

Praxishandbuch

KI-VO

Künstliche Intelligenz rechtskonform im privaten
und öffentlichen Bereich einsetzen

» Hier geht's
direkt
zum Buch

DIE LESEPROBE

1

Was ist KI und wie unterscheiden sich Datenwissenschaft und Datenanalytik?

Gabriele Bolek-Fügl

Künstliche Intelligenz steht im Mittelpunkt einer Revolution, die nicht nur die Art und Weise verändert, wie Maschinen denken und lernen, sondern auch unser Verständnis von Intelligenz selbst herausfordert. Diese Revolution wird angetrieben durch große Datenmengen und die außerordentliche Rechenleistung moderner Computer, die komplexe statistische Berechnungen in einer Geschwindigkeit und Präzision ausführen können, die lange Zeit unvorstellbar war.

Dabei hat KI viele Gesichter. Von einfachen Programmen, die einzelne Aufgaben mit einer Genauigkeit ausführen, die menschliches Können übertrifft, bis hin zu komplexen Systemen, die lernen, sich anzupassen und Entscheidungen zu treffen, die an den menschlichen Geist erinnern.

Die zunehmende Digitalisierung kann uns von repetitiven und ermüdenden Aufgaben befreien und uns Freiräume geben, kreativer und strategischer zu agieren. Wer in einem Bereich nur wenig Talent oder Erfahrung hat, kann dies mit virtuellen Assistenten ausgleichen. Die KI bietet Möglichkeiten, unsere Fähigkeiten zu erweitern und unsere Grenzen zu verschieben. Um die Vorteile jedoch voll auszuschöpfen, müssen wir lernen, effektiv mit KI zu arbeiten.

Doch was genau macht eine Maschine „intelligent“? Ist es ihre Fähigkeit, menschliche Sprache zu erkennen und darauf zu reagieren? Ihre Effizienz in Millisekunden Entscheidungen zu treffen, die ein Mensch nur mit Mühe und nach langem Abwägen treffen könnte? Oder ist es die Fähigkeit, aus Erfahrungen zu lernen und sich über die Zeit zu verbessern?

Tatsächlich umfasst KI ein Spektrum von Technologien, die so vielfältig sind wie die Definitionen, die sie zu erfassen suchen. Beginnen wir also den Weg in die Welt der KI mit einem Blick auf die grundlegenden Komponenten und die verschiedenen Techniken, die unter dem Begriff „Künstliche Intelligenz“ zusammengefasst werden.

1.1 Die Bausteine der KI

Im Bereich der KI werden oft Namen und Begriffe verwendet, ohne die genauen Definitionen und Hintergründe zu kennen. Doch in der Welt der Mathematiker und Mathematikerinnen sowie Informatiker und Informatikerinnen, den Architekten und Architektinnen als auch Baumeistern und Baumeisterinnen der KI, ist Präzision unerlässlich. Es werden klare, eindeutige Definitionen benötigt, um die Grenzen des Möglichen zu erweitern und die nächste Generation intelligenter Systeme zu entwickeln.

Bei der Entwicklung jedes KI-Systems gibt es Komponenten, die unabhängig von der spezifischen Technik oder dem Anwendungsfall notwendig sind. Diese Bausteine bilden das Fundament, auf dem komplexere KI-Algorithmen aufgebaut werden können. Sehen wir uns daher diese Komponenten von KI-Systemen an und definieren die zugehörigen Details:

Tabelle 1.1 Übersicht wichtiger KI-Komponenten

Komponente	Beschreibung	Wichtigkeit	Komplexität	Anpassbarkeit
Daten	Informationen für das Lernen und die Entscheidungsfindung	Hoch	Variabel	Hoch
Algorithmen	Verfahren oder Methoden zur Verarbeitung von Daten	Hoch	Hoch	Hoch
Rechenleistung	Zur Verarbeitung großer Datenmengen und komplexer Berechnungen	Hoch	Mittel	Mittel
Speicher	Notwendig für die Speicherung von Daten, Modellen und Ergebnissen	Mittel	Niedrig	Mittel
Messung und Modelloptimierung	Wichtig für die Bewertung der Effektivität von KI-Modellen und deren Optimierung	Hoch	Hoch	Hoch
Schnittstellen zur Interaktion	Ermöglichen die Interaktion zwischen Menschen und KI-Systemen	Mittel	Mittel	Hoch

Komponente	Beschreibung	Wichtigkeit	Komplexität	Anpassbarkeit
Sicherheit und Datenschutz	Schützen Daten vor unbefugtem Zugriff und Missbrauch, gewährleisten die Einhaltung rechtlicher Standards	Hoch	Hoch	Mittel

Wichtigkeit: Die Beurteilung „Hoch“ bedeutet, dass diese Komponente sehr wichtig im Kontext auf das richtige und leistungsstarke Arbeiten eines KI-Systems ist. Mittel und Niedrig bezeichnen weniger wichtige Kriterien.

Komplexität: Die Bewertung bezieht sich auf den Schwierigkeitsgrad, mit dem zuverlässig richtige oder performante Ergebnisse von der KI erzeugt werden.

Anpassbarkeit: Die Bewertung zeigt an, wie flexibel ein Unternehmen die Komponenten einer KI an seine eigenen Bedürfnisse und Aufgabenstellungen anpassen kann.

Betrachten wir nun die einzelnen Komponenten genauer. Die folgenden Ausführungen stellen nur die wichtigsten Aspekte dar und haben keinen Anspruch auf Vollständigkeit.

1.1.1 Daten

Daten sind der Brennstoff für KI-Modelle. Ähnlich wie bei einem Fahrzeug würde das KI-Modell ohne Daten keinen Nutzen bringen können. Sie liefern die benötigten Informationen, um Muster zu erkennen, Entscheidungen zu treffen und aus Erfahrungen zu lernen.



Daten sind das Lebenselixier der KI. Ohne Daten hätte die KI keinen Bezugspunkt, um zu analysieren, zu lernen oder sich zu verbessern. Der Qualität der eingesetzten Daten kommt daher eine besondere Bedeutung zu! Ist die Qualität nicht ausreichend hoch, werden keine guten Ergebnisse erzielbar sein.

Von einfachen Datensätzen bis zum heutigen Big-Data-Ansatz

Die Verfügbarkeit und die Qualität von Daten haben in den letzten Jahren stetig zugenommen. Die Entwicklung des Internets und digitaler Technologien hat eine exponentielle Zunahme der Datenmenge und Vielfalt bewirkt, die KI-Systeme heute zum Lernen und Analysieren nutzen.

Daten müssen immer im Kontext ihrer Verwendung betrachtet werden, um ihre Bedeutung und ihren Wert zu verstehen. Der Kontext gibt Aufschluss darüber, was Daten repräsentieren, in welchem Rahmen sie gesammelt wurden und wie sie interpretiert werden sollen. Zum Beispiel könnte die Zahl 9 je nach Kontext eine Temperatur (in Grad Celsius), eine Länge (in Zentimetern), eine Zeitdauer (in Minuten) oder eine Vielzahl anderer Messungen darstellen.

Bei der Datenanalyse und in der künstlichen Intelligenz ist der Kontext ausschlaggebend, um korrekte Schlussfolgerungen zu ziehen und zutreffende Prognosen zu erstellen. Modelle erfordern ein Training mit Daten, die nicht nur genau und bereinigt sind, sondern auch in einem passenden Rahmen dargeboten werden. Der Kontext beeinflusst daher direkt das Ergebnis eines KI-Modells:

Tabelle 1.2 Ziele und Herausforderungen für den Kontext der KI-Ergebnisgenerierung

Bereich	Ziel	Herausforderung	Beispiel
Datenauswahl	Auswahl von Daten, die relevant und repräsentativ für das Problem sind.	Unzureichende oder verzerrte Daten können zu schlechten Modelleleistungen führen.	Auswahl von Patientendaten für ein KI-Modell zur Vorhersage von Diabetes.
Feature-Engineering	Entwicklung und Transformation von Merkmalen, um die Informationsdarstellung für das Modell zu verbessern.	Überanpassung (Overfitting) auf Trainingsdaten und Vernachlässigung der Generalisierbarkeit.	Erstellung von Zeitfenstern für das Kaufverhalten von Kunden in einem KI-Modell zur Vorhersage von Verkäufen.
Modellinterpretation	Analyse und Erklärung der von einem KI-Modell produzierten Ergebnisse.	Die Nachvollziehbarkeit und Überprüfung der Entscheidungsprozesse von komplexen Modellen wie tiefe neuronale Netzwerke sind oft schwer durchzuführen.	Interpretation der Kreditwürdigkeitsbewertung in einem Finanzmodell.

Je nachdem, in welchem Format die Daten vorliegen, kann der Kontext besser oder schlechter erkannt und verarbeitet werden.

- **Strukturierte Daten** sind in einem vordefinierten Format organisiert, meist in Tabellenform, was ihre Analyse und Verarbeitung erleichtert. Beispiele hierfür sind Datenbanken und Excel-Tabellen.
- **Unstrukturierte Daten** hingegen haben kein vordefiniertes Format oder Struktur. Sie umfassen Texte, Bücher, Bilder, Videos und mehr. Die Verarbeitung dieser Daten erfordert komplexere Methoden, da zuerst eine Strukturierung notwendig ist.

Für das Training von KI-Modellen sind umfangreiche und gut aufbereitete Datensätze erforderlich. Diese Datensätze werden in drei Hauptkategorien unterteilt:

1. **Trainingsdatensätze** sind die Daten, auf denen das Modell trainiert wird. Sie bilden die Grundlage für das Lernen von Mustern und Zusammenhängen.
2. **Validierungsdatensätze** werden verwendet, um die Leistung des Modells während des Trainings zu bewerten und die Parameter anzupassen.
3. **Testdatensätze** dienen dazu, die endgültige Leistung des Modells zu evaluieren. Sie müssen unabhängig von den Trainings- und Validierungsdaten sein, um eine objektive Bewertung zu gewährleisten.

Zudem ist die Qualität der Daten entscheidend für die Leistungsfähigkeit von KI-Modellen. Hochwertige Daten sind akkurat, vollständig, konsistent und aktuell. Mangelhafte Datenqualität führt zu unzuverlässigen Modellen, die verzerrte oder falsche Ergebnisse liefern. Oft merkt man erst an den Ergebnissen nach der Trainingsphase, dass die Datenqualität nicht ausreichend für den beabsichtigten Anwendungszweck ist, und muss die Daten dann nachbearbeiten.

Eine weitere große Herausforderung stellt die gesetzeskonforme Verarbeitung von sensiblen und personenbezogenen Daten sowie Daten von Minderheiten dar. Lösungsansätze wie Anonymisierung von Daten und Methoden zur Bias-Minderung müssen in jedem Projekt diskutiert und implementiert werden. Möglichkeiten und Kriterien dazu sind Gegenstand in den nachfolgenden Kapiteln.

1.1.2 Algorithmen

Ein Algorithmus ist eine präzise definierte Vorschrift oder eine Reihe von Anweisungen zur Lösung eines Problems oder zur Durchführung einer Aufgabe. Ganz grundsätzlich kann man sich einen Algorithmus wie ein Rezept in einem Kochbuch vorstellen: Er gibt eine Schritt-für-Schritt-Anleitung, wie man von einem Anfangszustand (rohe Kartoffel) zu einem gewünschten Endzustand (Pommes frites) gelangt.

In der Informatik besteht ein Algorithmus aus einer endlichen Folge von wohldefinierten, ausführbaren Schritten, die eine Aufgabe oder ein Problem lösen. In der KI sind Algorithmen die Verfahren oder Methoden, die genutzt werden, um aus Daten zu lernen, eine Aufgabe zu lösen, Vorhersagen und Entscheidungen zu treffen oder Muster in Daten zu erkennen.

Ein momentan besonders populärer Bereich der KI ist das maschinelle Lernen bzw. Machine Learning. Im Gegensatz zu einem Programmierer oder einer Programmiererin, der bei regelbasierter Software jede Anweisung explizit ausformulieren muss, lernt ein KI-System beim Machine Learning aus den bereitgestellten Daten und identifiziert selbstständig Muster, um die zugrunde liegenden Regeln daraus abzuleiten.

Hier sind einige Schlüsselkonzepte, die beim maschinellen Lernen häufig angewendet werden:

- **Überwachtes Lernen (Supervised Learning)**

Dies ist eine der gängigsten Kategorien von KI-Algorithmen, bei denen das Modell aus einem Datensatz lernt, der bereits die richtigen Antworten (Labels) enthält. Ziel ist es, Vorhersagen für neue, unbekannte Daten zu treffen (z. B. Spam-Erkennung oder Preisvorhersage).

- **Unüberwachtes Lernen (Unsupervised Learning)**

Hierbei lernt das Modell aus Daten ohne vorgegebene Labels, um Muster oder Strukturen selbstständig zu erkennen. Typische Anwendungen sind die Gruppierung von Datenpunkten basierend auf Ähnlichkeiten (z. B. Kundensegmentierung).

- **Halbüberwachtes Lernen (Semi-supervised Learning)**

Eine Mischform, bei der das Modell mit einem Datensatz trainiert wird, der sowohl markierte als auch unmarkierte Daten enthält. Diese Technik ist nützlich, wenn markierte Daten begrenzt oder teuer zu erhalten sind.

- **Verstärkendes Lernen (Reinforcement Learning)**

Bei diesem Ansatz lernt ein Agent, wie er in einer Umgebung handeln soll, um eine Belohnung zu maximieren. Es basiert auf dem Prinzip von Versuch und Irrtum und findet Anwendung in Bereichen wie Spiele (z. B. Schach) und autonome Fahrzeuge.

- **Deep Learning**

Ein Subset des maschinellen Lernens, das auf tiefen neuronalen Netzen basiert. Diese Netze sind in der Lage, komplexe Muster in großen Datenmengen zu erkennen. Anwendungen umfassen Bild- und Spracherkennung sowie natürliche Sprachverarbeitung (NLP).

- **Transfer Learning**

Hierbei wird Wissen aus einem Bereich auf einen anderen übertragen, was besonders nützlich ist, wenn für die Zielanwendung nicht genügend Trainingsdaten verfügbar sind. Dieses Konzept hat in Deep Learning, insbesondere bei der Verarbeitung von Bildern und Sprache, an Bedeutung gewonnen.

1.1.3 Rechenleistung

Moderne KI-Techniken, insbesondere Deep Learning, erfordern erhebliche Mengen an Rechenleistung. Diese Leistung ist notwendig, um die großen Mengen an Daten zu verarbeiten und die komplexen Berechnungen der mathematischen Methoden durchzuführen, die für das Training von KI-Modellen erforderlich sind.

Die Rechenleistung ist über die letzten Jahrzehnte exponentiell angestiegen, getrieben durch Fortschritte in der Halbleitertechnologie (bekannt als Mooresches Gesetz¹⁾) und Innovationen in der Computerarchitektur. Von den frühen Tagen der Großrechner bis zu den heutigen hochleistungsfähigen GPUs (Grafikprozessoren) hat die zunehmende Verfügbarkeit von Rechenleistung die Komplexität der KI-Modelle, die wir heute trainieren können, drastisch verbessert.

Zudem lässt sich für das Training von KI-Modellen Cloud-Technologie nutzen, um die Anschaffung eines eigenen Servers zu vermeiden. Auf diese Weise ermöglicht es die flexible Inanspruchnahme von Rechenleistung genau nach aktuellem Bedarf. Diese ist beim Training wesentlich höher als im Betrieb.

1.1.4 Speicher

Speicher findet in allen Bereichen der künstlichen Intelligenz Anwendung; angefangen von der Speicherung großer Trainingsdatensätze für maschinelles Lernen bis hin zur Speicherung von Zustandsinformationen in Echtzeitanwendungen wie autonomen Fahrzeugen.

Um Daten zu speichern, auf sie zuzugreifen und sie zu verarbeiten, benötigen KI-Systeme eine angemessene Speicherkapazität sowohl für den kurzfristigen Speicher (für die aktuelle Datenverarbeitung) als auch für den langfristigen Speicher (für die Speicherung von Modellen und Ergebnissen). Die Weiterentwicklung von Speichertechnologien hat die Möglichkeiten von KI-Systemen daher maßgeblich beeinflusst. Während bei magnetischen Bändern die Zugriffszeiten sehr langsam waren, ist bei modernen Solid-State-Drives (SSDs) und In-Memory-Datenbanken die Speicherkapazität erhöht und der Zugriff beschleunigt.

¹ Mooresches Gesetz: Formuliert von Gordon Moore, Mitbegründer von Intel, im Jahr 1965. Es besagt, dass sich die Anzahl der Transistoren auf einem Mikrochip etwa alle zwei Jahre verdoppelt, was eine entsprechende Zunahme der Rechenleistung bedeutet (siehe „Algorithmen für Dummies“ von John Paul Mueller, Luca Massaron[1]).

1.1.5 Messung und Modelloptimierung

Die Messung der Leistungsfähigkeit und die Optimierung von KI-Modellen sind entscheidende Schritte im Entwicklungsprozess künstlicher Intelligenz. Sie gewährleisten, dass Modelle nicht nur präzise Vorhersagen treffen, sondern auch effizient und unter realen Bedingungen einsetzbar sind. Es ist wichtig, die Leistung von KI-Modellen zu bewerten, um ihre Genauigkeit, Effizienz und Fairness sicherzustellen. Durch die Bewertung können Schwachstellen identifiziert und Verbesserungen vorgenommen werden. Optimierungstechniken sind notwendig, um die Modelle zu verbessern und ihre Leistung auf spezifische Anforderungen anzupassen.



Die kontinuierliche Messung und Optimierung sind auch unverzichtbar für die Entwicklung verantwortungsvoller KI-Modelle. Nur so kann sichergestellt und nachgewiesen werden, dass KI-Systeme ihre Aufgaben auf faire und nachvollziehbare Weise erfüllen.

Nachweise werden durch vorher definierte Kontrollhandlungen festgestellt und archiviert.

1.1.6 Schnittstellen zur Interaktion

Benutzerschnittstellen (UI) und Programmierschnittstellen (API) sind notwendig, um den Menschen die Interaktion mit KI-Systemen zu ermöglichen. Sie erleichtern die Eingabe von Daten, die Konfiguration von Modellen und den Zugriff auf die Ergebnisse der Datenverarbeitung oder Entscheidungsfindung durch die KI.

Die Interaktionsschnittstellen sollten für die jeweilige KI-Anwendung und die individuelle Benutzergruppe angepasst sein. Beim persönlichen Assistenten oder Kundenservice-Bot für eine nicht technisch ausgebildete Kunden- und Kundinnengruppe ist die UI intuitiv zu halten, während bei komplexen Steuerungssystemen für Robotik und Smart-Geräten für Techniker und Technikerinnen mehr Vorkenntnisse erwartet werden können. Diese Beispiele veranschaulichen, wie unterschiedliche Schnittstellen die Zugänglichkeit und Benutzerfreundlichkeit von KI-Systemen verbessern und helfen, Fehler zu vermeiden. Außerdem sollten die folgenden Aspekte zusätzlich Berücksichtigung finden:

- **Barrierefreiheit:** Gewährleistung, dass KI-Schnittstellen für alle Nutzer zugänglich sind, einschließlich Menschen mit Handicap und besonderen Bedürfnissen.
- **Natürlichkeit der Interaktion:** Entwicklung von Schnittstellen, die eine natürliche und intuitive Kommunikation ermöglichen.
- **Datenschutz und Sicherheit:** Sicherstellung, dass die Interaktion mit KI-Systemen die Privatsphäre der Nutzer und Nutzerinnen schützt.

Zukünftige Entwicklungen könnten immersive Technologien wie Virtual Reality (VR) und Augmented Reality (AR) umfassen, die neue Arten der Interaktion ermöglichen. Fortschritte könnten zu individuelleren und adaptiven Schnittstellen führen, die sich an die Präferenzen und Bedürfnisse des jeweiligen Benutzers bzw. der Benutzerin anpassen.

1.1.7 Sicherheit und Datenschutz

Da KI-Systeme oft mit sensiblen oder personenbezogenen Daten arbeiten, sind robuste Sicherheits- und Datenschutzmaßnahmen unerlässlich. Diese Maßnahmen schützen die Informationen vor unbefugtem Zugriff und Missbrauch und gewährleisten die Einhaltung rechtlicher und ethischer Standards.

Grundlegende Risiken und gesetzliche Anforderungen sind:

- **Datenschutz:** Es müssen Konzepte wie Datenminimierung, Zweckbindung und Transparenz umgesetzt werden, damit sichergestellt ist, dass personenbezogene Daten geschützt und vertraulich behandelt werden.
- **Urheberrecht:** Es sollte der Nachweis geführt werden, dass die verwendeten Daten und Werke nicht urheberrechtlich durch andere Personen geschützt sind und die Organisation die Nutzungsrechte für die Verwendung besitzt.
- **Sicherheitsrisiken in KI-Systemen:** Es müssen Maßnahmen gesetzt werden, um Bedrohungen und Angriffsszenarien gegen KI-Systeme zu bemerken und zu verhindern. Ein Beispiel dafür sind die Adversarial Attacks, bei denen Eingabedaten durch die Benutzer bzw. Benutzerinnen manipuliert eingegeben werden, um falsche Vorhersagen oder Entscheidungen zu provozieren. Dieses Risiko muss durch die Analyse der Eingabedaten minimiert werden.
- **Verschlüsselung und Anonymisierung:** Es müssen Techniken zum Schutz von Daten, sowohl bei der Übertragung als auch bei der Speicherung, implementiert werden. Für personenbezogene und sensible Daten sollte eine Anonymisierung²⁾ oder Pseudonymisierung³⁾ der Daten in Betracht gezogen werden.

² Hierbei wird die Verlinkung zwischen einer Person und ihren Daten irreversibel gelöscht. Anhand der Daten kann niemand identifiziert werden.

³ Bei der Pseudonymisierung werden die Namen und andere Identifikationsmerkmale durch ein Kennzeichen ersetzt, damit eine Zuordnung zwischen einer Person und seinen Daten nur durch berechtigte Personen möglich ist. Beispiele hierfür sind ein Autokennzeichen oder eine Kundennummer.



Wichtig ist es zu dokumentieren, welche Schutzmaßnahmen zu welchem Zweck implementiert wurden und wie der Schutz während des gesamten Lebenszyklus der KI gewährleistet bleibt. Dafür sind Kontrollhandlungen zu definieren und deren Einhaltung nachvollziehbar zu archivieren.

Sicherheit und Datenschutz sind entscheidende Aspekte, die den Erfolg und die Akzeptanz von KI-Technologien prägen. Durch die kontinuierliche Anpassung an neue Bedrohungen und die Einhaltung strenger Datenschutzstandards können KI-Systeme vertrauensvoll und verantwortungsbewusst eingesetzt werden.

1.2 Datenwissenschaft und Datenanalytik

Begriffe wie „Datenwissenschaft“ und „Datenanalytik“ werden oft austauschbar verwendet, obwohl es spezifische Unterschiede gibt. Während die Datenwissenschaft ein breiteres Feld ist, das sich mit der Erstellung von Modellen und Algorithmen zur Datenanalyse beschäftigt, konzentriert sich die Datenanalytik mehr auf die Interpretation konkreter Daten, um unmittelbare Fragen zu beantworten. In der folgenden Tabelle werden die Unterschiede anhand grundsätzlicher Kriterien dargestellt:

Tabelle 1.3 Unterscheidung von Datenwissenschaft und Datenanalytik

Kriterium	Datenwissenschaft	Datenanalytik
Primäres Ziel	Entdeckung von Mustern und Vorhersage zukünftiger Trends und Verhaltensweisen durch Analyse von Daten.	Interpretation von historischen Daten, um konkrete Fragen zu beantworten oder Probleme zu lösen.
Schwerpunkte	Datenaufbereitung, -exploration, -modellierung, maschinelles Lernen, Vorhersageanalysen.	Datensammlung, Datenbereinigung, deskriptive Analyse, Berichterstattung.
Methoden und Techniken	Maschinelles Lernen, (Deep Learning) Lernalgorithmen, statistische Modelle, Datenmining.	Statistische Analysen, explorative Datenanalyse, Visualisierungswerkzeuge, Process Mining.
Anwendungsbeispiele	Entwicklung neuer Produktideen, Betrugserkennung, Risikomanagement, personalisierte Empfehlungssysteme.	Verkaufs- und Marktanalysen, Kundenverhaltensanalysen, operative Effizienz, Qualitätskontrolle.
Benötigte Skills	Statistik, maschinelles Lernen, Programmierung (z. B. Python, R), Datenbankkenntnisse, Geschäftsverständnis.	Statistik, analytisches Denken, Kenntnisse in BI-Tools (z. B. Tableau, Power BI), Grundkenntnisse in Programmierung.

Die Datenwissenschaft beschäftigt sich also mit der Erhebung, Bereinigung und Analyse von Daten, um Muster zu erkennen, Vorhersagen zu treffen und Entscheidungen auf der Grundlage großer Datensätze zu treffen. Datenwissenschaftler und -wissenschaftlerinnen verwenden dazu komplexe Algorithmen, maschinelles Lernen und statistische Methoden, um mit Daten zu arbeiten. Sie suchen nach Antworten auf Fragen, die noch nicht gestellt wurden, und schaffen neue Wege zur Datennutzung.

Die Datenanalytik konzentriert sich mehr auf die Verarbeitung und Analyse bestehender Daten aus den Verarbeitungsprozessen einer Organisation, um konkrete Fragen zu beantworten, die das Geschäftsfeld betreffen. Datenanalytiker und -analytikerinnen verwenden statistische Werkzeuge und Techniken, die Unternehmen dabei helfen sollen, operative Entscheidungen zu treffen. Im Gegensatz zur Datenwissenschaft, die oft explorativer und zukunftsorientierter ist, ist die Datenanalytik in der Regel zielgerichteter und auf die Lösung spezifischer Probleme ausgerichtet.

Es gibt aber keine „fixen“ Definitionen dieser Begriffe, da das Feld der Datenwissenschaft und -analytik sich schnell weiterentwickelt und die Grenzen zwischen den Disziplinen oft fließend sind. Unternehmen sollten bei Stellenausschreibungen und Positionsbeschreibungen die präzisen Aufgabenbereiche und Einsatzgebiete angeben, damit es nicht zu unerfüllbaren oder enttäuschenden Erwartungen führt.

1.3 Entwicklung von KI in KMUs

Der Einsatz von künstlicher Intelligenz bietet gerade für kleine und mittelständische Unternehmen (KMUs) sowie Start-ups erhebliche Vorteile und Chancen. Mit KI-Technologien können sie Prozesse automatisieren, Effizienzen steigern und innovative Produkte oder Dienstleistungen entwickeln, die einen Wettbewerbsvorteil darstellen können.

Für KMUs eröffnet der Einsatz von KI die Möglichkeit, Ressourcen effizienter zu nutzen, indem der Belegschaft ein KI-Assistent zur Seite gestellt wird, der zeitaufwendige Aufgaben übernimmt und Entscheidungsfindungsprozesse durch datengetriebene Einsichten unterstützt. Dies kann zu einer signifikanten Kostensenkung und einer verbesserten Kundenbetreuung führen.

Start-ups, die mit KI experimentieren, können maßgeschneiderte Lösungen entwickeln, die auf spezifische Branchenbedürfnisse zugeschnitten sind, und so neue Märkte erschließen oder bestehende Märkte disruptiv beeinflussen.

Darüber hinaus fördert der Einsatz von KI in beiden Unternehmensarten eine Kultur der Datenfokussierung und Innovation. Dies ist für das langfristige Wachstum und die Anpassungsfähigkeit an sich schnell ändernde Marktbedingungen unerlässlich. In einer Welt, in der Daten zur neuen wichtigsten Ressource geworden sind, bieten KI-Technologien KMUs und Start-ups die Werkzeuge, um diese wertvollen Ressourcen zu nutzen und zukunftssichere Geschäftsmodelle zu entwickeln.

Der Einsatz von KI in KMUs und Start-ups birgt also erhebliches Potenzial, stellt diese jedoch auch vor signifikante Herausforderungen. Eine der Hauptproblematiken ist die Rekrutierung und Bindung von qualifiziertem Personal, das die Entwicklung und Implementierung von KI-Technologien beherrscht. Weitere Herausforderungen sind:

■ **Rekrutierung von Experten und Expertinnen**

- **Fachkräftemangel:** Es gibt einen globalen Mangel an Experten und Expertinnen im Bereich KI und maschinelles Lernen. Dies macht es für KMUs und Start-ups besonders schwierig, talentierte Fachkräfte zu gewinnen, die mit den neuesten Technologien und Methoden vertraut sind.
- **Wettbewerb mit Großunternehmen:** Start-ups und KMUs stehen im Wettbewerb mit großen, etablierten Unternehmen um KI-Talente. Großunternehmen können oft attraktivere Gehälter, Benefits und Karriereentwicklungsmöglichkeiten bieten.

■ **Finanzielle Ressourcen**

- **Hohe Kosten:** Die Entwicklung und Implementierung von KI-Lösungen können teuer sein, insbesondere wenn es um die Anschaffung der notwendigen Hardware oder Software geht. Für Start-ups und KMUs mit begrenzten Budgets kann dies eine erhebliche Barriere darstellen.
- **ROI-Unsicherheit:** Die Investition in KI birgt Unsicherheiten bezüglich des Return on Investment (ROI), besonders wenn das Unternehmen neu in diesem Bereich ist. Dies kann zu Zögern bei der Allokation von Ressourcen führen.

■ **Daten und Datenschutz**

- **Datenqualität und Verfügbarkeit:** Effektive KI-Systeme erfordern große Mengen qualitativ hochwertiger Daten. Viele KMUs und Start-ups haben jedoch Schwierigkeiten, Zugang zu solchen Daten zu erhalten oder diese in einer nützlichen Weise zu sammeln und zu verarbeiten.
- **Datenschutzbestimmungen:** Die Einhaltung von Datenschutzgesetzen wie der DSGVO kann für Unternehmen, die mit personenbezogenen Daten arbeiten, eine Herausforderung darstellen. Die Entwicklung von KI-Systemen, die diese Anforderungen erfüllen, erfordert spezifisches Wissen und Ressourcen.
- **IT Security:** Die Implementierung und kontinuierliche Gewährleistung von IT Sicherheitsmaßnahmen in der Infrastruktur der Organisation benötigt erfahrene Spezialisten, teure Software und zeitliche Ressourcen. Dies gilt auch für die Integration von Sicherheitsmerkmalen in Software bzw. KI-Produkte.

■ **Technologische Integration**

- **Prozessintegration:** Die Integration von KI-Lösungen in bestehende Geschäftsprozesse und -systeme kann komplex und herausfordernd sein. Kompatibilitätsprobleme und der Bedarf an maßgeschneiderten Lösungen können die Implementierung verzögern oder verteuern.