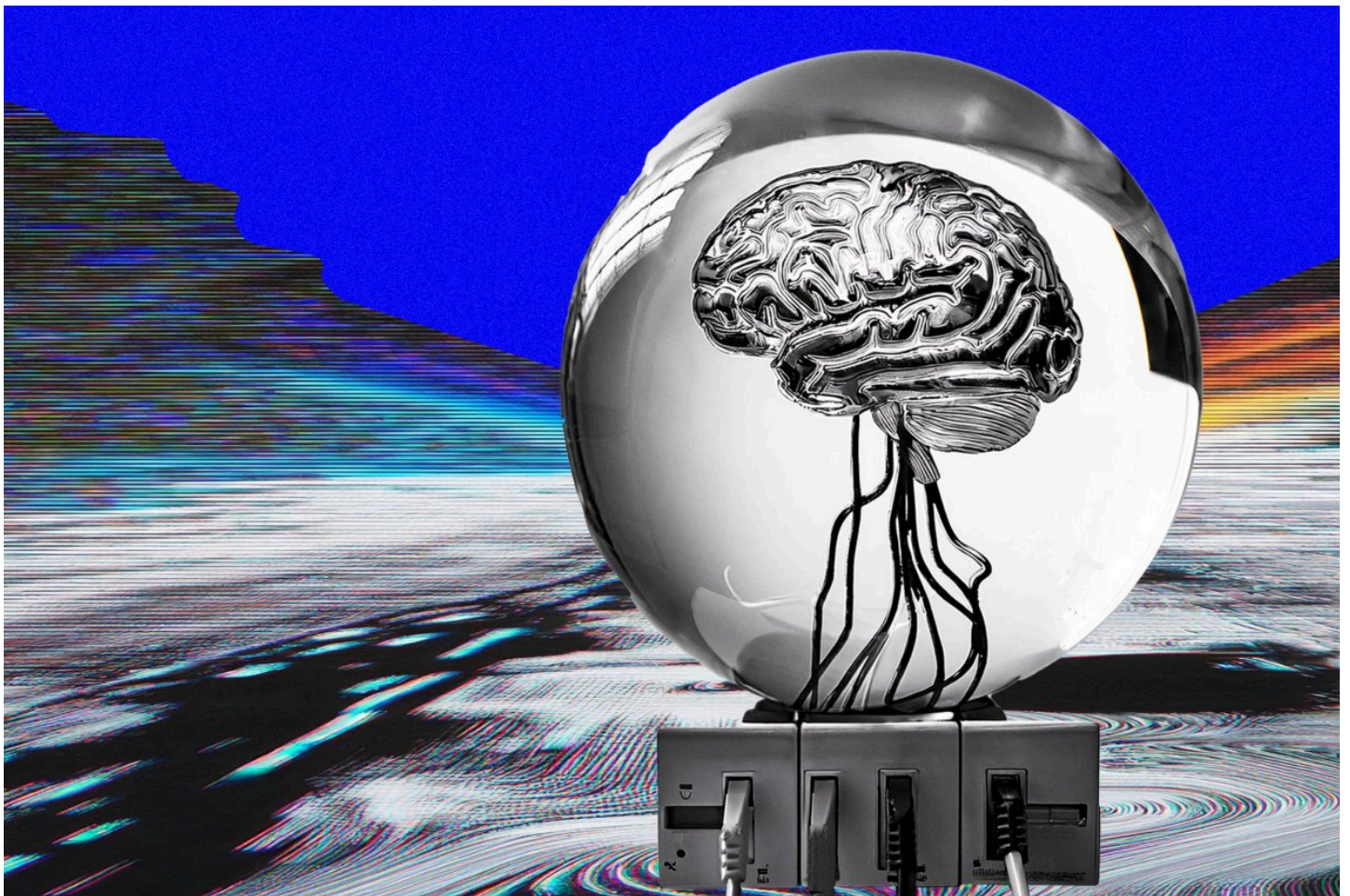


Was ist KI und welche Formen von KI gibt es?

Philipp Ladwig

05.12.2024 / 8 Minuten zu lesen

Was ist eigentlich künstliche Intelligenz genau? Wie funktioniert sie und was ist Deep Learning im Gegensatz dazu?



Das Gehirn im Tank? Wie wird die KI zukünftig unsere Realität bestimmen? Dafür muss erstmal die "Verschaltung" verstanden werden. Hier ist auch das Gehirn im Tank KI generiert. | Illustration: www.leitwerk.com (© bpb)

Künstliche Intelligenz, kurz KI, ist ein Teilbereich der Informatik. Sie beschäftigt sich mit der Entwicklung von Algorithmen, die menschliche kognitive Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität imitieren. KI kann große Mengen an Daten wie Texte, Bilder oder Audiosignale analysieren, darin Muster erkennen und daraus Erkenntnisse gewinnen. Der Begriff KI wurde erstmals von amerikanischen Wissenschaftlern verwendet. Die englische Übersetzung für KI ist AI und steht für „Artificial Intelligence“. In den 1950er Jahren wurden die theoretischen Grundlagen für die heutige Computertechnologie gelegt und mit ihr die ersten Überlegungen angestellt, wie Algorithmen erstellt werden können, die selbstständig Problemlösungen erlernen, z.B. durch ein sich wiederholendes „Versuch-und-Irrtum“-Verfahren.

Um KI zu verstehen, ist der Begriff des Algorithmus wichtig. Ein Algorithmus beschreibt eine klar definierte Abfolge von endlich vielen Schritten oder Anweisungen, um ein bestimmtes Problem zu lösen oder eine Aufgabe zu erfüllen. Ein Kochrezept ist ein anschauliches Beispiel für einen Algorithmus aus dem Alltag. Man gibt Zutaten hinein, verarbeitet diese und erhält ein Ergebnis – in diesem Fall das fertige Gericht. In der Informatik erhalten Algorithmen Eingabedaten, verarbeiten diese nach einer bestimmten Vorschrift und geben Ausgabedaten zurück.

Das Besondere an KI ist, dass KI eine Gruppe von Algorithmen darstellt, die bis zu einem gewissen Grad automatisiert weitere Algorithmen (Kochrezepte) erstellen können und somit festgelegte Abläufe zur Lösung bestimmter Probleme generieren können. Während man früher mit KI nur sehr spezifische Probleme aus einer Domäne lösen konnte, werden die Algorithmen der KI immer anpassungsfähiger und können auch domänenübergreifend eingesetzt werden. Dieselbe KI kann heute beispielsweise Städtetrips planen, programmieren oder Hausaufgaben lösen. In den letzten 75 Jahren haben sich verschiedene Formen der KI entwickelt, wobei sich bestimmte Ansätze durchgesetzt haben. Zwei Fachbegriffe, die in den letzten Jahren viel Aufmerksamkeit in den Medien erhalten haben, sind maschinelles Lernen und Deep Learning bzw. tiefes Lernen. **Maschinelles Lernen (ML)** ist eine Teildisziplin der KI. In den 1950er Jahren war die Idee der KI eher ein abstraktes und theoretisches Modell, das jedoch in den 1960er Jahren durch die Konzepte des maschinellen Lernens erstmals in die Praxis umgesetzt werden konnte. Dabei werden Algorithmen eingesetzt, die anhand eines Datensatzes trainiert werden und Aussagen über statistische Wahrscheinlichkeiten treffen. Diese Aussagen basieren

auf der Erkennung von Mustern in den Daten. Die Muster werden nicht von einem Programmierer vorgegeben, sondern der Algorithmus findet sie selbst. Nehmen wir das Beispiel der Erkennung von Katzen in einem Bild. Ein ML-System kann beispielsweise mit einem Datensatz trainiert werden, in dem Menschen zuvor Bilder mit der Information versehen haben, ob auf einem Bild eine Katze zu erkennen ist oder nicht. Das ML-System sucht nach Mustern, in diesem Fall nach typischen optischen Merkmalen einer Katze, und gibt nach der Analyse eines Bildes einen Zahlenwert über die Wahrscheinlichkeit zurück, ob auf dem Bild eine Katze zu sehen ist oder nicht. Beim Training eines solchen Systems müssen viele tausend Katzenbilder verarbeitet werden, damit das System zuverlässige Aussagen treffen kann.

Das **tiefe Lernen** (englisch Deep Learning) ist wiederum eine Teildisziplin des maschinellen Lernens und der Grund für den heutigen Erfolg von z.B. ChatGPT. Beim tiefen Lernen werden tiefe (daher „deep“) künstliche neuronale Netze trainiert. Das Wort „tief“ bedeutet in diesem Zusammenhang, dass viele künstliche Neuronen miteinander verschaltet werden. Ihre Funktionsweise ist dem menschlichen Gehirn nachempfunden und wurde bereits theoretisch in den 1940er und 50er Jahren beschrieben. Im Bereich des maschinellen Lernens waren neuronale Netze lange Zeit ein Werkzeug unter vielen, doch durch Forschungsergebnisse, insbesondere aus den letzten zehn Jahren, konnte ihr Potenzial bewiesen werden und breite Anwendung finden. Ein Durchbruch war es, dass man die Netze softwareseitig beliebig groß skalieren konnte. Die Skalierung wird heute nur noch durch Limitierungen der Hardware wie z.B. Speichergröße eingeschränkt. Bezugnehmend auf das Beispiel mit dem Erkennen von Katzen in Bildern, sind heutige Systeme des tiefen Lernens (also neuronale Netze) den bisherigen Ansätzen aus dem Bereich der KI in der Genauigkeit der Aussagen und der Geschwindigkeit der Antwortgenerierung überlegen. Neuronale Netze können aber nicht nur Katzen in Bildern erkennen, sondern auch Texte über Katzen schreiben, Bilder oder sogar Videos von Katzen generieren und vieles mehr. Wenn heute in den (sozialen) Medien von (generativer) KI die Rede ist, sind meist die Ergebnisse neuronaler

Dieser Inhalt muss geladen sein, um ihn drucken zu können. Zum Laden scrollen Sie auf der Website zu dem Inhalt und drucken Sie erneut.

Abbildung 1: Das tiefe Lernen (Deep Learning) ist eine Teildisziplin des maschinellen Lernens. Dies ist wiederum eine Teildisziplin der künstlichen Intelligenz.

Netze gemeint. Auch ChatGPT ist ein neuronales Netz (auch Large Language Model, kurz LLM genannt), das aus mehreren Milliarden Neuronen besteht und mit einer großen Menge an Textdaten trainiert wurde. KI ist jedoch nicht immer ein neuronales Netz. Es gibt viele Anwendungen, in denen einfachere Lernalgorithmen oder statistische Methoden ausreichen und etabliert sind, z.B. in Bereichen, in denen die Datenmengen überschaubar sind und Entscheidungen durch eine Folge von Wenn-Dann-Regeln vorhergesagt werden können.

Zusammenfassend lässt sich sagen, dass KI und insbesondere neuronale Netze ein umfassendes Spektrum an Möglichkeiten bieten. KI ist in der Lage, Muster in großen Datenmengen zu erkennen, zu klassifizieren, neue Daten zu synthetisieren und kann flexibel auf eine Vielzahl von Problemen angewandt werden. Dabei wird eine Generalisierung der Eingangs- und auch Ausgangsdaten erreicht, die den ursprünglichen Trainingsdaten zwar ähneln, aber nicht mit ihnen identisch sind. Die Anwendungsmöglichkeiten sind daher enorm vielfältig.

Die Vielseitigkeit von KI

Bemerkenswert ist, dass die Art der Daten, die neuronale Netze verarbeiten, im Prinzip irrelevant ist. Ob Text, Bild, Ton oder Video - alles ist letztlich in Bits und Bytes kodiert, und die Architektur des neuronalen Netzes muss lediglich in der Lage sein, die jeweilige Datenmenge zu verarbeiten. So kann ein Netz, das Text erzeugt, für die Erzeugung von Bildern angepasst werden. Ein einzelnes Wort in einem Text ist für eine KI das, was ein einzelner Pixel (Bildpunkt) in einem Bild ist. Dabei berechnet die KI Wahrscheinlichkeiten dafür, an welcher Stelle im Text bzw. im Bild ein bestimmtes Wort bzw. ein bestimmter Pixel am besten zur Lösung des Problems beiträgt. Diese Analogie lässt sich beliebig auf andere Medien und Datentypen übertragen. Das macht KI und insbesondere neuronale Netze sehr anpassungsfähig und ist der Grund, warum diese Technologie bereits heute in vielen Bereichen unseres Alltags präsent ist. Textgenerierung, Bildsynthese, autonomes Fahren und Deepfakes sind nur wenige der heutigen Anwendungsbereiche von KI.

Wie funktionieren künstliche neuronale Netze?

Als vereinfachte Betrachtungsweise, können wir ein künstliches Neuron als eine simple Rechenoperation betrachten. Zum Beispiel kann ein Neuron eine Zahl, sagen wir 55, empfangen, diese Zahl mit 2 multiplizieren und das Ergebnis, in diesem Fall 110, an die nächsten Neuronen weitergeben. Diese Neuronen führen dann ihre eigenen Rechenoperationen mit anderen Multiplikationswerten durch und so weiter. Die Fähigkeit, statistische Aussagen bzw. zuverlässige Vorhersagen über eine Eingabe zu machen, erhält das neuronale Netz durch das Training mit einem Trainingsdatensatz. Ein Datensatz kann z.B. aus Paaren von Eingabedaten und korrekten Ergebnissen bestehen. Nimmt man vereinfacht an, dass jedes Neuron nur eine Multiplikation ist, dann wird in vielen Trainingsdurchläufen nach und nach der Multiplikationsfaktor ermittelt, also mit welchem Wert die Eingabezahl multipliziert werden muss, um am Ausgang des Netzes das richtige Ergebnis zu erhalten. In jedem Trainingslauf wird die Vorhersage des Netzes mit dem korrekten Ergebnis aus dem Trainingsdatensatz verglichen. Weicht das Ergebnis ab, werden die Neuronen entsprechend angepasst. Die Berechnung dieses Fehlers wird als Fehlerfunktion bezeichnet und dient als Feedback-Mechanismus für das Lernen. Ein Grund für den heutigen Erfolg der KI ist, dass die Forschung immer effektivere Verfahren gefunden hat, um mit mathematischen Methoden den richtigen Multiplikationswert der einzelnen Neuronen zu bestimmen.

Fehlerhafte Vorhersagen und Vorurteile

Die Qualität der Vorhersage von KI-Modellen ist nur so gut wie die Daten, mit denen sie trainiert werden. Werden in unserem Beispiel Hunde fälschlicherweise als Katzen deklariert, kann dies zu Fehlinterpretationen des Systems führen. Die Trainingsdatensätze für heutige Modelle werden jedoch nicht immer minutiös von Hand ausgewählt, sondern KI-Systeme erhalten riesige Datenmengen, in denen sie Muster finden sollen. Die Verarbeitung großer Datenmengen und das Finden von Zusammenhängen wurde bisher auch als „Big Data“ bezeichnet. Dieser Begriff weicht vermutlich immer mehr dem Begriff der KI. Obwohl diese riesigen Datensätze vorher automatisch gefiltert werden, also anstößige Aussagen oder Bilder entfernt werden, können weniger auffällige Inhalte dennoch versteckte Muster und Vorurteile enthalten. Warum ist das so?

Es ist wichtig zu verstehen, dass KI statistische Wahrscheinlichkeiten abbildet und Strukturen in dem zuvor verarbeiteten Trainingsdatensatz repliziert. Diese Strukturen können sehr komplex sein, und je komplexer der selbstlernende Mechanismus ist, desto schwieriger ist es für Menschen, die Schlussfolgerungen des Systems nachzuvollziehen.

Um dies besser zu verstehen, ist es hilfreich zwei etablierte Trainingsverfahren näher zu betrachten: „**Überwachtes Lernen**“ und „**Unüberwachtes Lernen**“. Das Beispiel mit den Katzen ist eine Form des überwachten Lernens, da die Datenpaare vorgegeben sind und das Ergebnis vorhersagbar, also ‚überwacht‘, ist. Wenn KI-Systeme automatisch nach Mustern und Zusammenhängen suchen sollen, ist das Ergebnis offen und nicht ‚überwacht‘. Unüberwachtes Lernen wird heute von Plattform-Betreibern von sozialen Medien häufig eingesetzt, um beispielsweise Nutzergruppen mit ähnlichen Interessen zu finden. Dabei kann nicht konkret nach einfachen Indikatoren wie „Katze im Bild“ und „keine Katze im Bild“ klassifiziert werden, da die Strukturen, die eine solche Nutzergruppe ausmachen, viel komplexer sind und vor allem vor dem Training nicht bekannt sind. Ein damit einhergehendes Problem des unüberwachten Lernens ist die Bildung und Verstärkung von Vorurteilen, was im folgenden Beispiel weiter verdeutlicht werden soll.

Ein Praxisbeispiel: Entscheidungssysteme

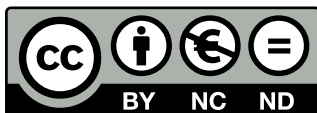
Entscheidungssysteme auf Social-Media-Plattformen nutzen historische Nutzerdaten, um Inhalte vorzuschlagen, die auf den Interessen der Nutzenden basieren. Dabei werden Eigenschaften von Beiträgen, Interaktionen und Nutzerdaten wie z.B. Alter und Geschlecht berücksichtigt. Diese Systeme lernen Muster in den Daten und optimieren ihre Vorhersagen über die mögliche Betrachtungsdauer von Inhalten. Durch unüberwachtes Lernen werden Nutzergruppen mit ähnlichen Interessen identifiziert. Es ist schwierig, diese Systeme des unüberwachten Lernens vorurteilsfrei zu trainieren, da die Methoden immer statistische Wahrscheinlichkeiten reproduzieren und somit z.B. Sexismus und Rassismus verstärken können, wenn die Datensätze nicht ausreichend gefiltert, divers und ausgewogen sind. Diese Verzerrungen können Vorurteile fördern, da sie die Nutzer in ihren bisherigen Interessen bestätigen, anstatt Vielfalt zu fördern. Daher ist es wichtig, diese Systeme verantwortungsbewusst zu entwickeln und die Nutzenden über mögliche Verzerrungen zu informieren.

Kritik aus technischer Sicht

Die Vorteile selbstlernender Systeme, die Muster in Daten erkennen und damit Probleme lösen können, sind enorm. Allerdings gibt es auch Kehrseiten, wie die bereits erwähnte mögliche Verstärkung von Vorurteilen, Datenschutzprobleme sowie der vermutlich global stark steigende Energieverbrauch durch KI. Sowohl das Training neuronaler Netze als auch die Inferenz, also das Ausführen des Schlussfolgerungsprozesses, benötigt elektrische Energie. Verschiedene Studien kommen zu unterschiedlichen Ergebnissen des zukünftigen Mehrverbrauchs, jedoch geht aus dem Nachhaltigkeitsbericht 2024 von Google heute schon hervor, dass der Konzern 48% mehr Treibhausgas als 2019 ausgestoßen hat. Google schreibt (aus dem Englischen übersetzt): „Mit der weiteren Integration von KI in unsere Produkte kann die Verringerung der Emissionen aufgrund des steigenden Energiebedarfs durch die höhere Intensität der KI-Berechnungen eine Herausforderung darstellen.“^[1]

Fußnoten

^[1] S. 31 in <https://www.gstatic.com/gumdrop/sustainability/google-2024-environmental-report.pdf>



Dieser Text ist unter der Creative Commons Lizenz "CC BY-NC-ND 4.0 - Namensnennung - Nicht kommerziell -

Keine Bearbeitungen 4.0 International" [Link: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>]

[Link: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>] veröffentlicht. Autor/-in: Philipp Ladwig für bpb.de

Sie dürfen den Text unter Nennung der Lizenz CC BY-NC-ND 4.0 und des/der Autors/-in teilen.

Urheberrechtliche Angaben zu Bildern / Grafiken / Videos finden sich direkt bei den Abbildungen.

Sie wollen einen Inhalt von bpb.de nutzen? [Link: <https://www.bpb.de/die-bpb/faq/184955/nutzungsrechte/>]