenten Robotern zu werden. Das Problem des autonomen Fahrens ist nicht einfach, vor allem dann nicht, wenn das Fahrzeug in der „echten Welt“ und nicht in einer kontrollierten Testumgebung fährt. Die Vielfalt der unterschiedlichen Situationen, denen das Fahrzeug begegnen kann, ist enorm. Daher stellen solche Fahrzeuge eine große Herausforderung für die Technologie dar. Vielleicht ist das Problem des autonomen Fahrens so schwierig wie das Abräumen eines Tisches mit Geschirr. Der Druck auf die Qualität der Lösung, d. h., keinen Fehler zu machen, ist ebenfalls hoch. Während unser manueller Geschirrspülerroboter im schlimmsten Fall ein paar Gläser oder Teller kaputt macht, trägt ein Autoroboter eine viel größere Verantwortung: Er ist für Menschenleben verantwortlich. Dies ist ein weiterer Grund, der ein erfolgreiches selbstfahrendes Auto zu einem schwierigen Ziel macht. Dennoch hat es in diesem Bereich einige Fortschritte gegeben. Autonome Fahrzeuge sind wohl die schlauesten und intelligentesten Roboter, die die Menschheit bisher gebaut hat. Und doch gibt es noch viel zu tun. Die Frage ist also: Was war nötig, um diese Maschinen intelligent zu machen? Und vor welchen Problemen und Hürden stehen diese Maschinen in Bezug auf Intelligenz noch? Ist das alles nur Data Science, um größere und intelligentere Modelle zu bauen, oder steckt da mehr dahinter?

Um diese Fragen zu beantworten, müssen wir zunächst feststellen, dass KI nicht gleichbedeutend mit einem maschinellen Lernmodell ist. Um das zu verstehen, hilft es, zwischen einem Produkt und einer kritischen Komponente zu unterscheiden, die für den Aufbau eines Produkts erforderlich ist. Ein Produkt ist viel mehr als nur seine kritischen Komponenten. Ein Messer ist mehr als nur eine Klinge, auch wenn die Klinge der entscheidende Bestandteil ist. Ein Monitor ist mehr als seine kritische Komponente, der Bildschirm. Ein Speicherstick ist mehr als ein SSD-Chip. Ein Fahrrad ist mehr als nur ein Paar Räder und Pedale. In all diesen Fällen stellen wir fest, dass ein Produkt mehr ist als seine entscheidenden Komponenten.

Wir können diesen Unterschied am Beispiel eines Autos erkennen. Ein Auto, das sich auf dem Markt verkaufen lässt und somit geeignet ist, einen Wert für den Kunden zu schaffen, ist mehr als ein Motor auf vier Rädern. Für ein Auto braucht man ein Lenkrad und Bremsen. Doch das ist noch kein vollständiges Produkt. Zu einem vollständigen Produkt gehören auch Scheinwerfer für Nachtfahrten, eine Windschutzscheibe, Türen, Fenster an den Türen und Scheibenwischer an der Windschutzscheibe. Man braucht auch eine vollständige Kabine mit Sitzen. Dann braucht man eine Heizung, eine Klimaanlage und ein Unterhaltungssystem. All dies muss in ein schönes Design verpackt werden, das dem menschlichen Auge gefällt. Erst wenn wir all dies zusammengefügt haben, haben wir ein vollständiges Produkt, das wir Auto nennen.

Eine KI ist wie ein vollständiges Produkt. Sie ist eine Maschine, die einen Dienst für einen Menschen leistet, und damit dieser Dienst zufriedenstellend erledigt werden kann, muss die Maschine vollständig sein. Man muss ein vollständiges Produkt schaffen. Ein maschinelles Lernmodell kann also eine entscheidende Komponente für eine KI sein, vielleicht das Äquivalent zu einem Motor für ein Auto. Wichtig ist jedoch, dass wir erst dann eine KI haben, wenn wir ein Produkt um diesen (maschinellen Lern-)Motor herum gebaut haben.

In der Praxis erfordert die Erstellung eines Produkts zumindest, dass das Modell in Produktion geht und eine Schnittstelle für die Erfassung der Eingaben, die in das Modell für maschinelles Lernen einfließen, eingerichtet wird und dann auch eine Form von Ausgabe erzeugt wird. Oft ist viel mehr erforderlich, um ein nützliches Produkt zu schaffen. Wie wir im Fall des autonomen Fahrzeugs gesehen haben, ist eine Menge Hardware erforderlich, um ein komplettes Auto zu bauen.

Aber es sind nicht nur „nicht intelligente“ Komponenten, die man zu maschinellen Lernmodellen hinzufügen muss, um eine KI zu schaffen. Ein tieferer Grund, warum maschinelles Lernen allein für KI nicht ausreicht, ist, dass KI-Lösungen oft sehr viel komplexer sind als das, was mit einem einzelnen maschinellen Lernmodell erreicht werden könnte. Betrachten wir zum Beispiel einen Chatbot. Nehmen wir an, dass wir außer der intelligenten Komponente nur eine minimale Schnittstelle schaffen müssen, die aus Textfeldern besteht, in die die Fragen der Benutzer eingegeben und die Antworten der Maschine ausgedruckt werden. Daraus könnte man schließen, dass es ausreichen sollte, zwischen diesen beiden Komponenten ein großes, gut trainiertes maschinelles Lernmodell zu platzieren, das das Chatten mit einem menschlichen Benutzer übernimmt. Leider funktioniert das so nicht. Jeder ausgeklügelte intelligente Chat-Assistent (man denke an Alexa, Siri, Cortana usw.) ist viel komplexer als ein einzelnes Deep-Learning-Modell.

Nachfolgend ist die Architektur der ursprünglichen Watson AI die Lösung – eine Maschine, die 2010 Geschichte schrieb, als sie das Spiel Jeopardy gegen die besten menschlichen

Spieler in diesem Spiel gewann. Es ist klar, dass die Organisation dieser KI viel aufwendiger war als ein einzelnes maschinelles Lernmodell. Viele ihrer Komponenten beruhen nicht einmal auf maschinellem Lernen, tragen aber dennoch zur Gesamtintelligenz von Watson bei. Es ist wichtig zu verstehen, dass nur die Maschine als Ganzes eine KI ist; keine einzelne Komponente allein ist eine. Ein großer Teil dieser Gesamtintelligenz kommt von der Architektur – wie der Rechenfluss organisiert ist und wie entschieden wird, welche Komponente wann ausgeführt wird. Es sind also nicht nur die Gewichte in den maschinellen Lernmodellen, die zur Gesamtintelligenz beitragen. Es gibt noch viel mehr, einschließlich der Regeln, nach denen die verschiedenen Modelle miteinander interagieren und sich gegenseitig helfen. Erst die vollständige Kombination aller Teile, also Watson, ist ein vollständiges Produkt und eine KI.

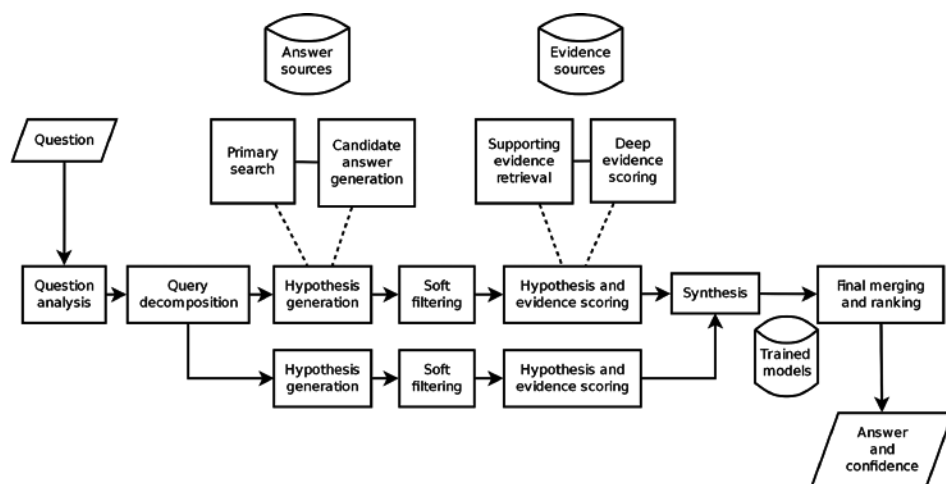


Bild 9.1 Die Architektur der ursprünglichen Watson-KI, die das Spiel Jeopardy gegen die besten menschlichen Konkurrenten gewann (<https://en.wikipedia.org/wiki/File:DeepQA.svg>)

Ähnliches gilt für die Intelligenz von autonomen Fahrzeugen. Die internen Architekturen der Algorithmen, die die Autos steuern, sind nicht einfacher als die von Watson. Außerdem werden die Autos mit der Zeit immer intelligenter und bessere Fahrer, sodass die Zahl der Komponenten und die interne Komplexität der KI-Gesamtlösungen tendenziell zunehmen.

Wichtig ist, dass viele der Komponenten solcher Lösungen auch keine Modelle des maschinellen Lernens sind, sondern andere Algorithmen verwenden. Diese anderen Komponenten können die Suche in Datenbanken, Brute-force-Ansätze zum Finden optimaler Lösungen, rein wissenschaftliche Berechnungen, regelbasierte Entscheidungsfindung und so weiter sein. Auch hier tragen alle diese Komponenten gemeinsam zur Gesamtintelligenz der KI bei.

Schließlich gibt es einen weiteren Grund, warum maschinelles Lernen und KI nicht dasselbe sind. Maschinelles Lernen wird oft für andere Zwecke als die Entwicklung intelligenter Maschinen eingesetzt. Maschinelles Lernen hat Verwendungszwecke, die über das hinausgehen, wofür KI gedacht ist. Insbesondere wird maschinelles Lernen oft als Werkzeug für die Datenanalyse eingesetzt. Der Autor dieses Kapitels hat maschinelles Lernen ausgiebig genutzt, um zu analysieren, wie das Gehirn sensorische Informationen speichert. Wir

haben maschinelle Lernmodelle trainiert, um Informationen aus Gehirnsignalen zu lesen. Entscheidend ist, dass es uns nicht darum ging, ein Produkt zu entwickeln. Vielmehr stellten wir Fragen über das Gehirn, z. B. wie lange speichert das Gehirn Informationen über ein Bild, das wir kurz auf dem Bildschirm gezeigt haben? Oder: Wie schnell kann diese Information durch einen neu dargebotenen Reiz wieder gelöscht werden? Auf diese Weise haben wir zahlreiche Erkenntnisse darüber gewonnen, wie das Gehirn sensorische Informationen aufbewahrt [1-3]. Für reine Engineers mag eine solche Anwendung des maschinellen Lernens überraschend sein. Für einen Data Scientist sollte dies jedoch nicht so unerwartet sein. Kein Wissenschaftler sollte zögern, Algorithmen des maschinellen Lernens als Analyserwerkzeuge einzusetzen. Der Einsatz von maschinellem Lernen bringt große Vorteile mit sich, insbesondere in Situationen, in denen die Daten komplex und Erkenntnisse mit herkömmlichen Analysemethoden nur schwer zu gewinnen sind.

Um die Beziehung zwischen maschinellem Lernen und KI zu verstehen, ist es üblich, Venn-Diagramme zu zeichnen, wie sie in Bild 9.2 zu sehen sind. Das Venn-Diagramm auf der linken Seite ist dasjenige, das in der KI-Literatur häufig zu finden ist. Das rechte Diagramm ist jedoch korrekter, da es auch die Tatsache berücksichtigt, dass maschinelles Lernen für andere Zwecke als KI eingesetzt werden kann.

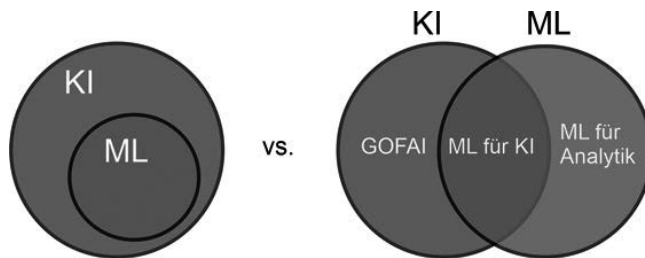


Bild 9.2 Die Beziehung zwischen KI und maschinellem Lernen (ML). Links: die Beziehung, wie sie in der Literatur häufig dargestellt wird. Rechts: eine realistischere Darstellung, die zeigt, dass maschinelles Lernen auch für andere Zwecke als KI verwendet werden kann, z. B. für die Analyse von Daten. GOF AI steht für „Good old-fashioned AI“ (gute altmodische KI), die kein maschinelles Lernen einsetzt.

■ 9.2 Eine kurze Geschichte der KI

Um zu verstehen, dass KI sich nicht ausschließlich auf maschinelles Lernen stützen muss, ist es am besten, einen Blick auf ihre Geschichte zu werfen. Die Geschichte der KI ist nämlich länger als die des maschinellen Lernens. Grob gesagt gab es bei der KI zwei große Phasen. In der ersten Phase lag der Schwerpunkt auf der Entwicklung von Algorithmen, die nichts mit maschinellem Lernen zu tun hatten. Stattdessen stützten sich diese frühen Algorithmen ausschließlich auf maschinelles Wissen, das manuell in die Maschine eingegeben wurde – von Menschen. In dieser ersten Phase galt zum Beispiel ein großer regelbasierter Entscheidungsbaum als ein hochmoderner Algorithmus für KI. Kennzeichnend für diese Form der KI ist, dass sie ihr Wissen nicht in Form von Zahlen speicherte und keine Schlussfolgerungen durch Anwendung von Gleichungen auf diese Zahlen zog. Der Grund dafür war

einfach: Es wäre für Menschen sehr schwierig gewesen, zahlenbasiertes Wissen, wie z. B. die Verbindungsgewichte von künstlichen neuronalen Netzen, in eine Maschine einzugeben. Stattdessen wurde das meiste Wissen in einer symbolischen Form gespeichert – einer Form, die für Menschen verständlich ist. Zum Beispiel konnte ein Wissensobjekt Symbole verwenden, um „wenn Fieber, dann Grippe“ darzustellen. Die Schlussfolgerungen wurden durch Anwendung logischer Regeln auf diese Symbole gezogen. Die Regeln waren wiederum für Menschen verständlich.

Wir bezeichnen diese Stufe der KI häufig als *symbolische* KI (siehe auch Kapitel 10). Ein anderer allgemein bekannter Begriff ist Good-old-fashioned-AI, abgekürzt GOFAI. Die Forschung im Bereich der symbolischen KI begann bereits in den 1950er-Jahren, wobei der offizielle Geburtsort die historische Dartmouth-Konferenz im Jahr 1956 ist. Der Organisator dieser Konferenz, John McCarthy – der den Begriff „Künstliche Intelligenz“ formulierte –, stellte auch die erste Programmiersprache vor, die Computern zu symbolischer Intelligenz verhelfen sollte: LISP. Die aus zwei Buchstaben bestehende Abkürzung „AI“ wurde erst nach Steven Spielbergs Film „A.I. Artificial Intelligence“ aus dem Jahr 2001 allgemein verwendet.

Das maschinelle Lernen trat erst später in den Bereich der KI ein und leitete die zweite – und immer noch –aktuelle – Phase ein. Tatsächlich hat sich das maschinelle Lernen erst in den 1970er-Jahren durchgesetzt, obwohl einige der Algorithmen schon viel früher existierten. Der Grund für diese Verzögerung war, dass man erst mit der Zeit erkannte, dass die symbolische KI ihre Grenzen hatte und ein anderer Ansatz erforderlich war. Ein Problem bestand darin, dass die symbolische KI nicht von sich aus lernen konnte; das Wissen musste von Menschen vermittelt werden. Dies stellte einen enormen Engpass dar, da die Menge an Wissen, die manuell aufgebaut werden musste, oft zu groß war. Daher wurde der GOFAI-Ansatz zur Steigerung der Intelligenz von Maschinen nicht mehr tragbar. Infolgedessen erreichten viele Projekte nicht die Nützlichkeitsebene und kamen über den anfänglichen Proof-of-Concept nicht hinaus; was in kleinem Maßstab gut funktionierte, konnte in größerem, nützlicherem Maßstab nicht umgesetzt werden.

Heute, in der zweiten Phase, verlassen wir uns überwiegend auf Algorithmen des maschinellen Lernens, um Wissen in Maschinen einzuspeisen und es aus großen Datensätzen in Matrizen von Modellparametern umzuwandeln. Diese Algorithmen stellen eine große Erleichterung für die manuelle Arbeit dar. Alles, was der Mensch tun muss, ist, Beispiele für intelligentes Verhalten zu liefern. Die Maschine ist dann in der Lage, die Regeln zu extrahieren, nach denen dieses Verhalten zustande kommt.

Offensichtlich haben wir auf diese Weise einen großen Fortschritt in unserer Fähigkeit erzielt, die Intelligenz von Maschinen zu steigern. Es ist jedoch falsch anzunehmen, dass sich die Welt von der symbolischen KI entfernt hat und dass GOFAI-Algorithmen der Vergangenheit angehören. Ganz und gar nicht. Der symbolische Ansatz ist immer noch lebendig und gut. Jede komplexe KI-Lösung, die heute entwickelt wird, ist ein Mischmasch aus maschinellem Lernen und GOFAI-Komponenten. Symbolische KI ist ein nicht minder wichtiger Bestandteil. Es ist nur so, dass GOFAI-Komponenten nicht beworben werden, was mehr mit dem aktuellen Hype und den Marketingstrategien zu tun hat als mit den Fakten, wie die Maschinen unter der Haube arbeiten. Symbolische KI ist weit verbreitet. Oft entscheidet GOFAI, welcher Deep-Learning-Algorithmus als Nächstes ausgeführt wird. In anderen Fällen erhält GOFAI die Ergebnisse der maschinellen Lernmodelle, um die nächste Entscheidung

dung zu treffen. In anderen Fällen unterstützt das maschinelle Lernen GOF AI bei der Suche nach einer optimalen Lösung. Oft sind die beiden Komponenten ineinander verschachtelt: Ein symbolischer Algorithmus ruft ein maschinelles Lernmodell auf, das wiederum eine andere GOF AI-Komponente um Hilfe bittet, die wiederum zum maschinellen Lernen zurückkehrt und so weiter. Die Möglichkeiten sind grenzenlos. Watson könnte ohne GOF AI-Komponenten kein Jeopardy-Spiel gewinnen. Ohne die Verwendung eines GOF AI hätte alphaGo das Go-Spiel gegen den Weltmeister Lee Sedol nicht gewinnen können (das Ergebnis war vier zu eins für die Maschine). Ein autonomes Fahrzeug beispielsweise kann nicht ohne altmodische KI-Komponenten fahren. Alexa, Siri und Co können sich ohne symbolische Teile ihrer allgemeinen Intelligenzarchitekturen nicht mit Ihnen unterhalten.

Was bedeutet das alles für einen Data Scientist, der heute mit der Entwicklung eines KI-Produkts betraut ist? Sehr wahrscheinlich wird Ihre Lösung viel mehr beinhalten müssen als nur ein Modell für maschinelles Lernen. Es wird eine Menge Technik außerhalb des maschinellen Lernens benötigt. Es wird schwierig sein, symbolische Komponenten zu vermeiden. Das bedeutet, dass Sie kluge architektonische Entscheidungen über die gesamte Lösung treffen müssen, und diese Entscheidungen werden viel mehr als nur maschinelles Lernen umfassen. Um ein effektives Produkt zu schaffen, benötigen Sie möglicherweise sogar Komponenten, die außerhalb der Technik liegen. Ein gutes Design der Schnittstelle für Ihre KI kann für den Erfolg ebenso entscheidend sein wie die Leistung des zugrunde liegenden Modells. Ähnlich wie man einen ergonomischen Griff an einer Klinge anbringen muss, um ein gutes Messer herzustellen, oder bequeme Sitze braucht, um ein gutes Auto zu bauen, muss sich Ihre KI in vielen verschiedenen Dimensionen entwickeln, um ein gutes Produkt zu präsentieren. Die Modelle des maschinellen Lernens sind nur ein Teil des Gesamtergebnisses und somit auch nur ein Teil des gesamten Kundenerlebnisses.

■ 9.3 Fünf Empfehlungen für die Entwicklung einer KI-Lösung

Auf dem Weg zur Entwicklung einer KI-Lösung muss ein Data Scientist eine Reihe von Entscheidungen treffen. Sie als Data Scientist müssen zwangsläufig eine Architektur erstellen, die Komponenten verschiedener Typen kombiniert, die interagieren und gemeinsam die Intelligenz Ihrer Maschine hervorbringen. Vielleicht werden Sie diese Architektur mit mehreren Kästchen und Pfeilen zeichnen, wie die Zeichnung der Watson-Architektur in Bild 9.1. Die Frage ist dann: Welche Strategien können Sie anwenden, und worauf sollten Sie achten, um bestimmte häufige Fehler zu vermeiden?

9.3.1 Empfehlung Nr. 1: Seien Sie pragmatisch

In den vorangegangenen Kapiteln dieses Buches haben Sie verschiedene Rezepte zur Lösung von Data-Science-Problemen kennengelernt. All dies wird Ihnen als Einzelteile präsentiert, zum Beispiel als einzelne Algorithmen für maschinelles Lernen. Außerdem wer-

den die Teile in einer idealisierten Welt gezeigt, unabhängig vom wirklichen Leben. Wenn Sie eine echte künstliche Intelligenz – ein komplettes Produkt – entwerfen, müssen Sie darüber nachdenken, wie Sie Algorithmen für eine unvollkommene Welt auswählen. Sie müssen darüber nachdenken, wie Sie sie kombinieren können. Außerdem werden Sie Algorithmen finden und verwenden müssen, die nicht in diesem Buch beschrieben sind. Es ist wichtig, dass Sie nicht bei einem Satz von Algorithmen bleiben, nur weil sie in der Vergangenheit für Sie funktioniert haben oder weil Sie sie kennen. Erweitern Sie Ihr Wissen, wenn Sie es brauchen. Wählen Sie die Algorithmen nach ihrer Eignung für ein bestimmtes Problem aus, nicht nach ihrer Bequemlichkeit. Denken Sie daran, dass Ihr neues Problem immer etwas anders sein wird als alles, was Sie bisher gesehen haben. Seien Sie eklektisch bei der Auswahl des Werkzeugs zur Lösung der Aufgaben. Wählen Sie aus einer möglichst großen Auswahl. Schränken Sie sich nicht ein.

Bleiben Sie außerdem pragmatisch. Ihre erste Sorge sollte das Erreichen des Ziels sein. Sie müssen nicht immer die neuesten Algorithmen, das heißeste und am meisten gehypte Werkzeug verwenden. Nehmen Sie stattdessen das, was für das jeweilige Problem am besten geeignet ist. Ich habe schon erlebt, dass sich Data Scientists in bestimmte Modelle „verliebt“ haben und dann Favoriten spielen. Aber Erfolg in Data Science stellt sich nicht ein, wenn man Favoriten spielt. Ich habe Leute erlebt, die versuchen, jedes Problem mit demselben Ansatz zu lösen. Es gibt Leute, die erwarten, dass alles mit Deep Learning gelöst werden muss. Ich habe auch schon eingefleischte Fans von bayesschen Ansätzen gesehen. Sicher, sowohl die bayesschen als auch die Deep-Learning-Methoden sind charmant und haben einige attraktive Eigenschaften, die ihnen einzigartige „Superkräfte“ verleihen. Aber beide haben auch Nachteile. Tatsächlich hat jeder Ansatz, für den Sie sich entscheiden, einige Vorteile gegenüber anderen und zwangsläufig auch einige Nachteile. Ihre Aufgabe ist es, beide Seiten zu berücksichtigen und die Vor- und Nachteile abzuwägen, um eine gute Wahl zu treffen.

Es ist von größter Wichtigkeit, sich sowohl der Vorteile als auch der Nachteile einer bestimmten Methode oder eines Algorithmus bewusst zu sein. Die Nachteile sind vielleicht schwieriger zu erkennen, weil die Autoren, die über ihre neuen Methoden schreiben, dazu neigen, sich auf die positiven Aspekte zu konzentrieren. Die rosigen Bilder sind es, die sie dazu motivieren, überhaupt zu forschen und Artikel zu schreiben. Wir sollten also ein gewisses Verständnis aufbringen. Dennoch muss man sich die Fähigkeit aneignen, „zwischen den Zeilen zu lesen“ und mögliche Einschränkungen und Fallstricke zu erkennen. Ein erfahrener Data Scientist wird in der Lage sein, mögliche Nachteile einer neuen Methode zu erkennen, auch wenn sie nicht so klar formuliert sind wie die Vorteile. Entwickeln Sie diese Fähigkeit, denn sie wird Ihnen viel Kraft geben, um gute Designentscheidungen für Ihre KI-Architekturen zu treffen. Ziel ist es, sich Wissen über eine Vielzahl von Algorithmen, Modellen und Optimierungstechniken anzueignen.

Das Sortiment an Tools, aus dem man wählen kann, ist riesig. Eine einzelne Person kann wahrscheinlich nie einen vollständigen Überblick über den Bereich der Data Science haben. Um umfassende Kenntnisse über Methoden des maschinellen Lernens und KI-Algorithmen zu erlangen, ist lebenslanges Lernen erforderlich. Und man ist nie fertig. Außerdem nimmt das Tempo, mit dem neue Algorithmen vorgeschlagen werden, rapide zu, da immer mehr Menschen an diesem Thema arbeiten, Universitäten neue Abteilungen für KI und Data Science eröffnen und Regierungen mehr Geld für die KI-Forschung bereitstellen. Mit all

diesen Entwicklungen Schritt zu halten ist eine Herausforderung. Man sollte nie aufhören zu lernen, aber auch nicht erwarten, alles zu wissen.

Um sich in diesem ständig wachsenden Wald neuer Werke zurechtzufinden, ist ein gründliches Verständnis von Algorithmen Voraussetzung. Sie werden einen neuen Algorithmus besser verstehen, wenn Sie bereits ein tiefes Verständnis für einen verwandten, bereits existierenden Algorithmus haben. Ein oberflächliches Verständnis von Methoden ist nicht annähernd so leistungsfähig. Das richtige Verständnis mehrerer verschiedener Algorithmen, die jeweils zu einer anderen Kategorie gehören, ist wahrscheinlich die beste Strategie, die man verfolgen kann, um das Gebiet der Data Science zu beherrschen. Neue Algorithmen sind oft mit den bestehenden verwandt. Es kommt selten vor, dass Forscher einen völlig neuen Ansatz zur Lösung eines Problems des maschinellen Lernens entwickeln (obwohl sie gelegentlich genau das tun). Wenn Sie einen Algorithmus gut verstehen, fällt es Ihnen leicht, das Wesen seiner Vettern zu erfassen – sie werden zu einer Variation des Themas. Wenn Sie dagegen einen Algorithmus nur oberflächlich verstehen, kann eine Variation dieses Algorithmus für Sie ein Rätsel sein, und Sie haben möglicherweise Schwierigkeiten zu entscheiden, ob diese neue Variation für Ihr neues Problem hilfreich ist oder nicht.

Man kann den Algorithmus immer an den Daten ausprobieren und sehen, was passiert. Es gibt auch Tools, mit denen man automatisch mehrere Algorithmen ausprobieren und den besten auswählen kann (als autoML bezeichnet). Aber damit kommt man nicht weit. Man kann ein autonomes Fahrzeug nicht entwickeln, indem man wahllos verschiedene Architekturen ausprobiert. Wenn man KI entwickelt, muss man das gute alte menschliche Denken einsetzen – und zwar eine ganze Menge. In diesem Fall wollen Sie die Entscheidungsfindung minimieren, indem Sie die Algorithmen an Ihren Daten ausprobieren. Natürlich werden Sie das irgendwann tun müssen, daran besteht kein Zweifel. Der Unterschied zwischen einem erfahrenen KI-Entwickler und einem unerfahrenen Entwickler besteht jedoch darin, dass ersterer die Aufgabe mit mehr Nachdenken und weniger Ausprobieren bewältigen kann. Erfahrene Menschen können die Möglichkeiten im Kopf durchgehen, ohne den Algorithmus auf den Daten trainieren zu müssen. Dank des erweiterten Wissens können sie erkennen, dass etwas nicht gut funktionieren wird, noch bevor sie es ausprobieren. Das spart eine Menge Zeit.

Was kann Ihnen noch helfen, gute Entscheidungen zu treffen? Eine gute Idee ist es, Ihre zukünftige Architektur zu zeichnen, bevor Sie mit der Programmierung beginnen. Legen Sie die Details fest und versuchen Sie, den Datenfluss im System mental zu simulieren. Stellen Sie sich bei jedem Schritt eine Frage: Sehe ich einen Grund, warum dieser Schritt scheitern oder Schwierigkeiten bereiten könnte? Wenn Sie mögliche Probleme sehen, gehen Sie diese Probleme sofort an. Pragmatisch ist es, die schwächsten Punkte zuerst anzugehen. Hoffen Sie nicht, dass ein Wunder geschieht, nachdem Sie sich mit dem einfachen Teil beschäftigt haben.

Es ist eine weit verbreitete Ansicht, dass mit genügend Rechenleistung und einer ausreichenden Menge an Daten alles möglich ist: dass alles von einer Maschine gelernt werden kann. Obwohl an dieser Aussage etwas Wahres dran ist, gibt es auch eine Menge Unwahrheiten. Auf einige dieser Punkte werde ich später in diesem Kapitel eingehen. Die Quintessenz ist, dass die blinde Verfolgung der Strategie „mehr Daten mit mehr Rechenleistung“ fast garantiert zu Problemen führen wird. Es ist viel besser, Ihre Algorithmen gründlich zu bereinigen, indem Sie Ihr Verständnis von Statistik, maschinellem Lernen und KI im Allgemeinen nutzen. Vertrauen Sie auf Big Data und Rechenleistung als Ihre letzte Ressource.

Sicherlich müssen Sie verschiedene Konzepte ausprobieren. Und Sie werden die Ergebnisse dieser Versuche als Feedback nutzen müssen. Sie werden Ihnen zeigen, wie Sie sich verbessern können. Es ist wichtig zu erkennen, dass Ihre Iterationen viel schneller und effektiver sein werden, wenn Sie besser verstehen, was Sie tun.

Das Denken ist vergleichsweise schwer. Das Codieren und Ausführen von Modellen ist vergleichsweise einfach. Wenn Sie sich jedoch nicht davor scheuen, den schwierigen Teil zu tun, werden Sie wahrscheinlich den Wettbewerbsvorteil erlangen, den Sie brauchen, um ein Produkt zu schaffen, das der Markt braucht und das ihm gefällt.

Vergessen Sie nicht, dass eine Person nicht alles wissen kann. Stellen Sie ein Team aus Personen mit unterschiedlichen Fachkenntnissen zusammen. Lassen Sie alle mitarbeiten; jeder sollte ein Mitspracherecht haben. Stellen Sie sicher, dass Sie das Talent jedes Einzelnen für Ihr Endprodukt nutzen können.

9.3.2 Empfehlung Nr. 2: Erleichtern Sie Maschinen das Lernen – schaffen Sie induktive Verzerrungen

Es gibt eine einfache Wahrheit über Algorithmen für maschinelles Lernen: Einige lernen schneller und besser als andere. In manchen Fällen genügen schon wenige Beispiele, um eine hohe Leistung zu erreichen. In anderen Fällen sind Millionen von Beispielen erforderlich. Es gibt zwar viele Gründe für diese Unterschiede, aber es gibt einen Grund, den Sie selbst beeinflussen können: Ein Faktor, der die Lerneffizienz eines Algorithmus bestimmt, sind seine induktiven Verzerrungen. Induktive Verzerrungen sind wie ein Stück Wissen, das einem Algorithmus hinzugefügt wird und es ihm ermöglicht, einige Lernschritte zu überspringen und schneller und sicherer zum Ziel zu kommen. Im wahrsten Sinne des Wortes ermöglichen induktive Verzerrungen Algorithmen, voreilige Schlüsse zu ziehen. Und wenn Sie die richtigen induktiven Verzerrungen eingefügt haben, wird Ihr Algorithmus auch die richtigen Schlussfolgerungen ziehen.

Was ist also eine induktive Verzerrung? Es handelt sich um eine Veranlagung, eine bestimmte Beziehung in den Daten zu finden (d. h. zu folgern, zu induzieren). Induktive Verzerrungen helfen dem Algorithmus, eine bestimmte Beziehung zu finden, selbst wenn die Beweise sehr schwach sind und andernfalls Millionen von Datenpunkten durchlaufen werden müssten. Induktive Verzerrungen sind eine Art Vorurteil, um eine bestimmte Art von Muster in den Daten zu erkennen.¹ Wenn Ihr mathematisches Modell beispielsweise aus Sinus- und Kosinusfunktionen besteht und Sie hauptsächlich die Parameter solcher Funktionen (z. B. Amplitude und Phase eines Sinus) anpassen, dann wird Ihr Modell wahrscheinlich in der Lage sein, solche Funktionen in den Daten zu finden, selbst bei kleinen Datenmengen. Mit anderen Worten: Das Modell neigt dazu, eine Sinuswelle zu finden.

Was die Leute dazu verleitet, die Bedeutung der induktiven Verzerrungen zu ignorieren, ist die Tatsache, dass man theoretisch dieselbe Art von Sinusmodell verwenden kann, um andere Funktionen als Sinuswellen anzunähern. Man könnte Millionen von Sinuswellen kombinieren, um eine Potenzgesetzfunktion genau zu approximieren. Aber das ist viel schwieriger. Man benötigt ein größeres Modell – d. h. eines mit einer größeren Anzahl elementarer

¹ Induktive Verzerrungen haben nichts mit Verzerrungen in den Daten zu tun, das ist ein ganz anderes Problem.

Sinuswellen und damit einer größeren Anzahl von Parametern –, und man benötigt mehr Daten zum Trainieren.² Diese Beziehung gilt für jedes Modell und für alle Daten. Mit ausreichend großen Deep-Learning-Algorithmen kann man fast alles approximieren. Ähnliches gilt für Entscheidungsbäume, die groß genug sind (siehe Abschnitt 6.2.3 über Entscheidungsbäume). Es gibt sogar ein mathematisches Theorem, das universelle Approximationstheorem³, das beweist, dass ein künstliches neuronales Netz mit nur einer verborgenen Schicht jede beliebige mathematische Funktion approximieren kann, sofern genügend Neuronen in der verborgenen Schicht vorhanden sind [4]. Wo liegt also das Problem, wenn wir alles approximieren können? Warum sollten wir uns Gedanken über die Hinzufügung induktiver Verzerrungen machen, wenn Modelle jede Funktion auch ohne sie approximieren können? Das offensichtlichste Problem habe ich bereits angedeutet: Wenn die induktiven Verzerrungen des Modells nicht gut mit den Daten übereinstimmen, braucht man viele Daten, ein großes Modell und viele Berechnungen. Das bedeutet auch, dass während des Trainings und der Erstellung des Modells mehr CO₂ in die Atmosphäre abgegeben wird. Das alles sind keine guten Nachrichten.

Andererseits können Sie die Modellgröße reduzieren, wenn Sie die richtigen induktiven Verzerrungen hinzufügen. Sie können es dann mit weniger Datenpunkten trainieren, da dieses schlankere Modell nicht so leicht in die lokalen Minima der Überanpassung fällt.⁴ Die Vorteile der induktiven Verzerrungen sind der Grund dafür, dass wir so viele verschiedene Modelle haben. Jedes Problem unterscheidet sich ein wenig von jedem anderen Problem und kann daher mit einem spezielleren Satz von Gleichungen optimal angegangen werden. Für jedes Problem gibt es theoretisch ein optimales Modell, das genau auf dieses Problem spezialisiert ist. Daher wird uns der Platz für die Erfindung neuer Modelle nie ausgehen. Die Liste aller möglichen Modelle ist unendlich; wir werden nie das Ende dieser Liste erreichen.

Ich lernte die Macht induktiver Verzerrungen in der Praxis bei einer Gelegenheit kennen, bei der mein Team und ich eine Überanpassung in neuronalen Netzen für tiefes Lernen herbeiführen wollten. Unser Ziel war es, einen Algorithmus zu testen, der die Überanpassung in einer Situation des One-shot Learnings reduziert, und unser Ansatz war folgender: Eine unbegrenzte Menge an Daten für das Training des One-shot-Learning-Algorithmus (siehe Kapitel 10)⁵ zu erzeugen, eine Überanpassung auf diesem Datensatz zu induzieren und dann das Netz mit unserem neuen Algorithmus vor der Überanpassung zu „retten“. Meine Idee war es, unsere „unbegrenzten Daten“ mit einem Deep-Learning-Netz mit einem zufälligen Satz von Gewichten zu erstellen und dann ein anderes naives Deep-Learning-Netz zu trainieren, um dieselben zufälligen Mappings zu lernen. Wir waren zuversichtlich, dass wir auf diese Weise eine Überanpassung erzeugen könnten, aber wir wurden eines Besseren belehrt: Wir reduzierten die Größe des Trainingsdatensatzes immer weiter, aber das neue Netz wollte nicht überanpassen. Die Leistung bei den Testdaten blieb gut, manchmal sogar mit nur zehn oder 20 Datenpunkten. Zunächst waren meine Kollegen und ich

² Die Fourier-Transformation ist ein Instrument, mit dem sich beurteilen lässt, wie komplex ein auf Sinuswellen basierendes Modell für eine Zeitreihe sein muss. Zeitreihen, die periodisch sind und den Formen von Sinuswellen ähneln, können durch einfache Modelle angenähert werden. Andere benötigen komplexe Modelle und viele Parameter.

³ https://en.wikipedia.org/wiki/Universal_approximation_theorem

⁴ <https://en.wikipedia.org/wiki/Overfitting>

⁵ Hier kann man etwas über One-shot Learning erfahren: https://en.wikipedia.org/wiki/One-shot_learning

verwirrt. Wie war das möglich? Es sollte sich um sehr schwer zu erlernende Daten handeln, mit komplexen zufälligen Beziehungen in einem mehrdimensionalen Raum. Wie konnte das Netz diese Beziehungen mit nur einer kleinen Anzahl von Beispielen lernen? Dieses Lernen war auch dann noch effizient, wenn wir die Architektur des Netzes, die Anzahl der Schichten und die Größe der einzelnen Schichten änderten. Die Fähigkeit zum effizienten Lernen der Daten war robust.

Es dauerte ein paar Tage, bis wir erkannten, dass das Modell, von dem wir hofften, es würde sich übermäßig anpassen, dazu „verdammt“ war, da es perfekte induktive Verzerrungen für die Daten aufwies. Wir verwendeten die gleichen ReLu- und sigmoide Transferfunktionen für die Generierung der Daten und für das Modell, das die Daten lernte, was die Arbeit des Lernmodells im Grunde sehr einfach machte. Dies veranschaulichte mir, wie mächtig induktive Verzerrungen sein können: Dasselbe Netz kann eine Million Beispiele benötigen, um etwas zu lernen, das für seine induktiven Verzerrungen kontraintuitiv ist, wie z. B. das Erkennen einer Blume auf einem Foto, und nur zehn Beispiele, um etwas zu lernen, das für jedes andere Modell hochkomplex ist, aber für dieses spezielle Netz vollkommen intuitiv ist. Das liegt daran, dass das Netz genau die richtigen induktiven Verzerrungen hat.⁶

Induktive Verzerrungen geben uns eine Menge Möglichkeiten, mit denen wir bei der Entwicklung von Modellen spielen können. Das Spiel ist zweidimensional. Die eine Dimension bezieht sich auf die Art der induktiven Verzerrungen: Sollen wir ReLu- oder Sigmoid-Übertragungsfunktionen verwenden, oder sollen wir Tangens oder sogar Sinuswellen verwenden? Auf diese Weise ändern wir die Annahmen, die das Modell über die Welt macht. Wir können eine Annahme durch eine andere ersetzen und dadurch die induktiven Verzerrungen ändern. Ein lineares Modell geht von einer bestimmten linearen Beziehung zwischen Daten aus. Ein Entscheidungsbaum geht von einer anderen Annahme aus. Und so weiter.

Die andere Dimension, in der wir mit induktiven Verzerrungen spielen können, ist die Frage, wie eng die Annahmen sind, die wir treffen wollen. Wir können die Annahmen lockerer fassen, was im Grunde bedeutet, dass wir ein Modell mit mehr Parametern haben. Wir können auch ein strengeres Modell mit weniger Parametern erstellen. Indem wir einem neuronalen Netz mehr Einheiten (Neuronen) hinzufügen, lockern wir seine Annahmen. Modelle, die für ein bestimmtes Problem gut geeignet sind, d. h. genau die richtige Menge an induktiven Verzerrungen haben, können oft mit nur einer Handvoll Parameter großartige Arbeit leisten. Die größten Modelle haben heute Milliarden von Parametern. Diese Modelle sind ziemlich entspannt: Es gibt eine ganze Menge verschiedener Dinge, die sie möglicherweise lernen können.

Wie wir bereits erwähnt haben, hat dies direkte Auswirkungen auf die Datenmenge, die zum Lernen benötigt wird. Ein strenges Modell ist natürlich in der Lage, aus nur wenigen Datenpunkten zu lernen, vorausgesetzt, die induktiven Verzerrungen sind korrekt. Wenn die induktiven Verzerrungen nicht korrekt sind, wird ein kleines Modell niemals gut passen, egal wie viele Datenpunkte Sie ihm zum Training geben. Die einzigen beiden Möglichkeiten zur Verbesserung sind entweder die Vergrößerung des Modells (mit einer entsprechenden Vergrößerung des Datensatzes) oder die Korrektur der induktiven Verzerrungen. Daher können Sie auch mit schlechten induktiven Verzerrungen Daten gut anpassen; alles,

⁶ Später erfuhr ich, dass jemand denselben Fehler wie wir gemacht und eine ganze Abhandlung veröffentlicht hatte, ohne sich des von uns entdeckten Problems der induktiven Verzerrung bewusst zu sein, was zu der falschen Schlussfolgerung führte, dass neuronale Netze nicht anfällig für Überanpassung sind [5].

was Sie brauchen, sind genügend Parameter und genügend Daten. Deep Learning fällt in die letztgenannte Klasse von Modellen, die nicht spezialisiert sind, lockere Annahmen haben und eine große Datenmenge erfordern. In Bild 9.3 finden Sie die Beziehung zwischen der Anzahl der erforderlichen Daten (ausgedrückt als „Trainingsaufwand“) und der Strenge des Modells (ausgedrückt als „Spezialisierung“) für verschiedene Arten von Modellen. Die strengsten Modelle sind die Gesetze der Physik. Für $E = mc^2$ gibt es z. B. nur einen Parameter, der angepasst werden muss, nämlich c . Dann kann man das „Modell“ verwenden, um E aus m vorherzusagen.

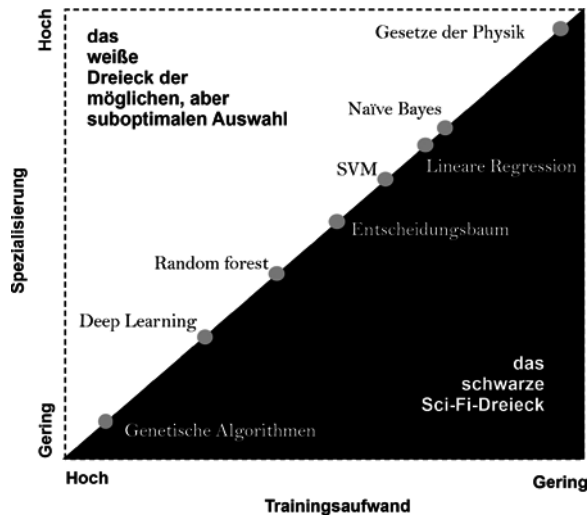


Bild 9.3 Verschiedene Modelle haben unterschiedliche Fähigkeiten zu lernen. Einige benötigen eine große Datenmenge und einen hohen Trainingsaufwand. Andere können mit nur wenigen Beispielen schnell lernen. Ein Modell ist optimal, wenn es irgendwo auf der Diagonalen liegt: In diesem Fall wurde das richtige Modell für die Aufgabe gewählt. Wenn die benötigte Datenmenge und der Trainingsaufwand für den gegebenen Spezialisierungsgrad zu groß sind, dann machen Sie etwas falsch, auch wenn Ihr Modell gut funktioniert (das weiße Dreieck). Es ist unmöglich, ein gut funktionierendes Modell zu haben, das gleichzeitig generisch ist und eine geringe Datenmenge zum Lernen benötigt. Das kann nur in der Fantasie passieren, und manchmal hoffen Data Scientists ganz naiv, ein solches Modell zu finden.

Wie können Sie sich also dieses Wissen über induktive Verzerrungen zunutze machen? Sie können solche Verzerrungen in Ihre Modelle einbauen, damit diese besser und schneller lernen. Auf diese Weise können Sie Modelle kleiner, schneller und zuverlässiger machen. Sie müssen nur die richtigen induktiven Verzerrungen finden. Manchmal müssen Sie auch das Gegenteil tun, nämlich das Modell vergrößern und damit seine Annahmen lockern. Sie müssen herausfinden, was der richtige Ansatz für Ihr Problem ist. Wenn Sie schon einmal eine Hyperparameter-Abstimmung durchgeführt haben⁷, dann haben Sie bereits erste Erfahrungen mit der Anpassung der induktiven Verzerrungen von Modellen gemacht. Wenn Sie über gut strukturierte Validierungs- und Trainingsdatensätze verfügen, haben Sie die

⁷ https://en.wikipedia.org/wiki/Hyperparameter_optimization

Rania Wazir

“All algorithms should be seen as untrustworthy until proven otherwise.”

Cathy O'Neil



Fragen, die in diesem Kapitel beantwortet werden:

- Wie sieht der derzeitige Rechtsrahmen für vertrauenswürdige KI aus, insbesondere in der EU?
- Wer sind die möglichen KI-Stakeholder?
- Was ist Fairness in der KI, und wie wird Bias definiert?
- Was sind die verschiedenen Metriken zur Messung der Auswirkungen von Algorithmen auf die Fairness?
- Was sind mögliche Techniken, um unerwünschte Verzerrungen abzuschwächen?
- Wie können Daten und Modelle dokumentiert werden, um ihre Transparenz, Nutzbarkeit und Sicherheit zu verbessern?
- Welche Methoden gibt es, um Modellentscheidungen zu erklären?

Die breite Klasse an Technologien, die unter den Begriff KI fallen – von Expertensystemen bis hin zu datenwissenschaftlichen Anwendungen und Lösungen, die auf maschinellem Lernen basieren –, revolutionieren die Industrie, durchdringen die meisten Wirtschaftssektoren und haben das Potenzial, der Wirtschaft, der Gesellschaft und der Umwelt zu nutzen. Wie sich in den letzten Jahren gezeigt hat, sind diese Technologien jedoch auch mit Risiken verbunden^{1 2 3}. Mit der Aufdeckung von Beispielen für Stereotypisierung und Diskriminierung, Bedenken hinsichtlich Arbeitnehmerrechte und nachteilige Auswirkungen auf demokratische Grundsätze und die Umwelt hat die Skepsis in der Öffentlichkeit zugenommen. Damit die KI-Technologien weiterhin eine rasch wachsende Akzeptanz finden und ihr nützliches Potenzial entfalten können, wird die Nachfrage nach KI-basierten Systemen, denen man vertrauen kann, steigen. Für die Anbieter von KI-Systemen führt dieses Vertrauen zu einer erhöhten Akzeptanz von Produkten, bei denen es vorhanden ist, und zu rechtlichen

¹ O'Neil, C., Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Broadway Books, 2017.

² Agentur der EU für Grundrechte (FRA), Getting the Future Right

³ Kate Crawford, AI Now Report 2019

und rufschädigenden Konsequenzen, wenn dieses Vertrauen verletzt wird. Im folgenden Kapitel werden wir in der Praxis untersuchen, was Vertrauen in KI-Systeme bedeutet, insbesondere im Zusammenhang mit maschinellem Lernen und datenwissenschaftlichen Lösungen, sowie wer die Interessengruppen sind, die berücksichtigt werden müssen. Hinzu kommen einige praktische Implementierungsschritte, die den Entwicklungsprozess begleiten können.

Unsere Aufgabe wird es sein, die vielen unterschiedlichen Anforderungen miteinander zu verweben, um ein kohärentes Bild zu schaffen, das den Entwicklungsprozess von KI-Systemen von Anfang bis Ende begleiten kann. Wir beginnen mit dem rechtlichen und Soft-Law-Rahmen, indem wir uns prominente Ethikrichtlinien sowie bestehende und künftige Vorschriften und Standards ansehen. Vertrauen hat für verschiedene KI-Stakeholder unterschiedliche Bedeutungen – daher machen wir einen kurzen Abstecher zur Identifizierung von KI-Stakeholdern, bevor wir uns auf die Fragen der Fairness in der KI und der Erklärbarkeit konzentrieren. Dieses Kapitel erhebt keinen Anspruch auf Vollständigkeit, sondern soll vielmehr Anbietern und/oder Nutzern von KI-Systemen, die vertrauenswürdige Produkte entwickeln bzw. einsetzen wollen, eine Orientierungshilfe bieten.

■ 18.1 Rechtlicher und Soft-Law-Rahmen

Seit 2016 gibt es eine explosionsartige Zunahme von sogenannten „Ethikrichtlinien“ für KI. Tatsächlich gab es 2019 bereits über 80 veröffentlichte Richtlinien.⁴ Von akademischen Forschungsinstituten bis hin zu großen Technologieunternehmen, von internationalen Nichtregierungsorganisationen bis hin zu staatlichen Regierungen – alle haben ihren Beitrag dazu geleistet, was „ethische“ KI ausmacht. Leider sind die meisten Richtlinien sehr allgemein gehalten und gehen bei den Grundsätzen, die sie für eine „ethische“ KI für notwendig erachten, weit auseinander. Der Untersuchung von Jobin et al.⁵ zufolge gibt es fünf allgemeine Grundsätze, auf die sich mindestens die Hälfte der Leitlinien bezieht: Transparenz, Gerechtigkeit und Fairness, Nicht-Malefizierung, Verantwortung und Schutz der Privatsphäre; ihre genaue Bedeutung und die entsprechenden Umsetzungsstrategien sind jedoch wiederum unterschiedlich.

Einige der wichtigsten internationalen Ethik-Leitlinien zur KI sind:

- OECD-Grundsätze zur künstlichen Intelligenz⁶
- UNESCO-Empfehlung zur Ethik der KI⁷
- UNICEF-Leitlinien zu künstlicher Intelligenz für Kinder⁸

⁴ Jobin, Anna, Marcello lenca und Effy Vayena. „The global landscape of AI ethics guidelines“. *Nature Machine Intelligence* 1.9 (2019): 389 – 399.

⁵ Jobin, Anna, Marcello lenca und Effy Vayena. „The global landscape of AI ethics guidelines“. *Nature Machine Intelligence* 1.9 (2019): 389 – 399.

⁶ <https://www.oecd.ai/ai-principles>

⁷ <https://unesdoc.unesco.org/ark:/48223/pf0000373434>

⁸ <https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>

- EU-HLEG-Leitlinien für vertrauenswürdige KI⁹
- EU-Whitepaper zu KI¹⁰

Eine vertrauenswürdige KI geht jedoch über ethische Aspekte hinaus. Eine offensichtliche zusätzliche Anforderung ist ein Qualitätsgebot: Das System sollte robust, zuverlässig und sicher sein. Die OECD-Prinzipien beispielsweise richten sich an Regierungen und andere staatliche Akteure und sollen als Leitfaden für die Förderung der Entwicklung vertrauenswürdiger KI dienen. Sie schlagen die folgenden fünf Hauptprinzipien vor:¹¹

1. **Integratives Wachstum, nachhaltige Entwicklung und Wohlstand.** Eine allgemeine Voraussetzung für den Nutzen vertrauenswürdiger KI ist die Verbesserung der menschlichen Fähigkeiten, der Abbau von Ungleichheiten und der Schutz der Umwelt.
2. **Auf den Menschen ausgerichtete Werte und Fairness.** Eine vertrauenswürdige KI muss die Rechtsstaatlichkeit und die Menschenrechte achten, einschließlich des Rechts auf Freiheit, des Rechts auf Würde und Autonomie, des Rechts auf Privatsphäre und Datenschutz sowie des Rechts auf Nichtdiskriminierung.
3. **Transparenz und Erklärbarkeit.** Verlangt eine verantwortungsvolle Offenlegung von Informationen über das KI-System, um das allgemeine Verständnis für solche Systeme zu fördern, die Betroffenen auf ihre Interaktionen mit einem KI-System aufmerksam zu machen und es den von einem KI-System Betroffenen zu ermöglichen, dessen Ergebnisse zu verstehen und anzufechten.
4. **Robustheit, Sicherheit und Schutz.** Rückverfolgbarkeit von Datensätzen, Prozessen und Entscheidungen sowie geeignete Maßnahmen zum Risikomanagement, um Risiken wie Sicherheit, IT-Sicherheit, Datenschutz und Verzerrungen zu vermeiden in jeder Phase des Lebenszyklus eines KI-Systems.
5. **Rechenschaftspflicht.** Alle Akteure, die an der Entwicklung, dem Einsatz oder dem Betrieb von KI-Systemen beteiligt sind, sollten entsprechend ihrer Rolle für das ordnungsgemäße Funktionieren der KI-Systeme verantwortlich gemacht werden, einschließlich der Gewährleistung, dass die oben genannten Anforderungen erfüllt werden.

Die hochrangige EU-Sachverständigengruppe für KI hat eine noch umfangreichere Liste von Anforderungen an eine vertrauenswürdige KI aufgestellt, die sich an Entwickler, Anbieter und Nutzer von KI-Systemen richtet. Eine vertrauenswürdige KI muss legal, ethisch und robust sein und sollte die folgenden Anforderungen erfüllen:¹²

1. **Menschliches Handeln und Aufsicht.** Einschließlich Grundrechte, menschliches Handeln und menschliche Aufsicht
2. **Technische Robustheit und Sicherheit.** Einschließlich Widerstandsfähigkeit gegen Angriffe, Sicherheit, Ausweichplan, Genauigkeit, Zuverlässigkeit und Reproduzierbarkeit

⁹ <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

¹⁰ https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

¹¹ <https://www.oecd.ai/ai-principles>

¹² High Level Expert Group on Artificial Intelligence set up by the European Commission, „Ethics Guidelines for Trustworthy AI“, April 2019, S. 14. Abgerufen von <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

3. **Datenschutz und Data Governance.** Einschließlich Achtung der Privatsphäre, der Qualität und Integrität der Daten und des Datenzugangs
4. **Transparenz.** Einschließlich Rückverfolgbarkeit, Erklärbarkeit und Kommunikation
5. **Vielfalt, Nichtdiskriminierung und Fairness.** Einschließlich der Vermeidung unfairer Voreingenommenheit (Bias), Zugänglichkeit und universellem Design sowie der Beteiligung von Interessengruppen
6. **Gesellschaftliches und ökologisches Wohlergehen.** Einschließlich Nachhaltigkeit und Umweltfreundlichkeit, sozialer Auswirkungen, Gesellschaft und Demokratie
7. **Rechenschaftspflicht.** Dazu gehören Überprüfbarkeit, Minimierung und Meldung negativer Auswirkungen, Abgleichungen und eventuelle Rechtsmittel.

Die HLEG-Leitlinien sind vielleicht eine der praktischsten Leitlinien, die es bisher gibt. Sie vermitteln ein klares Verständnis der den Anforderungen zugrundeliegenden Überlegungen und geben Informationen darüber, wie sie in der Praxis umgesetzt werden können. Auf der Grundlage des Leitfadens hat die Gruppe auch die Bewertungsliste für vertrauenswürdige KI (ALTAI) entwickelt,¹³ ein Instrument, das Anbietern, Entwicklern und Nutzern von KI-Systemen helfen soll zu bewerten, inwieweit ihr KI-System die sieben Anforderungen an eine vertrauenswürdige KI erfüllt.

18.1.1 Normen

Der Weg von den Leitlinien zur praktischen Umsetzung ist lang, und die Regulierung und internationale Normen sind notwendige Zwischenschritte. Mehrere internationale Normungsorganisationen arbeiten aktiv an der Erstellung der für die Gewährleistung vertrauenswürdiger KI erforderlichen Normen:

- **IEEE Ethically Aligned Design:** <https://ethicsinaction.ieee.org/#series>. Das IEEE hat eine eigene Reihe von ethischen Leitlinien, die fast 300 Seiten umfassen.¹⁴ Diese werden durch die Normenreihe 7000 ergänzt, in der besondere Aspekte der ethischen KI spezifiziert werden. Die ersten beiden, die veröffentlicht wurden, behandeln allgemeine Grundsätze des ethischen Designs und Spezifikationen für die Messung der Auswirkungen von autonomen und intelligenten Systemen auf das menschliche Wohlbefinden.
- **ISO/IEC-Normen zu KI und vertrauenswürdiger KI:** <https://www.iso.org/committee/6794475.html>. ISO und IEC haben einen gemeinsamen Ausschuss eingerichtet, der sich mit künstlicher Intelligenz befasst. Mehrere Normen und technische Berichte sind bereits veröffentlicht worden, und viele weitere sind in Vorbereitung. Insbesondere der kürzlich veröffentlichte ISO/IEC TR 24028: Overview of trustworthiness in artificial intelligence (Überblick über die Vertrauenswürdigkeit künstlicher Intelligenz)¹⁵ gibt einen Überblick über die Anforderungen und Fallstricke bei der Entwicklung und dem Einsatz eines vertrauenswürdigen KI-Systems und kann als Fahrplan für künftige Normenspezifikationen angesehen werden.

¹³ <https://altai.insight-centre.org/>

¹⁴ <https://ethicsinaction.ieee.org/#ead1e>

¹⁵ <https://www.iso.org/standard/77608.html>

- **NIST-Standards für vertrauenswürdige und verantwortungsvolle KI:** <https://www.nist.gov/programs-projects/trustworthy-and-responsible-ai>. Das Projekt des NIST umfasst Standards für verschiedene Schlüsselaspekte der vertrauenswürdigen KI, einschließlich eines kürzlich veröffentlichten Entwurfs zur Abschwächung schädlicher Verzerrungen¹⁶ sowie bereits veröffentlichte Standards zu Erklärbarkeit und Sicherheit.
- **CEN-CENELEC-Ausschuss für künstliche Intelligenz:** <https://www.cenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/>. CEN und CENELEC haben das neue gemeinsame Komitee als Reaktion auf das Whitepaper der Europäischen Kommission zur künstlichen Intelligenz und die deutsche Normungs-Roadmap für künstliche Intelligenz gegründet.¹⁷

18.1.2 Verordnungen

Insbesondere in der EU wurde die Entwicklung einer digitalen Strategie vorangetrieben, die über Leitlinien hinausgeht und der KI-Branche eine gewisse Regulierung auferlegt. Der erste Rechtsakt in dieser Richtung war die Allgemeine Datenschutzverordnung (DSGVO), die 2018 in Kraft getreten ist. Weitere Verordnungen sind in Vorbereitung – zum Beispiel das Gesetz über digitale Dienste (DSA) und das Gesetz über digitale Märkte (DMA), deren Ziel es ist, den „Gatekeeper“-Effekt sehr großer Online-Plattformen zu verringern und den Nutzern und Verbrauchern mehr Transparenz und Wahlmöglichkeiten gegenüber diesen Plattformen zu schaffen (DSA) und kleineren Akteuren den Eintritt in die Plattformökonomie und den Wettbewerb zu ermöglichen (DMA). Während diese Verordnungen jedoch Elemente mit direkten Auswirkungen auf die Datenerhebung und die Transparenz von KI-Systemen enthalten, ist die zentrale Verordnung, die sich mit KI befasst, die EU-KI-Verordnung, die im April 2021 als Entwurf veröffentlicht wurde.

- **Digitale Strategie der EU:** https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en
- **DSGVO (GDPR):** https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en
- **DSA:** https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en
- **DMA:** https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en
- **EU-Entwurf eines KI-Gesetzes:** <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

Der Entwurf des KI-Gesetzes richtet sich an alle KI-Systeme, die in der EU in Verkehr gebracht oder in Betrieb genommen werden. Er verfolgt bei der Regulierung von KI einen risikobasierten Ansatz, bei dem das Risiko nicht nur physische oder psychische Schäden, sondern auch Risiken für die Grundrechte umfasst. Der Entwurf des KI-Gesetzes definiert den Begriff „KI“ absichtlich weit und schließt viele Algorithmen ein, deren Einordnung als

¹⁶ <https://doi.org/10.6028/NIST.SP.1270-draft>

¹⁷ <https://www.din.de/en/innovation-and-research/artificial-intelligence>

„KI“ heftige Diskussionen ausgelöst hat: nicht nur Algorithmen des maschinellen Lernens, sondern auch logikbasierte Methoden und Expertensysteme, statistische und Bayesische Verfahren, Optimierung und Suche. Die vollständige Liste ist in Anhang I des Entwurfs des KI-Gesetzes zu finden.

Der Entwurf zum KI-Gesetz nennt derzeit vier Arten von Anwendungen, die verboten sind. Dazu gehören unterschwellige Manipulationen, soziales Scoring und Gesichtserkennung:

- KI-Systeme, die Menschen manipulieren und sie zu Verhaltensweisen verleiten können, die ihnen selbst oder anderen physisch oder psychisch schaden.
- KI-Systeme, die die Schwächen bestimmter Gruppen aufgrund ihres Alters oder einer geistigen oder körperlichen Behinderung ausnutzen und zu einem Verhalten führen können, das für sie selbst oder andere physisch oder psychisch schädlich ist.
- Soziales Scoring durch öffentliche Stellen
- Die Verwendung von biometrischen Echtzeit-Fernerkennungssystemen in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken (dieses Verbot ist jedoch mit mehreren Ausnahmen verbunden).

Der Hauptinhalt der vorgeschlagenen Verordnung ist jedoch für Anwendungen mit hohem Risiko bestimmt. Diese sind in Anhang II – der eine Liste von Anwendungen enthält, die bereits einer sektoralen Regulierung unterliegen und für die der Rechtsakt zusätzliche Verpflichtungen vorsieht – und in Anhang III aufgeführt, in dem acht neue Anwendungsbereiche genannt werden, wobei in jedem Bereich spezifische Anwendungsfälle mit hohem Risiko genannt werden. Anhang II umfasst u. a. KI-Systeme, die in Spielzeug, Maschinen, medizinischen Geräten, in der Luftfahrt, in Kraftfahrzeugen und anderen Verkehrsmitteln eingesetzt werden. Die in Anhang III aufgeführten Anwendungsbereiche sind:

1. Biometrische Identifizierung und Kategorisierung von natürlichen Personen
2. Verwaltung und Betrieb von kritischen Infrastrukturen
3. Allgemeine und berufliche Bildung
4. Beschäftigung, Arbeitnehmermanagement und Zugang zur selbstständigen Arbeit
5. Zugang zu und Inanspruchnahme von wesentlichen privaten und öffentlichen Diensten und Leistungen
6. Strafverfolgung
7. Verwaltung von Migration, Asyl und Grenzkontrollen
8. Rechtspflege und demokratische Prozesse

Das Neue an Anhang III ist, dass der Entwurf des KI-Gesetzes der Kommission das Recht vorbehält, neue Anwendungsfälle in den Anhang aufzunehmen, wenn diese zu einem der acht Anwendungsbereiche gehören und sich herausstellt, dass sie ein hohes Risiko für Sicherheit, Gesundheit oder Grundrechte darstellen. Dies ermöglicht es der Kommission, erneute parlamentarische Verhandlungen über eventuelle Änderungen zu umgehen, und verschafft ihr ein gewisses Maß an Flexibilität, um auf neue Erkenntnisse über Schäden zu reagieren.

Die vorgeschlagene Verordnung stellt einige Anforderungen an Anbieter von KI-Systemen mit hohem Risiko, wenngleich in den meisten Fällen keine externe Prüfung erforderlich ist und eine Selbstbewertung ausreicht. Die wichtigsten Anforderungen beziehen sich auf

Datenqualität und Governance (Artikel 10), Risikobewertung und Risikomanagementsysteme (Artikel 9), Modellleistungstests (Artikel 15) und Modelldokumentation (Artikel 11, Anhang IV).

■ 18.2 KI-Stakeholder

KI-Systeme sind in komplexe Ökosysteme eingebettet, an denen ein breites Spektrum von Akteuren beteiligt ist. Um die Risiken für Bias in der KI-Anwendung zu verstehen und sie zu mindern, muss man die verschiedenen Akteure, ihre Rollen und ihre Bedürfnisse erfassen. Die folgende Liste kann als Leitfaden dienen, ist aber keineswegs erschöpfend.

- **Datenprovider:** Organisation/Person, die die vom KI-Anbieter verwendeten Daten sammelt, verarbeitet und bereitstellt.
- **KI-Provider:** Organisation/Person, die KI-Systeme entwickelt. Innerhalb der Organisation können spezifische zusätzliche Rollen identifiziert werden.
 - Management und Vorstand
 - Rechtsabteilung/Abteilung für Unternehmensverantwortung
 - Datenschutzbeauftragte
 - SystemarchitektInnen, DateningenieurInnen
 - EntwicklerInnen, IngenieurInnen für maschinelles Lernen, DatenwissenschaftlerInnen
 - Qualitätssicherung
- **KI-Nutzer:** Organisation/Person, die ein KI-System einsetzt. Innerhalb der Organisation können spezifische zusätzliche Rollen identifiziert werden.
 - Management und Vorstand
 - Rechtsabteilung/Abteilung für Unternehmensverantwortung
 - Qualitätssicherung
 - Datenschutzbeauftragte
 - SystemarchitektInnen, DateningenieurInnen
 - Personalwesen
 - Beschaffung
 - Personen, die direkt mit dem neuen KI-System arbeiten müssen oder deren Arbeitsplätze durch das neue KI-System ersetzt werden
- **KI-Subjekt:** Organisation/Person, auf die sich die Ergebnisse/Vorhersagen des KI-Systems beziehen.
- **Zertifizierungsstelle:** Organisation, die die Einhaltung festgelegter Standards bescheinigt.
- **Regulierungsbehörde:** Behörde, die Leistungskriterien für KI festlegt, die in ihrem Zuständigkeitsbereich eingesetzt wird.

- **Die breitere Gesellschaft, z. B. Menschenrechtsorganisationen, Verbraucherschutzorganisationen, Umweltschutzorganisationen und Medien:** Sie müssen über die Anforderungen an vertrauenswürdige KI informiert werden und sollten in der Lage sein, deren Einhaltung zu einzufordern.

■ 18.3 Fairness in der KI

Was ist ein fairer Algorithmus? Die Diskussionen finden hauptsächlich auf Englisch statt, weshalb wir uns des Weiteren auf die englischen Begriffe und ihre Definitionen beziehen. Hierzu ein Zitat aus dem Oxford English Dictionary:

Fairness¹⁸: *Impartial and just treatment or behaviour without favouritism or discrimination.*

Diese Definition ist noch nicht umsetzbar – um zu bestimmen, ob ein KI-System fair ist, muss das Konzept irgendwie quantifiziert werden. Fairness ist jedoch ein soziales Konstrukt, das vom Kontext und von kulturellen/gesellschaftlichen Normen abhängt. Dies hat dazu geführt, dass es viele verschiedene Definitionen von Fairness gibt (21 und mehr),¹⁹ jede mit ihrer eigenen mathematischen Formulierung (Fairness-Metrik) – wie im Folgenden beschrieben wird. Um die Verwirrung noch zu vergrößern, werden in der einschlägigen (meist englischen) Literatur die Begriffe „unfair Algorithm“ und „biased Algorithm“ häufig synonym verwendet.

Bias (laut Oxford English Dictionary)²⁰: *Inclination or prejudice for or against one person or group, especially in a way considered to be unfair.*

1.1 *A concentration on or interest in one particular area or subject.*

1.2 *A systematic distortion of a statistical result due to a factor not allowed for in its derivation.*

Diese Vermengung von unfair und bias mag natürlich erscheinen, wenn man die Hauptdefinition von Bias betrachtet. Dennoch ist es wichtig zu bedenken, dass jedes Klassifizierungsmodell einen Bias haben muss, um zu funktionieren. Nehmen wir zum Beispiel einen Klassifizierer, der zwischen Bildern von Säugetieren und Vögeln unterscheiden muss. Es muss eine Tendenz haben, Bilder von Tieren mit Flügeln als Vögel zu klassifizieren. Wäre es hingegen völlig frei von Bias (Voreingenommenheit), könnte es überhaupt keine Unterscheidung treffen und würde alle Objekte in dieselbe Kategorie einordnen. Daher ist eine erste Klarstellung erforderlich: Algorithmen müssen *unerwünschten* Bias vermeiden – also Vorurteile, die auf einem geschützten Merkmal oder einer falschen Korrelation beruhen und für die jeweilige Aufgabe nicht relevant sind.

¹⁸ Unparteiische und gerechte Behandlung oder Verhalten ohne Bevorzugung oder Diskriminierung.

¹⁹ Verma, S. und Rubin, J. (2018), „Fairness Definitions Explained“, Proceedings of the International Workshop on Software Fairness (FairWare), S. 1 – 7.

²⁰ Neigung oder Vorurteil für oder gegen eine Person oder Gruppe, insbesondere in einer Weise, die als ungerecht empfunden wird.

1.1 Konzentration auf oder Interesse an einem bestimmten Bereich oder Thema.

1.2 Systematische Verzerrung eines statistischen Ergebnisses durch einen bei der Herleitung nicht berücksichtigten Faktor.

Darüber hinaus gibt es in den Ingenieurs- und Statistikkreisen bereits eine bestimmte Art von unerwünschtem Bias: Bias im Sinne der Definition 1.2 (statistische Verzerrung). Dies führt häufig zu Verwirrung und Missverständnissen bei der Diskussion über Bias im maschinellen Lernen: Einfach ausgedrückt kann ein „fairer“ Algorithmus immer noch statistische Verzerrungen aufweisen, während ein System, das frei von statistischem Bias (Verzerrungen) ist, dennoch unfair sein kann.

Der Kern des Problems liegt in der Definition: Eine „systematische Verzerrung eines statistischen Ergebnisses“ setzt voraus, dass eine „Grundwahrheit“ (oder ein „wahrer Wert“ bekannt ist, sodass eine systematische Verzerrung durch Vergleich festgestellt werden kann. Aber was ist diese „Grundwahrheit“? Wenn es sich dabei, wie bisher üblich, um den aktuellen Parameterwert der Bevölkerung handelt, dann sollte es nicht überraschen, dass z. B. ein Einstellungsalgorithmus für eine Ingenieurstelle, der auf der Grundlage historischer Beschäftigungsdaten trainiert wurde, Frauen benachteiligen würde, eben weil er den Status quo genau widerspiegelt (und daher keine statistische Verzerrung aufweist). Dies ist nicht nur eine bloße Hypothese – man denke nur an das von Amazon verworfene, auf maschinelles Lernen gestützte Rekrutierungstool.²¹ Umgekehrt könnte es als notwendig erachtet werden, statistische Verzerrungen in den Algorithmus einzubauen, um mehr Geschlechtergerechtigkeit zu erreichen und „fairness“ in den Algorithmus einzubauen. Natürlich könnte dieser Widerspruch zwischen statistischer Verzerrung und Fairness nicht aufkommen, wenn die „Grundwahrheit“ als ein idealisiertes Ziel angesehen würde (d. h. die ideale Geschlechterverteilung bei den Ingenieuren). Dies ist jedoch ein umstrittenes Thema, und eine Änderung der Terminologie würde immer noch das grundlegende Problem, wie denn die ideale Verteilung aussehen sollte, nicht lösen. Aus diesem Grund vermeiden viele aktuelle Fairness-Kennzahlen die Verwendung einer „Grundwahrheit“ als Referenzparameter.

Um Verwirrung zu vermeiden, werden wir in diesem Kapitel den Begriff Bias verwenden, um die Eingaben in ein maschinelles Lernmodell (oder allgemeiner ein KI-System) oder dessen Eigenschaften zu beschreiben. Fairness hingegen wird verwendet, um die Auswirkungen von modellbasierten Ergebnissen oder Vorhersagen auf verschiedene geschützte Bevölkerungsgruppen zu beschreiben. Dies steht auch im Einklang mit einer wachsenden Anzahl an Veröffentlichungen, in denen versucht wird, Quellen von Bias in KI-Systemen zu identifizieren und zu reduzieren, und in denen Fairness-Metriken zur Bewertung der Auswirkungen von Modellen verwendet werden.

18.3.1 Bias

Bias kann in vielen Formen auftreten und in den Lebenszyklen von maschinellem Lernen und Data Science in verschiedenen Phasen auftreten. Es können vier Hauptphasen identifiziert werden:

1. Der Bias kann in den Trainings- oder Testdaten liegen. Eine große Datenmenge befreit Datensammler nicht automatisch von den traditionellen statistischen Datenfehlern. Stichprobenverzerrung, Auswahlverzerrung und Antwortausfall sind nur einige der häufigsten.

²¹ Dastin, J. (2018), „Amazon scraps secret AI recruiting tool that showed bias against women“, Reuters, 11. Oktober 2018.

figsten Fallen, die Daten für Unwissende bereithalten. Wie das obige Beispiel für das Training eines Einstellungsalgorithmus anhand historischer Daten zeigt, kann aufgrund menschlicher Vorurteile Bias selbst dann in den Daten bestehen, wenn das Verfahren zur Beschaffung der Daten statistisch korrekt war. Die Einstellungsdaten, die zum Trainieren des Algorithmus verwendet werden, könnten den Status quo genau widerspiegeln – und somit die derzeitige gesellschaftliche Voreingenommenheit gegenüber Frauen im Ingenieurwesen kodieren und aufrechterhalten. Worteinbettungen und Sprachmodelle sind ein weiteres Beispiel für diese Art von Bias – der Text, der zum Trainieren dieser Modelle verwendet wird, ist voller gesellschaftlicher Vorurteile, sodass die Worteinbettungen nicht nur allgemeine semantische Muster, sondern auch geschlechtsspezifische²² und ethnische²³ Stereotypen und Vorurteile widerspiegeln.

2. Bias kann auch beim Entwurf des Algorithmus in das System einfließen – so könnte ein Klassifizierungssystem aufgrund der Kategorien, für die es entwickelt wurde (schwarz/weiß, männlich/weiblich),²⁴ verzerrt sein; Bias kann bei der Entwicklung von Merkmalen auftreten (einige Merkmale könnten für einige Gruppen aussagekräftiger sein als für andere, und die Auswahl von Merkmalen auf der Grundlage der Gesamtgenauigkeit könnte dazu führen, dass das Modell für einige Gruppen schlechter abschneidet) oder bei der Wahl des zu verwendenden Algorithmus (zu einfache Algorithmen können beispielsweise die Daten nicht richtig erfassen und zu Verzerrungen in den Modellen führen). Eine besonders heimtückische Form von Bias kann in den Algorithmusentwurf einfließen, wenn versucht wird, ein Konzept zu modellieren, das nicht vollständig quantifizierbar ist – z. B. bei der Zulassung an einer Universität, wenn Aufzeichnungen über zuvor zugelassene Studenten verwendet werden, um ein Modell zur Erkennung erfolgreicher Kandidaten für ein Doktorandenprogramm zu trainieren²⁵ (in Wirklichkeit werden dadurch einfach die Präferenzen und Vorurteile früherer Zulassungsausschüsse modelliert); oder bei der Verwaltung der Krankenhausversorgung, wenn die Kosten für die Gesundheitsversorgung als Indikator für die Schwere der zu behandelnden Krankheit verwendet werden.²⁶
3. Bias kann auch post hoc in das System einfließen, zum Beispiel bei der Interpretation der Modellergebnisse. Oder Entscheidungen, die auf Modellvorhersagen beruhen, könnten sich auf Daten auswirken, die dann wieder in einen Online-Lernalgorithmus eingespeist werden, was zur Bildung von unkontrollierbaren Rückkopplungsschleifen führt²⁷ und bestehenden Bias in den Daten oder im Modell noch verstärkt.

²² Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., Kalai, A. (2016), „Man is to computer programmer as woman is to homemaker? debiasing word embeddings“, Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS 2016, S. 4356 – 4364.

²³ Manzini, T., Yao Chong, L., Black, A. W., Tsvetkov, Y. (2019), „Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings“, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, S. 615 – 621.

²⁴ Leufer, D. (2021), „Computers are binary, people are not: how AI systems undermine LGBTQ identity“, Access Now, April 2021.

²⁵ Burke, L. (2020), U of Texas will stop using controversial algorithm to evaluate Ph.D. applicants, Inside Higher Ed, 14 December 2020.

²⁶ Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019), „Dissecting racial bias in an algorithm used to manage the health of populations“, Science, Vol. 366, S. 447 – 453.

²⁷ Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C. und Venkatasubramanian, S. (2018), „Runaway feedback loops in predictive policing“, Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, Vol. 81, S. 160 – 171.

4. Und schließlich ist der Einsatz von KI auch anfällig für Bias: von der zeitlichen Abweichung über die unangemessene Nutzung (in einem anderen als dem beabsichtigten Kontext) und von feindlichen Angriffen (man denke nur an den berüchtigten Chatbot Tay von Microsoft²⁸) bis hin zum selektiven Einsatz (z. B. Verwendung von Vorhersagemodellen zur Ermittlung von Noten für Kinder in größeren Klassen, aber Verwendung menschlicher Bewertungen zur Ermittlung von Noten für Kinder in kleineren Klassen)²⁹.

Es ist nicht möglich, alle erdenklichen Arten von Bias aufzulisten, die in ein Modell für maschinelles Lernen einfließen können; wir beschreiben aber im Folgenden kurz einige der häufigsten Formen von Bias.³⁰

- **Menschlicher kognitiver Bias:** Jede Art von Voreingenommenheit, die bei der Verarbeitung und Interpretation von Informationen durch Menschen auftreten kann
- **Gesellschaftlicher Bias:** Voreingenommenheit und Vorurteile, die sich aus einem sozialen, kulturellen oder historischen Kontext ergeben
- **Confirmation Bias:** Eine Tendenz, Modellvorhersagen zu akzeptieren, die mit den eigenen bereits bestehenden Überzeugungen übereinstimmen
- **Group Attribution Bias:** Tritt auf, wenn angenommen wird, dass das, was für ein Individuum in einer Gruppe gilt, auch für alle Mitglieder dieser Gruppe gilt
- **Automation Bias:** Eine Tendenz, sich zu sehr auf die Ergebnisse eines Vorhersagemodells zu verlassen
- **Zeitliche Verzerrung:** Bias, der dadurch entsteht, dass Unterschiede in den beobachteten/gemessenen Größen im Zeitverlauf nicht berücksichtigt werden
- **Stichprobenverzerrung (Sampling Bias):** Tritt auf, wenn die Daten nicht nach dem Zufallsprinzip aus der vorgesehenen Grundgesamtheit gezogen werden, sodass einige Personen mit größerer Wahrscheinlichkeit in die Stichprobe aufgenommen werden als andere
- **Repräsentationsverzerrung:** Tritt auf, wenn Einzelpersonen oder Gruppen in einer Studie systematisch von der Population von Interesse abweichen. Dies kann zwar den Fall der Stichprobenverzerrung einschließen, ist aber ein breiteres Konzept. Selbst wenn die Daten nach dem Zufallsprinzip aus der Gesamtbevölkerung entnommen werden, kann der Stichprobenumfang bzw. die Datenqualität für bestimmte Untergruppen gering sein, was zu Ergebnissen führt, die sich nicht gut auf diese Untergruppen verallgemeinern lassen.
- **Measurement Bias:** Diese Art von Messfehlern kann auftreten, wenn die im Modell verwendeten Merkmale und/oder Bezeichnungen Stellvertreter (Proxys) für die tatsächlich interessierende Größe sind, wodurch möglicherweise systematische Fehler zwischen dem, was beabsichtigt ist, und dem, was tatsächlich gemessen wird, entstehen (wie im oben genannten Beispiel der Verwendung von Gesundheitskosten zur Messung der Schwere einer Krankheit)³¹.

²⁸ The Guardian (2016), „Microsoft ‘deeply sorry’ for racist and sexist tweets by AI chatbot“, 26. März 2016.

²⁹ Elbanna, A., Engesmo, J. (2020), „A-level results: why algorithms get things so wrong – and what we can do to fix them“, The Conversation, August 19, 2020.

³⁰ Suresh, H. und Guttat, J. (2021), „A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle“, arXiv preprint, <https://arxiv.org/pdf/1901.10002.pdf>

³¹ Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019), „Dissecting racial bias in an algorithm used to manage the health of populations“, Science, Vol. 366, S. 447 – 453.



Stefan Papp ist Unternehmer. Er hilft Organisationen beim Aufbau von Datenarchitekturen und bei der Migration von On-Premise-Lösungen in die Cloud. Stefan Papps Hauptaugenmerk liegt auf Lösungen im Bereich Climate Action. Er und sein Team entwickeln Lösungen zur Reduzierung von Kohlenstoffemissionen, wobei auch der Kohlenstoffhandel ein wichtiges Thema ist. Stefan Papp ist dabei, ein Kompetenzzentrum für Climate Action in Armenien einzurichten.

Stefan ist auch Berater für vielversprechende armenische Start-ups. Er ist einer von vielen Business Angels, die investieren, um Armenien dabei zu helfen, das „Silicon Valley des Ostens“ zu werden.



Wolfgang Weidinger ist Data Scientist und hat in einer Vielzahl von Branchen und Sektoren wie Start-ups, Finanzen, Beratung und Großhandel gearbeitet. Dort leitete er Data-Science-Teams und trieb deren Rolle als Speerspitze der digitalen und datengetriebenen Transformation voran.

Er ist Präsident der Vienna Data Science Group (www.vdsg.at), einer Non-Profit-Vereinigung von und für Data Scientists. Diese bringt sowohl Forschung als auch Praxis in einem breiten Spektrum von Branchen zusammen. Die VDSG ist eine schnell wachsende internationale Gemeinschaft, deren Ziel es ist, über Data Science und seine Teilbereiche wie maschinelles Lernen und künstliche Intelligenz sowie deren Auswirkungen auf die Gesellschaft aufzuklären.

Wolfgang interessiert sich besonders für die gesellschaftlichen Auswirkungen von Data Science und KI sowie für die Etablierung von interdisziplinären Data-Science-Teams in Unternehmen und deren disruptive Auswirkungen auf Geschäftsmodelle.



Katherine Munro ist Data Scientist und Data Science Ambassador im Bereich E-Commerce. Sie forscht und hält Unternehmensschulungen in den Bereichen KI, maschinelles Lernen, Natural Language Processing und Data Science.

Mit einem Hintergrund in Computerlinguistik und maschinellem Lernen hat Katherine in der Forschung und Entwicklung für Mercedes-Benz und das Fraunhofer-Institut gearbeitet und sich dabei auf Benutzeroberflächen und natürliches Sprachverständnis spezialisiert. Sie hat auch als Universitätsdozentin und Englischlehrerin gearbeitet und hält jetzt Vorträge, ist Education Lead für Women in AI Upper Austria, ehrenamtliche Mentorin bei Female Coders Linz und Trainerin für LinkedIn Learning.



Bernhard Ortner arbeitet derzeit als Data Architect Lead und Enterprise Architect bei den Wiener Linien, wo er das Unternehmen durch den Aufbau der nächsten Mobilitätsplattform umgestaltet. Zu seinen Tätigkeiten gehören die Anpassung bestehender Prozesse rund um Big Data und die Etablierung von Big-Data-Standards und Best Practices. Außerdem hält er Vorlesungen zu ausgewählten Data-Engineering-Themen an der Dualen Hochschule Baden-Württemberg und ist in verschiedenen Open Source Communitys aktiv. Seine Hauptmotivation ist es, neue Wege und Möglichkeiten für die Anwendung von Big Data zu finden.



Annalisa Cadonna ist Statistikerin und Beraterin für Data Science. Sie promovierte in angewandter Mathematik und Statistik an der University of California, Santa Cruz. Annalisa hat statistische und Machine-Learning-Methoden für Projekte in der Finanz-, Energie- und Medizinbranche eingesetzt. Derzeit ist es ihr berufliches Ziel, die Lücke zwischen Zeitreihenforschung und industriellen Anwendungen zu schließen, indem sie probabilistische Programmierung und Cloud-Technologien einsetzt. Annalisa ist bestrebt, Statistik und ML als Mittel zur Erreichung der Ziele für nachhaltige Entwicklung einzusetzen und sich aktiv an der Entwicklung von Werkzeugen und Frameworks für verantwortungsvolle künstliche Intelligenz zu beteiligen. Sie ist auch eine der Organisatorinnen von R-Ladies Vienna.



Georg Langs ist Professor für Maschinelles Lernen in der medizinischen Bildgebung an der Medizinischen Universität Wien, wo er das Computational Imaging Research Lab am Department für Biomedizinische Bildgebung und bildgeführte Therapie leitet. Er ist Mitbegründer und Chief Scientist des Spin-offs contextflow GmbH, das Software für KI-basierte Bildsuche entwickelt. Georg Langs studierte Mathematik an der Technischen Universität Wien und Informatik an der Technischen Universität Graz und war Research Scientist am MIT Computer Science and Artificial Intelligence Lab, wo er immer noch Research Affiliate ist.



Roxane Licandro ist Postdoktorandin an der Medizinischen Universität Wien und Forschungsstipendiatin am Massachusetts General Hospital und der Harvard Medical School. Sie schloss ihr Studium der Medizinischen Informatik an der TU Wien ab, wo sie als Universitätsassistentin am Computer Vision Lab arbeitete. Sie erhielt ein Marie Skłodowska-Curie-Stipendium und absolvierte Forschungsaufenthalte an der Charité Berlin, dem Kinderspital Zürich und dem University College London. Sie arbeitete am Kunsthistorischen Museum Wien und bei Agfa Healthcare. Ihr Forschungsschwerpunkt liegt auf der Suche nach neuen Möglichkeiten zur rechnerischen Modellierung und Vorhersage dynamischer Prozesse in Raum und Zeit, der pädiatrischen und fötalen Gehirnentwicklung, der statistischen Musteranalyse in der Krebsforschung und der geometrischen Formanalyse von anatomischen und kulturellen Objekten.



Mario Meir-Huber ist Head of Data bei UNIQA, dem führenden Versicherungsunternehmen in Mittel- und Osteuropa. Hier arbeitet er mit seinem Team daran, das Unternehmen datengetrieben zu machen. Vor seinem Eintritt bei UNIQA war er in ähnlichen Positionen bei einem führenden Telekommunikationsunternehmen sowie bei großen Technologieanbietern wie Microsoft tätig. Darüber hinaus ist er Keynote-Speaker bei verschiedenen internationalen Veranstaltungen wie der GITEX oder der London Tech Week. Mario hat bereits mehrere Bücher zu den Themen Cloud und (Big) Data veröffentlicht. Sein Blog ist unter *cloudvane.net* zu erreichen.



Dr. Danko Nikolić ist Experte für Hirnforschung und KI. Viele Jahre hat er ein elektrophysiologisches Labor am Max-Planck-Institut für Hirnforschung geleitet. Als Experte für KI und Machine Learning leitet er ein Data-Science-Team und entwickelt kommerzielle Lösungen auf der Grundlage von KI-Technologie.

Er erfand den KI-Kindergarten – ein Konzept für das Training der KI der Zukunft, um eine Intelligenz auf nahezu menschlichem Niveau zu erreichen. Er leistete auch Pionierarbeit beim Einsatz von ML, um „Gedanken“ aus den elektrischen Signalen des Gehirns zu lesen; er und sein Team konnten allein durch die Analyse der Gehirnsignale rekonstruieren, was ein Tier sah. Er führte das Konzept der Ideasthesie („Konzeptwahrnehmung“) in die Neurowissenschaften ein und ist der Autor einer Theorie namens Practopoiesis, die beschreibt, wie biologische Systeme Intelligenz erreichen. Er hat einen Abschluss in Psychologie und Bauingenieurwesen von der Universität Zagreb, Kroatien, und einen Dokortitel von der Universität Oklahoma, USA. Von 2014 bis 2019 war er Honorarprofessor an der Universität Zagreb.



Zoltan C. Toth ist Data Engineering Architect, Dozent und Unternehmer. Mit einem Hintergrund in Informatik und Mathematik hat er Datenarchitekturen, Big-Data-Technologien und den Betrieb von ML für Fortune-500-Unternehmen weltweit unterrichtet. In den letzten zwei Jahrzehnten hat er als Solution Architect mit mehreren großen Unternehmen zusammengearbeitet und dabei Datenanalyseinfrastrukturen implementiert und diese bis zur Verarbeitung von Petabytes an Daten skaliert. Außerdem ist er Dozent an der Central European University. Er gründete Datapao, ein Beratungsunternehmen für Data Engineering, das zum europäischen Professional Services Center von Databricks und zu einem Microsoft Gold Partner für Data Science wurde.



Barbora Vesela ist Data Scientist und Software Engineer und arbeitet bei Frequentis, einem Unternehmen, das in einem sicherheitskritischen Kommunikations- und Informationsumfeld tätig ist. Ihr Hintergrund ist ein Studium der Biophysik an der Masaryk-Universität in Brunn und der Biomedizintechnik an der FH Technikum Wien und der Technischen Universität Brunn. Sie interessiert sich für verschiedene Themen, die Data Science und Signal- und Bildverarbeitung in verschiedenen Bereichen wie Medizin, Forschung und Luftverkehrsmanagement kombinieren.



Rania Wazir ist Mathematikerin und Data Scientist mit den Schwerpunkten Trustworthy KI, Natural Language Processing und Social Media Monitoring. Sie ist stellvertretende Vorsitzende des österreichischen Normenausschusses für KI und österreichische Delegierte in der ISO-Arbeitsgruppe für vertrauenswürdige KI; außerdem ist sie Koordinatorin der data4good-Initiative des VDSG, die mit gemeinnützigen Organisationen an datenbasierten Projekten arbeitet. Sie leitete ein Konsortium von Experten aus den Bereichen maschinelles Lernen, Recht und Sozialwissenschaften, das vor Kurzem eine Untersuchung über Voreingenommenheit in Algorithmen für die EU-Grundrechteagentur abgeschlossen hat, und ist derzeit die technische Leiterin eines dreijährigen Projekts zur Entwicklung eines fairen KI-Entwicklungsprozesses, das von der österreichischen Forschungsagentur finanziert wird.

Dr. Wazir ist gemeinsam mit der Open-Innovation-Expertin Dr. Gertraud Leimüller Mitbegründerin des kürzlich gegründeten Start-ups leiwand.ai, dessen Ziel es ist, Unternehmen und Organisationen, die mit der Entwicklung oder dem Einsatz von KI-Systemen befasst sind, die notwendigen Werkzeuge und das Know-how zur Verfügung zu stellen, um die Vertrauenswürdigkeit ihrer Systeme zu gewährleisten.



Günther Zauner ist langjähriger Mitarbeiter der dwh GmbH, Mathematiker und Experte auf dem Gebiet der Modellierung und Simulation, Parametrisierung und Prognosemodellierung. Er arbeitet sowohl in Industrieprojekten als auch in Forschungsprojekten (z.B. EU FP7 CEPHOS-LINK, Horizon 2020 RheumaBuddy). Er ist spezialisiert auf die Entwicklung von Modellierungskonzepten, die Integration von Routinedaten und das Verhalten der Bevölkerung. Er ist Mitglied des VDSG, der Society of Medical Decision Making (SMDM) und Mitglied des Vorstands der International Society for Pharmacoeconomics and Outcomes Research Austria (ISPOR Austria). Darüber hinaus ist er Gutachter für mehrere Fachzeitschriften und promoviert im Bereich Public Health unter der Leitung von Professor Majdan an der Universität Trnava.

Index

Symbole

1644 424

A

- Ableitung 180
- A/B-Test 138
- Accuracy 155, 214
- Actions 243
- Adam Optimizer 293
- Ad-hoc-Entscheidung 452
- Adversarial Training 255
- Adversary 255
- Agent 242
- Agentenbasierte Modellierung 403, 409
 - Agent 404
 - Akteur 404
 - Covid-19 409
 - entstehendes Verhalten 404
 - Flexibilität 405
 - natürliche Beschreibung 405
- AGI 295
- Agile Analytics 466
- Agiles Manifest 552
- AlexNet 249
- Aliasing-Effekt 355
- Allgemeine Datenschutzverordnung (DSGVO) 146
- alphaGo 268
- Amazon Redshift 68
- Analyse 277
- Analytics-Abteilung 468
- AnalyticsOps 139
- Analytische Kompetenz 451
- Anonymisierung 481
 - Arten 481
- Antipatterns 556
- Apache Airflow 104
- Apache Atlas 151
- Apache Hadoop 38, 496
- Apache Hive 92
- Apache Parquet 35
- Apache Spark 94, 142
- Arbeitsgedächtnis 295
- Architektur 265
- Area Under the Curve 216
- Armenien 554
- Artificial General Intelligence 295
- Artificial Muse 522
- Artificial Neural Networks (ANN) 237
- Assoziationsdiagramm 428
- Athena 92
- Aufmerksamkeit erregen 429
- Auftragsverarbeitung 479
- Auftragsverarbeitungsvertrag 479
- Autoencoder 253
- Automatisierung 133, 144, 453
- Automobilindustrie 498
- Autonomes Fahren 263, 500
- Autonome Fahrzeuge 265
- AWS 64
 - AWS Elastic Beanstalk 69
 - AWS Redshift Spectrum 92
 - AWS S3 66
- Azure 64
 - Azure Data Factory 104
 - Azure Synapse Serverless 92
 - Azure VM 67

B

Bagging 233
 Balanced Scorecard 510
 Balkendiagramm 425, 433
 Basislinie 444
 Baumdiagramm 437
 Bayer-Muster 357
 Bayes-Theorem 194
 Bayesscher Ansatz 269
 Bereitgestellte IOPS 40
 Bernoulli-Verteilung 189
 Betrugserkennung 510
 Bevölkerung 410
 BI-Abteilung 457
 BI-Engineer 457
 Bi-modale IT 557
 Bias 230 f., 236, 567, 569
 – Definition 572
 – Erklärbarkeit 585
 – Überprüfung 580
 BigQuery 92
 Bildauflösung 354
 Bildformate 358
 Bildhelligkeit 361
 Bildregistrierung *siehe* Registrierung
 Bildschärfe 353
 Bildverzerrung 365
 Binomialverteilung 189
 Biologische Neuronen 236
 Blende 354
 Bootstrapping 217
 Box Plot 436
 Brennweite 353
 Brute-force 265
 Brute-force-Suche 293
 Bullwhip-Effekt 397, 415
 Business-Analyst 13, 456
 Business Data Owner 457
 Business Intelligence 496

C

Camera Obscura *siehe* Lochkamera
 CCD 355
 Center for Internet Security (CIS) 158
 Central Limit Theorem 191
 Change Data Capture 134

Charge-Coupled Device *siehe* CCD
 Chief Data Officer 457
 China 553
 Cholesky-Zerlegung 179
 Churn Rate 536
 Cilium 42
 Cloud 466
 Cloud-Anbieter 65
 Cluster 239
 Color Filter Array 356
 Completeness 155
 Computational Photography 363
 Computer Vision 348, 350, 361 f., 364, 366
 Computergestütztes Modell 387
 Configuration Drift 59
 Confusion-Matrix 138
 Consistency 155
 Container 141 f.
 Containerization 140
 Continuous Delivery (CD) 139
 Convolutional Neural Network (CNN) 246
 Covid-19 409
 – Modell-Implementierung 414
 – Modell-Kalibrierung 413
 – Parametrisierung 413
 – Struktur und Zeitplanung 411
 CPU 34, 39, 60
 Cross-Validation 138
 CSV 82
 Customer Journey 25, 510

D

DAG 103
 Dagster 104
 Dartmouth-Konferenz 267
 Databricks 92, 101, 149
 – Unity Catalog 149
 Data Discovery 151
 Data Engineer 14 f.
 DataFrame API
 – Apache Spark 95, 98
 – MLlib 100
 Data Governance 147
 Data Governance Council 154
 Data Lake 91, 116
 DataLakeHouse 118

- Data Lineage 18
- DataOps 139
- Data Ownership 147
- Data Pipeline 113, 122, 134
 - Daten bereinigen 135
 - Datenerhebung 134
 - Delivery 120
 - Design 120
 - Exploration 134
 - Modellierung 134
 - Visualisierung 137
- Data Science Lab 468
- Data-Science-Projekte 11
- Data Scientist 456
- DataSet API
 - Apache Spark 95
- Data Steward 16
- Data Swamp 146, 467
- Data Warehouse 87, 467, 496
- Daten
 - Verfügbarkeit 382
 - verwalten 429
- Datenarchitekt 16
- Datenarchitektur
 - Skalierbarkeit 77
 - Verlässlichkeit 77
 - Wartbarkeit 77
- Datenbanken 115
- Datenerhebung 134
- Datenexport 485, 488
- Daten-Governance 484
- Datenkatalog 149
- Datenklassifizierung 159
- Datenladeszenarien 134
- Datenlöschung 485
- Datenmanagement 147
- Datenmenge 273
- Datenplattform 19, 23, 29
- Datenprogramm 494
- Datenqualität 154
- Datenschutz 32, 555
- Datenschutzbeauftragter 487
- Datensicherheit 489
- Datenstrategie 9, 23, 30, 453
- Datentypen 439
- De-Biasing 580
- Decision Tree 274

- Deep Blue 280
- Deep Fakes 375
- Deep Learning 6, 274, 370
- Deep-Learning-Modelle 122
- Defense in Depth 166
- Delta Lake 86, 92
- DenseNets 249
- Depth of Field 354
- Deskriptive Statistik 277
- Determinante 177
- DevOps 16, 139
- Dezentralisierte Teams 458
- Digitales Bild 357
- Digitale Fotografie 355
- Digitale Gesellschaft 547
- Digitale Transformation 494
- Diskrete Ereignissimulation 400
 - dynamische diskrete Systeme 401
 - Ereignisliste 401
 - Prioritäten 401
- Diskriminator 255
- Diversity 548
- Docker 42, 128
- Dokumentation 390
 - Quellcode 390
 - textlich 390
- Domänenanpassung 373
- Domänenexperte 456
- Drei V von Big Data 80
- Driver
 - Apache Spark 97
- Dropout 276
- DSGVO 474, 476
 - Grundsätze 477
- du 45
- DW 87
- DynamoDB 68

E

- EC2 (Elastic Compute Cloud) 67, 70
- Eckendetektor 363
- Effektanordnung bei der Datendarstellung 424
- Effizienzgrenze 277
- Eigenvektoren 176
- Eigenwerte 176
- Eigenwertzerlegung 177

Eingeschränkt schützenswerte Daten 475
 Einwilligungserklärung 477f., 484f., 487
 – Beispiel 478
 Einzelhandel 532
 Elastic Compute Cloud 67
 Elastizität 63
 Embedding Space 241
 End-to-End-Prozess 19
 Energiesektor 506
 Ensemble-Methoden 233
 Entscheidungsbaum 266
 Environment 243
 ePrivacy-Verordnung 476, 479, 487
 Erklärbare KI 375
 Erklärbarkeit 567f., 585
 – Modell 582
 Erwartungswert 192
 Ethikrichtlinien 566
 ETL 78, 113
 ETLT 80
 Europa 555
 Executor
 – Apache Spark 97
 expectation maximization 240
 Exploration 137
 Exponentielles Wachstum 282
 Externer Validierungssatz 260
 eXtreme Programming 552

F

Fachexperte 556
 Fail-Fast-Ansatz 466
 Fairness 566 ff., 572 f.
 – Definition 572, 580
 – kausale Zusammenhänge 579
 Fairness-Kennzahlen 573, 577f.
 – Modelldokumentation 584
 Fairness-Metriken 576 f.
 FAIR-Prinzip 158
 Feature Engineering 276
 Fertigung, Massenproduktion 523
 Finanzinstitute 509
 Flächendiagramm 433
 F-Score 214
 Function as a Service (FaaS) 64
 Funktion 180

G

Gaussian distribution 191
 Gaussian Mixture Model (GMM) 241
 Gauß'sche Verteilung 292
 Gehirn 265
 Generalised Linear Model (GLM) 278
 Generative Adversarial Networks (GAN) 255
 Genetic Algorithms 274
 Geografisches Profiling 517
 Git 55
 Gleichverteilung 190
 Globale Merkmale 361
 Globales Minimum 293
 GOFAI 266 f., 294
 Google Cloud 64
 Google Cloud AI Platform 64
 Google Compute Engine 67
 GoogLeNet 249
 GPT-3 296
 GPU 35, 39
 Gradient 182, 248
 Gradientenabstieg 184, 248, 293
 – stochastisch 185
 Grafana 151
 Grenzen 386
 grep 52

H

Haltbarkeit 32
 HDR 363
 Heat Map 436
 Hessian 182
 Heterogene Daten 429
 Hidden Layer 237
 High Dynamic Range Imaging *siehe* HDR
 Histogramm 426
 Homografie 365
 Homoskedastizität 292
 htop 44
 Hybrid Cloud 466
 Hybride Modellzerlegung 390
 Hyperparameter-Abstimmung 274

I

IAM (AWS Identity and Access Management) 70
Idempotenz 59
Identity Providers 157
Image Morphing 365
ImageNet-Datensatz 283
Image-Retrieval-Systeme 367
Image Stitching 366
Induktive Verzerrung 271
Informatik 428
Informationssicherheit 158
Infrastructure as a Service (IaaS) 64
Inpainting 364
Input Layer 237
Integrity 156
Interest Points 361
I/O 35
iperf3 50
iTerm2 42

J

Jenkins 123
Jeopardy 264
JPEG 359
jps 54
JSON 84
jumbo packets 34

K

Kalibrierung 389
Kanban 552
K-Anonymität 482
Kantenerkennung 361
Kapitalallokationslinie 277
Kasparov, Garry 280
Keiretsu 551
Key Management System 165
Keypoints *siehe* Interest Points
KI für das Gesundheitswesen 280
KI-Stakeholder 565 ff.
– Kennzeichnung der Daten 583
– Rollen 571
– Transparenz 581
KISS 36

Klassifikationsmodell 221
Klassifizierungsmodelle 198
K-Means 316
k-means clustering 240
Kommandozeilen-Tools
– awk 57
– grep 57
– logcheck 57
– logwatch 57
Komprimierung 359
Konfigurationsmanagement 63
Konfusionsmatrix 213
Korrelation 192, 277
Kovarianz 192
Krankheit 410
Kreisdiagramm 425, 435
Kreuzvalidierung 217
Kryptowährungen 277
Kubernetes
– Pods 142
Kumulative Variablen 278
Kundenbindung 26
Kundenzufriedenheit 504
Kunst 520
Künstliche Intelligenz (KI) 8, 32, 220, 262, 431
Kurzzeitgedächtnis 295

L

Lakehouse 92
Lasso 276
Laws of Physics 274
l-diversity 162
Lebenszyklus 386
Legacy-Systeme 35
Lego 288
LeNet 249
Lernen, few-shot 342
Lernen, one-shot 343
Lernen, zero-shot 341, 343
Lieferantenverwaltete Bestände 414
Lieferkette 415
Lime 484
Lineare Beziehungen 292
Lineare Regression 198, 274
Lineare Transformationen 175
Liniendiagramm 425, 432

Link Aggregation 40
 Linsen-Formel 354
 LISP 267
 Loan Prediction 511
 Lochkamera 352
 Logistische Regression 209
 Logstash 57
 Lokale Merkmale 361
 Long-Range-Korrelationen 275
 Long Short-Term Memory (LSTM) 251, 275, 293
 Isof 52
 Luftfahrt 502
 LU-Zerlegung 178

M

Machine Learning 375, 377
 Machine learning perpetuum mobile 292
 Makroskopische Methoden 395
 Manifolds 241
 MapReduce 39
 Markdown 55
 Marketing-Segment-Analyse 25
 Marvin Minsky 280
 Maschinelles Lernen 220, 367
 Maslowsche Bedürfnishierarchie 76
 Matrix, positiv definit 177
 Matrixmultiplikation 173
 Matrix-Vektor-Multiplikation 175
 Matrizen 169

- diagonale 172
- Identitätsmatrix 172
- Kehrwert 172
- Transponierung 171

 Maximum Likelihood 241
 McCarthy, John 267
 Mean Absolute Error 204
 Mean Squared Error 200
 Menschliche Intelligenz 295
 Merkmalsextraktion 361
 Merkmalsvektor 361
 Mersenne Twister 291
 Mikroskopische Methoden 395
 Minard 426
 MIPS 39
 MNIST 282
 Modell

- abstrakt 382

- epidemisch 396
- Falsifikation 384
- Implementierung 391
- Konzept 382
- konzeptionell 386, 388
- Lebenszyklus 386 f.
- Output 387
- qualitatives 390
- reproduzierbar 384
- Skalierung 143
- Stochastik 413
- Vergleich 394

Modellierung

- Blackbox 384
- diskriminative 340
- dynamisch 380
- Eisenbahnnetze 406
- generative 340
- iterativ 386
- iterativer Prozess 383
- statisch 380
- Whitebox 385

Modell-Update 141

Model Specialization 274

Module 410

- Maßnahmen 410

Modulare Modellierung 409

Monetarisierung von Daten 501

Moore-Penrose Pseudoinverse 179

Mosaicing *siehe* Image Stitching

Mosaikdiagramm 428

Multi-Cluster Shared Data Architecture 69

Multilayer Perceptrons 236, 246

Multiple lineare Regression 206

N

Nagios 57

Naive Bayes 274

Natural Language Generation 300

- GPT 341

Natural Language Processing (NLP) 299, 524

- Hidden Markov Models 318

- maschinelle Übersetzung auf Phrasenbasis 321

- Naive Bayes 314

- regelbasiert (symbolisch) 311

- statistisches maschinelles Lernen 314

- statistische Sprachmodellierung 319
- Stimmungserkennung 314
- symbolische Bedeutungsdarstellungen 313
- Text-Clustering 316
- Natural Language ToolKit (NLTK) 301
- Natural Language Understanding 300
 - Domain Detection 300
 - Intent Detection 300
 - Slot Filling 300
- NDA (Non-Disclosure Agreement) 557
- netstat 52f.
- Netzwerk 34
- Neuronen 237
- Nightingale, Florence 427
- NLP-Datenvorbereitung
 - Abschneiden und Auffüllen 309
 - Bag-of-Words (BOW) 308
 - Lemmatisierung 304
 - Named Entity Recognition (NER) 307
 - POS Tagging 303
 - Stemming 304
 - Stoppwort-Entfernung 306
 - TF-IDF 309
- NLP, neuronal 323
 - Convolutional Neural Networks 323
 - Decoder 325, 340
 - Encoder 325
 - Feedforward Neural Network 328
 - Long Short-Term Memory Networks 324
 - neuronale Aufmerksamkeit 327
 - Recurrent Neural Networks 324
 - Seq2Seq 324
- NLP Transferlernen 329
 - BERT 338
 - ELMO 334
 - GloVe 331
 - GPT 340
 - M2M100 344
 - Masked Language Modelling 339
 - Multi-headed Attention 337
 - Next Sentence Prediction 339
 - Self-Attention 337
 - transformer 335
 - transformer attention 336
 - Word2Vec 330
- nmap 51
- nmon 53
- No free lunch theorem 291

O

- Objektidentifizierung 366
- ODBC 81
- Omnichannel-Prozess 510
- One-shot Learning 272
- ONNX 142
- OPC UA 81
- OPC Unified Architecture 81
- openAI 283
- Open Innovation 20
- Open Source 465
- Operational Applications 453
- Operationalisierung 472
- Optimierung 180
 - eingeschränkt 186
- Optische Achse 353
- Optische Täuschungen 351
- Ordinary Differential Equations 396
- Output Layer 237
- Overfitting 231
- Overplotting 441

P

- Paketmanager
 - apk 44
 - apt 44
 - brew 44
 - yum 44
- Parameter 387
- Parametrisierung 389
- Parquet 85
- Partial Differential Equations 396
- Partielle Differentialgleichung 381, 396
- Pearsons Korrelationskoeffizient 278
- Perceptron 236
- Performance 32
- Personal Identifiable Information (PII) 146
- Personenbezogene Daten 476
- Perspektivische Projektion 353
- Physische Sicherheit 32
- Platform as a Service (PaaS) 64
- PMML 142
- PNG 360
- Poisson-Verteilung 189
- Polardiagramm 427
- policy 243

Polyglotte Datenspeicherung 497
 Potenzgesetz 284
 Pragmatism 268
 Precision 214
 Predictive Analytics 26
 Predictive Maintenance 25
 – Energieanlagen 508
 – Öl und Gas 528
 – Telekommunikation 537
 – Transportwesen 539
 Prefect 104
 Prescriptive Analytics 494
 Presto 92
 Principal Component Analysis 177
 Privacy by Default 486
 Privacy by Design 159 f., 486
 probability density function 190
 probability mass function 188
 Prometheus 57, 151
 Proof of Concept 452
 Provisioning Tools 59
 Pseudo-Anonymisierung 481
 Pseudozufallszahlengenerator 291
 PuTTY 42
 PyLint 124
 Python 125, 141

Q

Quadratische Verlustfunktion 200
 Qualitative Modellierung 390
 Qualitative Variablen 198
 Qualitätsoptimierung 524
 Qualitätssicherung 139
 Quantitative Variablen 197

R

R, Programmiersprache 125
 Radiometrische Auflösung 355
 Random Forests (RF) 232, 274, 511
 random initialization 240
 Randomization 482
 Randomization-Kette 483
 RDD API
 – Apache Spark 95
 Recall 214
 Receiver Operating Characteristic 215

Recommendation Engines 521
 Recommender Engine 26
 Regelbasierte Entscheidungsfindung 265
 Regierung 515
 Registrierung 372
 Regressionsmodelle 198, 221, 508
 Regularisierung 276
 Regulierung 568 ff.
 Reinforcement Learning (RL) 242, 407, 414
 Relevancy 156
 ReLu 273, 290
 Remote Work 561
 Reproduzierbarkeit
 – Dokumentation 384, 390
 – transparent 384
 – Verifizierung und Validierung 384
 – Visualisierung 384
 ResNets 249
 REST API 81
 reward 243
 Ridge 276
 Risikoanalyst 457
 Risikofolgeabschätzung 480
 Rohstoffpreise 524
 Root Mean Squared Error 200
 Rosendiagramm 427

S

S3 70
 Sarbanes-Oxley-Gesetz 146
 Säulendiagramm 433
 Scale-out 37
 Scale-up 37
 Scheinbare Verallgemeinerung 290
 Schema Evolution 86
 Scrum 552
 Secret Managers 165
 – Ansible Vault 165
 – HashicorpVault 165
 Selbstfahrende Autos 366, 370
 Self-Service Analytics 472
 Sensitive Daten 476
 Sensitivitätsanalyse 394
 Sensitivity 214
 Sensorauflösung 355
 Sensordaten 500

Serverless 117
Serverless Computing
– AWS Lambda 72
– Azure Functions 72
– Google Cloud Functions 72
Shared-Nothing-Architektur 69
Shared Responsibility Model 159
Sharpe Ratio 277
Sicherheitsexperte 17
Siebdiagramm 428
SIFT 363
sigmoid 273
Silicon Valley 553
similarity function 239
Singular Value Decomposition (SVD) 178
Singulärwert-Zerlegung 178
Skalierende Intelligenz 281
Skalierte Korrelation 278
Skalierungsfalle 279
Skip-Verbindungen 253
Smart Cities 518
Snowflake 92
Snow, John 426
Software as a Service (SaaS) 64
SonarCube 124
Spaltenorientiertes Speicherformat 85
Spam-Erkennung 314
Specificity 214
Speicher 34
Speicherdienste 66
– Azure Storage 67
– Google Storage 67
Spezifikation 391
Spur 172
SSH Key 48
Stammdaten 156
Standardabweichung 192
Standards 568, 571
State 243
Strategie 11, 24
Strenges Modell 273
Streudiagramm 426, 432
Structured Streaming
– Apache Spark 99
Subject Matter Expert (SME *siehe* Fachexperte)
Supply Chain 24
Support Vector Machine (SVM) 274

Symbolisch
– Beschreibung 381
Symbolische KI 267
Systemdynamik 397
– Ebene 398
– Flow 398
– hypothetische Beziehungen 397
– Kausalschleifendiagramm 399
– Top-down-Ansatz 400
Szenarien 387

T

Tagged Image File Format *siehe* TIFF
t-closeness 162
Telekommunikationsanbieter 534
Terraform 59 f.
Testdaten 272
Tests der Datenvalidität 393
Teufelkreis 297
Theatrum Orbis Terrarum 424
TIFF 360
Timeliness 156
Time Travel 86
tmux 47
Trainingsaufwand 274
Transferlernen 248
Transparenz 566 ff.
– Dokumentation 582
– Gesetze 569
– KI-Systeme 581
Transport Layer Security (TLS) 164
Triangulation 424
Trustworthy AI 565
Tuft 428
Turing, Alan 280

U

Überanpassung 231, 272
Unabhängigkeit der Stichproben 292
Underfitting 231
U-Netze 253
Uniqueness 156
Unteranpassung 231

V

Validierung 391
 – Gesichter 393
 – Theorien testen 393
 Validity 156
 van Langren 424
 Varianz 192, 230 f.
 Variety 18
 Vektoren 169
 Vektormultiplikation 174
 Velocity 18
 Vendor Lock-in 467, 496
 Veracity 18
 Verallgemeinerungsfälle 289
 Verborgene Zusammenhänge 423
 Verfügbarkeit 32
 Verifizierung 391
 – Cross-Check 392
 – doppelte Implementierung 393
 – formale Methoden 393
 – strukturelle Analyse 392
 – strukturierte Code-Walk-troughs 392
 – Unit Testing 393
 Verkehrsanalyse 517
 Verlust 283
 Verschlüsselung 163
 Vertrauenswürdige KI 567 f.
 – Gesellschaft 572
 Verzeichnisdienste 70
 – Active Directory 70
 – Google Cloud Identity 70
 Verzerrung 230
 Verzerrungen *siehe* Bias
 Vim 55
 Vision Cycle 376
 Visualisierung
 – Datenanalyse 391
 – Modellierungsstruktur 391

Visuell 423
 Visuelles System 350
 Volume 18
 Von-Neumann-Architektur 37
 Vorhersagen 220, 430
 Vorurteil *siehe* Bias

W

Wahrscheinlichkeitsdichtefunktion 190
 Wahrscheinlichkeits-Masse-Funktion 188
 Wahrscheinlichkeitsrechnung 187
 Watson AI 264
 Weak Learner 230, 233
 Werte
 – endogene 382
 – exogene 382

X

XaaS 64
 XML 83

Y

YAGNI 36

Z

Zentraler Grenzwertsatz 191
 Zentrale Organisation 458
 Zielvariable 221
 zsh 43
 Zufallsvariablen
 – diskrete 188
 – kontinuierliche 190
 – unabhängige 193
 Zuständigkeitskonflikte 452