



Bernhard G. Humm, Peter Buxmann und Jan C. Schmidt

2.1 Historie von KI

Der Begriff „Artificial Intelligence“ wurde auf dem Dartmouth Workshop 1956 geprägt. Teilnehmer dieses Workshops waren Pioniere der KI, u. a. John McCarthy, Marvin Minski, Allen Newell und Herbert Simon. Im Workshop Proposal wurde folgende Vermutung formuliert: „[E]very 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“ (McCarthy et al. 1955).

Die Grundlagen der KI wurden jedoch schon deutlich früher gelegt, vor allem durch Alan Turing (1912–1954) mit seinen grundlegenden Arbeiten über Berechenbarkeit und dem so genannten „Turing-Test“ zur Überprüfung des intelligenten Verhaltens einer KI-Anwendung.

Dieses Kapitel entstand durch intensive Diskussion und Zusammenarbeit mit der interdisziplinären Projektgruppe „Digitale Arbeitswelten“, der Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt und Indra Spiecker genannt Döhmann angehörten.

B. G. Humm (✉)

Hochschule Darmstadt, Darmstadt, Deutschland

E-Mail: bernhard.humm@h-da.de

P. Buxmann

Fachgebiet Wirtschaftsinformatik, Technische Universität Darmstadt, Darmstadt, Deutschland

E-Mail: buxmann@is.tu-darmstadt.de

J. C. Schmidt

Department of Social Sciences, Hochschule Darmstadt, Darmstadt, Deutschland

E-Mail: Jan.Schmidt@h-da.de

In den 1960–1980er-Jahren folgte ein KI-Hype ohnegleichen mit enthusiastischen Versprechungen von schnellen und fantastischen KI-Erfolgen. Dieser Hype führte zu enormen finanziellen Förderungen der KI-Forschung, vor allem in den USA. Es passierte, was passieren muss, wenn unrealistische Erwartungen aufgeblasen und Versprechungen maßlos überzogen werden: die Blase platzte. Bildverarbeitungssysteme für autonomes Fahren konnten nicht zwischen geöffneten Kanaldeckeln und Schatten unterscheiden; Logik-Anwendungen konnten in Punkto Allgemeinwissen nicht mit kleinen Kindern mithalten. Es folgte eine Desillusionierung, verbunden mit einem Kollaps des KI-Markts. Diese Zeit zwischen den späten 1980er- und frühen 2000er-Jahren wird manchmal auch als „KI-Winter“ bezeichnet.

Allerdings wurden in dieser Zeit KI-Techniken kontinuierlich weiterentwickelt, häufig ohne Nennung des damals diskreditierten Begriffs KI. Vor allem die führenden IT-Unternehmen waren hier wegbereitend. Beispielsweise sagte der Google-Mitbegründer Larry Page 2006: „We want to create the ultimate search engine that can understand anything. Some people could call that artificial intelligence“ (Sarathi 2014) KI gewann dadurch implizit enorm an Bedeutung und wurde allgegenwärtiger Bestandteil des täglichen Lebens: Navigationsgeräte mit Sprach-Ein- und Ausgabe, digitale Assistenten auf Smartphones, Bilderkennung in Kameras, KI-Module in Computerspielen etc. Erst in den letzten Jahren wurde dafür auch wieder der Begriff KI verwendet.

Heute sehen wir einen ähnlichen Hype wie in den 1970er-Jahren. Wieder gibt es überzogene Erwartungen bzw. Befürchtungen. Beispielsweise prognostiziert Nick Bostrom in seinem Buch „Superintelligenz: Szenarien einer kommenden Revolution“ (Bostrom 2014), dass innerhalb eines Menschenlebens so genannte Superintelligenzen, KI-Systeme, die der menschlichen Intelligenz weit überlegen sind, entstehen, was das Ende der Menschheit bedeuten könnte. Erwartungen von KI-Systemen, die sich selbst, unabhängig vom Menschen außerhalb ihres ursprünglichen Anwendungszwecks weiterentwickeln (wie z. B. Bostrom es postuliert) werden auch als *starke KI* bezeichnet.

Ob es jemals eine solch starke KI geben kann, ist eine spekulative Frage, die niemand mit Sicherheit beantworten kann. Bis heute sind keine Ansätze für starke KI bekannt. Stand heute gibt es also keine Evidenz dafür, dass starke KI überhaupt jemals möglich sein wird. Die allerwenigsten KI-Forscher/-Forscherinnen und -Praktiker/-Praktikerinnen rechnen auf absehbare Zeit mit einer solchen. Allerdings nehmen Vorstellungen von einer starken KI in der öffentlichen Debatte einen großen Raum ein – ähnlich zum früheren KI-Hype. Es ist durchaus möglich, dass sich hier die Geschichte wiederholt und eine folgende Phase der Desillusionierung zu einem neuen KI-Winter führt. Wir empfehlen die Lektüre der fundierten Diskussion des Robotik-Pioniers Rodney Brooks zu Prognosen über die Zukunft der KI (Brooks 2017), sowie den Blogeintrag von Bernhard Humm zu diesem Thema (Humm 2017).

In dieser Studie verwenden wir nicht den Ausdruck „Künstliche Intelligenzen“, da er eine Personifizierung der Maschine suggeriert. Stattdessen sprechen wir von *KI-Anwendungen* („Apps“), besonders wenn sie auf Standard-Hardware wie PCs oder Smartphones laufen. Wir verwenden auch den Ausdruck *KI-Systeme*, wenn die Hardware inhärenter Systembestandteil ist, wie z. B. bei Robotern.

Alle KI-Anwendungen, die heute existieren, werden der *schwachen KI* zugeordnet. Solche KI-Anwendungen sind von Ingenieuren speziell auf einen Anwendungszweck hin entwickelt worden, wie z. B. Gesichter erkennen, Schach spie-

len, autonom fahren, etc. Auch wenn Verfahren des Maschinellen Lernens eingesetzt werden, ändert sich der Anwendungszweck nicht, solange Ingenieure nicht die Software anpassen. Ein selbstfahrendes Auto wird niemals von sich aus das Schachspiel lernen und ein Schachprogramm wird niemals von sich aus Gesichter erkennen.

Solche maßgeschneiderten KI-Anwendungen haben mittlerweile eine hohe Verbreitung im Privat- und Geschäftsleben. Nachfolgend einige Beispiele:

- Automatische Sprach-Eingabe und -Ausgabe in Autos, Mobiltelefonen, Helpdesks etc.
- Gesichtserkennung in Kameras, Fotoarchiven, Flughäfen, etc.
- Spam-Klassifikation mittels Maschinellern Lernen in E-Mail-Programmen
- KI-Komponenten in Computerspielen
- Semantische Suche im Internet
- Maschinelle Übersetzung zwischen Sprachen wie Deutsch und Englisch als Web-Service
- Business-Intelligence-Auswertungen von Geschäftsdaten
- Sentiment-Analyse von Einträgen in sozialen Medien, z. B. für bestimmte Produkte
- Roboter-Anwendungen, z. B. in der industriellen Fertigung
- Anwendungen für Militär, Luft- und Raumfahrt, z. B. Drohnen und autonome Fahrzeuge wie Mars Rover.

Diese Anwendungen haben oft eine erstaunliche Leistungsfähigkeit, die manchmal sogar die Fähigkeit menschlicher Experten übertrifft. Das gilt nicht nur für Spiele wie Schach oder Go. Auch in der Medizin übertreffen in eingeschränkten Bereichen bildgebende KI-Verfahren zur Tumorerkennung die Vorhersagegenauigkeit von Medizinern.

Allerdings gibt es auch Negativbeispiele. So können neuronale Netze zur Bildverarbeitung durch leicht modifizierte Bilder fehlgeleitet werden. Solche Modifikationen, für den Menschen kaum zu erkennen, provozieren Falschklassifikationen. Wird beispielsweise ein Stoppschild so manipuliert, dass ein Vorfahrtsschild erkannt wird, dann kann dies fatale Auswirkungen bei selbstfahrenden Autos haben. Ebenso kann eine Verzerrung (Bias) in Trainingsdaten, z. B. aufgrund von Diskriminierungen in der Vergangenheit, eine solche Diskriminierung für die Zukunft fortschreiben. Solche Beobachtungen wurden in Machine-Learning-Anwendungen für Arbeitsämter oder im Rechtswesen gemacht. Die scheinbare Objektivität von Daten kann in diesem Zusammenhang trügen.

2.2 Grundlagen der KI

2.2.1 *Die Landkarte der KI*

In Kap. 1 haben wir Künstliche Intelligenz definiert als die Fähigkeit von Computersystemen, Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern. Als Beispiele haben wir genannt: Wahrnehmen (Sehen, Hören, Fühlen

etc.), Lernen, Wissen, Denken, Kommunizieren (Sprechen, Schreiben etc.) und Handeln. Wir verwenden nun diese Aufgaben und Fähigkeiten, um die Teilgebiete der KI strukturieren. Siehe hierzu eine Landkarte der KI in der folgenden Abb. 2.1.

Bewusst haben wir die Fähigkeiten in Anführungszeichen gesetzt, um nicht zu suggerieren, dass KI-Anwendungen wirklich lernen, denken, wissen und handeln können wie Menschen. Dennoch kann man gewisse funktionale Äquivalente von KI-Funktionen und menschlichen Fähigkeiten sehen, die die sechs Begriffe rechtfertigen. Dargestellte Überlappungen zwischen Kreisen sollen illustrieren, dass diese Fähigkeiten miteinander in Beziehung stehen.

- **„Wahrnehmen“:** Das KI-Teilgebiet *Computer Vision* beschäftigt sich mit der Verarbeitung von Bildern. Ein Anwendungsbeispiel ist die automatische Erkennung von Krankheiten wie Krebs mittels bildgebender Verfahren wie MRT, CTG, Ultraschall oder Röntgen.

- **„Kommunizieren“:** *Natural Language Processing* (NLP) behandelt die Verarbeitung natürlicher Sprache. Beispiel ist die Spracherkennung und Sprachgenerierung in digitalen Assistenten wie Apple Siri oder Amazon Alexa.

Information Retrieval (Informationsgewinnung) umfasst die Suche nach Dokumenten basierend auf einem Informationsbedarf. Anwendungen sind Suchmaschinen wie Google, Yahoo und Bing.

- **„Lernen“:** *Machine Learning* (Maschinelles Lernen) umfasst mathematische Verfahren zur Verallgemeinerung von beobachteten Beispielen. Es wird verwendet, um Vorhersagen zu treffen. Ein Beispiel ist die Klassifikation von Emails als Spam.

Data Mining umfasst mathematische Verfahren zum Erkennen von Mustern in (großen) Datenmengen. Ein Beispiel ist die Gruppierung von Kundendaten.

- **„Wissen“:** *Knowledge Representation* ist das KI-Teilgebiet, welches sich mit der expliziten Modellierung von Wissen beschäftigt. Ein Beispiel ist der Google Knowledge Graph, die Informationsquelle für Infoboxen, welche passend zu Google Suchanfragen angezeigt werden.

- **„Denken“:** *Logische Programmierung* umfasst spezielle Programmiersprachen, mit denen komplexe Regelwerke implementiert werden können, beispielsweise für die Versicherungsbranche.

Probabilistic Reasoning befasst sich mit Schlussfolgern bei unsicherem Wissen, wie dies beispielsweise in der Robotik nötig ist.

Complex Event Processing befasst sich mit der Verarbeitung von fortlaufenden Ereignissen, z. B. Kreditkartenbuchungen. Dabei sollen komplexe Muster in Ereignisfolgen erkannt werden, z. B. zur Betrugserkennung.

- **„Handeln“:** Das KI-Teilgebiet *Planen* beschäftigt sich mit der Optimierung von Handlungsschritten zur Erfüllung von Zielen, z. B. in der industriellen Fertigung.

Agenten-Technologie wird eingesetzt bei komplexen KI-Anwendungen wie autonomem Fahren, die viele KI-Gebiete umfassen. Intelligente Agenten verbinden Komponenten für Computer Vision, Natural Language Processing, Machine Learning, Knowledge Representation etc.

Bei der *Robotik* geht es um die Entwicklung von stationären oder mobilen Robotern sowie deren Programmierung.

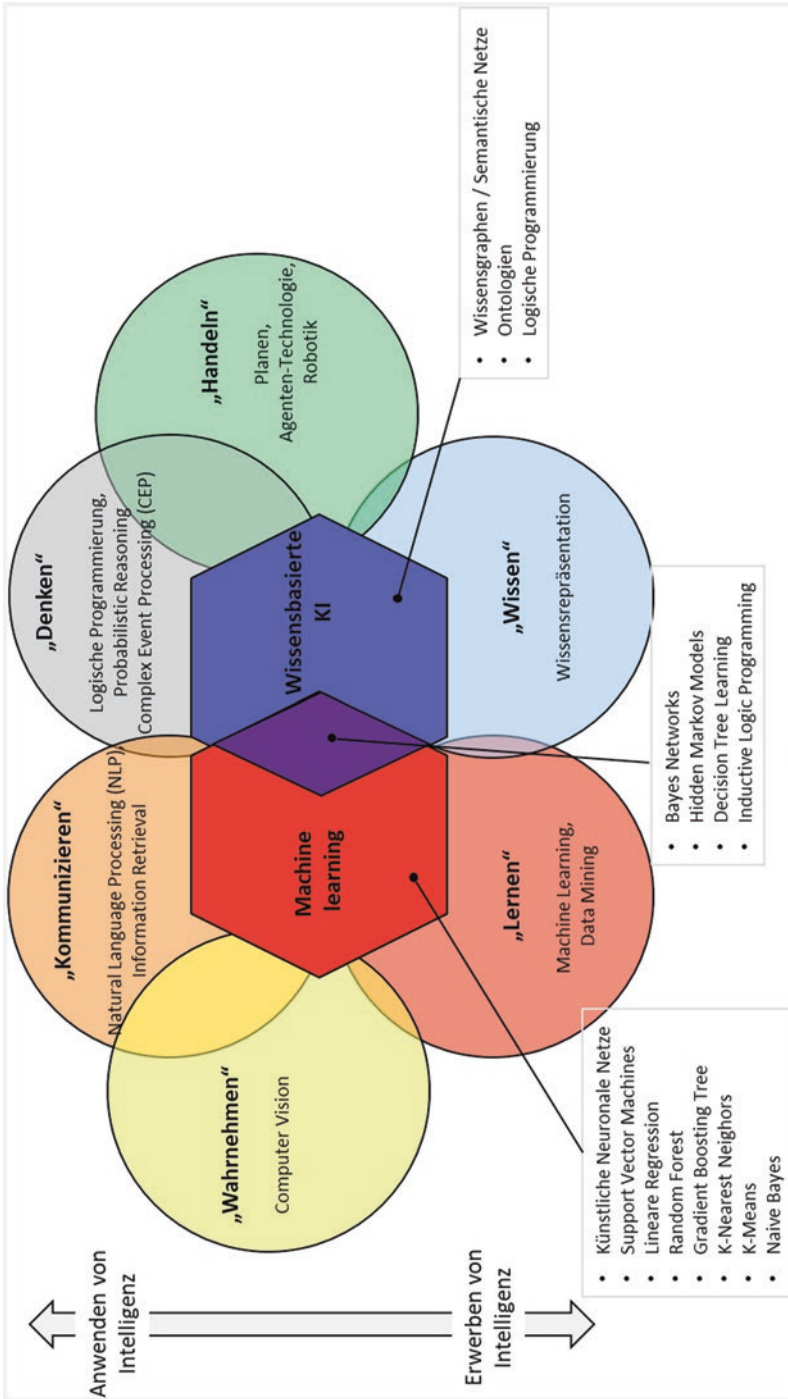


Abb. 2.1 Landkarte der KI (nach Humm 2020)

Die von uns als Teilgebiete der KI dargestellten Aspekte werden in der Fachliteratur nicht einheitlich bezeichnet. Was KI ist, ist – wie in jedem jungen und dynamischen Forschungsfeld – nicht ad hoc eineindeutig definierbar.

Im Laufe der KI-Geschichte haben sich zwei Hauptströmungen von *KI-Verfahren* herausgebildet, welche in der Grafik als Sechsecke dargestellt sind:

- **Wissensbasierte KI:** Hier wird Wissen von menschlichen Experten explizit modelliert in einer Form, die sowohl von Menschen gelesen und verstanden als auch von Computern interpretiert werden kann. Bekannte Verfahren sind *Wissensgraphen / Semantische Netze*, *Ontologien* und *Logische Programmiersprachen* (in der Grafik als Notizen in Rechtecken dargestellt).
- **Machine Learning:** Hier werden Aussagen anhand von numerischen, z. B. statistischen Verfahren berechnet. Häufig kann eine hohe Vorhersagegenauigkeit erzielt werden, aber die Ergebnisse sind für Menschen in der Regel nicht erklärbar.¹ Verfahren sind *Künstliche Neuronale Netze*, *Support Vector Machines*, *Lineare Regression* etc.

Hybride Verfahren vereinen die Vorteile von wissensbasierter KI und Machine Learning. Sie sind in der Grafik durch die Schnittmenge zwischen den beiden Sechsecken dargestellt (violette Raute). Sie umfassen Verfahren wie *Bayes Networks*, *Hidden Markov Models*, *Decision Tree Learning* etc.

In allen Teilgebieten der KI kommt wissensbasierte KI und Machine Learning zum Einsatz. Allerdings überwiegen in den KI-Teilgebieten für die Fähigkeiten Wahrnehmen, Kommunizieren und Lernen die Machine Learning Verfahren, während in den KI-Teilgebieten für Wissen, Denken und Handeln die wissensbasierten Verfahren meist bevorzugt werden.

In den nächsten beiden Abschnitten stellen wir die beiden KI-Hauptströmungen, Machine Learning und wissensbasierte KI, kurz vor.

2.2.2 *Machine Learning*

Dass das Interesse an der Entwicklung und Nutzung von KI in den letzten Jahren stark zugenommen hat, liegt insbesondere an den Fortschritten, die im Bereich des Machine Learning (ML) erzielt wurden. Im Allgemeinen umfasst Machine Learning Methoden, die Zusammenhänge in Datensätzen erkennen, um darauf basierend Vorhersagen zu treffen (Murphy 2012). Die fast universelle Anwendbarkeit von ML-Algorithmen macht die Nutzung interessant für Wirtschaft und Wissenschaft. Ebenfalls kann KI einen Beitrag zur Lösung gesellschaftlicher Herausforderungen leisten. So kann ML die Diagnose von Krankheiten unterstützen,

¹ Gängige ML-Verfahren wie künstliche neuronale Netze arbeiten assoziativ, nicht kausal. Das bedeutet, dass Korrelationen erkannt, aber nicht begründet werden können. Derzeit gibt es intensive Bestrebungen, auch ML-Verfahren mit Erklärungskomponenten anzureichern. Dies wird auch als erklärbare KI (explainable AI) bezeichnet.

Therapievorschläge entwickeln oder bei der Entwicklung von Medikamenten helfen. Daher sprechen Erik Brynjolfsson und Andrew McAfee vom MIT auch von der wichtigsten „general purpose technology“ unseres Zeitalters (Brynjolfsson und McAfee 2017).

Im Laufe der Aktivitäten der Arbeitsgruppe wurde von eingeladenen Gästen häufig das Argument vorgebracht, dass Künstliche Intelligenz bzw. Machine Learning nichts Neues seien, da viele der auch heute genutzten Methoden, wie z. B. Künstliche Neuronale Netze oder Support Vector Machines, schon seit Jahrzehnten existieren. Aus diesen Gründen den Schluss zu ziehen, dass es sich nicht lohnt, sich mit KI oder ML zu beschäftigen, wäre aber ein Fehler. Es gibt eine Vielzahl von Beispielen, die zeigen, dass es häufig an den Rahmenbedingungen liegt, ob sich eine Technologie durchsetzt oder nicht – und dass es nicht die Technologie „an sich“ ist, die zu betrachten ist. Das prominenteste Beispiel ist vielleicht das Internet.

In den letzten Jahren haben sich die Rahmenbedingungen für die Nutzung von KI- bzw. ML-Algorithmen stark geändert. Diese Rahmenbedingungen unterstützen dabei die Durchsetzung und Verbreitung dieser Typen von Algorithmen. Insbesondere sind die folgenden vier Punkte zu nennen:

- Erstens sind Daten heute in einer nie gekannten Menge verfügbar – sowohl im Internet als auch in Unternehmen. Diese Daten sind die Grundlage für den Einsatz von ML-Verfahren. Darüber hinaus existieren Datenplattformen, wie z. B. Kaggle, die Daten für eine Vielzahl von möglichen KI-Anwendungen bereitstellen.
- Zweitens sind Rechenleistung und Speicherplatz so kostengünstig wie noch nie zuvor und können von Cloud-Anbietern problemlos bezogen werden. Möglicherweise werden Entwicklungen im Bereich des Quanten-Computing in der Zukunft weiteren rechenintensiven KI-Anwendungen zum Durchbruch verhelfen.
- Drittens hat sich die Performance von ML-Algorithmen in den letzten Jahren verbessert. Insbesondere gilt das für die Technik des Deep Learning, welche mehr und mehr Anwendung findet.
- Viertens existieren viele kostenlos verfügbare Toolkits und Bibliotheken zur Entwicklung von KI-Anwendungen. Beispiele sind Scikit-learn, Apache Spark MLlib, Keras, CNTK, PyTorch oder TensorFlow (siehe hierzu auch Buxmann und Schmidt 2019). Die meisten dieser Werkzeuge stehen unter einer Open-Source-Lizenz. Diese Werkzeuge erleichtern die Entwicklung von ML-Anwendungen enorm. So können mit Hilfe von Tools wie Tensorflow oder Scikit-learn beispielsweise Methoden wie Künstliche Neuronale Netze vergleichsweise einfach in eine Software eingebunden werden. Darüber hinaus existieren Frameworks, wie RapidMiner, die den gesamten Entwicklungsprozess, einschließlich Aufgaben wie die Modellierung sowie die Aufarbeitung, Bereinigung und Visualisierung von Daten, unterstützen.

Zusätzlich haben sich die Möglichkeiten zur Nutzung von ML-Algorithmen auch dadurch vereinfacht, dass Anbieter wie beispielsweise Google, IBM, Microsoft oder Amazon mittlerweile KI-Services auf Basis eines Pay-per-Use-Zahlungsmodells anbieten. Das bedeutet, Anwender können Dienste, wie z. B. zur Umwandlung von Stimme in Text oder zur Erkennung von Objekten über ein

Software-as-a-Service-Modell beziehen. Es entwickeln sich also Geschäftsmodelle rund um den ML-Einsatz, was die Nutzung und Verbreitung zukünftig weiter befördern wird.

Methoden zur Messung der Qualität von ML-Algorithmen

Natürlich führt die Nutzung von ML-Algorithmen nicht immer zu guten Ergebnissen. Die Qualität ist abhängig von dem gewählten Verfahren sowie der Datenbasis, d. h. der Trainingsdaten. Vor diesem Hintergrund stellt sich die Frage nach der Anwendung von Verfahren zur Messung der Qualität. Das Grundprinzip dieser Messung soll am Beispiel von Klassifikationsproblemen erläutert werden. Hierzu zählen beispielsweise Objekterkennung auf Bildern (Dalal und Triggs 2005) oder medizinische Diagnosen (Kononenko 2001). Um zu prüfen, wie gut ML-Lösungen für solche Probleme funktionieren, kann die so genannte *Confusion Matrix* aufgestellt werden.

Dies lässt sich am einfachsten anhand eines Beispiels erklären (Buxmann und Schmidt 2021): Es soll eine trainierte ML-basierte Anwendung evaluiert werden, die auf Grundlage von Gesundheitsdaten ausgibt, ob ein Patient an Krebs erkrankt ist oder nicht. Die Anwendung gibt folglich an, in welche der Klassen – „krebskrank“ versus „nicht krebskrank“ – ein Patient eingeordnet wird. Um zu verstehen, wie gut diese Klassifizierung funktioniert, werden echte Diagnosen als Vergleichswerte herangezogen, in denen vermerkt ist, ob ein Patient wirklich an Krebs erkrankt ist oder nicht. Es stehen folglich für jeden Patienten des Testdatensatzes jeweils die tatsächliche Klasse und die von der Anwendung vorhergesagte Klasse zur Verfügung. Werden anschließend die beiden Klassen eines Patienten verglichen, liegt stets einer der folgenden vier Fälle vor (Fawcett 2006):

- Richtig-Positiv (RP): Der Patient ist tatsächlich krebserkrankt und wird auch so von der Anwendung klassifiziert.
- Richtig-Negativ (RN): Der Patient ist tatsächlich nicht krebserkrankt und wird auch so von der Anwendung klassifiziert.
- Falsch-Positiv (FP): Der Patient ist tatsächlich nicht krebserkrankt, aber die Anwendung klassifiziert den Patienten fälschlicherweise als krebserkrankt.
- Falsch-Negativ (FN): Der Patient ist tatsächlich krebserkrankt, aber die Anwendung klassifiziert den Patienten fälschlicherweise als nicht krebserkrankt.

Anschließend wird gezählt, wie oft diese vier Fälle in den klassifizierten Testdaten vorkommen, um die Confusion Matrix zu erstellen. Abb. 2.2 a) zeigt den allgemeinen Aufbau einer Confusion Matrix. Sie besteht aus den Dimensionen „tatsächlicher Wert“ und „vorhergesagter Wert“, die die Daten in die obigen vier Fälle unterteilen. Jede der vier resultierenden Zellen spiegelt die Anzahl des jeweiligen Falls wider. Durch das Ablesen der vier Werte kann grundlegend eingeschätzt werden, wie gut die Anwendung Klassen richtig zuordnen kann (Fawcett 2006). Aus Abb. 2.2 b) wird beispielsweise ersichtlich, dass das entsprechende Modell relativ viele Patienten als krebserkrankt klassifiziert hat, die tatsächlich nicht krebserkrank sind (150 Patienten wurden richtigerweise als krebserkrank klassifiziert, während 60 fälschlicherweise als krebserkrank klassifiziert wurden). Dafür wird aber auch er-

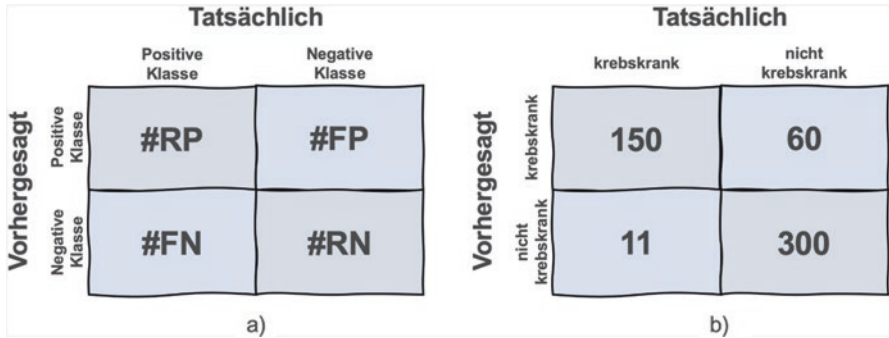


Abb. 2.2 (a) Allgemeiner Aufbau einer Confusion Matrix; (b) beispielhafte Confusion Matrix. (Quelle: Buxmann und Schmidt 2021)

sichtlich, dass das Modell nicht krebskranke Patienten relativ gut klassifizieren kann (300 Patienten wurden richtigerweise als nicht krebskrank klassifiziert während nur 11 fälschlicherweise als nicht krebskrank klassifiziert wurden).

Die Confusion Matrix kann also eingesetzt werden, um einen Einblick in die Verteilung von tatsächlichen und vorhergesagten Werten zu erhalten. Unternehmen können die Informationen nicht nur nutzen, um zu verstehen, wie viele Fehler eine Anwendung insgesamt produziert, sondern vor allem, um potenzielle Fehlerarten steuern zu können. So können je nach Relevanz ML-Algorithmen während der Entwicklung so parametrisiert werden, dass eine Art der potenziellen Fehler reduziert wird, was jedoch gewöhnlich zu einer Steigerung der anderen Fehlerart führen wird. Wenn dem Unternehmen beispielsweise wichtiger ist, dass wenige FP erzeugt werden, dann kann der Algorithmus stärker auf diese Fehlerart fokussiert werden, was jedoch zu einem erhöhten Aufkommen von FN führen kann.

Eine Confusion Matrix repräsentiert allerdings noch keinen konkreten Zielwert, auf dessen Grundlage eine ML-Lösung optimiert werden kann. Basierend auf den Werten der Confusion Matrix können konkretere Qualitätsmaße aufgestellt werden. Hierzu zählen vor allem Accuracy, Precision, Recall, F1 Score sowie die ROC-Curve, welche unterschiedliche Vor- und Nachteile mit sich bringen (Fawcett 2006; Powers 2011).

Bias im Machine Learning

Bei der Nutzung von ML-Anwendungen ist zu berücksichtigen, dass die Ergebnisse verzerrt sein können. Diese Verzerrungen sind in der Regel auf die verwendeten Trainingsdaten zurückzuführen und können in Label-, Feature-, und Sample Bias unterschieden werden (Buxmann und Peters 2020).

Diese Verzerrungen können beispielhaft anhand eines Systems für die Vorhersage von Rückfallkriminalität veranschaulicht werden. Um die Wahrscheinlichkeit vorherzusagen, dass ein Strafgefangener rückfällig wird, werden z. B. in den USA intelligente Systeme eingesetzt, deren Einschätzungen auf historischen Verhaftungsstatistiken beruhen (Corbett-Davies und Goel 2018). Das System analysiert die Merkmale des Strafgefangenen (eingehende Variablen), z. B. demografische Daten

oder seine kriminelle Vergangenheit (z. B. historische Verhaftungsdaten) und trifft dann eine Vorhersage darüber, wie hoch das Risiko eines kriminellen Rückfalls ist (Zielvariable). Auf Basis dieser Vorhersage kann dann beispielsweise die Entscheidung getroffen werden, ob der Gefangene auf Bewährung freigelassen wird.

Beim Label Bias treten Verzerrungen in der vorherzusagenden Zielgröße auf. Oftmals ist die gewählte, beobachtbare Zielgröße eine fehlerhafte Näherungsvariable für die tatsächlich gewünschte Zielgröße. Im oben genannten Beispiel ist die tatsächlich gewünschte Zielgröße eine Risikoeinschätzung über die Durchführung einer erneuten Straftat im Falle einer Freilassung. Da die Vorhersage auf historischen Verhaftungsdaten beruht, wird jedoch implizit die Wahrscheinlichkeit einer erneuten Verhaftung im Falle der Freilassung prognostiziert statt dem Risiko über die Durchführung einer neuen Straftat. In Studien wurde jedoch gezeigt, dass Verhaftungsstatistiken für Drogenkriminalität verzerrt sein können, beispielsweise zum Nachteil von Minderheiten wie Hispano- und Afroamerikanern (Lum und Isaac 2016), sodass die tatsächliche Kriminalitätsrate nicht ausreichend genau abgebildet wird. Die Verzerrungen beruhen beispielsweise darauf, dass Stadtviertel, in denen bestimmte Minderheiten stärker vertreten sind, von der Polizei stärker überwacht werden und somit die Wahrscheinlichkeit der Verhaftung für diese Minderheiten steigt.

Neben dem Label Bias kann im gewählten Beispiel auch ein Feature Bias auftreten. Dieser liegt beispielsweise dann vor, wenn eine im Trainingsdatensatz enthaltene Variable unterschiedliche Vorhersagekraft für bestimmte Bevölkerungsgruppen hat. Die Entwickler des Systems aus dem obigen Beispiel könnten auf Basis allgemeiner Kriminalitätsstatistiken beispielsweise annehmen, dass Langzeitarbeitslosigkeit die Wahrscheinlichkeit einer zukünftigen Straftat erhöht. Wird der Beschäftigungsstatus nun als Variable in das Modell aufgenommen, wird der Algorithmus diese scheinbar allgemeingültige Korrelation verwenden, um eine Vorhersage darüber zu treffen, wie hoch die Wahrscheinlichkeit einer erneuten Verhaftung von Strafgefangenen ist. Nehme man jetzt an, dass die getroffene Annahme für eine bestimmte Bevölkerungsminderheit, z. B. Hispanoamerikaner, in der Wirklichkeit nicht zutrifft. Das gelernte Modell wird die gelernte Korrelation jedoch trotzdem anwenden, arbeitslosen Hispanoamerikanern fälschlicherweise ein höheres Rückfallrisiko zuweisen und sie dadurch benachteiligen.

Der Sample Bias wird dadurch hervorgerufen, dass ungenügende Daten für das Training ausgewählt werden. Dieser Bias entsteht, wenn die Population in den Trainings-Daten die tatsächliche Population im Einsatzbereich nicht adäquat repräsentiert, sondern mindestens eine relevante Gruppe über- oder unterrepräsentiert ist (Heckman 1979). Im Falle des vorangegangenen Beispiels würde dies bedeuten, dass ein System zur Vorhersage möglicher Kriminalitätsrückfälligkeit von Strafgefangenen in einem spezifischen US-Bundesstaat trainiert wurde, jedoch in der gesamten USA eingesetzt wird. Hier liegt das Problem darin, dass sich die Demografie der einzelnen Bundesstaaten grundlegend voneinander unterscheiden kann. Letztendlich kann ein intelligentes System nur dann richtig funktionieren, wenn es zuvor genügend Daten zu allen Fällen bekommen hat, die es später klassifizieren soll.

Label-, Feature-, und Sample Bias werden zusammenfassend als Data Bias bezeichnet. Data Bias ist aber nur eine von mehreren Formen von Bias im Machine Learning (siehe Abb. 2.3).

Data Bias ist also durch Verzerrungen in den Daten verursacht. Diese können beispielsweise aus der unterschiedlichen Aktivität von Benutzergruppen resultieren (Activity Bias). So stammen die meisten verfügbaren Medizindaten aus westlichen Industrienationen, die jedoch nur einen kleinen Prozentsatz der Weltbevölkerung ausmachen. Damit verwandt ist der Sampling Bias, nämlich die Auswahl an existierenden Daten für das Trainieren von ML Anwendungen. Algorithmic Bias repräsentiert Verzerrungen, die durch ML-Algorithmen hervorgerufen werden. Da alle ML-Verfahren Näherungsverfahren sind, kann es verzerrungsfreie ML-Algorithmen prinzipiell nicht geben. In der Interaktion zwischen ML-Anwendung und Anwender kann es Bias geben, besonders wenn die inhärente Unsicherheit von ML-Verfahren dem Anwender nicht deutlich gemacht wird und Prognosen der ML-Anwendung als Tatsachen suggeriert werden. Schließlich kann die Nutzung von ML-Anwendungen neue Daten generieren, welche möglicherweise wiederum verzerrt sind. Im schlimmsten Fall kann dies zu einem Teufelskreis führen (Second Order Bias).

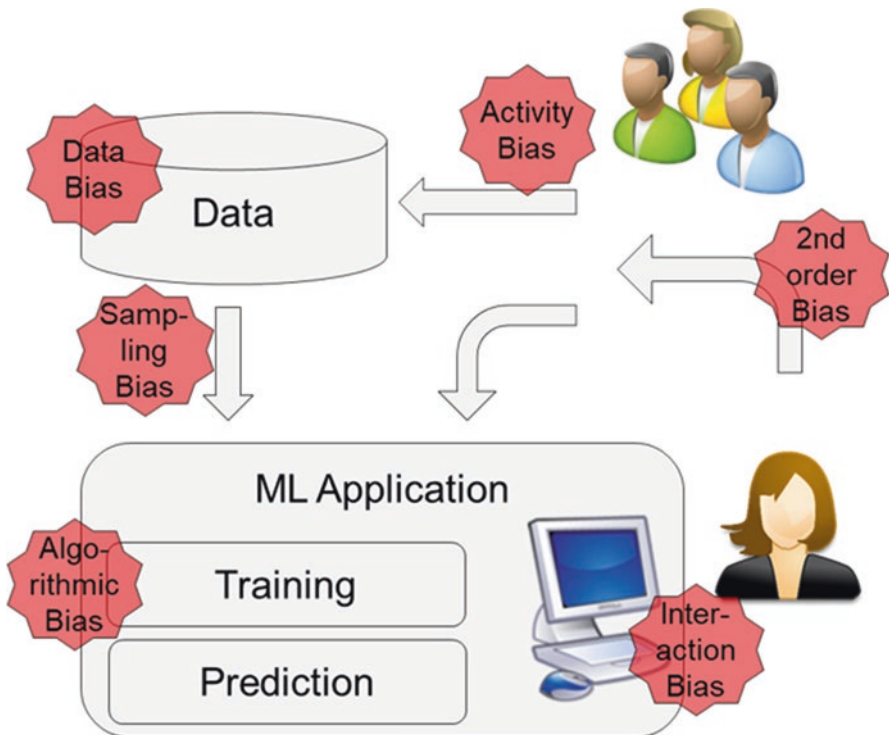


Abb. 2.3 Bias im Machine Learning (aus Humm 2020)

2.2.3 Wissensbasierte KI

In der Populärliteratur wird KI heute manchmal mit Machine Learning gleichgesetzt. Das ist eine stark verkürzte Sicht, denn wissensbasierte KI hat eine ebenso große Bedeutung in Forschung und Praxis.

Expertensysteme

Erste praktische Relevanz für wissensbasierte KI hatten die so genannten Expertensysteme der 1970er- und 1980er-Jahre, noch lange bevor Machine Learning eine Bedeutung im Alltag erlangte. Noch heute unterstützen die Expertensysteme beispielsweise bei medizinischen Diagnosen oder der Fehlersuche in IT-Systemen. Ein Expertensystem löst für Menschen komplexe Probleme in einem begrenzten Fachgebiet. Es agiert quasi als Experte und unterstützt durch Handlungsempfehlungen. Die Problemlösungen und Handlungsempfehlungen leitet es aus einer Wissensbasis ab (Luber und Litzel 2019).

Expertensysteme kommen in vielen verschiedenen Bereichen und Fachgebieten zum Einsatz. Typische Anwendungen finden sich dort, wo es an Experten fehlt oder die vorhandenen Experten von der Auswertung großer Datenmengen entlastet werden sollen. Darüber hinaus erhöhen Expertensysteme durch die unmittelbare Bereitstellung von Lösungen die Sicherheit in kritischen Situationen oder verbessern die Qualität eines Produktes. In der Medizin unterstützen Expertensysteme bei Diagnosen oder der Auswertung von Röntgenaufnahmen. Ein weiterer Anwendungsbereich ist in der Chemie zu finden. Dort analysieren Expertensysteme die Struktur chemischer Verbindungen oder unterstützen bei organischen Synthesen. Ebenfalls Anwendungen für Expertensysteme sind geologische Erkundungen, die militärische Aufklärung, Erdölbohrungen, Erdbebenvorhersagen, die Umweltentwicklung oder die Überwachung und Steuerung von Kernreaktoren (Luber und Litzel 2019).

Die Wissensbasis repräsentiert das formalisierte Expertenwissen meist in Form von Wenn-dann-Regeln. Expertensysteme sind in der Lage, das aus Fakten und Regeln bestehende Wissen zu interpretieren und Schlussfolgerungen abzuleiten. Hierfür sind sie mit einer Inferenzmaschine ausgestattet, die entscheidet, wie und in welcher Reihenfolge oder Form die Regeln zur Lösung eines Problems herangezogen werden. Über eine sogenannte Erklärungsmaschine macht das Expertensystem das Zustandekommen der Problemlösungen und Handlungsempfehlungen dem Menschen gegenüber verständlich. Moderne Expertensysteme bestimmter Fachgebiete sind in ihrem Problemlösungsverhalten mit den Leistungen menschlicher Experten vergleichbar oder übertreffen sie in Teilbereichen sogar. Typische Aufgaben für Expertensysteme sind:

1. Interpretation von Daten durch den Vergleich von Soll- und Ist-Werten
2. Klassifikation von Ereignissen
3. Konfiguration komplexer Systeme unter Berücksichtigung verschiedener Bedingungen
4. Erkennen von Fehlerursachen und Reduzierung von Arbeitsfehlern
5. Beseitigung kritischer Zustände durch das Einleiten von Aktionen

6. Planung einer Folge von Aktionen zur Erreichung eines bestimmten Ziels
7. Dialogorientierte, fachspezifische Beratung von Menschen
8. Vorhersage von Ereignissen auf Basis bestimmter Geschehnisse.

Ein Expertensystem besteht u. a. aus folgenden Komponenten: eine Wissenserwerbskomponente, die Wissensbasis, eine Inferenzmaschine, eine Erklärungskomponente, sowie die Schnittstelle für den Benutzerdialog. Eine zentrale Rolle übernehmen die Wissensbasis und die Inferenzmaschine. In der Wissensbasis ist das komplette problembezogene Wissen gespeichert. Auf Basis dieses Wissens generiert die Inferenzmaschine mittels Schlussfolgerungen Problemlösungen. Sie verknüpft hierfür Fakten und Regeln und generiert daraus neues Wissen. Dieses präsentiert das System über den Benutzerdialog dem Anwender und erklärt ihr Zustandekommen mithilfe der Erklärungskomponente. Die Wissenserwerbskomponente gestattet die Erweiterung der Wissensbasis. Neue Fakten und Regeln lassen sich zum vorhandenen Wissen hinzufügen. Die Wissenserwerbskomponente kann darüber hinaus die Vollständigkeit und Konsistenz des gespeicherten Wissens prüfen (Luber und Litzel 2019).

Wissensnetze

Die Wissensbasen von wissensbasierten KI-Systemen werden häufig als sog. Wissensnetze (Knowledge Graph)² dargestellt. Wissensnetze spezifizieren formal die wesentlichen Konzepte eines Fachgebiets, sowie deren Beziehungen zueinander. Abb. 2.4 zeigt beispielhaft einen Ausschnitt aus einem Wissensnetz für die Medizin.

In diesem Wissensnetz-Ausschnitt wird dargestellt, dass Krebs und Allergien Krankheiten sind, sowie Chemotherapien und Immuntherapien Behandlungen sind. Des Weiteren wird dargestellt, dass Chemotherapien für Krebserkrankungen eingesetzt werden können, während Immuntherapien sowohl für Allergien als auch für Krebserkrankungen eingesetzt werden.

Neben solchen Wissensnetzen für spezielle Fachgebiete gibt es auch umfassende Wissensnetze. Mit ca. 70 Mrd. Fakten (Vincent 2016) ist eines der umfangreichsten

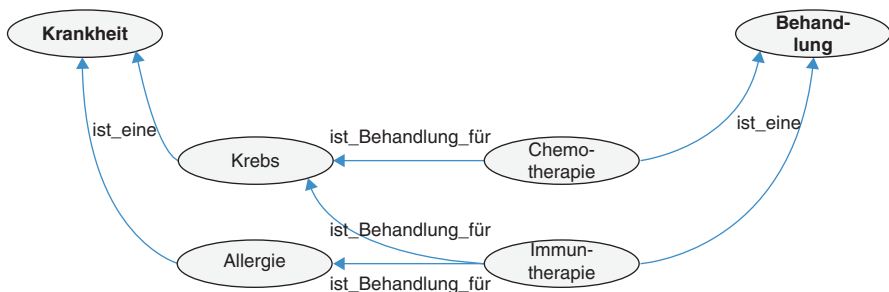


Abb. 2.4 Ausschnitt eines Wissensnetzes für die Medizin (aus Hoppe 2020)

²Für die genaue Abgrenzung von Knowledge Graph, Ontologie, Thesaurus, WordNet und kontrolliertem Vokabular siehe Hoppe 2020, Kap. 4.

und am meisten genutzten Wissensnetze der Google Knowledge Graph. Er ist die Grundlage für alle Info-Boxen, die bei Google-Suchen angezeigt werden. Er wird auch verwendet, um die Suche zu verbessern, sowie allgemeine Fragen zu beantworten. Versuchen Sie mal, in die Google-Suche eine Frage einzugeben wie z. B. „Wann wurde die Mauer gebaut?“; siehe Abb. 2.5 für einen Screenshot der Antwort.

Die Google-Suche assoziiert das Wort „Mauer“ mit der Berliner Mauer und gibt als Ergebnis der Anfrage ein konkretes Datum, den 13. August 1961 zurück. In einer Infobox auf der rechten Seite ist eine kurze Beschreibung der Berliner Mauer, inklusive wichtiger Daten wie der Baubeginn, sowie ein Link auf den Wikipedia-Eintrag angegeben. Die Wissensbasis für all diese Informationen ist der Google Knowledge Graph.

Linked Open Data

Neben kommerziellen Wissensnetzen wie dem Google Knowledge Graph, dem Facebook Graph oder dem Microsoft Graph, existiert eine große Menge freier Wissensnetze. Linked Open Data (LOD) ist eine Initiative des World Wide Web Consortiums (W3C), welche Organisationen Wissensnetze im Web zur freien Verfügung bereitstellt. Durch gegenseitige Verweise zwischen Wissensnetzen soll deren Nutzen noch gesteigert werden. Abb. 2.6 zeigt eine Übersicht über öffentlich verfügbare Wissensnetze, die sogenannte LOD Cloud.

Jeder Knoten in der LOD Cloud (repräsentiert durch einen Kreis in Abb. 2.6) beschreibt ein Wissensnetz, oft mit hunderttausenden Fakten. Verbindungen zwischen Knoten zeigen an, dass die entsprechenden Wissensnetze über Hyperlinks miteinander verbunden sind. Die Wissensnetze stammen aus so unterschiedlichen

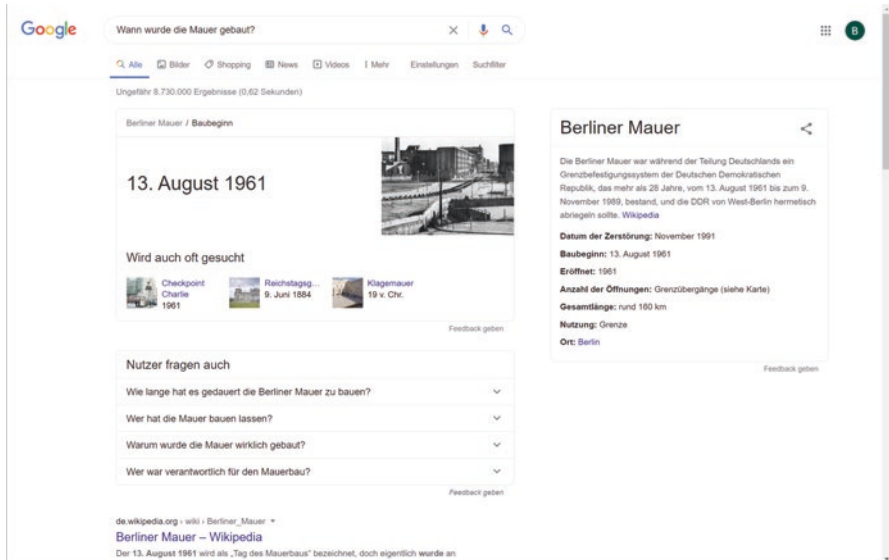


Abb. 2.5 Google Resultat auf die Frage „Wann wurde die Mauer gebaut?“ (Google 2020)

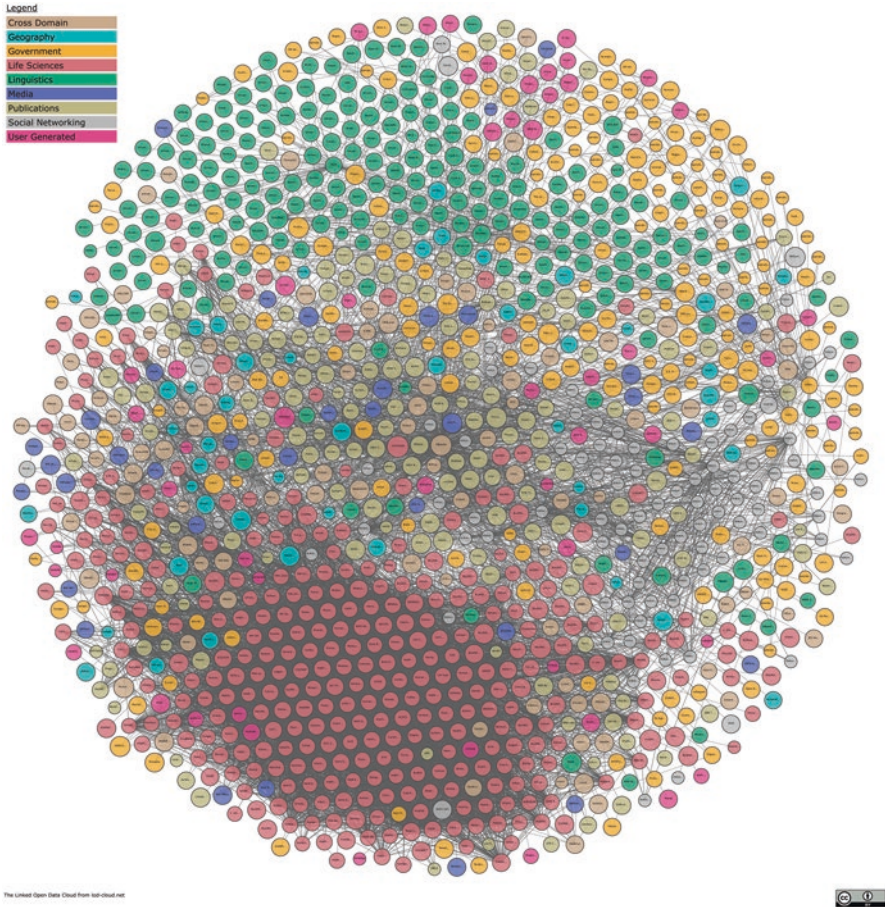


Abb. 2.6 Linked Open Data Cloud (Insight Centre for Data Analytics 2020)

Domänen wie Lebenswissenschaften, Geografie, öffentliche Verwaltung, Medien etc. oder sind domänen-übergreifend.

Besonders im Bereich der Lebenswissenschaften existieren viele, umfangreiche Ontologien zu Krankheiten, Symptomen, Körperteilen, Behandlungen, Medikamenten, Genen, u. v. m. Beispiele sind der NCI (NCI Thesaurus), MeSH (Medical Subject Headings) und die GO (Gene Ontology, Ashburner et al. 2000; The Gene Ontology Consortium 2019). Zahlreiche umfangreiche domänen-übergreifende Ontologien existieren ebenfalls, z. B. Wikidata, DBpedia oder YAGO. In diesen Ontologien sind Konzepte aus Wikipedia formalisiert, die Verwendung in domänen-übergreifenden KI-Anwendungen finden (Siehe auch Hoppe 2020, Kap. 4).

2.2.4 Zusammenspiel von wissensbasierter KI und Machine Learning

Wissensbasierte KI und Machine Learning ergänzen sich gegenseitig. Faktenwissen wird mit Wissensnetzen repräsentiert. Für Problemfelder, in denen keine gesicherte Faktenlage verfügbar ist, wo Unsicherheit und Rauschen vorliegt, wo Entscheidungen dynamisch an Nutzerverhalten angepasst werden soll, da spielt Machine Learning seine Stärken aus. Der Nachteil der meisten Machine-Learning-Verfahren ist jedoch, dass sie wie eine Black Box funktionieren und Vorhersagen meist nicht begründet werden können. Hier liegt wiederum die Stärke von wissensbasierten Verfahren, da Aussagen aufgrund der Wissensbasis begründet werden können. Jedoch ist die Erstellung und Pflege einer Wissensbasis durch Experten aufwändig.

Auch die KI unterliegt Moden und Strömungen. Während in den 1970–1990er-Jahren die wissensbasierte KI dominant war und Machine Learning eher eine Nischenrolle einnahm, hat sich das in den letzten beiden Jahrzehnten eher umgedreht. In der Populärliteratur wird sogar manchmal KI mit Machine Learning gleichgesetzt. Aber in letzter Zeit werden zunehmend auch Probleme mit Machine Learning in den Blick genommen, wie z. B. Bias (s. o.) oder die fehlende Erklärbarkeit von Entscheidungen.

So bildet sich aktuell ein neuer Forschungsschwerpunkt „Explainable AI“ (Erklärbare KI) heraus, in dem die Vorteile beider KI-Ansätze kombiniert werden sollen. Wir sehen solche hybriden Ansätze als zukunftsweisend an, weil sie die jeweiligen Stärken der beiden Ansätze verbindet, sowie deren Schwächen adressiert.

2.2.5 Die Rolle von Algorithmen in KI-Systemen

Neben der unpräzisen Gleichsetzung von KI und Machine Learning stellt auch die populäre Verwendung des Begriffs „Algorithmen“ für KI-Systeme eine verkürzte Sicht dar, die klärungsbedürftig ist: Ein Algorithmus ist eine Handlungsvorschrift zur Lösung eines Problems oder einer Klasse von Problemen, z. B. das Sortieren von Elementen einer Liste. Ein Algorithmus ist vergleichbar mit einem Kochrezept, in dem Schritt für Schritt beschrieben wird, wie ein Gericht hergestellt werden soll. Aber genauso wenig, wie ein Kochrezept das Gleiche wie ein Abendessen ist, ist ein Algorithmus das Gleiche wie ein KI-System. Ein Kochrezept macht nicht satt. Für ein Abendessen benötigt man Köche, welche viele Zutaten gemäß der einer Vielzahl von Kochrezepten mit Hilfe von Werkzeugen wie Messern, Töpfen, Backöfen und Herden und mit Einsatz von Energie zubereiten und anrichten. Erst dann kann gegessen werden.

Um ein KI-System zu entwickeln, verwenden Software-Entwickler sogenannte Software-Bibliotheken: konkrete Software-Implementierungen von hunderten und tausenden von Algorithmen mittels konkreter Programmiersprachen. Für eine konkrete KI-Anwendung wird eine Software-Architektur entwickelt: ein Diagramm,

welches beschreibt, wie verschiedene Implementierungen aus Software-Bibliotheken für einen bestimmten Anwendungszweck zusammengeschaltet werden sollen. Basierend auf der Software-Architektur wird die KI-Anwendung implementiert und anschließend auf Computer-Hardware ausgeführt. Danach muss die KI-Anwendung mit Daten trainiert oder um eine Wissensbasis angereichert werden. Erst dann kann sie auf einer Hardware (z. B. einem Roboter) ausgeführt werden und kann Nutzen für Anwender stiften. Ein Algorithmus stiftet hingegen für einen Endanwender keinerlei Nutzen – genauso wie ein Kochrezept nicht satt macht.

Wir vermeiden daher in dieser Studie die Verwendung des Begriffs „Algorithmen“ (außer im Sinne der Definition) und sprechen von KI-Anwendungen bzw. -Systemen, wie oben bereits eingeführt.

2.3 Anwendungen von KI

Nachfolgend stellen wir beispielhaft KI-Anwendungen im Umfeld der Forschung von drei Wissenschaftsdomänen vor: Physik, Klimaforschung und Medizin.

2.3.1 *Teilchenphysik*³

KI, Machine Learning und Big Data sind in der Physik und in der physikalischen Grundlagenforschung seit Jahrzehnten wohletabliert. Die Teilchen- bzw. Hochenergiephysik bildet ein weitreichendes Anwendungsfeld, in dem diese Methoden und Instrumente in Kooperation mit Informatikern nicht nur angewendet, sondern auch weiterentwickelt werden. Dieses Feld der Physik ist deshalb so beispielhaft für Forschung allgemein, weil in Experimenten große Datenmengen anfallen, deren Informationsgehalt zunächst unklar ist. Hier unterstützen KI- und Machine Learning-Verfahren seit über drei Dekaden die Grundlagenforscher/-forscherinnen dabei, aus diesen Daten Informationen zu gewinnen, Theorien zu entwickeln und Wissen zu generieren. Insofern zukünftig auch in anderen Wissenschaftsdisziplinen und Forschungsfeldern große Datenmengen erzeugt und zugänglich werden, die zur Erkenntnisgewinnung genutzt werden können, kann die Teilchenphysik als paradigmatisch angesehen werden: allgemeine Entwicklungstrajektorien kristallisieren sich vermutlich in kaum einem anderen Feld so prägnant heraus wie hier.⁴

³Dieser Abschnitt basiert in Teilen auf dem Vortrag von P. Mättig, Universität Bonn, zum Thema „Big Data und maschinelles Lernen am LHC“ beim Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 02.10.2019.

⁴Auch aus historischer Perspektive gilt dies für die Physik, aber auch insbesondere für das CERN: Am CERN sind von Physikern grundlegende Aspekte des World Wide Web entwickelt worden (Gruppe um Tim Berners-Lee u. a.): Sprache HTML, Transferprotokoll http, die URL Signaturen, der erste robuste Browser, u. a.

Die weltweit größten und aufwendigsten Experimente der Naturwissenschaften werden am Large Hadron Collider (LHC) am Europäischen Kernforschungszentrum CERN bei Genf durchgeführt. In dem Teilchenbeschleuniger werden sehr hohe Energien (mit fast Lichtgeschwindigkeit) erzeugt, um in die grundlegende Struktur der Materie einzudringen und so fundamentale Erkenntnisse über den Aufbau, Struktur und Entwicklung der Materie zu erhalten. Derart hohe Energien waren auch kurz nach dem Urknall präsent; der Teilchenbeschleuniger simuliert mithin experimentell Aspekte des Urknalls. Damit ist die Teilchenphysik auch für die Wissenserzeugung der Kosmologie und für ein grundlegendes Verständnis des Universums relevant. Im Zentrum der Teilchenphysik steht seit einigen Jahrzehnten das so genannte Standardmodell der Elementarteilchen. Dieses Modell kann als Theoriekern der Teilchenphysik angesehen werden. Es bündelt alle wesentlichen Erkenntnisse der Physik in einem kompakten Modellrahmen, insbesondere die Existenz und Eigenschaften der Elementarteilchen. Diese umfassen (a) die Fermionen, also „Materieteilchen“ wie Leptonen, z. B. das bekannte Elektron, sowie die Quarks, aus denen z. B. Protonen aufgebaut sind, und (b) die Bosonen, also „Feldteilchen“ wie die Eichbosonen, d. h. „Fundamentalkräfteteilchen“, sowie das Higgs-Boson. Die Teilchenphysik rückte vor 10 Jahren erneut in die Schlagzeilen, als am Large Hadron Collider im Jahre 2012 das schon in den 1960er-Jahren vorhergesagte Higgs-Boson experimentell nachgewiesen werden konnte. Durch diesen Fund – genauer: durch Rekonstruktion zentraler Eigenschaften des Higgs-Bosons als spezielle Signaturen in den großen Datenmengen der Detektorsignale – gilt das Standardmodell im Prinzip als experimentell hinreichend bestätigte, vollständige Beschreibung der Grundlagen der Materie und der subatomaren Welt. Das Standardmodell beschreibt fast alle beobachteten subatomaren Phänomene und kann als eine sehr genau getestete Theorie angesehen werden.

Doch es gibt offene Fragen, die die Physik beschäftigen – und die, unter Anerkennung der Leistung des Standardmodells, gleichwohl zu einer Suche nach einer „Physics beyond the Standard Model“ geführt haben. Eine der offenen Fragen betrifft die so genannte Dunkle Materie, die vom Standardmodell nicht erfasst wird, obwohl sie mehr als 20 % des Kosmos ausmacht. Bisher ist sie allerdings nur durch Gravitationseffekte, also durch ihre Gravitations-Wechselwirkung, nachgewiesen. Nicht ausgeschlossen ist, dass Dunkle Materie auch nicht-gravitativ wirkt und etwa im Rahmen schwacher Wechselwirkung charakterisierbar ist. Fragen nach dem Anteil und den Eigenschaften Dunkler Energie schließen sich hier an.⁵ Zudem ist mit dem Standardmodell noch keine hinreichende Vereinheitlichung der vier fundamentalen Kräfte, d. h. der vier Wechselwirkungen erreicht. Dies betrifft teilweise die Zusammenführung von starker und elektroschwacher Kraft, aber insbesondere die Integration der Gravitationskraft in einen umfassenden und widerspruchsfreien

⁵ Auch wenn man beides als Falsifikation des Standardmodells ansehen könnte, wird das Standardmodell von Physikern keinesfalls verworfen, sondern es wird daraus ausgegangen, dass dieses in einen neuen und größeren Theorierahmen, d. h. in eine umfassendere Theorie, eingebettet werden kann, ähnlich wie Newtonsche Gravitationstheorie in die Allgemeine Relativitätstheorie.

Theorierahmen. Dies zeigt, dass die fast als klassisch zu bezeichnende Problematik der Physik, nämlich die Zusammenführung von Quantenmechanik bzw. Quantenfeldtheorien mit der Allgemeinen Relativitätstheorie (als zentraler Gravitationstheorie) noch nicht erreicht ist. Im Prinzip war die Vereinheitlichung im Sinne der Entwicklung eines kohärenten und konsistenten Theoriekorpus stets ein grundlegendes Erkenntnisideal der Physik, das das physikalische Forschungshandeln geprägt hat. Heute sind indes auch kritische Stimmen zu hören, die das Erkenntnisideal einer vereinheitlichten Theorie hinterfragen und eine neue Ära einer „Prediction without Explanation“ heraufziehen sehen.

Um eine „Physik jenseits des Standardmodells“ zu entwickeln, ist es – jenseits der Existenz und einiger Eigenschaften der dunklen Materie – notwendig, weitere empirische Abweichungen vom Standardmodell zu finden und so den ersten Schritt über das Standardmodell hinauszugehen. Auf theoretischer Seite gibt es unterschiedliche Ansätze, eine Vielzahl von Modellen und Theoriefragmenten, wie eine solche neue Physik aussehen könnte. Doch bis dato sind zwischen den Ableitungen bzw. Prognosen des Standardmodells einerseits sowie jenen der Konkurrenten der neuen Physik experimentell kaum Differenzen zu beobachten. Das könnte sich natürlich zukünftig ändern. Auf dem weiteren Weg zu einer „Physik jenseits des Standardmodells“, darin scheinen sich die meisten Physiker einig zu sein, wird KI, Machine Learning und Big Data eine kaum zu überschätzende Rolle spielen. Das betrifft in erster Linie die Entdeckung und Rekonstruktion von Signaturen in sehr großen Datenmengen, die auf neue Teilchen und ihre Eigenschaften hindeuten könnten – und weniger das Ziel, theoriebasierte Erklärungen zu liefern.

Nun wurden KI- und Machine Learning-Verfahren von Teilchenphysikern auch bisher vielfach eingesetzt, etwa zur Entdeckung des Higgs-Bosons. Derartige Elementarteilchen liegen nämlich nicht an sich vor und müssen nur noch vermessen werden. Vielmehr müssen sie durch hohe Energien zugänglich gemacht bzw., besser gesagt, erst erzeugt werden. Dazu werden Kollision- und Streuexperimente durchgeführt: bei sehr hohen Geschwindigkeiten werden Teilchen zur Kollision gebracht und Streuungen induziert, wobei Elementarteilchen hervortreten oder generiert werden. Die Kollisionen werden von hochtechnischen Detektoren vermessen, die entsprechende Signale auffangen. Dabei entstehen große Datenmengen, die in der Größe von Millionen Terabytes, d. h. Exabytes, liegen. Entscheidend ist die anschließende Datenanalyse mit dem Ziel, Informationen zu erhalten, was bei der jeweiligen Kollision passierte und welche Elementarteilchen in welcher Form beteiligt waren. Es geht dabei um die Rekonstruktion der Existenz und des Verhaltens der Elementarteilchen durch Identifikation ihrer Eigenschaften, d. h. die Elementarteilchen sind operational durch Ereignis-Signaturen definiert, die sich als Muster in den Datenreihen finden (vgl. Aaboud et al. 2018). Diese Muster zu erkennen und mit Bedeutung zu versehen, ist herausfordernd. In praktischer Hinsicht ist eines der zentralen Probleme die Trennung von gesuchtem Elementarteilchen-Signal einerseits und anderen, unspezifischen Signalen sowie dem Hinter- und Untergrund(rauschen) andererseits. Machine Learning-Verfahren können diese Mustererkennung unterstützen und helfen, die relevanten Ereignis-Signaturen zu identifizieren. Dieses Klassifikationsproblem ist ein klassisches Problem, welches ähnlich in vielen empirischen Wissenschaftsdisziplinen auftritt.

Dazu werden im Rahmen von Machine Learning-Verfahren Algorithmen, die in diesem Fall auf künstlichen Neuronalen Netzen basieren, anhand von modellbasierten, simulierten Teilchenkollisionen trainiert. In diesem Verfahren des überwachten Lernens („supervised Machine Learning“) wird gelernt, irrelevante von relevanten Ereignis-Signaturen zu unterscheiden. Damit verbunden ist ein Klassifikationsschema. Erst im Anschluss werden die mit Theorie-Modellen trainierten Machine Learning Algorithmen eingesetzt, um in experimentell erzeugten Datenreihen Muster, wie etwa die Ereignis-Signatur des Higgs-Teilchens, zu identifizieren.

Ein verwandter Fall von Machine Learning, in dem sich die Relevanz des überwachten Lernens in der Teilchenphysik zeigt, bezieht sich auf mathematische Modellierungen und modellbasierte Simulationen im Zusammenhang des Standardmodells. Um konkrete, vor dem Theorie-Hintergrund des Standardmodells entwickelte teilchenphysikalische Modelle zu optimieren, wird Machine Learning zum Auffinden von Parametern oder von Raum- und Zeitdimensionen sowie zum Test von wissenschaftlicher Robustheit verwendet. Ferner unterstützt Machine Learning das Auffinden von raumzeitlichen Skalierungseigenschaften und die Auflösung von Substrukturen, was für die Analyse von experimentellen Datenreihen von Interesse ist.

Es gibt noch ein weiteres, deutlich avancierteres Feld der Teilchenphysik, in dem Machine Learning – und zwar in der Variante des „unsupervised Machine Learning“ (unüberwachten Lernens) – eine Rolle spielt. Insofern sich Grenzen des Standardmodells, wie oben beschrieben, zeigen, gibt es Bestrebungen zu einer „Physik jenseits des Standardmodells“. Dabei weiß niemand, wie diese aussehen könnte. Unklar ist, wonach man in Datenreihen suchen soll. Machine Learning ermöglicht hier im Prinzip, modell- und theorieunabhängig Datenreihen auf Signaturen zu untersuchen – auch solche, die im Standardmodell nicht vorkommen. Vieles ist hier noch Zukunftsmusik, aber es zeigt sich ein zunehmendes Interesse von Teilchenphysikern, diese Verfahren zumindest heuristisch anzuwenden, um neue Entdeckungen zu machen und auf weitere Ideen zu kommen.

Zusammengenommen stellen KI und Machine Learning für die Teilchenphysik methodische Verfahren bereit, die in der Erzeugung von neuem Wissen eine erkenntnisgenerierende Funktion besitzen („Context of Discovery“). Diese Verfahren fördern die Effizienz und Effektivität des physikalischen Erkenntnishandelns. (1) Sie bieten die Möglichkeit, neue physikalische Objekte zu identifizieren, zu klassifizieren bzw. gar diese zu definieren sowie Substrukturen oder Eigenschaften von bekannten und neuen Objekten besser auflösen zu können. (2) Entscheidend ist, dass die informatischen Verfahren eine bessere Trennbarkeit der Signale vom Untergrund ermöglichen und die Signifikanz von Observablen, d. h. der Signalmessgrößen erhöhen. (3) Zudem werden die Verfahren auch im Rahmen von Parametrisierungen und Optimierungen von Simulationen eingesetzt, also zu Bestimmung und Verbesserung der Größen, die in die Modellbildung und anschließende numerische Simulation eingehen. So bilden sie auch eine entscheidende Grundlage für Computerexperimente der Teilchenphysik.

Trotz der breiten Anwendung und des weitreichenden Erfolgs der informatischen Verfahren bleiben Teilchenphysiker insgesamt kritisch gegenüber KI und Machine Learning: Sie verwenden daher einen Teil ihrer Forschungszeit auf die kritische Validierung dieser Verfahren. So bleibt es dabei, dass Korrelationen, Klassifikationen oder Prognosen, wie man sie über KI- und Machine Learning-Verfahren generieren kann, für die grundlagenforschende Physik zwar hilfreich, aber am Ende (als Resultate) zu wenig sind. Es geht (nach wie vor) um kausale Erklärungen und um Einordnung neuer Ereignisse und Phänomene in einen gegebenen Theorierahmen, d. h. um deduktiv-nomologische Erklärungen. Die Physik möchte, ganz traditionell, die Ordnung, Struktur und Entwicklung der Materie und des Universums verstehen.

2.3.2 Klimaforschung⁶

Die Klimaforschung ist traditionell ein Teilgebiet der Physik und der Angewandten Mathematik: Ihr Ursprung liegt in der Kontinuumsmechanik, d. h. in der Fluid- oder Hydrodynamik, die Mitte des 19. Jahrhunderts entstand, und sie verwendet zudem Erkenntnisse der Thermodynamik. Doch auch wenn ihr Ursprung in der Physik und Mathematik liegt, ist sie heute interdisziplinär ausgerichtet. Sie umfasst nicht nur zentrale Fragestellungen der Chemie und den Geowissenschaften, sondern integriert insbesondere die Informatik und einige Technikwissenschaften (u. a. wegen Sensor-, Mess- und Detektortechnik). Und insofern aktuelle Klimaforschung im Horizont des so genannten Anthropozäns betrieben wird und anthropogene Ursachen und Folgen des Klimawandels untersucht, reicht ihre Forschung bis in die Biologie und die Gesellschaftswissenschaften hinein.

Die Klimaforschung ist wissenschaftlich äußerst herausfordernd. Das globale Klimasystem ist nicht – wie üblicherweise in den exakten Naturwissenschaften – abgrenzbar und unter Laborbedingungen kontrollierbar. Phänomene sind selten reproduzierbar. Ein traditionelles Experimentieren ist nicht möglich, ähnlich wie in der Kosmologie und in Teilbereichen von Biologie und Ökologie. Zudem ist das Klimasystem, wie die allgemeine Grundgleichungen der Klimaforschung (Navier-Stokes-Gleichungen) nahelegen, hochgradig nichtlinear und damit vielfach auch sensitiv und instabil. Kleinste Veränderungen von Variablen und Parameter, wie beispielsweise lokale Drücke, Temperaturen oder Strömungsgeschwindigkeiten sowie ihre jeweiligen Gradienten, können große Folgen nach sich ziehen. Systemdynamiken kippen oder bifurkieren: der Golfstrom kann zum Erliegen kommen, Hitze- und Dürreperioden können entstehen. Damit verbunden sind mathematische und numerische Probleme, aber auch Herausforderungen adäquater Messung und Datenerhebung. Allgemein ist für die Klimaforschung herausfordernd, die relevanten Variablen oder

⁶ Dieser Abschnitt basiert teilweise auf dem Vortrag von Hans von Storch zum Thema „Künstliche Intelligenz & Big Data in der Klimaforschung“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 2. Juli 2019.

Zustandsgrößen zu bestimmen, die in einem bestimmten Raumzeitbereich die Klimadynamik bestimmen. Mit anderen Worten, es ist jeweils offen, wie hochdimensional der Zustandsraum ist, in dem die Dynamik beschrieben und sodann, auf Basis der Numerik, durch Computersimulation dargestellt werden kann. Ferner sieht sich die Klimaforschung angesichts voranschreitender Sensorik und der Implementierung sehr vieler neuer Messstationen mit sehr großen und sehr unterschiedlichen Zeitreihen an Daten und ihren jeweiligen Typen konfrontiert.

Angesichts der grundlegenden Nichtlinearität und Instabilität des Klimasystems einerseits und der großen Datenmengen andererseits stellen KI-, Machine Learning- und Big Data-Verfahren ein zukunftsweisendes Instrumentarium zur Verfügung, die wissenschaftlichen Herausforderungen anzugehen. Dies kann anhand von drei Fällen beispielhaft dargestellt werden:

Rekonstruktion des Zustandsraumes, Identifikation lokaler Unterräume und Entwicklung von Prognosen

Wie oben angedeutet, weist das Klimasystem eine große Anzahl an Freiheitsgraden auf. Für die Klimaforschung sind damit eine Vielzahl möglicher Zustandsgrößen – also Variablen, die das Klimasystem beschreiben – verbunden, die schwer zu überschauen sind. Zudem besitzen diese Zustandsgrößen zumeist komplizierte Skalierungseigenschaften: die raumzeitliche Skalenabhängigkeit ist empirisch und mathematisch schwer in den Griff zu kriegen. So ändern sich beispielsweise die für die Klimaforschung relevanten Differenzialgleichungen mit fortschreitender Gitterauflösung, d. h. mit Skalierung strukturell. Das heißt die Differenzialgleichungen, die jeweils partikuläre Modelle für Teile des Klimasystems darstellen, sind nicht immer strukturell stabil. Approximationsmöglichkeiten sind dadurch effektiv reduziert, verbunden mit wesentlichen Prognoseunsicherheiten. Die durch Nichtlinearität und Instabilität/Sensitivität des Klimasystems sowie der dieses repräsentierende mathematische Modell (Differenzialgleichungen) erzeugten Probleme der Approximation zeigen sich beispielsweise bei atmosphärischen Konvektionen oder bei Wolkenbildungsphänomenen. Aufgrund der großen Anzahl der Freiheitsgrade ist, wie angedeutet, der Zustandsraum des Klimasystems grundsätzlich als hochdimensional anzusehen. Im Detail ist das wissenschaftliche Wissen effektiv begrenzt; detaillierte Modelle sind schwer zu entwickeln und zu validieren, auch wenn die Klimaforschung in den beiden letzten Jahrzehnten deutliche Fortschritte verzeichnen kann. So ist es bis dato schwierig, quantitative Zusammenhänge für viele der Zustandsgrößen aufzufinden und zu rechtfertigen. Eine gewisse Abhilfe leistet hier das Konzept der „Principal Interaction Patterns“, das dynamisch weitgehend abgeschlossene niederdimensionale Unterräume bzw. lokale Zustandsräume des Klimasystems zu definieren ermöglicht, die an ihren „Rändern“ mit dem Rest des Zustandsraums stochastisch oder ggf. auch deterministisch-konditioniert in Wechselwirkung treten (Hasselmann 1988). Die Modellparameter bzw. Zustandsgrößen, inklusive der Skalierungseigenschaften dieser Unterräume sind empirischen Verfahren der Quantifizierung ihrer internen Dynamiken zugänglich, mit denen man z. B. bereits die langfristigen stochastischen Oszillationen im südlichen Ozean durch kontinuierliche Frischwasserzufuhr erklären kann (Zhang et al. 2019). Siehe Abb. 2.7.

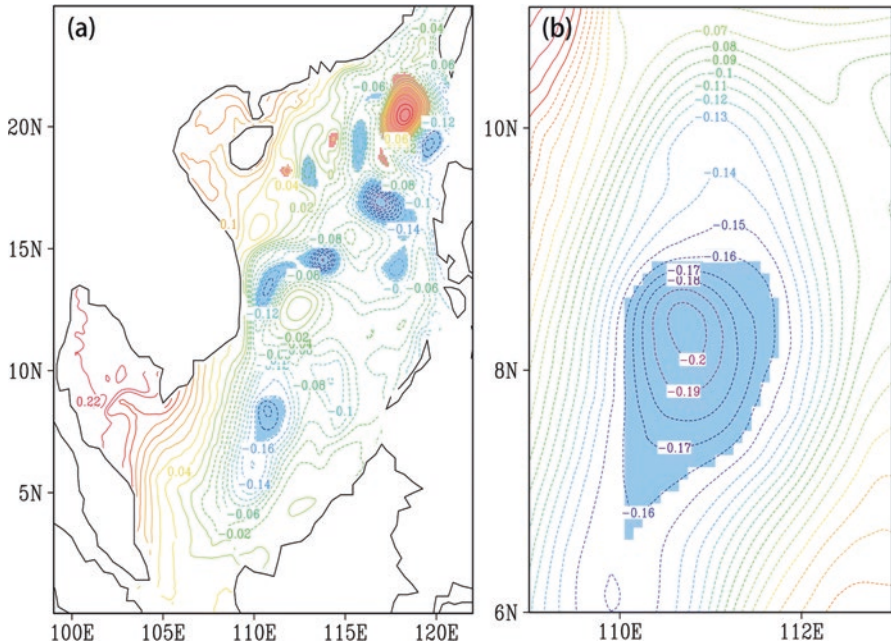


Abb. 2.7 Mustererkennung für Strömungswirbel im Südchinesischen Meer (aus Zhang et al. 2019)

In diesem Beispiel geht es um Mustererkennung für Wirbelbildungen im Südchinesischen Meer. Gesucht wird nach einer Statistik der jährlichen Häufigkeit, Intensität und Größe von wandernden Wirbeln in diesem Meeresabschnitt. Eine entsprechende KI-Anwendung detektiert und charakterisiert anhand des täglich schwankenden Meeresspiegels im Mittel pro Jahr 28 antizyklonische Wirbel und 54 zyklonische Wirbel in dieser Region. Diese Erkenntnisse sind für das Verständnis der regionalen und überregionalen Klimadynamik von zentraler Bedeutung.

Mit der zunehmenden Verfügbarkeit und Verarbeitung großer Datenmengen in Verbindung mit neuen informatischen Verfahren lassen sich nun weitere Unterräume des Klimasystems besser beschreiben und verstehen. Das trifft insbesondere für das Verständnis solcher Unterräume zu, welche auf die Interpretation großer Datenmengen mit hoher technischer Rechen- und Speicherkapazität angewiesen sind. Ein Beispiel hierfür ist das noch ungenaue Verständnis der nordatlantischen Oszillation und der ihr zugrunde liegenden nichtlinear-instabilen Dynamik von dominanter Westwinddrift und ihrer zeitweisen Blockade. Hier könnten vertiefende niederdimensionale Analysen, d. h. die Analyse der Unterräume von der Anwendung geeigneter Big-Data- und Machine-Learning-Verfahren profitieren. Ein anderes Beispiel ist die Modellierung und Simulation des Golfstroms, die bislang nur durch Kopplung mit einem stark vereinfachten Atmosphärenmodell in grober Form gelingt. Erst die heute verfügbaren Daten und Computer-Leistungen ermöglichen Kopplungen mit wesentlich realitätsnäheren Atmosphärenmodellen, in denen auch wetterbedingte Fluktuationen und ihre Stochastik einfließen, um die Dynamik des

Golfstroms auch im Detail verstehen zu können. Entsprechend abgeleitete und abgesicherte Vorhersagen der Dynamik von Meeresströmungen und Atmosphäre in bzw. über dem (Nord-)Atlantik sind neben dem wissenschaftlichen Erkenntnisgewinn auch für die mittel- bis langfristige Klimavorsorge in Europa von signifikanter Bedeutung.

Entwicklung von Modellen bzw. Zerlegung in Teilmodelle

Wie angedeutet, zielt die Klimaforschung darauf ab, mathematische Modelle, basierend auf partiellen Differenzialgleichungen, zu entwickeln. Dies sind so genannte Zustandsraummodelle, mit denen die klimatischen Zustandsgrößen oder Zustandsvariablen in einem gewählten raumzeitlichen Ausschnitt des (als Ganzes unzugänglichen) Klimasystems auf Basis von zwei oder mehreren (interagierenden) Teilmodellen gegeben sind. Dabei berechnet ein dynamisches Teilmodell – von einem bekannten bzw. definierten Anfangszustand aus – die weitere Entwicklung der Zustandsgrößen über gewünschte Zeitskalen hinweg. Diese Vorhersagen werden schrittweise mit beobachteten Daten und dem so genannten Beobachtungsmodell abgeglichen. Bei diesem werden die modellierten mit den bekannten Zustandsgrößen desselben Zustandsraums in Beziehung gesetzt und ggf. bei zu hoher Abweichung über geeignet Korrekterterme jeweils angepasst. So erhält man konsistente Zeitreihenverläufe in den gewählten Zustandsräumen.

Entsprechende algorithmischen Verfahren werden beispielsweise – jeweils von den aktuellen atmosphärischen Zuständen ausgehend – mit großem Erfolg für die Wettervorhersage verwendet. Darüber hinaus eignen sie sich auch für widerspruchsfreie Re-Analysen von langfristigen Klimaentwicklungen sowie für das konsistente „Down-Scaling“ großskaliger Strukturen im Klimasystem. In diesen Bereichen der Zustandsraumanalyse und -vorhersage weisen KI-, Machine Learning- und Big Data-Verfahren bis dato das größte Potenzial für die Klimaforschung auf, das noch weiter zu nutzen ist.

Signal-Rauschen-Unterscheidung zur Identifikation exzeptioneller Ereignisse

Eine Machine-Learning- und Big-Data-gestützte Kernaufgabe der Klimaforschung ist die Unterscheidung von bedeutsamen Signalen einerseits und interner Variabilität („Rauschen“) andererseits im hochdimensionalen Zustandsraum eines lokal betrachteten Klimateilsystems. Erst wenn sich dort Strukturen, d. h. bedeutsame Signale detektieren lassen, die nicht durch das interne Rauschen dieses Systems generiert werden, schließt man auf die Existenz eines außerhalb dieses lokalen Klimasystems liegenden Signals. Hieraus ergibt sich dann die Frage nach der externen Ursache dieses Signals an. Dieser Ansatz wird heute standardmäßig auf den Nachweis von Klimawirkungen erhöhter Treibhausgaskonzentrationen angewandt sowie auf deren Unterscheidung von natürlichen Schwankungen, etwa wie die der wechselnden Sturmtätigkeit in Nordeuropa oder der süd pazifischen „El-Nino“-Oszillation und ihrer Wirkungen auf die Umwelt.

Ein ambitioniertes Ziel ist die sogenannte „Event Attribution“, bei welcher einzelne Ereignisse wie Stürme oder Hitzewellen auf charakteristische Eigenschaften hin untersucht werden, um zu prüfen, ob diese noch unter natürlichen Klimaverhältnissen plausibel sind oder ob sie bereits den Einfluss des anthropo-

genen Klimawandels widerspiegeln. Für derart kurzfristige Analysen und Interpretationen entsprechender Extremereignisse werden geeignete Algorithmen und neue informatische Verfahren eingesetzt.

Zwischenfazit

In der Erforschung des Erdklimas wurde schon lange und mit wechselndem Erfolg nach wirksamen Methoden und Verfahren gesucht, die eine Beschreibung von Klimasystemkomponenten und -prozessen (Zustandsgrößen, relevante Unterräume und lokale Modelle) ermöglichen. Nur auf dieser Basis können vereinfachte, effiziente und automatisierbare Klimavorhersagen entwickelt werden. Inzwischen sind KI- und Big-Data-Verfahren so weit, dieses Ziel zu erreichen: Sie werden bereits routinemäßig zur konsistenten raumzeitlichen Modellierung von Klima und Wetter, ihrer Dynamik und daraus abgeleiteter Prognosen eingesetzt. Der Einsatz neuronaler Netze eignet sich darüber hinaus auch für die automatisierte Analyse von dynamisch relevanten Teilen des Phasenraums im Klimasystem. Unter anderem kann KI dabei auch die Trennung von signifikanten Signalen von Hintergrundrauschen erleichtern, was eine notwendige Bedingung für entsprechend belastbare wissenschaftliche Aussagen sowie deren Erklärung darstellt.

Auch wenn KI-, Machine Learning- und Big Data-Verfahren damit zum besseren Verständnis der Komponenten und Prozesse des Klimasystems beitragen können, betonen Klimaforscher, dass entsprechende Erkenntnisbildung durch künstliche Intelligenz alleine – also automatisiert und ohne Zutun von Forschenden – prinzipiell nicht erreichbar ist. Der Klimaforschung geht es nicht nur um Prognosen und Klassifikationen allein, sondern um ein Verständnis des Klimasystems und von wissenschaftlichen Erklärungen. Eine gute Prognostik ist in der Klimaforschung – ganz traditionell – abhängig von wissenschaftlichem Wissen, welches freilich in Verfahren von KI und Big Data Eingang findet.

2.3.3 Medizin⁷

KI und insbesondere Machine Learning haben das Potenzial, Ärzte bei der Diagnose von Krankheiten und mit Vorschlägen zu Therapien zu unterstützen. Beispielsweise können Symptome, Röntgenbilder und Krankenakten in wenigen Sekunden überprüft und mit anderen Datenbeständen, beispielsweise aus verschiedenen Krankenhäusern, abgeglichen werden (Jäkel 2018). Auf diese Weise können deutlich größere medizinische Wissensbestände bei den Analysen berücksichtigt werden, als dies ein einzelner Arzt könnte. So kann Spezialwissen aus den verschiedenen medizinischen Bereichen (zum Beispiel Anatomie, Chirurgie) auch für Hausärzte

⁷Dieser Abschnitt basiert teilweise auf dem Vortrag von Thomas Lengauer zum Thema „Data Analysis in Medicine“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 02.07.2019, sowie auf dem Vortrag von Karl-Heinz Jöckel zum Thema „Künstliche Intelligenz in den Medizinischen Wissenschaften“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 06.09.2019.

zugänglich gemacht werden. Seltene Krankheiten können schneller und zu geringen Kosten identifiziert werden (Kuhn et al. 2018). Vor diesem Hintergrund haben KI-Methoden bereits viele Anwendungszwecke in der Medizin, so zum Beispiel:

- Erkennung von pathogenen Mustern, z. B. in der Bildgebung von Organen und Geweben
- Erkennung von Gesetzmäßigkeiten, z. B. in der biomedizinischen Forschung
- Übersetzung zwischen unterschiedlichen Sprachen oder Terminologien, z. B. Übersetzung pathologischer Befunde von natürlicher Sprache in die medizinische Klassifikation ICD-O
- Prädiktion für Individuen, z. B. in der präzisionsmedizinischen Differenzialdiagnostik
- Prädiktion für Gruppen/Populationen, z. B. zur Einschätzung von Epidemien
- Unterstützung menschlichen Handelns in konkreten Entscheidungssituationen, z. B. für Therapieempfehlungen
- Unterstützung bei abstrakten Planungsprozessen, z. B. bei der Studienplanung
- „Intelligente“ Interaktion mit Menschen, z. B. bei automatisierter Durchführung von Befragungen
- Entwicklung von Medikamenten

Ein Beispiel ist die *personalisierte Therapie* von HIV. So existieren auf der einen Seite Millionen von HIV-Varianten. Auf der anderen Seite gibt es mehrere Dutzend Medikamente für HIV mit hunderten von Therapie-Kombinationen. Für eine beim Patienten beobachtete Population von Virenvarianten eine möglichst effektive Therapie aus zwei bis drei Medikamenten zu finden, stellt eine kombinatorische Herausforderung dar (Deutsches Ärzteblatt 2018, vgl. Shah et al. 2019). Diese Herausforderung kann mittels KI-Methoden auf der Basis von konkreten Therapiebeispielen angegangen werden. Das Ergebnis sind maßgeschneiderte Behandlungen für individuelle Patienten mit verbesserten Erfolgchancen.

Die Chancen der Nutzung von KI in der Medizin liegen auch im Fortschritt von Algorithmen in der Bilderkennung begründet. Die folgende Abb. 2.8 zeigt, wie sich die Fähigkeiten von KI-Anwendungen im Vergleich zu Menschen in den letzten Jahren verbessert hat.

Damit sind sie grundsätzlich gut geeignet, auch eine Analyse von CT- oder MRT-Bildern durchzuführen und Ärzte bei der Diagnose von Krankheiten zu unterstützen. Das Potenzial wird auch deutlich aus einem Gespräch, das einer der Autoren mit Prof. Dr. Jürgen Schäfer (Mediziner am Universitätsklinikum Marburg für seltene Krankheiten) geführt hat: „Zum Beispiel gibt es jetzt schon so gute Daten in der Pathologie, dass eine KI-Anwendung sehr präzise Gewebeschnitte analysieren und tumorsuspekte Zellen erkennen kann. Ein Pathologe, der 10.000 Gewebeschnitte befunden hat, der hat natürlich mehr Erfahrung als jemand, der nur 500 Befunde erstellt hat. Aber wenn man dann sieht, dass eine KI-Anwendung nicht nur 10.000, sondern Millionen Befunde im Speicher hat, dann bekommt man eine Vorstellung, wie viel profunder solch eine Befundung in Zukunft sein mag. Allerdings wird KI nicht so schnell einem Patienten die Hand halten und mit menschlichem

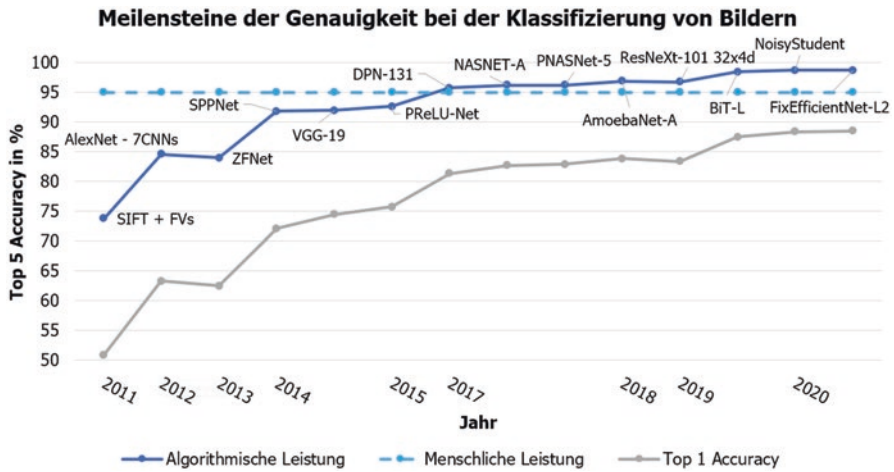


Abb. 2.8 Meilensteine der Genauigkeitsverbesserung bei der Klassifizierung von Bildern. (Quelle: ImageNet 2020)

Einfühlungsvermögen begegnen“ (Buxmann 2020). Solange dies der Fall ist, brauchen sich Ärzte keine Sorgen machen, dass sie vom Computer verdrängt werden.

Eine wichtige zukünftige Herausforderung besteht darin, dass KI-Anwendungen in die Lage versetzt werden, ihre Entscheidungen zu erklären. Kaum ein Arzt oder kaum ein Patient würde beispielsweise Therapieempfehlungen einer KI-Anwendung akzeptieren, bei denen nicht nachvollziehbar sind, wie sie entstanden sind. Hierbei gibt es weltweit gerade eine Vielzahl von Forschungsarbeiten, die auch unter dem Stichwort „Explainable AI“ (Erklärbare KI) zusammengefasst werden.

2.4 Fazit

KI hat nicht nur eine hohe Bedeutung für Wirtschaft und Gesellschaft. Auch in Forschung und Wissenschaft wird KI erfolgreich eingesetzt, wie beispielhaft in der Physik, der Klimaforschung und der Medizin beleuchtet. KI zeigt sich damit als ein mächtiges Werkzeug, welches die Wissensarbeit verändert:

1. *Höhere Effizienz:* Durch die Automatisierung von Routine-Tätigkeiten können (Forschungs-)Arbeiten beschleunigt werden. Allerdings ändern sich dadurch auch Tätigkeitsbereiche und Stellenprofile. Beispiele hierfür sind KI-gestützte Laborauswertungen in den Lebenswissenschaften.

Die Erhöhung des Automatisierungsgrades kann auch indirekt bewirken, dass Forschende einen immer weniger tiefen Einblick in die zu untersuchende Materie bekommen; Forschende würden dann Ergebnissen von KI-Anwendungen vertrauen, ohne sie nachvollziehen und kritisch hinterfragen zu können.

2. *Größere Genauigkeit:* Vor allem durch die Verfügbarkeit von großen Datenmengen können Forschungsfragen in der Regel präziser beantwortet werden. Beispiele hierfür sind KI-gestützte Auswertungen von Versuchen in der Physik. Falsch eingesetzt können Ergebnisse jedoch auch falsch interpretiert werden. Garantien gegen Falschaussagen gibt es nicht.
3. *Neue Möglichkeiten:* KI und verwandte Technologien erlauben Dinge, die bislang unmöglich waren, z. B. Genomanalyse und darauf aufbauende Forschungen. Besonders die Hypothesenbildung aufgrund von KI-Techniken eröffnet in vielen Wissenschaftsdisziplinen ganz neue Möglichkeiten.

Eine mögliche Folge von datengetriebenen KI-Methoden in der Wissenschaft könnte ein Paradigmen-Shift sein, bei welchem statistisch valide Korrelationen an Bedeutung zunehmen und erklärbare Kausalzusammenhänge an Bedeutung abnehmen (Näheres hierzu wird in Kap. 4 dieses Bandes beschrieben). Ob und welche Folgen das für die künftige Rolle der menschlichen Forscher/Forscherinnen hat, erläutert das Kap. 3.

Literatur

- Aaboud M et al (2018) Measurements of Higgs boson properties in the diphoton decay channel with 36 fb⁻¹ of *pp* collision data at $s\sqrt{=13}$ TeV with the ATLAS detector. Phys Rev D 98(5):052005. <https://doi.org/10.1103/PhysRevD.98.052005>
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29. <https://doi.org/10.1038/75556>
- Bostrom N (2014) Superintelligenz: Szenarien einer kommenden Revolution. Suhrkamp, Berlin
- Brooks R (2017) The seven deadly sins of predicting the future of AI. <https://rodnebrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/>. Zugegriffen am 13.10.2020. Deutsche Übersetzung: Algorithmenethik. 2017. Die sieben Todsünden der Prognosen über die Zukunft der KI. <https://algorithmenethik.de/2017/11/14/die-sieben-todsunden-der-prognosen-ueber-die-zukunft-der-ki>. Zugegriffen am 13.10.2020
- Brynjolfsson E, McAfee A (2017) The business of artificial intelligence. Harv Bus Rev 1–20. <https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence>. Zugegriffen am 13.10.2020
- Buxmann P, Peters F (2020) Ethik der Künstlichen Intelligenz. In: Canibol H-P, Hossenfelder J, Lünendonk J (Hrsg) *Handbuch Consulting 2020*. Lünendonk und Hossenfelder GmbH, Mindelheim, S 92–103
- Buxmann P, Schmidt H (2019) Open source tools Für Künstliche Intelligenz & Maschinelles Lernen. <https://www.ki-business.de/tools>. Zugegriffen am 25.11.2020
- Buxmann P (2020) Künstliche Intelligenz – Wirtschaft und Innovation. Studie im Auftrag des Ministeriums für Digitales für das Land Hessen. Wiesbaden (im Erscheinen)
- Buxmann P, Schmidt H (2021) Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. In: Buxmann P, Schmidt H (Hrsg) *Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg*, 2. Aufl. Springer Gabler, Berlin
- Corbett-Davies S, Goel S (2018) The measure and mismeasure of fairness: a critical review of fair machine learning. arXiv preprint arXiv:1808.00023
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Bd 1, San Diego, S 886–893

- Deutsches Ärzteblatt (2018) Die Forderung nach transparenten Algorithmen kann ihre Vorhersagekraft schwächen. <https://www.aerzteblatt.de/nachrichten/95778/Die-Forderung-nach-transparenten-Algorithmen-kann-ihre-Vorhersagekraft-schwaechen>. Zugegriffen am 26.11.2020
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Google (2020) Wann wurde die Mauer gebaut? – Google Suche. <https://www.google.de/search?source=hp&q=Wann+wurde+die+Mauer+gebaut%3F&oq=Wann+wurde+die+Mauer+gebaut%3F>. Zugegriffen am 10.09.2020
- Hasselmann K (1988) PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns. *J Geophys Res* 93(D9):11015. <https://doi.org/10.1029/JD093iD09p11015>
- Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47(1):153–161
- Hoppe T (2020) Semantische Suche. Grundlagen und Methoden semantischer Suche von Textdokumenten. Mit Beiträgen von B.G. Humm. Springer Vieweg, Wiesbaden
- Humm BG (2017) „Technische Superintelligenzen werden die Menschheit bedrohen“ – Wie kann man eine so absurde Behauptung aufstellen? <https://www.cognitiveagent.org/2017/05/05/technische-superintelligenzen-werden-die-menschheit-bedrohen-wie-kann-man-eine-so-absurde-behauptung-aufstellen/>. Zugegriffen am 13.10.2020
- Humm BG (2020) Applied Artificial Intelligence. An Engineering Approach, 2. Aufl. Unabhängig veröffentlicht bei Amazon KDP, Seattle, USA 2020. ISBN: 979-8635591154. Online <https://www.amazon.de/Applied-Artificial-Intelligence-Engineering-Approach/dp/B086Y6H7ZL>
- ImageNet (2020) Image Classification on ImageNet. <https://paperswithcode.com/sota/image-classification-on-imagenet>. Zugegriffen am 01.10.2020
- Insight Centre for Data Analytics (2020) The linked open data cloud. <https://lod-cloud.net/>. Zugegriffen am 12.03.2020
- Jäkel L (2018) Wie Künstliche Intelligenz unser Leben berührt. <https://orange.handelsblatt.com/artikel/53316>. Zugegriffen am 13.02.2019
- Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 23(1):89–109. [https://doi.org/10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X)
- Kuhn S, Jungmann SM, Jungmann F (2018) Künstliche Intelligenz für Ärzte und Patienten: „Googeln“ war gestern. <https://www.aerzteblatt.de/archiv/198854/Kuenstliche-Intelligenz-fuer-Aerzte-und-Patienten-Googeln-war-gestern>. Zugegriffen am 25.11.2020
- Luber S, Litzel N (2019) Was ist ein Expertensystem? Big Data Insider. <https://www.bigdata-insider.de/was-ist-ein-expertensystem-a-819539>. Zugegriffen am 09.09.2020
- Lum K, Isaac W (2016) To predict and serve? *Significance* 13(5):14–19
- McCarthy J, Minsky ML, Rochester M, Shannon CE (1955) A proposal for the Dartmouth summer research project on artificial intelligence. <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>. Zugegriffen am 02.02.2021
- Medical Subject Headings. U. S. National Library of Medicine. <https://www.nlm.nih.gov/mesh/meshhome.html>. Zugegriffen am 13.10.2020
- Murphy KP (2012) Machine learning. A probabilistic perspective. MIT Press, Cambridge
- NCI Thesaurus. U. S. National Cancer Institute. <https://ncithesaurus.nci.nih.gov>. Zugegriffen am 13.10.2020
- Powers DMW (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol* 2(1):37–63
- Sarathi P (2014) 12 famous quotes on Artificial Intelligence (AI) by Google founders. Big Data Made Simple. <http://bigdata-madesimple.com/12-famous-quotes-on-artificial-intelligence-by-google-founders>. Zugegriffen am 30.09.2019
- Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, Ringel M, Schork N (2019) Artificial intelligence and machine learning in clinical development: a translational perspective. *npj Digit Med* 2(1):69. <https://doi.org/10.1038/s41746-019-0148-3>
- The Gene Ontology Consortium (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* 47(D1):D330–D338. <https://doi.org/10.1093/nar/gky1055>

- Vincent J (2016) Apple boasts about sales; Google boasts about how good its AI is. The Verge. <https://www.theverge.com/2016/10/4/13122406/google-phone-event-stats>. Zugegriffen am 12.03.2020
- Zhang M, von Storch H, Chen X, Wang D, Li D (2019) Temporal and spatial statistics of travelling eddy variability in the South China Sea. *Ocean Dyn* 69(8):879–898. <https://doi.org/10.1007/s10236-019-01282-2>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

