



---

Behind the buzzwords –

---

# Was bedeutet Algorithmus, Data Science und Machine Learning?

---

DGQ-Whitepaper

[www.dgq.de](http://www.dgq.de)

**DGQ+**

Deutsche Gesellschaft  
für Qualität

# Behind the buzzwords – Was bedeutet Algorithmus, Data Science und Machine Learning?

Begriffe wie Machine Learning, Data Mining oder Smart Data begegnen uns immer häufiger und halten auch Einzug in Trainings, Fachbeiträgen und Blogposts der DGQ. Um etwas mehr Klarheit in diese Buzzwords zu bringen, hat die DGQ ein Glossar erstellt, welches die wichtigsten Begriffe rund um die Digitalisierung verständlich erklärt.

---

## Algorithmus

Algorithmen begleiten Menschen überall und sind wesentlicher Bestandteil des modernen Alltags. Der Begriff beschreibt eine Kette von Handlungsanweisungen, mit denen ein bestimmtes Problem gelöst wird. Wenn Menschen beispielsweise einer Aufbauanleitung für einen Schrank folgen, führen sie streng genommen einen Algorithmus aus, um das Problem „Schrank aufbauen“ zu lösen. Computer und Maschinen folgen stets in ihrer Programmierung festgelegten Algorithmen, um die an sie gestellten Aufgaben zu erfüllen. Ihre Kapazitäten und Fähigkeiten sind daher auf die implementierten Algorithmen beschränkt. Künstliche Intelligenz erlaubt es Computern und Programmen, durch maschinelles Lernen eigene Algorithmen zur Lösung neuer Probleme zu entwickeln.

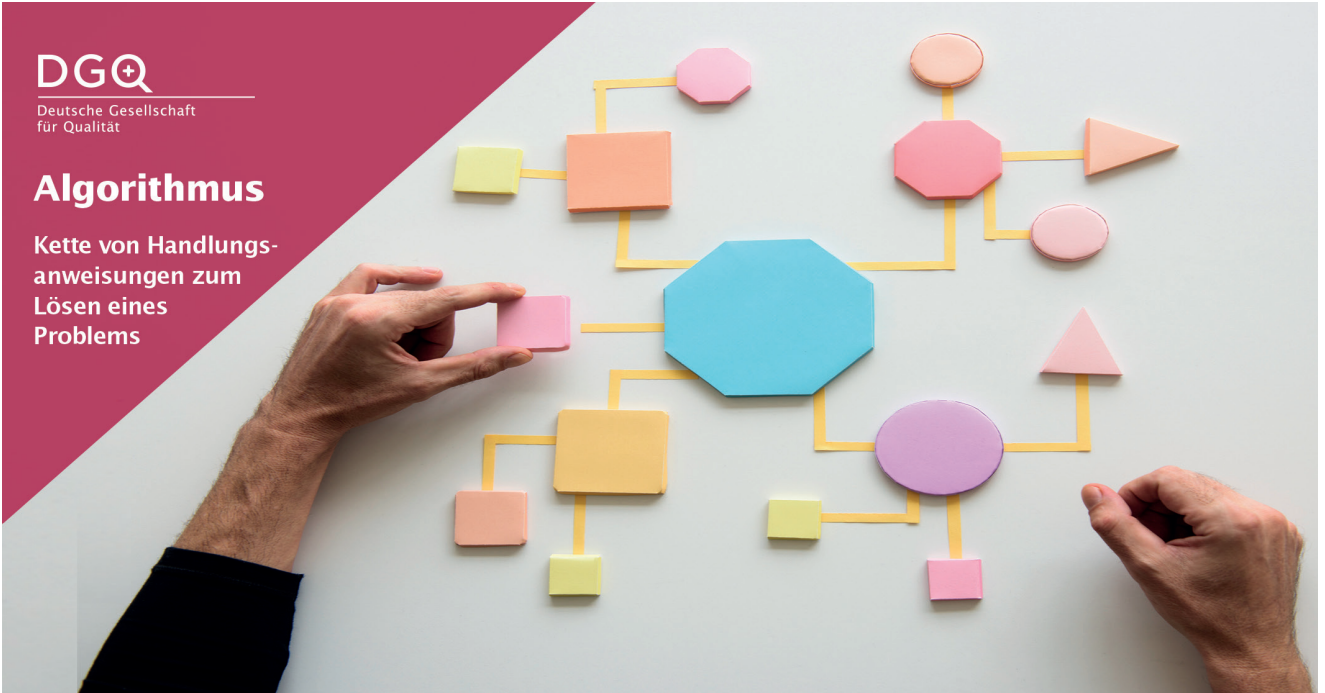
Algorithmen müssen bestimmte Eigenschaften erfüllen, damit ihre Funktionalität sichergestellt ist. Dazu gehört die Eindeutigkeit der Befehle ebenso wie die Ausführbarkeit dieser. An jeder Stelle in der Abfolge der Handlungsanweisungen muss eindeutig sein, welche Aktion auszuführen ist. Außerdem müssen Algorithmen endlich sein. Denn eine unendliche Befehlsabfolge wird niemals eine Lösung

liefern. Ebenso müssen die Befehle des Algorithmus immer so eindeutig formuliert sein, dass seine Ausführung stets das gleiche Ergebnis ausgibt. Der Wert der Lösung oder des Ergebnisses ist selbstverständlich immer unterschiedlich, die Art des Ergebnisses ändert sich aber niemals. Ein Algorithmus, der die täglichen Besucher einer Website zählt, wird zwar jeden Tag andere Zahlen liefern, aber immer Besucherzahlen der Website.

Der Begriff Algorithmus für eine feste Kette an Befehlen oder Aktionen ist vor allem in der Mathematik und Informatik verbreitet. Bei Computerprogrammen sind sie dafür zuständig, Eingabedaten in Ausgabedaten umzuwandeln. Ein Navigationsgerät beispielsweise erhält Daten zum aktuellen Standort, dem Straßennetz und dem gewünschten Ziel. Daraus errechnet es mit Hilfe des programmierten Algorithmus die schnellste Route. Mit steigender Rechenleistung von Computern können immer umfangreichere Algorithmen geschrieben werden. So ist es möglich, Big Data nach Mustern und Zusammenhängen zu durchsuchen, die Aufschluss über Nutzerverhalten und Kaufentscheidungen geben.

## Algorithmus

Kette von Handlungs-  
anweisungen zum  
Lösen eines  
Problems



### Augmented Reality

Der Begriff Augmented Reality, kurz AR, bedeutet "erweiterte Realität" und beschreibt eine computergenerierte Erweiterung realer Wahrnehmungsbilder. Hierbei wird also die Darstellung eines realen Ereignisses um virtuelle Informationen oder Objekte ergänzt. Grundsätzlich kann die Erweiterung dabei auf jede Sinneswahrnehmung des Menschen abzielen. In der Praxis basiert AR-Technologie bislang aber in erster Linie auf eine Erweiterung von auditiven und allen voran visuellen Informationen.

Als Beispiel für die praktische Anwendung von Augmented Reality können Sportereignisse aufgeführt werden. Hier gehört das AR-Prinzip bei TV-Übertragungen längst zum Standard. So werden den Zuschauern bei Fußballübertragungen beispielsweise virtuelle Abseitslinien eingeblendet, die die Abseitsposition eines Spielers verdeutlichen bzw. widerlegen sollen. Bei Wettläufen kann dagegen etwa die Platzierung und die Zeit der einzelnen Läufer in Echtzeit eingeblendet werden. Bei American-Football-Spielen wird die First-Down-Linie virtuell eingeblendet.

Abseits von Sportübertragungen sind unter anderem auch moderne Smartphones und Tablets in der Lage, Kameraaufnahmen der realen Umgebung virtuell zu erweitern. Das Ziel der Anwendung von AR-Technologie variiert hier stark und kann die Installation entsprechender Apps voraussetzen oder macht spezielle Hardware erforderlich. Google Glass ist beispielsweise eine bekannte AR-Hardware. Der als Brille getragene Minicomputer ermöglicht die Erweiterung realer Bilder durch virtuelle Informationen. Wegbeschreibungen von Google Maps können direkt im echten Sichtfeld des Nutzers eingeblendet werden, Sehenswürdigkeiten oder Restaurantbewertungen sind mit einem Blick direkt wahrnehmbar.

Die erweiterte Realität wird oftmals mit dem Prinzip der virtuellen Realität, kurz VR, gleichgesetzt. Es handelt es sich dabei allerdings um gegensätzliche Konzepte. Während bei AR die Wahrnehmung der tatsächlichen Realität lediglich um bestimmte virtuelle Informationen oder Objekte erweitert wird, wird dem Anwender bei der VR-Technologie eine komplett virtuelle Umgebung präsentiert. AR ergänzt die Realität also virtuell, während VR künstliche Realitäten schafft.

## Automatisierung

Die Automatisierung beschreibt Arbeitsprozesse, die von computerbasierten Maschinen weitestgehend eigenständig durchgeführt werden. Der Begriff wird dabei sowohl für den Prozess selber, als auch für das Resultat bzw. das Produkt des automatisierten Prozesses verwendet. Der automatisierte Arbeitsprozess kann als Gegenteil der manuellen Arbeitsweise angesehen werden. Während die Automatisierung also vordergründig auf computerisierten Maschinen basiert, wird die manuelle Arbeit größtenteils durch Menschen getätigt.

Das Prinzip der Automatisierung verfolgt immer das Ziel, Arbeitsabläufe hinsichtlich Schnelligkeit, Effizienz, Sparsamkeit und Sicherheit zu optimieren. Sowohl mechanisch-manuelle Arbeitsabläufe werden automatisiert als auch digitale Prozesse. In diesem Zusammenhang spielt die künstliche Intelligenz eine große Rolle, welche der Maschinen oder Programme des Automatisierungssystems die Durchführung autonomer Prozesse erlaubt. Automatisierte Druckausleitungen von Versandkatalogen auf Grundlage von Online-Shop-Sortimenten sind heute genauso gängig, wie automatisierte Buchhaltungs- und Rechnungssysteme. Im sozialwissenschaftlichen Kontext wird die Automatisierung auch unter dem Gesichtspunkt der Arbeitspolitik diskutiert. Da Automatisierungssysteme effizientere Arbeitsprozesse versprechen und damit zu generell besseren Unternehmensresultaten führen, verdrängt die Automatisierung zunehmend die manuelle Arbeitsweise und macht somit viele menschliche Arbeitsplätze obsolet.

aber von Menschenhand gesteuert. Für die Erreichung des Ziels wird also weiterhin vordergründig die Verrichtung von manueller Arbeit benötigt. Im Gegensatz dazu wird beim vollständig automatisierten Arbeitsprozess der gesamte Arbeitsprozess maschinell durchgeführt. Allenfalls für die Überwachung des Automatisierungssystems kann ein menschlicher Input benötigt werden.

## Big Data

Mit Big Data sind alle digitalen Daten gemeint, die in quantitativ riesigen Mengen gesammelt werden. Große Datenmengen werden in fast jedem unternehmerischen Kontext gewonnen und von Unternehmen zunehmend strategisch genutzt. Die Datenmenge ist in der Regel groß, unstrukturiert und komplex.

In Abgrenzung zu anderen Datenmengen muss Big Data nach der allgemein akzeptierten Definition folgende drei Attribute erfüllen, um tatsächlich als Big Data zu gelten:

- High-Volume (Großes Datenvolumen)
- High-Velocity (Hohe Geschwindigkeit der Datengenerierung und des Datentransfers)
- High-Variety (Große Bandbreite an verschiedenen Datentypen und Datenquellen)

Aufgrund ihres Volumens, ihrer Komplexität und Unstrukturiertheit eignen sich die Daten nicht für die Auswertung durch konventionelle Methoden der Datenverarbeitung. Um die Massendaten auswerten zu können werden digitale und automatisierte Prozesse benötigt.



DGQ

Deutsche Gesellschaft  
für Qualität

## Automatisierung

Arbeitsprozesse, die  
von computerbasierten  
Maschinen  
eigenständig  
durchgeführt  
werden

Die Automatisierung wird manchmal fälschlicherweise mit der Mechanisierung gleichgesetzt. Die Mechanisierung beschreibt allerdings eine Art von manueller Arbeit, bei der ein menschlicher Arbeiter eine Maschine beim Arbeitsprozess verwendet. Das Arbeitsziel wird hierbei also ebenfalls mit maschineller Hilfe erreicht, die Maschine wird

Die Daten können aus verschiedensten Quellen stammen. In der Vergangenheit waren vor allem statistische Ämter und Regierungsbehörden für die Sammlung von Daten und Informationen verantwortlich. Heutzutage entstehen die größten Datenströme im Internet. Die Massendaten entwickeln sich etwa durch die Nutzung von Internet-basierten

Kommunikations- und Bezahlssystemen oder durch die Verwendung von speziellen Apps, die solche Daten aussenden. Auch die sozialen Netzwerke oder Fertigungsprozesse von vernetzten Produktionsanlagen stellen eine wichtige Quelle für Big Data dar.

Durch die digitale Datenverarbeitung lässt sich Big Data entsprechend auswerten und aufbereiten. Hierdurch können die nun äußerst wertvollen Daten für unterschiedliche Zwecke nutzbringend verwendet werden. Mit dem Wissen, welches durch die Nutzung von Daten im Fertigungsprozess generiert wird, lassen sich Produkte und Prozesse verbessern. Darüber hinaus entstehen völlig neue Geschäftsmodelle – unabhängig von der jeweiligen Branche. Produkte oder Services können z. B. anhand von Nutzungsdaten oder Kundenfeedback in sozialen Netzwerken maßgeschneidert entwickelt werden. So ist es möglich, Kundenbedürfnisse frühzeitig zu erkennen und zu analysieren.

Die Gefahr des Missbrauchs der oftmals personenbezogenen Daten ist allerdings hoch. Die unregulierte Verwendung von Big Data zur privatwirtschaftlichen Gewinnmaximierung stößt bei Datenschützern daher auf Kritik.

### Blockchain

Auf die Frage „Was ist eine Blockchain?“ gibt es eine zunächst eine einfach klingende Antwort: eine dezentrale Datenbank.

Aus technischer Perspektive ist eine Blockchain eine dezentrale Datenbank, die auf viele verschiedene Computer verteilt ist. Diese sind über das Internet vernetzt und bilden ein Blockchain-Netzwerk. In der so geteilten Datenbank werden Informationen hinterlegt, die für jeden Teilnehmer dieser Blockchain sichtbar sind. Anschaulich erklären lässt sich die Funktionsweise der Blockchain an der bekanntesten Anwendung: Bitcoin. Durch eine Transaktion entsteht ein neuer Datensatz („Block“). Ein Datenblock enthält alle wesentlichen Informationen über die Art der Transaktion, die Transaktionsdaten und einen exakten Zeitstempel. Die im Block gebündelten Informationen werden nun verschlüsselt. Hierzu werden kryptographische Verfahren angewendet, die aus diesen Informationen einen einzigartigen „Hash“ machen. Durch einen Konsensmechanismus wird der Block verifiziert und die Transaktion im Netzwerk dezentral von allen Teilnehmern bestätigt. Ein großer Vorteil der Blockchain besteht darin, dass Intermediäre – beispielsweise Banken – überflüssig werden, weil die Bestätigung einer Transaktion über das dezentrale Netzwerk funktioniert. Dadurch lassen sich beispielsweise Transaktionskosten drastisch senken. Ein weiterer Vorteil liegt in der hohen Manipulationssicherheit der Blockchain. Da in einer Blockchain alle jemals getätigten Bitcoin-Transformationen und deren Reihenfolge in den Datenblöcken vorliegen und diese sich dezentral auf allen Computern im Netzwerk gleichzeitig befinden, sind sie vor Manipulation geschützt. Dieser Schutz funktioniert, weil in seine techni-

sche Lösung ein ökonomisches Anreiz-System implementiert wurde. Wer dem Bitcoin-Netzwerk Rechenleistung zur Verfügung stellt und dadurch hilft, es gegen Manipulation zu schützen, wird finanziell belohnt. So ist es für den Einzelnen lukrativer sich dem Bitcoin-Netzwerk anzuschließen, anstatt es anzugreifen oder zu manipulieren. Dieses dezentrale Konsensverfahren stellt somit einen Hauptsicherheitsfaktor der Blockchain dar.

### Cloud Computing

Unter Cloud Computing versteht man die Auslagerung von zentraler IT-Infrastruktur in eine dezentrale Cloud. Während früher alle Computer, Speicherplätze, Programme oder Dateien vor Ort sein mussten, um den Zugriff zu gewährleisten, ist heute die Nutzung „fremder“ Anwendungen und Rechenleistung problemlos über eine Cloud möglich. Den Anfang machte dabei dezentraler Speicherplatz, der von Anbietern im Internet bereitgestellt wurde. Dateien, die in einer solchen Cloud gespeichert sind, stehen für verschiedene Anwender zeitgleich an unterschiedlichen Orten bereit.

Industriezweige mit rechenintensiven Anwendungen benötigen nicht mehr zwingend eigene Rechenzentren vor Ort, sondern können die Ressourcen von Cloud-Dienstleistern nutzen. Auch Software wird mittlerweile häufig als Cloud-Dienstleistung angeboten. Die Anschaffung und Installation eines physischen Produkts auf nur einem Gerät ist oft nicht mehr nötig. Anstatt ein Programm für einen festen Preis zu kaufen, erwerben die Nutzer eine Lizenz zur Anwendung der Software für einen bestimmten Zeitraum. Da die Software auf einem Server des Cloud-Netzwerks installiert ist, sparen Anwender Speicherplatz auf Ihren eigenen Geräten. Auch müssen sich Unternehmen, die Clouds für Anwendungen oder Speicherplatz nutzen, nicht um die Instandhaltung und Pflege der Server kümmern.

Beim Cloud Computing wird unterschieden zwischen öffentlichen und privaten Angeboten. Die Dienstleistungen der Public Cloud sind für jedermann verfügbar und können für eine bestimmte Zeit gemietet und genutzt werden. Der Preis richtet sich dabei in der Regel nach den benötigten Ressourcen und für wie lange diese gemietet werden. Eine Private Cloud ist eine Plattform, die ausschließlich für ein einzelnes Unternehmen oder eine Organisation betrieben und auch nur von dieser genutzt werden kann. Hier zahlen diese für den gesamten Aufbau und Unterhalt der Infrastruktur, wodurch die Kosten konstant und nicht mehr abhängig von der tatsächlichen Nutzung sind. Neben zahlreichen Vorteilen des Cloud Computings, wie Kostenersparnis und Standortunabhängigkeit müssen auch mögliche Risiken beachtet werden. Vor allem die Frage nach der angemessenen Sicherheit sollte geklärt werden. Werden sensible oder personenbezogene Daten in einer dezentralen Cloud gespeichert ist der Standort des Anbieters relevant. Befindet dieser sich im EU-Ausland, gelten beispielsweise keine europäischen Datenschutzrichtlinien.

## Cyber-physische Systeme

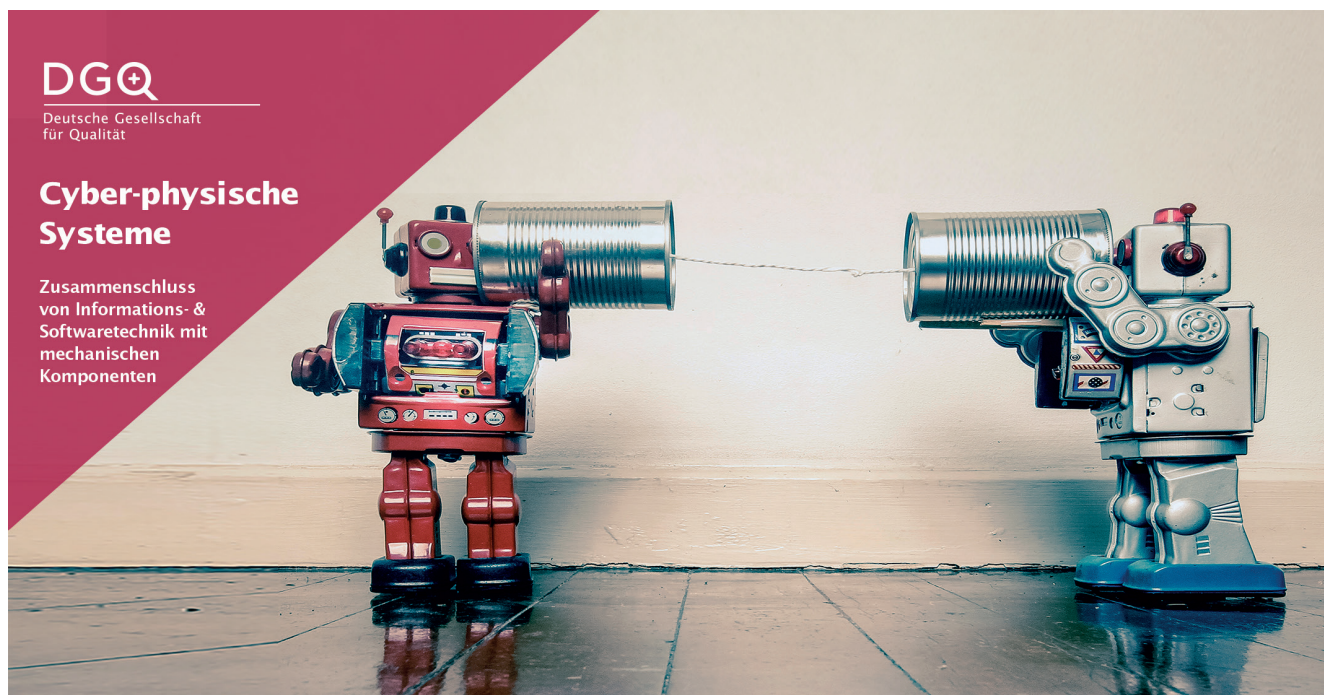
Als cyber-physisches System (CPS) wird der Zusammenschluss von Informations- und Softwaretechnik mit mechanischen Komponenten bezeichnet. Ein einfaches Beispiel für ein CPS sind dabei etwa automatisch öffnende Türen, die über optischen Sensor die Ankunft eines Menschen registrieren und den Öffnungsmechanismus in Gang setzen. Die Übertragung aus den Informations- und Sensorsystemen wird dabei in Echtzeit verarbeitet und übertragen, sodass solche Systeme völlig autonom arbeiten können.

Meist ist die Wahrnehmung und Datenverarbeitung von cyber-physischen Systemen dabei nicht auf ihre unmittelbare Umgebung beschränkt. Stattdessen sind sie mit dem Internet oder anderen Systemen verbunden und verfügen dadurch über eine große Menge unterschiedlicher Daten,

Auch im privaten Bereich gewinnen cyber-physische Systeme an Bedeutung. Im sogenannten Internet der Dinge kommunizieren unterschiedlichste Geräte miteinander, um ihren Nutzern ein optimales Anwendungserlebnis zu garantieren. Diese müssen sich dadurch weniger Gedanken über die Wartung oder Instandhaltung machen, sondern werden beispielsweise über spezielle Apps oder Webinterfaces benachrichtigt, wenn ihr Eingreifen erforderlich ist. Das autonome Fahren ist eine der bekanntesten Anwendungen für cyber-physische Systeme.

## Data Lake

Viele Konzerne sind auf die Erhebung und Auswertung von Big Data angewiesen, um Produkte und Dienstleistungen erfolgreich bewerben und verkaufen zu können. Big Data aber sind besonders schnelllebige Massendaten, die weltweit im Sekundentakt durch jede Aktion und Interaktion



DGQ

Deutsche Gesellschaft  
für Qualität

## Cyber-physische Systeme

Zusammenschluss  
von Informations- &  
Softwaretechnik mit  
mechanischen  
Komponenten

die jeweils in Echtzeit aktualisiert werden. Der Mensch greift in diese Systeme entweder über Schnittstellen wie Web-Applikationen, oder über Schnittstellen an der physischen Komponente ein.

Immer größere Wichtigkeit erhalten cyber-physische Systeme in der Industrie 4.0. Sie sind ein wesentlicher Bestandteil von smarten Produktionsketten, die intelligent auf Veränderungen in der Nachfrage oder den Produktionsbedingungen reagieren können. Auch in den Bereichen Logistik und Mobilität übernehmen CPS zunehmend die Steuerung von Abläufen und deren Anpassung an unterschiedliche Bedingungen. Ihr Vorteil ist, dass sie rund um die Uhr verfügbar und trotz hoher Anschaffungskosten günstig im Unterhalt sind. Auch arbeiten cyber-physische Systeme in vielen Bereichen genauer, effizienter und mit einer geringeren Fehlerquote als Menschen.

im Internet generiert werden. Unternehmen, die an der Auswertung von Big Data interessiert sind, brauchen eine Möglichkeit, diese Masse an Daten in unveränderter Rohform zu speichern. Data Lakes (auch Datenseen genannt) sind Datenspeicher, die die massenhafte Speicherung von generierten Big Data im Ursprungsformat erlauben. Die Daten müssen dabei nicht strukturiert sein und können entweder auf einer Text- oder Zahlengrundlage gespeichert werden, oder rohe Video- und Bilddateien sein. Sobald die Analyse und Auswertung bestimmter Daten aus dem Datensee erfolgen soll, können Anwender gezielt auf gespeicherte Daten zugreifen.

Der größte Vorteil von Datenseen ist, dass gesammelte Daten nicht wie in der unternehmerischen Praxis oft üblich auf vereinzelt Computern und Festplatten in dezentralen Datensilos liegen, wo sie unzugänglich für andere Nutzer sind. Stattdessen werden Daten im zentral zugänglichen Data Lake vorgehalten und können jederzeit und je nach tatsächlichem Bedarf von Nutzern ausgewertet, strukturiert

## Data Mining

Methodische Anwendung  
statistischer Methoden  
auf große  
Datenbestände  
zur Generierung  
von neuem  
Wissen



riert und formatiert werden. Da Daten im Data Lake im ursprünglichen Rohformat gespeichert werden, erfordern Speicherprozesse keine große Rechenleistung.

Als Möglichkeit zur massenhaften Datenspeicherung sind Data Lakes ähnlich den sogenannten Data Warehouses. Auch im Datenwarenspeicher werden große Mengen Daten gespeichert und zur Auswertung vorgehalten. Der größte Unterschied ist aber, dass ein Data Warehouse gespeicherte Daten aus unterschiedlichen Quellen automatisch in auswertungsfertigen Formaten und Strukturen speichert. So überführt das Data Warehouse Daten direkt in für unterschiedliche Nutzer sofort verwertbare Formate. Daten im Data Lake werden hingegen weder strukturiert noch kategorisiert oder formatiert.

### Data Mining

Data Mining bezeichnet das methodische Anwenden statistischer Methoden auf große Datenbestände um Zusammenhänge herzustellen und Erkenntnisse zu erlangen. Die eigentliche Datenverarbeitung geschieht computergestützt und mit Hilfe von modernen Machine-Learning-Algorithmen und Künstlicher Intelligenz. Der Begriff des Data Mining wird irrtümlich häufig mit Datenabbau gleichgesetzt, was implizieren würde, dass es sich dabei um einen Prozess zur Generierung von Daten handelt. Dies ist nicht korrekt, denn Data Mining beschreibt den Wissensabbau aus vorhandenen Daten. Die Abbauprodukte sind demnach Wissen, Erkenntnisse und Zusammenhänge, auf deren Grundlage unternehmerische Entscheidungen getroffen werden können.

Data Mining wird von Data Scientists betrieben und beruht auf mathematischen, statistischen und informationstechnologischen Erkenntnissen. Um aus Daten Wissen zu extrahieren, müssen Data Scientists die vorhandenen Daten

bearbeiten. Dazu werden entsprechend des gewünschten Erkenntnisinteresses bestimmte Zieldatensätze ausgewählt und mit Methoden der Data Science bereinigt, formatiert, strukturiert und transformiert, bis Muster und Trends erkennbar sind, die belastbare Schlussfolgerungen zulassen. Wenn aus massenhaften Rohdaten durch diese Methoden begründete Erkenntnisse gewonnen werden konnten, war der Wissensabbau erfolgreich.

Data Mining spielt besonders im Zusammenhang mit unstrukturierten und schnelllebigem Big Data eine bedeutende Rolle, wird aber auch für alle anderen Arten von Datensätzen und Datenblöcken verwendet. Eine Sonderform des Data Mining ist das Text Mining. Hier werden Informationen aus einer großen Zahl an Textdaten gewonnen. Mithilfe von intelligenten Algorithmen und maschinellem Lernen ist es möglich, Aussagen und Kernthesen innerhalb kürzester Zeit aus Texten zu extrahieren. Informationsgewinn aus Text Mining besitzt ein enormes wirtschaftliches Potenzial, denn ein Großteil aller geschäftlich relevanten Informationen und Kommunikation liegt unternehmensintern als Text in Form von E-Mails, Dokumenten und Handlungsanweisungen vor. Die sekundenschnelle Wissensextraktion aus diesen Texten kann unternehmerische Prozesse maßgeblich optimieren.

### Data Science

Data Science beschreibt Datenwissenschaft die dazu dient, Erkenntnisse, Einblicke und Wissen aus großen Datenmengen zu gewinnen. Die Data Scientists von heute sind keine Wissenschaftler im eigentlichen Sinne des Wortes, sondern vielmehr hochqualifizierte Datenexperten mit IT-, Mathematik- und Informatik-Hintergründen. Sie sind branchenübergreifend in Unternehmen angestellt und lösen komplexe Datenprobleme. Warum aber setzen Konzerne auf Erkenntnisse aus Data Science?

Digitale Informationen und Daten werden heute weltweit und zu jeder Zeit generiert. Dies geschieht durch Online-Einkäufe, bargeldlose Bezahlungen, Suchmaschinenabfragen und natürlich auch durch Interaktionen von Millionen Nutzern auf Social-Media-Plattformen. Riesige Ströme digitaler Daten sind die unsichtbaren Begleiterscheinungen des modernen Lebens. Data Scientists sind darauf spezialisiert, aus rohen, unstrukturierten und schnelllebigem Big Data nutzbare Erkenntnisse und gewinnbringende Informationen zu extrahieren. So wandeln sie Big Data in Smart Data um und machen verborgene Trends, Querverbindungen und Potenziale, die in den globalen Datenströmen versteckt sind, sichtbar.

Zur Verarbeitung und Analyse von Daten nutzen Data Scientists Machine-Learning-Algorithmen, konfigurieren Künstliche Intelligenzen und schaffen generelle Dateninfrastrukturen in Unternehmen. So können je nach Branche und unternehmensspezifischem Erkenntnisinteresse datengestützte Optimierungen des Geschäfts vorgenommen werden. Statistisch fundierte Entscheidungen auf Grundlage belastbarer Datenauswertungen haben das Potenzial, Unternehmensprozesse effektiver, kostensparender und profitabler zu machen.

## Digitalisierung

Dieser Beitrag des DGQ-Glossars greift den Überbegriff „Digitalisierung“ auf. Diesen oft benutzten und vielseitig verwendeten Begriff erläutert unser Experte Dr. Benedikt Sommerhoff. Er ist Leiter Innovation & Transformation der DGQ und beschäftigt sich in seiner alltäglichen Arbeit mit den grundlegenden Paradigmen und Dynamiken hinter der Digitalisierung. Vor allem geht es ihm darum, aus Sicht der Qualitätssicherung und des Qualitätsmanagements der Frage nachzugehen, wie Unternehmen die richtigen Schlüsse ziehen und zukunftsfähige Entwicklungspfade betreten können.

Der Begriff digital ist seit Jahrzehnten im Gebrauch. Digital kommt vom lateinischen digitus für Finger und bezeichnet ein Signal mit abgestuften Werten. Ein Sonderfall ist das binäre System, das mit zwei Werten operiert und auf dem die Computertechnik basiert. Analog, altgriechisch für verhältnismäßig, bezeichnet kontinuierliche Verläufe. Digital und analog sind Schlüsselbegriffe für neue und alte Welt geworden. So ist „digital natives“ die Bezeichnung für die Generation geworden, die von Kindheit an mit digitaler Technik (Niveau Smartphone) groß geworden sind. Dann gibt es noch die „digital immigrants“, ältere Jahrgänge, die



DGQ

Deutsche Gesellschaft  
für Qualität

## Digitalisierung

Transformationsprozess  
von klassischer zu  
moderner Technik  
gestützt auf  
digitale  
Technik

Bei der Data Science geht es nicht generell um die Auswertung und Analyse aller von Unternehmen gesammelten Daten, sondern per Definition nur um die Extraktion von Wissen aus komplexen und unstrukturierten Datenmengen. Zur Auswertung bereits strukturierter Google-Analytics-Daten zum Besucherverhalten auf einer Unternehmenswebsite bedarf es keiner Data-Science-Kenntnisse. Die sinnvolle Verknüpfung von millionenfachen unstrukturierten Massendaten aus verschiedenen Quellen und in unterschiedlichen Formaten zur Ableitung von Entscheidungsempfehlungen hingegen, ist ohne Data Scientists nicht möglich.

sich als Erwachsene die Nutzung digitaler Technik angeeignet haben und die „digital ignorants“, von denen die meisten fälschlicherweise behaupten, ohne digitale Technik leben zu können, nicht wissend oder leugnend, dass diese ohnehin auch ihren Alltag - oft im Verborgenen - schon prägt.

Digitalisierung im engeren Sinne ist der Prozess der Transformation von klassischer zu moderner Technik gestützt auf digitale Daten. Ergänzend dazu gibt es auch die typischen Begriffskombinationen digitale Transformation, digitaler Wandel, digitale Disruption. Die eigentliche Digitali-

sierung erfolgte mit Beginn der Schrift und Mathematik vor tausenden Jahren und im oft eher gemeinten Sinne einer Computerisierung und der zunehmenden Nutzung digitaler Signale verstärkt ab den 1960er Jahren. Wenn seit einigen Jahren erneut von Digitalisierung gesprochen wird, dann ist damit eher nicht eine Erstdigitalisierung oder Erstcomputerisierung gemeint, sondern die Nutzung der digitalen Technik der neuesten Generation mit großer Leistungsfähigkeit und somit das Überschreiten einer neuen, deutlich höheren Digitalisierungsschwelle. Begriffskombinationen wie Digitalisierung der Gesellschaft, der Schule, der Lehre, der Kirche, des Gesundheitswesens usw. zeigen, dass die Menschen den Begriff Digitalisierung bei weitem nicht nur technisch oder technologisch, sondern gesellschaftsbezogen verwenden.

Es gibt drei Stufen der Digitalisierung

1. Digitisierung (digitization): Die Umwandlung analoger in digitale Daten
2. Digitalisierung (digitalization):

Nutzung fortgeschrittener digitaler Techniken in den Prozessen, wodurch sich die Prozesse gegenüber nichtdigitalen Vorläufern signifikant verändern

3. Digitale Transformation (digital transformation): Die Schaffung neuartiger Geschäftsmodelle, die sich maßgeblich auf digitale Techniken stützen.

Unter Digitisierung ist nicht die Umwandlung in irgendein digitales Format (z. B. Word, Excel) gemeint, sondern in ein Format, das mit modernen Datensystemen unmittelbar weiterverarbeitet werden kann.

Eine basale Digitalisierung der Prozesse findet seit Jahrzehnten statt. Die eigentliche Digitalisierung nutzt signifikant fortschrittliche Technologien, mit denen andersartige Prozesse möglich sind. Die digitale Transformation ist die Königsdisziplin, hier geht es um die Existenzsicherung des Unternehmens.

### Internet der Dinge

Das Internet der Dinge, kurz IdD, ist auch unter dem englischen Begriff „Internet of Things“, kurz IoT, bekannt. Der Begriff umfasst Dinge des täglichen Gebrauchs, die sowohl untereinander als auch mit ihren Nutzern kommunizieren. Menschen und Geräte sind durch das Internet der Dinge miteinander vernetzt und können so zusammenarbeiten. Das IdD lässt sich auf zahlreiche Lebensbereiche ausweiten, zum Beispiel in Unternehmen oder im Privatbereich.

Die konkrete Anwendung wird in die industrielle und private Nutzung des Internet der Dinge unterschieden. Bei der privaten Nutzung des Internet der Dinge ist folgendes Szenarien denkbar: Ein Smartphone-Nutzer befindet sich zufällig in der Nähe eines Supermarktes. Die GPS-Daten des Smartphones des Nutzers werden automatisch an den intelligent vernetzten Kühlschrank im Haus des Anwenders geleitet. Dieser stellt ein Inventar der vorrätigen Lebensmittel auf und informiert den Nutzer, dass bestimmte Lebensmittel bald nachgekauft werden sollten.

In der Industrie bezieht sich das IdD vor allem auf die Kooperation von Menschen und Produktionsanlagen. Es stellt eine Grundlage der Industrie 4.0 dar, indem es die Kommunikation von Fachkräften und Maschinen ermöglicht. Ein Beispiel dafür könnte eine Fertigungsanlage sein, die feststellt, wenn Produktionsmaterialien zur Neige gehen. Ist ein kritischer Niedrigstand erreicht, der möglicherweise Produktionsengpässe oder -ausfälle nach sich ziehen könnte, erhält ein kooperierender Mitarbeiter über eine Schnittstelle die Aufforderung, neue Materialien zu bestellen. So werden Produktionsabläufe optimiert und Unterbrechungen in der Fertigung minimiert.

Kritik erfährt das IdD vor allem auf Grund von Sicherheitsrisiken. Hackerangriffe und Überwachungsszenarien sind möglich. Sichere Passwörter, geschützte Verbindungen, umfangreiche Verschlüsselungsmechanismen sowie gut geschultes Personal sind deswegen unabdingbar.

### Industrie 4.0

Industrie 4.0 ist ein Kunstbegriff, mit dem die vierte industrielle Revolution Entscheidungsträgern in Politik und Wirtschaft ausgerufen wurde. Dabei geht es kurz gesagt um die Digitalisierung von industriellen Abläufen. Industrielle Prozesse werden eng mit zukunftsweisenden Technologien aus der Informations- und Kommunikationstechnik verbunden und sollen so den Standort Deutschland und deutsche Unternehmen auch in Zukunft wettbewerbsfähig machen.

Mithilfe von Industrie 4.0 soll es ermöglicht werden, Produkte und Prozesse von der Idee bis hin zur Fertigung und darüber hinaus, zu optimieren. Die intelligente Vernetzung von Menschen und Maschinen liefert unterschiedliche Vorteile. Die Unternehmen, die an der Fertigung eines Produktes beteiligt sind, können durch Vernetzung einzelne Arbeitsschritte schneller und vorausschauender aufeinander abstimmen. Produkte werden individualisiert und besser auf den Konsumenten zugeschnitten. Einerseits durch die Auswertung von Nutzerdaten, andererseits durch eine erhöhte Wirtschaftlichkeit in der Produktion. Produktionsanlagen sind im Rahmen der Industrie 4.0 z. B. in der Lage, zu kommunizieren, wenn sie neue Arbeitsmaterialien benötigen oder melden frühzeitig, wenn Wartungsarbeiten notwendig sind. So werden nahtlose Produktionsprozesse möglich, da menschliche Servicekräfte eingreifen können, bevor es zum Ausfall der Produktionsanlage kommt. Maschinenausfälle, Kosten und Wartezeit werden dank Informationstransparenz minimiert. Doch Industrie 4.0 bezieht sich nicht ausschließlich auf den industriellen Bereich, sondern umfasst auch die Felder Gesundheit, Mobilität und Klima. Beispiele dafür sind selbstfahrende Autos, Roboter, die Fachkräfte in der Pflege unterstützen sowie elektronische Patientenakten.

Kritikpunkte an Industrie 4.0 sind unter anderem moralische Fragen wie die Substitution von Fachkräften durch Maschinen sowie das Risiko, dass Maschinen komplexe Vorgänge fehlinterpretieren oder Angriffen durch Hacker zum Opfer fallen könnten. Zusätzlich werden der hohe Kostenaufwand sowie die Gefahr des 'gläsernen Bürgers' als Kritik genannt.

## Künstliche Intelligenz

Als Künstliche Intelligenz, kurz KI, bezeichnet man den Versuch, menschliches Denken, menschliche Wahrnehmung und menschliche Entscheidungsstrukturen durch Computer nachzubilden.

Die älteste Definition einer KI stammt aus den 1950er Jahren und wurde von dem britischen Mathematiker und Informatiker Alan Turing formuliert. Seiner Definition nach ist eine Maschine dann intelligent, wenn ein menschlicher Gesprächspartner nicht eindeutig sagen kann, ob sein Gegenüber ein Mensch oder eine Maschine ist. Dieser Test ist auch als Turing-Test bekannt.

Zunächst ist eine große Menge an Daten notwendig. Auf diese Daten kann die Künstliche Intelligenz dann zurückgreifen, wenn sie Entscheidungen treffen muss. Mit jeder getroffenen Entscheidung hat die KI dazu gelernt und ist „intelligenter“ geworden. Man unterscheidet zwischen einer schwachen und einer starken KI. Die schwache KI kann klar definierte Probleme auf Basis mathematischer Algorithmen lösen. Die Herangehensweise an Probleme bleibt dabei immer gleich. Die starke KI besitzt vergleichbare intellektuelle Fähigkeiten wie der Mensch. Sie handelt aus eigenem Antrieb heraus intelligent. Bislang ist es jedoch nicht gelungen, eine solche starke KI zu entwickeln. Es gibt Wissenschaftler, die dies für unmöglich halten. Die Mehrheit der Wissenschaftler ist sich jedoch einig, dass eine starke KI entwickelt werden kann und dass dies in den nächsten 20 bis 40 Jahren geschehen wird.

Für Künstliche Intelligenz gibt es bereits viele Anwendungsbeispiele. Die bekanntesten sind sicherlich Suchmaschinen im Internet oder die Spracherkennung von Smartphones. In der Automatisierung von Arbeitsprozessen spielen KI eine besondere Rolle und ermöglichen effiziente Prozessoptimierungen. Börsenkursentwicklungen werden bisweilen durch KI-basierte Programme analysiert; Data-

Mining macht auf Grundlage von KI die Destillation von Kerninformationen aus riesigen Text- und Datenblöcken möglich; Autonomes Fahren basiert im Kern auf den intelligenten Entscheidungsauswertungen von KI-Systemen.

## Machine Learning

Als Machine Learning, oder auch maschinelles Lernen, bezeichnet man die digitale Bildung von Wissen aus Erfahrung. Hierfür wird ein Computersystem mit Informationen und Regeln versorgt und „lernt“ mit Hilfe komplexer Algorithmen, diese Informationen zu verallgemeinern. Anschließend ist das System in der Lage, selbstständig Lösungen für unbekannte Probleme zu finden. Machine Learning ist somit ein Teilgebiet der Künstlichen Intelligenz.

Grundsätzlich werden drei Lernkategorien unterschieden. Das überwachte Lernen dient dazu, aus einem Trainingsdatensatz mit vorgegebenen Ein- und Ausgaben eine Funktion abzuleiten, mit der dann bei späteren Eingaben die Ausgaben selbst berechnet werden können. Beim unüberwachten Lernen wird hingegen keine Ausgabe vorgegeben. Hier muss der Computer selbstständig die Daten analysieren und auf Basis von vorhandenen Mustern gruppieren. Wie diese Gruppen schließlich aussehen, ist vorher nicht bekannt. Das bestärkende Lernen ist dem menschlichen Lernen sehr ähnlich. Hierbei wird das System belohnt, wenn eine Zuordnung richtig erfolgt ist und bestraft, wenn es einen Fehler gemacht hat.

Da für das Machine Learning große Datensätze notwendig sind, kommen verteilte Rechnerstrukturen und künstliche neuronale Netze zum Einsatz. Die Anwendungsmöglichkeiten für Machine Learning sind sehr vielfältig und finden längst großflächige Anwendung in der digitalen Welt. Beispielsweise erkennt der Spam-Filter eines Mailprogramms auf Basis der bereits markierten E-Mails andere Nachrichten ebenfalls als Spam. Auch Produkt- oder Filmempfehlungen bei Online-Shops oder Streaming-Diensten oder die



Gesichtserkennung in sozialen Netzwerken basieren auf Machine Learning-Algorithmen. Besonders tiefgehende maschinelle Lernprozesse werden als Deep Learning bezeichnet. Hier sorgen komplexe innere Strukturen künstlicher neuronaler Netze mit vielen verschiedenen Lernschichten dafür, dass tiefgehende Lernerkenntnisse erreicht werden können.

### **Mensch-Maschine-Interaktion**

Als Mensch-Maschine-Interaktion wird die Interaktion bzw. Kommunikation von einem Menschen mit einer Maschine, heute in der Regel mit einem Computer bezeichnet. Diese Kommunikation erfolgt über eine sogenannte Schnittstelle. Ein ganz einfaches Beispiel hierfür ist ein Schalter. Wird dieser betätigt, wird dadurch eine Aktion ausgelöst. Die Norm ISO 9241 „Ergonomie der Mensch-System-Interaktion“ beschreibt Anforderungen an die Arbeitsumgebung, Hard- und Software. Sie enthält unter anderem „Leitlinien für die Individualisierung von User Interfaces“, „Grundsätze und Anforderungen für physikalische Eingabegeräte“ oder „Empfehlungen für die taktile und haptische Interaktion“. Hauptziel beim Design solcher Schnittstellen ist es, die Kommunikation mit dem Computer für den Nutzer möglichst intuitiv und einfach zu gestalten. Sie kann z. B. per Texteingabe über eine Tastatur, einen Touchscreen oder per Sprachsteuerung erfolgen. Daraus ergibt sich, dass es sich bei der Mensch-Maschine-Interaktion um ein interdisziplinäres Feld aus Informatik, Sprachwissenschaften, Psychologie, Verhaltenswissenschaften, Design, Ergonomie und anderen handelt.

Ein ganz aktuelles Produktbeispiel sind die in den letzten Jahren immer weiter entwickelten Datenbrillen. Dabei handelt es sich um kleine, am Kopf getragene Computer, die Informationen zum Sichtfeld des Trägers hinzufügen. Außerdem kann man mit ihnen telefonieren oder Fotos und Videos aufnehmen. Eine Datenbrille kann durch Sprache oder ein Touchpad am Brillenbügel gesteuert werden. Mit Hilfe moderner, intuitiver Schnittstellen werden auch hochkomplexe Systeme, wie beispielsweise Produktionsstraßen einfach bedienbar. Eine gute Kommunikation zwischen Mensch und Maschine ist wichtig, um Missverständnisse und damit Fehler zu vermeiden. Denn bislang ist ungeklärt, wer die Verantwortung trägt, wenn es durch Probleme in der Mensch-Maschine-Interaktion zu Unfällen kommt.

### **Mensch-Roboter-Kooperation**

Die Mensch-Roboter-Kooperation, kurz MRK, bezeichnet die Zusammenarbeit einer menschlichen Arbeitskraft und eines Roboters. Welchen Zweck erfüllt diese Art der Zusammenarbeit? In nahezu jedem größeren Betrieb fallen Arbeiten an, die nur wenige Mitarbeiter wirklich gerne übernehmen – sei es, weil sie zu repetitiv, zu anstrengend oder zu gefährlich sind. Diese Arbeitsprozesse können deutlich angenehmer, effektiver und sicherer gestaltet werden, wenn die Mitarbeiter Unterstützung durch einen Roboter erhalten. Die vollständige Arbeitsübernahme durch Roboter ist aus Kostengründen und technischen Begrenzungen oft nicht möglich. An diesem Punkt ist eine Mensch-Roboter-Kooperation eine sinnvolle Alternative.

Die MRK bietet dabei gleich mehrere Vorteile. Einerseits haben die Mitarbeiter erhöhte Kapazitäten, sich anderen Aufgaben zu widmen. Andererseits führen besonders monotone Arbeiten häufig zu einer reduzierten Konzentrationsfähigkeit seitens der Mitarbeiter. Sinkt die Konzentration, steigt das Verletzungsrisiko. Dieses Risiko wird durch den unterstützenden Einsatz von Robotern erheblich vermindert. Des Weiteren bietet der Teileinsatz von Robotern den menschlichen Mitarbeitern die Möglichkeit, einzugreifen, sollte es zu Systemfehlern kommen.

Die Zusammenarbeit zwischen Mensch und Maschine existiert in unterschiedlichen Formen und wird daher auch in der Begriffsbezeichnung unterschieden. Im Vergleich zur Mensch-Maschine-Interaktion teilen sich bei der MRK Mensch und Roboter einen Teilbereich des Arbeitsraumes. Die Interaktion zwischen Mensch und Roboter ist dabei zwar nicht zwangsläufig erwünscht, aber dennoch denkbar. Bei anderen Formen der Mensch-Maschinen-Zusammenarbeit ist keine Interaktion gewünscht. Im Gegenteil: Sie wird durch Sicherheitsvorrichtungen und strikt getrennte Arbeitsräume zwangsläufig verhindert. Doch es gibt auch eine Art der Zusammenarbeit von Mensch und Roboter, die eine noch engere Zusammenarbeit vorsieht, die sogenannte Mensch-Roboter-Kollaboration. Dabei arbeiten Mensch und Roboter gemeinsam an einem Projekt, wobei Kontakt und Interaktion ausdrücklich vorgesehen sind.

### **Neuronale Netze**

Im Bereich der Neurowissenschaften versteht man unter neuronalen Netzen durch Synapsen miteinander verbundene Nervenzellen. Diesen biologischen Aufbau macht sich die Neurowissenschaft zu Nutze, indem sie diese Netze aus künstlichen Neuronen nachbildet. Die Rede ist dann von künstlichen neuronalen Netzen, kurz KNN. Künstliche neuronale Netze gehören in den Bereich der künstlichen Intelligenz, da sie unterschiedliche Probleme computerbasiert lösen.

KNN werden in technisch und biologisch orientierte Simulationen unterschieden. Während biologisch simulierte KNN neuronale Eigenschaften so realistisch wie möglich wiedergeben sollen, dienen technisch orientierte Simulationen vorrangig der Lösung technisch-mechanischer Problemstellungen. Das neuronale Netz besteht aus mehreren Neuronen, die in der Lage sind, Informationen aufzunehmen. Die Informationen werden extern bezogen, von den Neuronen im Netzwerk verteilt und entsprechend modifiziert. Anschließend wird eine Problemlösung ausgegeben. Ein KNN besteht üblicherweise aus Input-Neuronen, Hidden-Neuronen und Output-Neuronen. Die Input-Neuronen dienen der Informationsaufnahme. Die Hidden-Neuronen bilden intern Informationen in Form von Informationsmustern ab und die Output-Neuronen geben die Informationen aus.

Bevor KNN einsatzbereit sind, müssen sie einem entsprechenden Training unterzogen werden. Dabei wird zwischen überwachtem und unüberwachtem Lernen unterschieden. Überwachtes Lernen impliziert die Vorgabe eines konkreten Ergebnisses für die jeweiligen Eingabemöglichkeiten. Das aktuelle Ergebnis wird kontinuierlich mit dem gewünschten Ergebnis verglichen, wodurch das künstliche neuronale Netz lernt, die Neuronen wie gewünscht zu verknüpfen. Beim unüberwachten Lernen wird dem KNN kein Ergebnis vorgegeben. Es nimmt die notwendigen Informationen und Veränderungen ausschließlich durch die unterschiedlichen Eingabemuster auf.

Künstliche neuronale Netze kommen in unterschiedlichen Bereichen zur Anwendung. Typische Einsatzgebiete sind die Spracherkennung, die Schrifterkennung, die Bilderkennung sowie Frühwarnsysteme und maschinelle Übersetzungen. Doch auch für komplexe Prognosen, zum Beispiel im Bereich der Wettervorhersage oder der Diagnostik medizinischer Fragestellungen werden sie angewandt.

sprechend bestimmter Fragestellungen aufgearbeitet wird. Damit diese großen Mengen an Daten verstanden werden können, müssen sie geordnet und analysiert werden. Smart Data ist das Resultat. Man spricht in diesem Zusammenhang auch von unstrukturierten und strukturierten Daten.

Smart Data findet in unterschiedlichen Bereichen Anwendung. So hat beispielsweise das Bundesministerium für Wirtschaft und Energie das Programm „Smart Data – Innovationen aus Daten“ ins Leben gerufen. Im Rahmen dessen werden Projekte gefördert, die innovative Dienstleistungen entwickeln und voranbringen. Schwerpunkt des Programms sind die Bereiche Mobilität, Industrie, Gesundheit und Energie. Durch den Einsatz von Smart Data können in diesen Bereichen zum Beispiel die Versorgung von Patienten oder die nachhaltige Verkehrsplanung verbessert werden. Weitere Beispiele für Smart-Data-Lösungen sind Smartwatches oder digitale Energiezähler.



### Smart Data

Smart Data umfasst digitale Informationen, aus denen Anwender direktes Wissen ziehen können. Diese Datensätze werden aus einer großen Datenmenge extrahiert und bereits so aufgearbeitet, geordnet und analysiert, dass sie Anwendern brauchbare Informationen liefern und von diesen verstanden werden. Im Vergleich dazu produziert Big Data immer größere und umfangreichere Datensätze. Dabei werden Daten aus diversen Quellen bezogen, um ein möglichst großes Datenvolumen zu erreichen. Diese große Menge an Daten ist jedoch nur brauchbar, wenn sie ent-

Auch im Online Marketing spielt Smart Data eine zentrale Rolle. Social-Media-Daten und Daten aus Onlinekäufen werden gesammelt und können mithilfe von Smart Data noch präziser analysiert und ausgewertet werden. Als Ergebnis lassen sich exakte Zielgruppenprofile erstellen, die es dem Online Marketing ermöglichen, effektiv zu werben. Aus Big Data destillierte Smart-Data-Erkenntnisse können dann strategisch eingesetzt werden, um spezifische Social-Media-Nutzergruppen gezielt mit maßgeschneiderter Werbung, Angeboten, Nachrichten oder Veranstaltungsinformationen zu versorgen.

---

Verstehen.

---

Verbessern.

---

Verantworten.

---

August-Schanz-Str. 21A  
60433 Frankfurt am Main  
T 069 95424-0  
F 069 95424-133  
info@dgq.de  
[www.dgq.de](http://www.dgq.de)



---

Deutsche Gesellschaft  
für Qualität