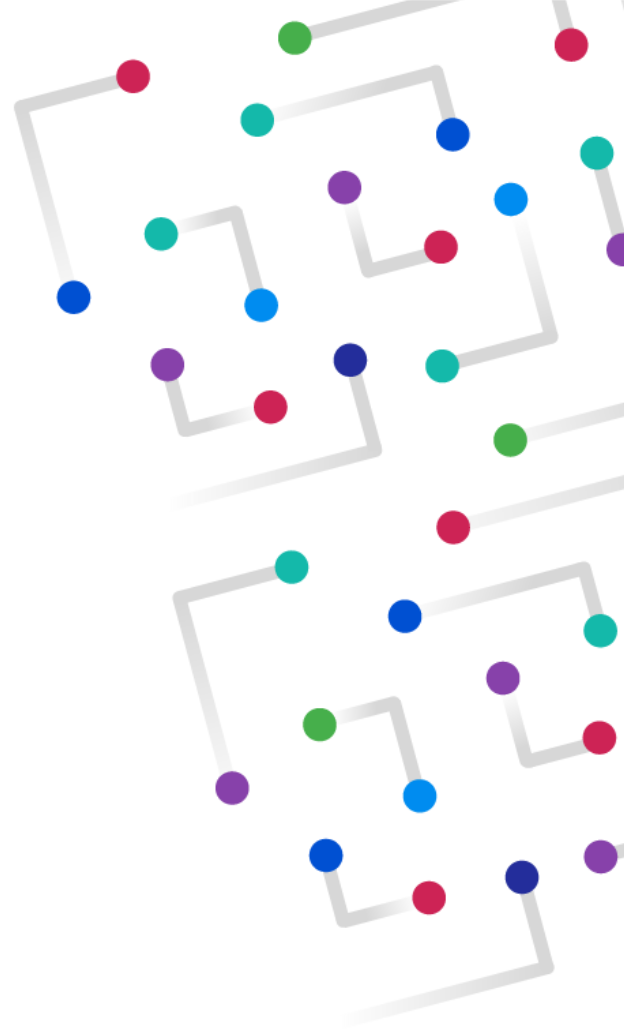


**Raspberry Pi**  
Foundation

# Experience AI

# KI – Glossar



# KI – Glossar

In diesem Glossar werden die wichtigsten Fachwörter aus den Bereichen künstliche Intelligenz (KI) und maschinelles Lernen (ML) erläutert, die in den [Experience AI-Lektionen](#) und darüber hinaus verwendet werden.

## [Glossar aufrufen](#)

Die Erklärungen wurden in erster Linie für Lehrkräfte und andere Pädagogen und Pädagoginnen konzipiert und zielen auf ein junges Publikum ab. Das Glossar soll Ihnen helfen, Ihre Kenntnis der Schlüsselbegriffe und Ihr technisches Wissen zu vertiefen.

Das Vokabular ist ein wichtiger Bestandteil des Lehrens und Lernens. Der richtige Gebrauch des Vokabulars kann Lernenden bei der Entwicklung ihres Verständnisses helfen. Eine uneinheitliche Verwendung kann hingegen zu abweichenden Auffassungen (Missverständnissen) führen und den Lernprozess beeinträchtigen. Mehr zu diesem Thema können Sie in [unserem Pedagogy Quick Read](#) (auf Englisch) nachlesen. Als Lehrkraft können Sie das konzeptionelle Verständnis der Lernenden fördern, indem Sie konsequent präzises Fachvokabular verwenden.

Beim Verfassen der Erklärungen haben wir uns am „Semantic Waves“-Konzept orientiert. Jede Erklärung besitzt die gleiche dreiteilige Struktur: Der erste Teil besteht aus einer eher abstrakten Erklärung des Worts, im zweiten Teil wird das Wort anhand eines gängigen Beispiels konkret erläutert („unpacking“) und im dritten Teil wird die anhand des Beispiels erklärte Bedeutung wieder abstrakter zusammengeführt („repacking“), um die Verbindung zu dem Fachterminus wiederherzustellen. Weitere Informationen enthält [unser Pedagogy Quick Read zu Semantic Waves](#) (auf Englisch).

Bei dem vorliegenden Glossar handelt es sich um die 1. Version. Das Glossar wird parallel zur Weiterentwicklung der Experience AI-Lektionen ergänzt, überarbeitet und aktualisiert.

# Inhaltsverzeichnis

[Bestärkendes Lernen](#)  
[Computer Vision](#)  
[Datenaufbereitung](#)  
[Datenbasiert](#)  
[Daten](#)  
[Datenvoreingenommenheit](#)  
[Entscheidungsbaum](#)  
[Entscheidungsbaumknoten](#)  
[Erklärbarkeit](#)  
[Genauigkeit](#)  
[Generative KI](#)  
[Gesellschaftliche Voreingenommenheit](#)  
[Kennzeichnung](#)  
[KI-Kompetenz](#)  
[Klasse](#)  
[Klassifizierung](#)  
[Künstliche Intelligenz](#)  
[Lebenszyklus von KI-Projekten](#)  
[Maschinelles Lernen](#)  
[Merkmal](#)  
[Modell](#)  
[Modellsteckbrief](#)  
[Regelbasiert](#)  
[Testdaten](#)  
[Trainieren](#)  
[Trainingsdaten](#)  
[Überwachtes Lernen](#)  
[Unüberwachtes Lernen](#)  
[Voreingenommenheit](#)  
[Vorhersage](#)  
[Wahrscheinlichkeitsschwellenwert](#)  
[Wahrscheinlichkeit](#)

## Bestärkendes Lernen

Bestärkendes Lernen ist eine Methode zum **Trainieren** von **Modellen** für **maschinelles Lernen (ML)**. Die Methode wird zur Lösung von Problemen mit einem klaren Ziel verwendet, wobei Belohnungen und Strafen eingesetzt werden, um dieses Ziel zu erreichen. Bestärkendes Lernen wird bei der Entwicklung selbstfahrender Autos oder beim Spielen komplizierter Spiele eingesetzt. Ein unter Einsatz von bestärkendem Lernen erstelltes Modell könnte beispielsweise zum Erstellen einer Anwendung zum Schachspielen verwendet werden. Das Modell wird darauf trainiert, **Vorhersagen**

über die Züge zu treffen, durch die, mit dem Ziel des Gewinnens, die Belohnungen maximiert und die Strafen minimiert werden. Beim bestärkenden Lernen werden Belohnungen und Strafen eingesetzt, um Strategien zum Erreichen eines festgelegten Ziels zu finden.

## Computer Vision

Computer Vision ist die Lehre von Systemen, die zur Verarbeitung von Informationen aus digitalen Bildern oder Videos entwickelt wurden. Anwendungsbeispiele für Computer Vision sind Gesichtserkennung, medizinische Bildgebung und Videoüberwachung. Computer Vision wird beispielsweise bei der Entwicklung selbstfahrender Autos eingesetzt, um Objekte zu erkennen und Kollisionen mit ihnen zu vermeiden. In den meisten Fällen werden in Computer Vision-Systemen **Modelle des maschinellen Lernens** eingesetzt, um Muster in Bild- und Videodaten zu erkennen. Computer Vision-Systeme sind nützlich, wenn Informationen aus digitalen Bildern oder Videos zur Lösung eines Problems genutzt werden können.

## Datenaufbereitung

Die Datenaufbereitung ist ein Schritt bei der Vorbereitung der **Daten**, die zum **Trainieren** eines **ML-Modells** verwendet werden. Bei der Datenaufbereitung werden Fehler in den Daten gesucht und korrigiert. Zwei einfache Arten der Datenaufbereitung sind beispielsweise das Korrigieren von Tippfehlern und das Entfernen von Duplikaten in Textdaten. Meistens sind Daten unstrukturiert und erfordern eine aufwändigere Aufbereitung, damit sie zum Trainieren von ML-Modellen verwendet werden können. Je nach Problem und Datentyp gibt es viele Möglichkeiten, Daten aufzubereiten. Die Verwendung aufbereiteter Daten ist für die Erstellung **genauer** ML-Modelle unerlässlich.

## Datenbasiert

Datenbasiert bezieht sich auf eine Methode zum Entwickeln von Systemen, bei der **Daten** anstelle von schrittweisen Anweisungen verwendet werden. Beispiel: Es ist schwierig, die Ursachen einer Krankheit festzustellen, doch es gibt viele Beispieldaten. Daher nutzen die Entwicklerinnen und Entwickler die medizinischen Daten betroffener Menschen, um die Krankheit zu diagnostizieren. Datenbasierte Systeme stehen im Gegensatz zu **regelbasierten** Systemen. Datenbasierte Systeme eignen sich zur Lösung von Problemen, bei denen es schwierig ist, Regeln zu erstellen, die alle Situationen abdecken. Stattdessen können genügend Beispiele als Informationsgrundlage für eine Lösung gesammelt werden.

## Daten

Daten sind Werte, Messungen, Fakten oder Beobachtungen, die eine für die Verarbeitung durch Computerprogramme geeignete Form besitzen. Es gibt viele Arten von Daten, beispielsweise Text-, Bild- oder Tondaten. Ein Beispiel für Textdaten sind Nachrichten, die Menschen auf digitalen Geräten mit Freunden austauschen. Beim **maschinellen Lernen (ML)** sind Daten die Beispiele, mit deren Hilfe **ML-Modelle trainiert** werden. Die Erhebung, **Aufbereitung** und Strukturierung sehr großer Datenmengen ist ein wichtiger Teil der Entwicklung von ML-Modellen.

# Datenvoreingenommenheit

Datenvoreingenommenheit ist eine **Voreingenommenheit** in den **Daten**, die zum **Trainieren** von **Modellen für maschinelles Lernen (ML)** verwendet werden. Eine Datenvoreingenommenheit kann dazu führen, dass ein ML-Modell auf eine Weise trainiert wird, dass es voreingenommene **Vorhersagen** erzeugt. Beispielsweise können Modelle zur Gesichtserkennung gegen Gesichter mit bestimmten Hauttönen voreingenommen sein, weil sie hauptsächlich mit Bildern von Gesichtern eines bestimmten Hauttons trainiert wurden. Datenvoreingenommenheit kann verschiedene Ursachen haben. Zu ihnen gehören unvollständige Daten oder Daten, die eine **gesellschaftliche Voreingenommenheit** widerspiegeln. Das Erkennen von Datenvoreingenommenheit ist wichtig, um zu vermeiden, dass ML-Modelle voreingenommene Vorhersagen generieren.

# Entscheidungsbaum

Ein Entscheidungsbaum für **maschinelles Lernen (ML)** ist eine Art von **ML-Modell**. Entscheidungsbäume werden verwendet, um Bedingungen zu strukturieren, auf deren Grundlage eine **Vorhersage** getroffen werden kann. Die Bedingungen werden aus in den **Daten** enthaltenen **Merkmale** abgeleitet. Ein Entscheidungsbaum kann beispielsweise zur Erstellung Empfehlungssysteme für Filme verwendet werden. Das Entscheidungsbaummodell wird anhand der Filmvorlieben vieler Menschen **trainiert**. Beim Training werden Bedingungen auf der Grundlage von Merkmalen wie Filmart, Länge oder Hauptdarsteller generiert. Basierend darauf, wie die Vorlieben einer Person den Bedingungen im Modell folgen, generiert das ML-Modell eine Vorhersage darüber, welchen Film die Person möglicherweise als Nächstes sehen möchte. Die Struktur von ML-Entscheidungsbäumen wird auf der Grundlage riesiger Datenmengen erzeugt und kann sich bei einem erneuten Training mit anderen Daten ändern.

# Entscheidungsbaumknoten

Ein **Entscheidungsbaum** für **maschinelles Lernen (ML)** besteht aus Knoten. Durch die Verbindung der Knoten entsteht eine Struktur, auf deren Grundlage eine **Vorhersage** generiert werden kann. Es gibt zwei Arten von Knoten: Entscheidungsknoten und Blattknoten. Ein Entscheidungsbaum kann beispielsweise zur Vorhersage der Arten von Sternen in unserem Sonnensystem erstellt worden sein. Die Entscheidungsknoten stehen für **Datenmerkmale** wie etwa die Temperatur, den Radius, die Farbe oder die Helligkeit von Sternen. Die Blattknoten stellen die Sternarten in Form von Vorhersage-**Kennzeichnungen** dar, beispielsweise „Roter Zwerg“, „Weißer Zwerg“ oder „Brauner Zwerg“. Entscheidungsbaumknoten bilden die Struktur, die ein **ML-Modell** zum Generieren einer Vorhersage benötigt.

# Erklärbarkeit

Erklärbarkeit ist das Ausmaß, in dem etwas verstanden werden kann. Beim **maschinellen Lernen (ML)** hilft die Erklärbarkeit zu verstehen, wie eine **Vorhersage** zustande kam.

**ML-Entscheidungsbaummodelle** sind beispielsweise erklärbar, da die **Knoten** auf eine für Menschen verständliche Weise analysiert werden können. Die meisten ML-Modelle sind nicht

vollständig erklärbar und manche sind erklärbarer als andere. Die Verbesserung der Erklärbarkeit eines Modells kann dazu beitragen, Probleme zu beheben und **Voreingenommenheit** zu beseitigen.

## Genauigkeit

Genauigkeit bezieht sich darauf, wie richtig etwas ist. Beim **maschinellen Lernen (ML)** ist Genauigkeit eine Methode, mit deren Hilfe gemessen werden kann, wie oft ein **ML-Modell** eine richtige **Vorhersage** macht. Ein Klassifizierungsmodell kann beispielsweise zur **Klassifizierung** von Äpfeln konzipiert worden sein. Von 100 Bildern von Äpfeln werden 90 richtig klassifiziert. Die Klassifizierungsgenauigkeit des Modells beträgt 90 %. Genauigkeit ist eine Methode zur Bewertung von ML-Modellen. Meist wird Genauigkeit zusammen mit anderen Messgrößen verwendet, um die Güte eines Modells zu bewerten.

## Generative KI

Generative KI ist eine Art der **künstlichen Intelligenz (KI)**, die entwickelt wird, um Inhalte wie etwa Text, Bilder oder Ton zu generieren. Generative KI wird für viele Anwendungen genutzt, etwa zur Erzeugung von Kunst oder Musik oder auch von Text für Chatbots. Eine generative KI-Kunstanwendung kann beispielsweise auf der Grundlage einer Aufforderung wie „Erstelle ein Bild von einem Drachen, der ein Buch liest“ ein Bild generieren. Generative KI-Kunst wird mithilfe von **Modellen** für **maschinelles Lernen** erzeugt, die mit Millionen von Bildern bestehender Kunst **trainiert** wurden. Die so entstehenden Bilder können den Stil eines Künstlers oder einer Künstlerin nachahmen, ohne dass diese Person davon Kenntnis hat oder es gutheißt. Anwendungen der generativen KI kommen immer häufiger zum Einsatz und oft ist nicht erkennbar, dass generative KI verwendet wurde.

## Gesellschaftliche Voreingenommenheit

Gesellschaftliche Voreingenommenheit entsteht durch die **Voreingenommenheit** (auch: Vorurteile) einer großen Gruppe von Menschen oder der Gesellschaft als Ganzes. Es gibt viele verschiedene Arten gesellschaftlicher Voreingenommenheit, etwa rassistische, geschlechtsspezifische oder ethnische Vorurteile. Ein Beispiel für geschlechtsspezifische Voreingenommenheit ist die Vorstellung, dass Frauen für den Ingenieursberuf weniger geeignet seien als Männer. Anhand großer Personengruppen erhobene **Daten** können eine gesellschaftliche Voreingenommenheit enthalten, die ihrerseits zu einer **Datenvoreingenommenheit** führen kann. Werden Daten, die eine gesellschaftliche Voreingenommenheit widerspiegeln, zum **Trainieren** von **Modellen** für **maschinelles Lernen (ML)** verwendet, kann dies dazu führen, dass die Modelle voreingenommene **Vorhersagen** generieren. Beim maschinellen Lernen ist es wichtig, gesellschaftliche Voreingenommenheit in den **Trainingsdaten** zu minimieren, um diskriminierende oder ungerechte Ergebnisse zu vermeiden.

## Kennzeichnung

Beim **überwachten Lernen** werden **Modelle** für **maschinelles Lernen (ML)** mit gekennzeichneten **Daten trainiert**. Jedes Datenelement ist mit einer oder mehreren Kennzeichnungen versehen, die

Informationen zu den Daten liefern. Ein Beispiel wäre ein ML-Modell, das den Gesang von Vögeln erkennen soll. Jeder Laut ist mit dem Namen des Vogels gekennzeichnet, der ihn erzeugt hat. Das ML-Modell wird mit den gekennzeichneten Lauten trainiert und kann die Kennzeichnung (den Namen des Vogels) neuer Laute **vorhersagen**. Daten werden meistens von Menschen gekennzeichnet, um genaue Beispiele für das Training von ML-Modellen bereitzustellen.

## KI-Kompetenz

KI-Kompetenz bezieht sich auf Fähigkeiten und Denkweisen, mit deren Hilfe Menschen konstruktiv mit **Anwendungen künstlicher Intelligenz (KI)** und Situationen, in denen in ihrem Umfeld KI-Anwendungen zum Einsatz kommen, umgehen können. Zu diesen Fähigkeiten gehören das Verständnis von KI, die Mitwirkung bei der Entwicklung von KI-Systemen und eine fundierte Ansicht über die Art und Weise, wie KI-Systeme in der Welt eingesetzt werden. Beispielsweise nutzen Menschen ihre KI-Kompetenz bei der Beurteilung der Genauigkeit von Informationen, die von einer KI-Chatbot-Anwendung bereitgestellt werden. Neben den verschiedenen Möglichkeiten des Umgangs mit KI-Anwendungen umfasst KI-Kompetenz auch das Potenzial der Menschen zur aktiven Beteiligung an Entscheidungen über die Nutzung von KI-Systemen in ihrem Umfeld.

## Klasse

Um **Klassifizierungsmodelle** zu trainieren, teilen Entwicklerinnen und Entwickler von **maschinellm Lernen (ML)** die **Daten** in Kategorien ein. Diese Kategorien werden „Klassen“ genannt. Klassen werden im Voraus auf der Basis als nützlich erachteter Kategorien definiert. Eine ML-Anwendung könnte beispielsweise den Zweck haben, Obst in einem Supermarkt zu erkennen. Die Daten könnten in Klassen von Äpfeln, Bananen, Orangen, Blaubeeren usw. eingeteilt sein. Eine Klasse ist eine Kategorie von Dingen, anhand derer Klassifizierungsmodelle Ähnlichkeiten in Daten erkennen.

## Klassifizierung

Unter Klassifizierung versteht man die Einteilung von Dingen in vordefinierte Kategorien, die sogenannten **Klassen**. Klassen werden im Voraus auf der Basis als nützlich erachteter Kategorien definiert. Ein Beispiel für ein Klassifizierungsproblem wäre die Stimmungsanalyse von Liedrezensionen. Ein **Klassifizierungsmodell für maschinelles Lernen (ML)** wird mit Rezensionen **trainiert**, die als „positiv“ oder „negativ“ **gekennzeichnet** sind. Nach dem Training kann das ML-Modell verwendet werden, um **vorherzusagen**, ob eine neue Rezension als „positiv“ oder „negativ“ klassifiziert werden muss. Ein Klassifizierungsmodell sagt eine oder mehrere Klassenkennzeichnungen vorher. Die Methode der Klassifizierung ist zum Lösen von Problemen nützlich, bei denen die Antworten in vordefinierte Gruppen fallen.

## Künstliche Intelligenz

Künstliche Intelligenz (KI) ist die Entwicklung und Untersuchung von Systemen, die intelligentes Verhalten nachzuahmen scheinen. Manche KI-Anwendungen basieren auf Regeln. Heute werden KI-Anwendungen häufiger unter Einsatz des **maschinellen Lernens** erstellt, dem „Lernen“ aus Beispielen in Form von **Daten**. Manche KI-Anwendungen sollen beispielsweise Fragen beantworten

oder bei der Diagnose von Krankheiten helfen. Andere wiederum können einen schädlichen Zweck haben, wie zum Beispiel die Verbreitung von Falschnachrichten. KI-Anwendungen denken nicht. KI-Anwendungen werden entwickelt, um Aufgaben auf eine scheinbar intelligente Weise auszuführen.

## Lebenszyklus von KI-Projekten

Als Lebenszyklus eines Projekts im Bereich der **künstlichen Intelligenz (KI)** werden die Schritte bezeichnet, die zum Entwerfen und Erstellen eines **Modells des maschinellen Lernens (ML)** erforderlich sein können. Zu diesen Schritten gehören die Definition des Problems, die Aufbereitung der **Daten**, das **Trainieren** des Modells, das Testen des Modells, das Bewerten des Modells und das Erklären des Modells. Wenn ein ML-Modell beispielsweise neue Song-Playlists generieren soll, überlegen die ML-Entwicklerinnen und Entwickler zunächst, welche Art von Playlist erstellt werden soll. Anschließend sammeln sie die Songdaten und bereiten sie auf. Das ML-Modell wird mit den Songdaten trainiert und getestet. Das ML-Modell wird bewertet, um zu sehen, ob es wie erwartet funktioniert. Zum Schluss wird das ML-Modell erklärt, damit andere es verwenden können. Normalerweise werden die Schritte des KI-Projekt-Lebenszyklus nicht nacheinander, sondern schrittweise wiederholend (iterativ) ausgeführt. Der Lebenszyklus eines KI-Projekts besteht aus einer Reihe iterativer Schritte zum Erstellen und Verbessern eines ML-Modells.

## Maschinelles Lernen

Maschinelles Lernen (ML) ist eine Methode zum Entwerfen und Erstellen von Systemen der **künstlichen Intelligenz (KI)**. Man sagt, dass ML anhand von Beispielen in Form von **Daten** „lernt“, anstatt schrittweise Anweisungen auszuführen. ML-Anwendungen sind also **datenbasiert**. Eine ML-Anwendung kann beispielsweise zur Spracherkennung eingesetzt werden. Sie basiert auf vielen Sprechbeispielen von Menschen mit unterschiedlichem Akzent und Tonfall. Zu ML-Anwendungen gehören auch solche zum Erkennen von Objekten in Bildern oder zum Spielen komplexer Spiele. Jede ML-Anwendung ist auf die Lösung eines bestimmten Problems ausgelegt.

## Merkmal

Beim **maschinellen Lernen (ML)** sind Merkmale mit den **Daten** verknüpfte Eigenschaften. Ein Musikdatensatz kann beispielsweise Merkmale wie Tempo, Tonhöhe, Energie oder Genre aufweisen. Manche ML-**Modelle** werden mithilfe von Merkmalen **trainiert**, um Ähnlichkeiten in den Daten zu finden. Andere treffen **Vorhersagen** über neue Merkmale in den Daten, die für Menschen nicht leicht erkennbar sind. Die Auswahl der Merkmale zum Trainieren eines ML-Modells hat Auswirkungen darauf, wie gut das Modell funktioniert.

# Modell

Ein Modell für **maschinelles Lernen (ML)** wird von einer ML-Anwendung verwendet, um eine Aufgabe auszuführen oder ein Problem zu lösen. Das ML-Modell ist eine Darstellung des zu lösenden Problems. ML-Entwickler verwenden riesige Mengen von **Daten**, die repräsentativ für ein Problem sind, um ein Modell für das Erkennen von Mustern zu **trainieren**. Durch das Training entsteht ein Modell, das zum Treffen von **Vorhersagen** über neue Daten im gleichen Zusammenhang verwendet wird. Beispielsweise werden selbstfahrende Autos mithilfe von ML-Modellen gebaut, um vorherzusagen, wann sie anhalten müssen. Die Modelle werden anhand von Millionen Beispielsituationen trainiert, in denen Autos anhalten müssen. Es gibt viele verschiedene Modelltypen, die mit unterschiedlichen Daten und unterschiedlichen Methoden trainiert werden. Alle ML-Modelle werden darauf trainiert, Muster in den **Trainingsdaten** zu erkennen und Vorhersagen über neue Daten zu treffen.

# Modellsteckbrief

Der Modellsteckbrief ist eine Methode zur strukturierten Dokumentation wichtiger Informationen über ein **Modell für maschinelles Lernen (ML)**. Modellsteckbriefe werden von ML-Entwickler sowohl für Fachleute als auch für Laien geschrieben. Eine ML-Anwendung kann beispielsweise zum Übersetzen von Sprachen entwickelt worden sein, etwa aus dem Arabischen ins Französische und umgekehrt. Der Modellsteckbrief enthält Informationen zur **Genauigkeit** der Übersetzung und zur Leistung des Modells im Hinblick auf Fachjargon, Slang und Dialekte. Ein Modellsteckbrief kann auch Informationen über den Modelltyp, verschiedene Leistungskennzahlen und sogar über bekannte **Voreingenommenheiten** enthalten. Modellsteckbriefe werden in der Erklärungsphase des **KI-Projektlebenszyklus** erstellt, um Informationen zu den Fähigkeiten und Grenzen des Modells auf leicht verständliche Weise darzustellen.

# Regelbasiert

Regelbasiert bezieht sich auf eine Methode zum Entwickeln von Systemen unter Verwendung einer Reihe vordefinierter Regeln. Ein Programm für das Spiel „Drei gewinnt“ wird beispielsweise anhand von Regeln erstellt, anhand derer die Spielzüge zum Gewinnen identifiziert werden. Die Regeln werden üblicherweise von Menschen festgelegt, die Fachleute auf dem Gebiet des zu lösenden Problems sind. Systeme **künstlicher Intelligenz (KI)**, die auf einem regelbasierten Konzept aufbauen, werden auch als „klassische KI“ bezeichnet. Regelbasierte Systeme stehen im Gegensatz zu **datenbasierten** Systemen, bei denen **Daten** als Beispiele zur Lösung des Problems verwendet werden. Regelbasierte Systeme eignen sich zur Lösung von Problemen, wenn Regeln für die meisten Situationen erstellt und befolgt werden können.

# Testdaten

Testdaten beim **maschinellen Lernen (ML)** sind **Daten**, die zum Testen und Bewerten **trainierter ML-Modelle** verwendet werden. Beispielsweise könnte ein Modell trainiert worden sein, um eine **Vorhersage** der Diagnose einer Krankheit zu treffen. Vor dem Einsatz in der Praxis wird das Modell anhand von Testdaten getestet und bewertet. Die Testdaten sind von den **Trainingsdaten** getrennt, mit denen ein ML-Modell trainiert wird. Testdaten werden verwendet, um die Leistung eines ML-Modells mit Beispielen zu messen, die über die Trainingsdaten hinausgehen.

# Trainieren

**Modelle** für **maschinelles Lernen (ML)** werden anhand von Beispielen in Form von **Daten** trainiert, um Muster zu finden und **Vorhersagen** zu treffen. Im Lauf des Trainings werden die Muster verfeinert, um die Vorhersagen zu verbessern. Ein Beispiel wäre ein Modell zur Empfehlung von Songs. Das Modell wird anhand der Songauswahl vieler Menschen trainiert, um Ähnlichkeiten zwischen den Lieblingsliedern verschiedener Menschen zu finden. Je vielfältiger die Songauswahl ist, anhand derer das Modell trainiert wird, desto besser wird wahrscheinlich die Empfehlungsvorhersage ausfallen. Es gibt viele Methoden zum Trainieren von ML-Modellen unter Verwendung unterschiedlicher Datentypen. Welche Art des Trainierens der oder die Entwickler:in auswählt, hängt von dem zu lösenden Problem und den zu dessen Lösung verfügbaren Daten ab. Die Qualität des Trainings hängt maßgeblich von der Güte der verwendeten Daten ab.

# Trainingsdaten

Trainingsdaten sind beim **maschinellen Lernen (ML)** Beispiele in Form von **Daten**, die zum **Trainieren** von **ML-Modellen** verwendet werden. ML-Entwickler:innen erstellen Modelle, um Muster in den Trainingsdaten herauszuarbeiten, die zum Generieren von **Vorhersagen** über neue Daten verwendet werden können. Ein Beispiel wäre eine Anwendung zur Spracherkennung. Die Trainingsdaten können viele Sprechbeispiele von Personen mit unterschiedlichem Akzent oder Tonfall enthalten. Je genauer die Trainingsdaten die Wirklichkeit abbilden, desto besser ist voraussichtlich die Leistung des Modells.

# Überwachtes Lernen

Überwachtes Lernen ist eine Methode zum **Trainieren** von **Modellen** für **maschinelles Lernen (ML)**. Beim überwachten Lernen werden große Mengen von **Daten** verwendet, die mit relevanten Informationen **gekennzeichnet** wurden. Eine Art des überwachten Lernens ist die **Klassifizierung**. Ein Beispiel für ein Klassifizierungsproblem wäre die Erkennung von Tigern auf Bildern. Die Daten bestehen aus vielen Bildern. Die Bilder, auf denen Tiger zu sehen sind, sind als solche gekennzeichnet.

Das ML-Modell wird mit den gekennzeichneten Bildern trainiert und trifft eine **Vorhersage** darüber, ob auf den Bildern ein Tiger zu sehen ist. Da die Bilder korrekt gekennzeichnet wurden, können die Entwickler:innen und Entwickler beurteilen, wie **genau** die Vorhersagen des Modells sind, und das Training entsprechend anpassen. Anschließend kann das ML-Modell verwendet werden, um bei

völlig neuen Bildern vorherzusagen, ob diese einen Tiger enthalten. Beim überwachten Lernen kommt es darauf an, genügend korrekt gekennzeichnete Daten zu haben, um genaue Vorhersagen zu ermöglichen.

## Unüberwachtes Lernen

Unüberwachtes Lernen ist eine Methode zum **Trainieren** von **Modellen** für **maschinelles Lernen (ML)**. Beim unüberwachten Lernen werden Modelle trainiert, um **Daten** basierend auf Ähnlichkeiten zu organisieren. Dabei werden in den Daten verborgene Muster gefunden. Eine Art des unüberwachten Lernens ist das Clustering. Ein Beispiel für ein Clustering-Problem wäre die **Vorhersage**, wie Gesundheitsdaten in Gruppen eingeteilt werden können, um die Diagnose von Krankheiten zu erleichtern. Solche Gruppen heißen „Cluster“ und sind im Voraus nicht bekannt. Mit dem ML-Modell kann vorhergesagt werden, ob neue Gesundheitsdaten in einen der Cluster fallen. Unüberwachtes Lernen kann bei der Lösung von Problemen hilfreich sein, bei denen Menschen möglicherweise nicht wissen, wonach sie suchen sollen.

## Voreingenommenheit

Voreingenommenheit bedeutet eine Vorliebe für oder eine Abneigung gegen etwas. Ein Beispiel wäre ein Schüler, dem der Englischunterricht mehr Spaß macht als der Mathematikunterricht und der deshalb mehr Zeit für seine Englisch-Hausaufgaben aufwendet. Dieser Schüler hat eine Voreingenommenheit für das Fach Englisch. Es gibt viele Arten der Voreingenommenheit, etwa die **gesellschaftliche Voreingenommenheit** und die **Datenvoreingenommenheit**. Entwicklerinnen und Entwickler von **maschinellern Lernen (ML)** müssen sorgfältig darüber nachdenken, ob die zum **Trainieren von ML-Modellen** verwendeten **Daten** eine Voreingenommenheit aufweisen. Voreingenommenheit kann dazu führen, dass einer Person, einer Gruppe oder einer Reihe von Ideen oder Überzeugungen ein Vorteil gegenüber anderen eingeräumt wird.

## Vorhersage

**Modelle** für **maschinelles Lernen (ML)** werden **trainiert**, um Vorhersagen zu treffen. Die von einem ML-Modell getroffene Vorhersage bezieht sich darauf, was die **Daten** darstellen oder was für eine Aufgabe nützlich sein könnte. Ein Modell kann beispielsweise trainiert werden, um basierend auf den Fernsehgewohnheiten einer Person vorherzusagen, welchen Film sie möglicherweise als Nächstes sehen möchte. Das Modell erzeugt eine Vorhersage, nachdem es anhand der Filmauswahl vieler Menschen trainiert wurde. Die Hauptaufgabe eines ML-Modells besteht darin, Vorhersagen zu treffen. Alle ML-Modelle treffen Vorhersagen, selbst wenn diese für die Nutzer und Nutzerinnen manchmal nicht offen zutage treten.

## Wahrscheinlichkeitsschwellenwert

Der Wahrscheinlichkeitsschwellenwert ist ein Wert, ab dem die Vorhersagen eines **ML-Modells** akzeptiert werden. Der Wahrscheinlichkeitsschwellenwert wird von dem/der ML-Entwickler:in beim Entwerfen des ML-Modells ausgewählt. Beispiel: Ein ML-Modell mit 50 % **Wahrscheinlichkeit** generiert eine Vorhersage, dass es morgen einen Schneesturm geben wird. Wenn der

Wahrscheinlichkeitsschwellenwert auf 60 % festgelegt wurde, wird die Vorhersage als ungenau angesehen. Anders ausgedrückt: Beträgt die Wahrscheinlichkeit einer Vorhersage nicht mindestens 60 %, wird sie nicht als **genau** akzeptiert. Der Wert wird entsprechend der Art des zu lösenden Problems festgelegt, wobei für medizinische Diagnosevorhersagen ein höherer Wahrscheinlichkeitsschwellenwert erforderlich ist als für Songempfehlungen. Durch die Wahl des Schwellenwerts wird das akzeptable Maß an Vorhersagewahrscheinlichkeit festgelegt.

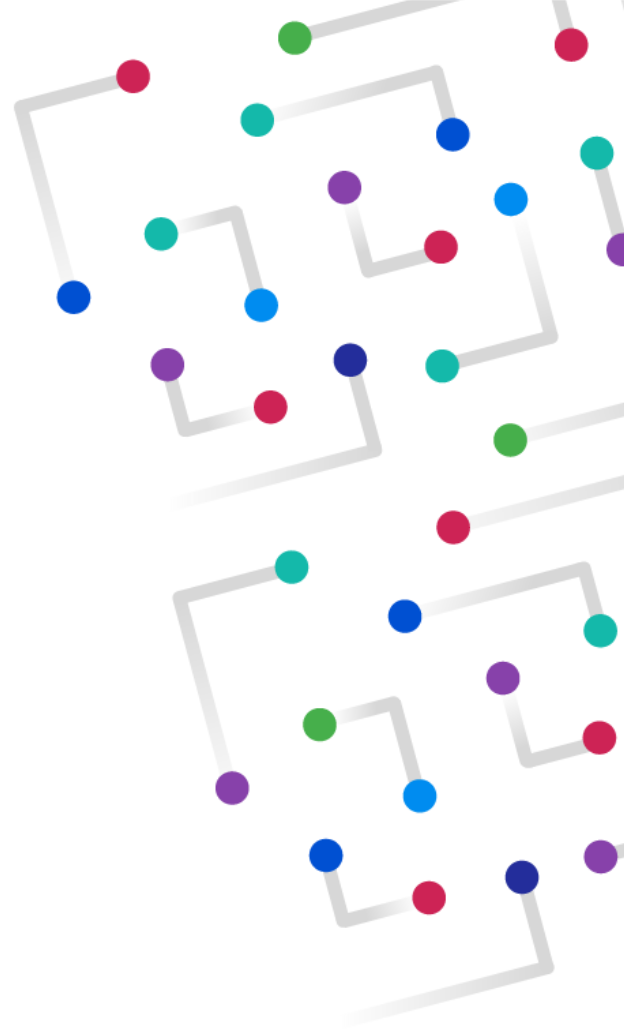
## Wahrscheinlichkeit

Wahrscheinlichkeit bezieht sich darauf, wie sicher etwas ist. Beim **maschinellen Lernen (ML)** ist die Wahrscheinlichkeit eine Methode zur Messung der Sicherheit einer **Vorhersage**. Beispiel: Ein **Klassifizierungsmodell** soll vorhersagen, ob es morgen regnet. Das Modell sagt mit 90-prozentiger Sicherheit voraus, dass es morgen regnen wird. Mit anderen Worten besteht eine 90-prozentige Sicherheit, dass es morgen regnet. Die Verwendung der Wahrscheinlichkeit zum Messen der Sicherheit von Vorhersagen hilft bei der Bewertung der Güte eines ML-Modells.



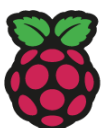
Diese Ressource wird von der [Raspberry Pi Foundation](https://www.raspberrypi.org/) unter der Creative Commons-Lizenz CC BY-NC-ND 4.0 angeboten, die das Vervielfältigen und Weiterverbreiten des Werkes, nicht jedoch seine Veränderung und seine kommerzielle Nutzung gestattet. Weitere Informationen zu dieser Lizenz finden Sie auf [creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/).

# Experience AI



Dieses Glossar wurde von der Raspberry Pi Foundation als Teil ihres Lernprogramms „Experience AI“ erstellt. Weitere Informationen auf [www.experience-ai.org](http://www.experience-ai.org).

Diese Ressource wird von der Raspberry Pi Foundation unter der Creative Commons-Lizenz CC BY-NC-ND 4.0 angeboten, die das Vervielfältigen und Weiterverbreiten des Werkes, nicht jedoch seine Veränderung und seine kommerzielle Nutzung gestattet. Weitere Informationen zu dieser Lizenz finden Sie auf [creativecommons.org/licenses/by-nc-nd/4.0](http://creativecommons.org/licenses/by-nc-nd/4.0).



**Raspberry Pi**  
Foundation