



Große Sprachmodelle

Ein Überblick

Herausgeber

Bitkom e.V.
Albrechtstraße 10
10117 Berlin
T 030 27576-0
bitkom@bitkom.org
www.bitkom.org

Ansprechpartner

Merle Uhl | Bereichsleiterin Künstliche Intelligenz
T 030 27576-242 | m.uhl@bitkom.org

Verantwortliches Bitkom-Gremium

AK Artificial Intelligence

Autorinnen und Autoren

Lorenz Lehmhaus | Aleph Alpha GmbH
Christopher Kränzler | Lengoo GmbH
Dr. Maria Börner | Westernacher Solutions GmbH

Layout

Anna Stolz | Bitkom e.V.

Titelbild

© Christina Wocintechchat – unsplash.com

Copyright

Bitkom 2023

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

1	Was sind große Sprachmodelle?	4
	Grundkonzept	4
	Weitere Konzepte	5
2	Funktionsweise	7
	Prozessschritte	7
	Die Richtigkeit der Antwort	8
3	Anwendungsbereiche	10
4	Wie Unternehmen profitieren können	13
5	Empfehlungen	14

1 Was sind große Sprachmodelle?

KI-Sprachmodelle oder Large Language Models (LLMs), eine Ausführung sogenannter »Foundation Models«, sind die neueste Entwicklung im Bereich der Künstlichen Intelligenz – sie haben die Fantasie der Öffentlichkeit entfacht und eine beispiellose Bekanntheit erlangt hat. Durch ihr Potenzial können zeitaufwändige menschliche Aufgaben im Bereich der Erstellung, Überarbeitung, Transformation und Analyse von Inhalten beschleunigt oder sogar komplett automatisiert werden. Gleichzeitig haben die Fähigkeiten der Modelle aber auch Bedenken geweckt.

Um eine fundierte Diskussion über die Konsequenzen dieser technologischen Entwicklung zu führen, soll dieser Leitfaden die technischen Grundlagen großer Sprachmodelle erläutern. Die Funktionsweise sowie zentrale Aspekte im Kontext der Modelle werden daher beleuchtet und allgemeine Anwendungsfälle skizziert. Diese werden zudem mit konkreten Use Cases beispielhaft dargestellt. Zuletzt werden Möglichkeiten diskutiert, wie Unternehmen von dieser technischen Entwicklung profitieren können und einige Empfehlungen für die Entwicklung und Nutzung der Modelle formuliert.

1.1 Grundkonzept

»GPT« ist ein bekannter Name für eine Reihe von großen Sprachmodellen, die von OpenAI entwickelt wurden. Das berühmteste GPT-Modell ist ChatGPT, welches es Nutzerinnen und Nutzern ermöglicht, über natürliche Spracheingaben mit der Technologie zu interagieren. Der Name »GPT« steht für »generative pre-trained transformer«. Das bedeutet, dass diese Modelle darauf trainiert wurden, automatisch Texte zu erzeugen und dabei auf vorherige Erfahrungen zurückzugreifen, die sie im Vorfeld gesammelt haben. ChatGPT hat seit seiner Veröffentlichung im November 2022 viel Aufmerksamkeit gewonnen und wird von Millionen von Nutzerinnen und Nutzern angewendet.

- **Generative** – Der »generative« Aspekt eines großen Sprachmodells ist seine grundlegende Fähigkeit, Text selbst zu generieren, indem es das beste nächste Wort in einer Antwort vorhersagt. Die Reihenfolge der Wörter und Phrasen in seinen Antworten basiert auf den Mustern, die das Modell aus den Trainingsdaten gelernt hat.
- **Pre-trained** – Große Sprachmodelle werden als »vortrainiert« bezeichnet, da die verwendeten Deep Learning-Algorithmen mit einer enormen Menge an Trainingsdaten versorgt werden, bevor eine spezifische Aufgabe für das neuronale Netz definiert wird. Mit anderen Worten: Die Lernalgorithmen suchen im Trainingsmaterial nach Informationen und Mustern, die ihnen helfen, fließenden Text zu generieren. Dabei versuchen sie nicht, diese Informationen und Muster an bestimmte Arten von Fragen oder Aufgaben anzupassen.

Diese Modelle können wiederum die Grundlage für explizit trainierte KI-Systeme (daher die Bezeichnung »Foundation Model«) bilden oder über natürliche Spracheingabe der Nutzerinnen und Nutzer zur Ausführung einer Vielzahl von Aufgaben genutzt werden.

- **Transformer** – Große Sprachmodelle basieren auf der sogenannten Transformer-Architektur für neuronale Netze. Diese Architektur wird bereits seit über fünf Jahren von Entwicklerinnen und Entwicklern maschineller Übersetzung genutzt und wurde in dieser Zeit kontinuierlich weiterentwickelt. Der Erfolg von Transformernetzen bei der Bewältigung einer der ältesten Herausforderungen der Informatik – der maschinellen Übersetzung – hat dazu beigetragen, dass sie inzwischen als Standard gelten. Ihr Erfolg beruht darauf, dass sie Sequenz-zu-Sequenz-Probleme effizient lösen, indem sie eine Eingabesequenz von Wörtern in einer Quellsprache in eine geeignete Sequenz von Wörtern in der Zielsprache umwandeln. Im Vergleich zu älteren neuronalen Netzwerkarchitekturen wie rückgekoppelten neuronalen Netzen (RNNs) oder den sogenannten Convolutional Neural Networks (CNNs) sind Transformer in ihrer Struktur einfacher aufgebaut. Durch Parallelisierung des Trainingsprozesses können sie wesentlich skalierbarer trainiert werden. Ein weiterer erheblicher Vorteil der Transformer-Architektur besteht darin, dass sie in der Lage sind, den Input nicht sequenziell, sondern gleichzeitig zu verarbeiten. Dies erlaubt es ihnen, entfernte Zusammenhänge im Input zu nutzen, z. B. das letzte Wort in einem langen Satz, das die Bedeutung des ersten Wortes beeinflusst.

1.2 Weitere Konzepte

Zwei weitere wichtige Konzepte gilt es hier zu berücksichtigen: **Multimodalität** und **neue Fähigkeiten**, die große Sprachmodelle entwickeln.

- **Multimodalität** – Wenn ein großes Sprachmodell mit Material trainiert wird, das mehrere Arten von Informationen (oder Modalitäten) enthält, wie Text, Bilder, Audio, Video und sogar Computercode, wird das resultierende Modell als multimodal bezeichnet. Solche Modelle haben die Fähigkeit, verschiedene Arten von Daten und Informationen gleichzeitig zu verarbeiten und zu verstehen, wodurch sie eine Vielzahl von Aufgaben ausführen können, die Kenntnisse über mehr als eine Art von Daten erfordern. Beispielsweise kann ein Modell, das sowohl mit Text als auch mit Bildern ausgebildet ist, beschreibende Bildunterschriften für Bilder generieren, indem es sein Verständnis des visuellen Inhalts mit seinen Kenntnissen der natürlichen Sprache kombiniert. Ein Modell, das sowohl mit Text- als auch mit Audiodaten ausgebildet ist, kann z. B. verwendet werden, um Sprache zu transkribieren, da es die Geräusche der gesprochenen Sprache mit den entsprechenden geschriebenen Wörtern in Verbindung bringen kann.
- **Neue Fähigkeiten** – Was vielleicht ein noch bemerkenswerterer Aspekt großer Sprachmodelle ist, sind neue Fähigkeiten und Verhaltensweisen, die entstehen, wenn der Umfang der Modelle eine bestimmte Schwelle überschreitet. Da diese Modelle mit massiven Mengen an Textda-

ten trainiert werden, können sie viel über die Struktur und Muster der menschlichen Sprache lernen. Durch diese Analyse baut ein großes Sprachmodell auch ein semantisches Mapping der Sprache auf und lernt damit gewissermaßen nicht nur die Sprache an sich, sondern auch den Inhalt der Sprache. So können die Modelle dieses Wissen verallgemeinern und auf neue Aufgaben und Bereiche anwenden, für die sie nicht speziell trainiert wurden.

Moderne große Sprachmodelle können dazu veranlasst werden, eine Vielzahl von Aufgaben zu erfüllen, die über die traditionelle Verarbeitung natürlicher Sprache hinausgehen, z. B. mathematische Berechnungen, Musikkomposition und sogar Computerprogrammierung.

Als zum Beispiel ChatGPT auf der Grundlage von GPT 3.5 aufgefordert wurde, die Fläche eines Dreiecks mit einer Höhe von 2,5 Metern und einer Basis von 3,7 Metern zu berechnen und seine Vorgehensweise zu erklären, gab es folgende Antwort:

Um die Fläche eines Dreiecks zu berechnen, müssen Sie die Formel verwenden:

$$\text{Fläche} = (\text{Basis} \times \text{Höhe}) / 2$$

In diesem Fall wird die Höhe des Dreiecks als 2,5 Meter und die Breite des Dreiecks (das auch die Basis ist) als 3,7 Meter angegeben. So können wir diese Werte in die Formel ersetzen, um zu erhalten:

$$\text{Fläche} = (3,7 \text{ Meter} \times 2,5 \text{ Meter}) / 2$$

$$\text{Fläche} = 4,625 \text{ Quadratmeter}$$

Somit beträgt die Fläche des Dreiecks 4,625 Quadratmeter.

Der Grund, warum wir in der Formel durch 2 teilen, ist, dass ein Dreieck die Hälfte eines Rechtecks ist. Wenn wir die Basis und die Höhe eines Dreiecks multiplizieren, erhalten wir die Fläche des Rechtecks, das das Dreieck umfasst. Aber da wir nur die Fläche des Dreiecks wollen, teilen wir durch 2, um die Hälfte der Fläche des Rechtecks zu erhalten. Deshalb ist die Formel für die Fläche eines Dreiecks

$$(\text{Basis} \times \text{Höhe}) / 2.$$

Eine neue Fähigkeit ist also eine Aufgabe, die zufällig und daher unzuverlässig Ergebnisse in immer größeren Iterationen eines Modells erzeugt, aber dann plötzlich signifikant verbesserte Ergebnisse liefert, nachdem das Modell eine bestimmte Größe erreicht hat. Basierend auf der schlechten Leistung einer solchen Aufgabe in sukzessiv größeren Modellen würde man keine plötzliche Verbesserung der Leistung bei dieser Aufgabe erwarten, wenn das Modell ein bestimmtes Level erreicht. [2]

Insgesamt sind neue Fähigkeiten großer Sprachmodelle ein Beweis für die Kraft des Deep Learnings und die Fähigkeit dieser Modelle, über Sprache zu lernen und aus riesigen Datenmengen eine Art Weltwissen aufzubauen. Da sich diese Modelle weiterentwickeln und verbessern, werden wir wahrscheinlich in Zukunft noch mehr Beispiele für neue Funktionalitäten sehen.

2

Funktionsweise

»Generative pre-trained Transformers« können mit einem großen, mehrdimensionalen Fischernetz verglichen werden, in dem jeder Knoten ein Wort ist und jedes Seil, das zwei Punkte verbindet, die Wahrscheinlichkeit ist, dass diese beiden Wörter nebeneinander erscheinen. Ein LLM sagt lediglich das wahrscheinlichste nächste Wort basierend auf den vorherigen Wörtern in einem Satz vorher. Trainingsmaterial wird aus dem gesamten Internet gesammelt – also von Menschen erzeugte Inhalte – und die Modelle lernen die Wahrscheinlichkeiten, dass Wörter nebeneinander vorkommen (auch Gewichtung genannt). Die Modelle werden außergewöhnlich gut darin zu imitieren, was für Menschen natürlich klingt, und dadurch wird eine Aura der Intelligenz erzeugt. Anhand dieser Analogie wird deutlich, dass diese Modelle nicht wirklich verstehen, was sie tun. Man könnte sogar so weit gehen zu sagen, dass frühere auf maschinellem Lernen basierte Ansätze, die mit sehr aufgabenspezifischen Daten trainiert wurden, um Muster zu finden und Vorhersagen über einzelne Ereignisse zu treffen, möglicherweise mehr »Intelligenz« besitzen als diese neuen Systeme.

Der wichtige Unterschied ist, dass LLMs die menschliche Sprache sehr gut imitieren und uns daher intelligent erscheinen, weil wir den Sprachgebrauch mit Intelligenz gleichsetzen. Es wäre jedoch genauer, LLMs als eine Art Sprachschicht darzustellen, die es uns ermöglicht, mit Daten in natürlicher Sprache zu interagieren. Und hier wird es interessant: Wir können diese Modelle mittels sogenanntem »Prompt Engineering« nicht nur steuern, sondern damit den Modellen auch Kontext geben.

2.1 Prozessschritte

Der Prozess, mit dem ein LLM eine Eingabe oder »Prompt« verarbeitet, kann in folgende Schritte unterteilt werden:

Tokenization der Eingabe

Das Modell untersucht den Satz und teilt ihn in kleine Stücke auf, die Tokens genannt werden. Jedes Token steht für ein Wort oder einen Teil davon. Zum Beispiel wird der Satz »Wie funktioniert ein großes Sprachmodell?« in Tokens wie »Wie«, »funktioniert«, »ein«, »großes«, »Sprachmodell« und »?« aufgeteilt.

Optimierung des Inputs für das LLM

Das Modell analysiert den Prompt und verwendet Informationen aus dem Trainingsmaterial sowie die Position der Wörter im Satz, um deren Bedeutung zu verstehen. Diese Informationen werden als Zahlen gespeichert. Durch die Kombination der Zahlen für alle Wörter entsteht eine Zusammenfassung des Satzes. Das Modell nutzt diese

Ein LLM sagt lediglich das wahrscheinlichste nächste Wort basierend auf den vorherigen Wörtern in einem Satz vorher.

Zusammenfassung, um den Satz in einer geeigneten Form für die weitere Verarbeitung zu organisieren.

Generieren einer Antwort

Das Ziel des LLM ist es, eine Antwort auf die Eingabe (Prompt) zu generieren. Das Modell schaut sich den Prompt an und versucht vorherzusagen, welches Wort oder welcher Satz als nächstes in der Antwort am besten passt. Dabei benutzt es eine Reihe von Transformerschichten, die ihm helfen, sich auf verschiedene Teile der Eingabe zu konzentrieren und Verbindungen zwischen ihnen herzustellen. Am Ende erzeugt das Modell eine Liste von möglichen Wörtern, die als nächstes in der Antwort verwendet werden können. Es wählt das wahrscheinlichste Wort aus und fügt es der Antwort hinzu. Dieser Prozess wird wiederholt, bis die Antwort fertig ist – zum Beispiel, wenn eine maximale Länge erreicht ist oder ein spezielles »Ende-der-Sequenz«-Wort generiert wurde.

Die Antwort in natürlicher Sprache lesbar machen

Das Modell verwandelt die ausgewählten Wörter oder Phrasen in einen Text, den Menschen lesen und verstehen können.

Ausgabe der Antwort

Das Modell stellt den erzeugten Text als Antwort auf den Eingabeaufruf zur Verfügung.

2.2 Die Richtigkeit der Antwort

Wahrscheinlichkeits-, nicht Wahrheitsmodelle

Sprachmodelle versuchen stets, eine Antwort zu geben – auch wenn sie nur begrenzt Informationen zu einem bestimmten Sachverhalt haben. Aufgrund dessen ist es für Nutzerinnen und Nutzer eine stetige Herausforderung die Herleitung einer Art nachvollziehen zu können. Es gibt bereits technische Möglichkeiten zur Erhöhung der Richtigkeit und Nachvollziehbarkeit durch eine kontinuierliche, kritische Evaluation des eigentlichen Nutzers oder der Nutzerin und durch eine Berücksichtigung des jeweiligen Nutzungskontexts.

Menschliches Feedback

Das Potenzial, menschliches Feedback zur Verfeinerung dieser Systeme zu nutzen, wurde bereits von OpenAI durch ihre Investition in die menschliche Beurteilung von GPT-3-Antworten gezeigt. Das gesammelte Feedback führte zu den erheblichen Verbesserungen in der Version GPT-3.5, durch die die Anwendung ChatGPT in ihrer ersten Veröffentlichung so viel Aufmerksamkeit erregte. Die Qualität dieser menschlichen Bewertungen hat maßgeblichen Einfluss auf die Antwortqualität des Modells. Fachkundige Nutzerinnen und Nutzer, die manchmal auch als Data Curators, Citizen Data Scientists oder Humans-in-the-Loop bezeichnet werden, werden künftig einen wichti-

gen Beitrag zum erfolgreichen Einsatz und zur kontinuierlichen Verbesserung künstlich intelligenter Systeme v. a. am Arbeitsplatz leisten.

Sprachmodelle sind keine Wahrheitsmaschinen. Kontext, Nachvollziehbarkeit und Erklärbarkeit können hier helfen.

Kontextualisierung & Finetuning

Der zweite wichtige Hebel liegt in der Fähigkeit dem Sprachmodell so viel relevanten Kontext wie möglich zur Verfügung zu stellen. Mit dem sogenannten Prompt Engineering kann man die Qualität der Antwort über optimierte Prompts erhöhen, in denen man den Kontext definiert, in dem das Modell agieren soll. Obwohl der Umfang des Kontexts begrenzt ist, den man mitgeben kann, zeigt sich schon heute deutlich das Potenzial. Der nächste logische Schritt dieses Konzepts wäre, den Kontext schon im Training des Modells mitzuverarbeiten. Hierbei kann eine deutlich größere Menge an Kontext genutzt werden. Es können heute schon maßgeschneiderte Modelle mit Daten trainiert werden, die ihre Branche oder Organisation zum Inhalt haben.

3 Anwendungsbereiche

Das nächste Wort in einer natürlichen, flüssigen Sprachausgabe vorherzusagen, klingt möglicherweise nicht besonders beeindruckend, bevor man sich die unzähligen realen Anwendungsbereiche vor Augen führt. Die Anzahl der Anwendungsfälle ist grundsätzlich nur dadurch begrenzt, wie kreativ wir Prompts erstellen. Im Folgenden sind potenzielle Anwendungsfälle der Technologie aufgeführt. Zum Ende der Publikation finden sich zudem konkrete Beispiele für den Einsatz der Modelle, bei denen insbesondere auf den nachhaltigen Wert im Unternehmen und den Return of Investment (ROI) eingegangen wird.

Textgenerierung

Ein großes Sprachmodell kann durch passende Eingaben dazu genutzt werden, Texte wie Artikel, Produktbeschreibungen, E-Mails und Social-Media-Posts zu erstellen und dadurch den Prozess der Textproduktion erheblich zu beschleunigen. Einige Anwendungen können sogar vollständig automatisiert werden, während für andere diese Texte als Entwurf verwendet werden und von einem Texter oder einer Texterin bei Bedarf weiterbearbeitet werden können. Zum Beispiel kann ein Sprachmodell aus einem Datensatz von Produktnamen, Stichworten und Spezifikationen ganze Produktbeschreibungen generieren, was die Erstellung eines Katalogs erheblich vereinfacht.

Textüberarbeitung

Ein großes Sprachmodell kann helfen, Texte zu verbessern, indem es Grammatik, Interpunktion und Rechtschreibung korrigiert und sie verständlicher für eine Zielgruppe macht. Es kann sogar spezifische Stilvorgaben eines Unternehmens berücksichtigen.

Ein Beispiel wäre, dass ein Sprachmodell einen komplexen Text in eine einfachere Version umformulieren kann, indem es den Text mit einfacheren Wörtern und Satzstrukturen neu schreibt. Dies erleichtert die effektive Kommunikation mit einem breiteren Publikum durch die Anpassung des Inhalts an die Bedürfnisse der Leserinnen und Leser.

Textzusammenfassung

Große Sprachmodelle können lange Texte in kürzere zusammenfassen, aber dabei trotzdem wichtige Informationen und den Zusammenhang bewahren. Dadurch können Mitarbeitende Zeit sparen und sich auf die für sie relevanten Informationen konzentrieren.

Wissensmanagement

Große Sprachmodelle können dabei helfen, Informationen in Unternehmen besser zu organisieren und zu strukturieren. Hierbei können sie ähnliche Daten oder Konzepte kategorisieren, zusammenfassen und verknüpfen. Oftmals haben Unternehmen viele verschiedene Informationsquellen, wie Wissensdatenbanken, Dokumentenarchive, gemeinsame Laufwerke oder Slack-Kanäle. Das Problem dabei ist oft, dass relevante Informationen mühsam gesucht werden müssen. Ein großes Sprachmodell könnte durch seine Fähigkeit, Informationen zu kategorisieren und zusammenzufassen, jedem Mitarbeitenden jeden Morgen wichtige Informationen zur Verfügung stellen, wie z. B. eine Zusammenfassung der letzten Besprechungsnotizen, Links zu relevanten Dokumenten oder eine Übersicht wichtiger Whitepaper.

Diese Modelle können auch Antworten auf direkte Fragen beantworten, indem sie die Anfrage analysieren und mit relevanten Informationen aus den Trainingsdaten oder auf Basis des zu Grunde liegenden Datensatzes beantworten.

Kundenservices

Mit Hilfe von KI-Sprachmodellen können Chatbots und virtuelle Assistenten erstellt werden, die Anfragen von Kundinnen und Kunden bearbeiten und rund um die Uhr Unterstützung bieten können. Werden Support-Chatbots basierend auf einer Wissensdatenbank gebaut, braucht es Trainingsdaten, um z. B. Anfragen von Kundinnen und Kunden auf bestimmte Informationen der Wissensdatenbank zuzuordnen. Die Trainingsdaten sind in dem Falle potenzielle Anfragen, welche von KI-Sprachmodellen schnell und einfach erweitert werden können. Der Aufbau eines Chatbots kann mittels LLMs effizienter und schneller erfolgen und die Anwendung anschließend die Arbeitsbelastung der menschlichen Kundendienstmitarbeitenden verringern und die Reaktionszeit verbessern.

Maschinelle Übersetzung

Große Sprachmodelle setzen seit Jahren neue Maßstäbe für Innovationen für maschinelle Übersetzungsanbieter. Sie haben in diesem Bereich bereits eine Reife erreicht, die sie für Leuchtturmprojekte in jedem Unternehmen prädestiniert. Sie zeigen nicht nur das enorme Potenzial, das in den Anpassungsmöglichkeiten solcher Modelle steckt, sondern bieten einen sich schnell amortisierenden Anwendungsfall, der nachfolgend hohe Budgeteinsparungen und einen signifikanten Strom an unternehmensspezifischen Sprachdaten bietet. Unternehmen, die international tätig sind, können von der Verwendung von großen, generativen KI-Sprachmodellen zur automatisierten Übersetzung von Inhalten profitieren. Dadurch können sie ihre Inhalte schneller und effizienter in verschiedene Sprachen übersetzen und ihre Reichweite auf globaler Ebene erweitern.

Die Anzahl der Anwendungsfälle ist grundsätzlich nur dadurch begrenzt, wie kreativ wir Prompts erstellen.

Automatisierte Kundenbetreuung

Durch die Verwendung von großen, generativen KI-Sprachmodellen können Unternehmen ihre Kundenbetreuung automatisieren. Chatbots und virtuelle Assistenten können schnell und effizient auf Anfragen von Kundinnen und Kunden zu reagieren und ihnen bei Bedarf helfen.

Personalisierte Produktvorschläge

Mit Hilfe von großen, generativen KI-Sprachmodellen können Unternehmen personalisierte Produktvorschläge für ihre Kundinnen und Kunden erstellen. Diese Vorschläge basieren auf dem Verhalten der Kundinnen und Kunden und können helfen, die Kundenzufriedenheit und den Umsatz zu steigern.

Automatisierte Content-Erstellung

Mit Hilfe von großen, generativen KI-Sprachmodellen können Unternehmen Inhalte automatisch erstellen lassen. Dies kann dazu beitragen, den Zeitaufwand und die Kosten für die Erstellung von Inhalten zu reduzieren und gleichzeitig die Qualität der Inhalte zu verbessern.

Marktforschung und Datenanalyse

Unternehmen können große, generative KI-Sprachmodelle nutzen, um Marktforschung und Datenanalyse durchzuführen. Durch die Analyse von Daten können sie Trends und Muster identifizieren, die ihnen helfen können, ihre Produkte und Dienstleistungen zu verbessern und ihre Kundinnen und Kunden besser zu verstehen.

Und darüber hinaus ...

Es gibt unzählige Möglichkeiten, wie Large Language Models eingesetzt werden können. Sie können zum Beispiel genutzt werden, um Daten zu kategorisieren oder spezielle Trainingsmaterialien zu erstellen. Auch das Schreiben von Computercode auf Basis von einfachen Beschreibungen ist möglich. Je mehr Erfahrung ein Unternehmen mit LLMs sammelt, desto mehr Anwendungsfälle werden sich ergeben, die speziell auf die Bedürfnisse des Unternehmens abgestimmt sind.

4 Wie Unternehmen profitieren können

Große Sprachmodelle können für viele verschiedene Zwecke eingesetzt werden. Ihre Anwendung bietet Unternehmen einige Vorteile, die im Folgenden näher erläutert sind.

Verbesserte Kundenbindung und Loyalität

Generative KI-Sprachmodelle können Unternehmen dabei helfen, besser mit ihren Kundinnen und Kunden in Kontakt zu treten, indem sie mehr natürlichsprachliche Interaktionen, personalisiertere Nachrichten und bessere Empfehlungen ermöglichen. Dies kann ihnen helfen, Kundinnen und Kunden zu binden und ihre Loyalität langfristig zu verbessern.

Gesteigerte Effizienz und Produktivität:

Mit Hilfe von generativen KI-Sprachmodellen können große Unternehmen viele Geschäftsprozesse automatisieren und rationalisieren, z. B. den Kundensupport, die Erstellung von Inhalten und Marketingkampagnen. Dies kann dazu beitragen, die Produktivität zu steigern, Kosten zu senken und wertvolle Ressourcen freizusetzen.

Verbesserte Entscheidungsfindung

Unternehmen können die von generativen KI-Sprachmodellen bereitgestellten Erkenntnisse und Analysen nutzen, um bessere Entscheidungen auf der Grundlage datengestützter Erkenntnisse zu treffen. Dies kann ihnen helfen, ihren Konkurrenten voraus zu sein und neue Wachstums- und Expansionsmöglichkeiten zu erkennen.

Wettbewerbsvorteil

Generative KI-Sprachmodelle können Unternehmen dabei helfen, sich von ihren Mitbewerbern abzuheben, indem sie ihren Kundinnen und Kunden einzigartige, personalisierte Erlebnisse bieten. Indem sie die neuesten KI-Technologien nutzen, können sie der Zeit voraus sein und sich als Marktführer in ihrer jeweiligen Branche etablieren. Werden proprietäre Unternehmensdaten ins Training der Modelle mit einbezogen, kann hieraus auch ein langfristiger Wettbewerbsvorteil geschaffen werden.

Innovation und Agilität

Durch den Einsatz von generativen KI-Sprachmodellen können Unternehmen eine Kultur der Innovation und Agilität fördern, die es ihnen ermöglicht, sich an veränderte Marktbedingungen anzupassen und in einer sich ständig weiterentwickelnden Geschäftslandschaft relevant zu bleiben. Dies kann ihnen helfen, ihren Konkurrenten voraus zu sein und langfristiges Wachstum und Erfolg zu fördern.

5 Empfehlungen

Gut aufgestellte Unternehmen in verschiedenen Branchen profitieren seit Jahren von der Steigerung der Produktivität und des Wettbewerbsvorteils durch Digitalisierung und Computerisierung von Aufgaben. Sie verstehen, dass Geschäft nicht an den technologischen Fortschritt verloren geht, sondern an Konkurrenten, die Technologie besser anwenden als sie.

Dies ist keine Zeit für Moratorien und Verbote, sondern für die Anwendung neuer Technologien, ihren sinnvollen und schnellen Einsatz und das Erlernen ihrer Stärken und Schwächen durch Anwendungen in der realen Welt. Wie bei jeder wesentlichen Geschäftsentscheidung ist es Aufgabe von Unternehmen, die ein großes Potenzial in LLMs sehen, einen ausgewogenen und verantwortungsvollen Ansatz zu verfolgen, der den Nutzen maximiert und die Risiken minimiert. Zu diesen Strategien zählen:

Verantwortungsvolle Entwicklung und Nutzung von KI

- Verantwortung wahrnehmen: Unternehmen sollten ethischen Überlegungen Priorität einräumen, wenn sie große Sprachmodelle entwickeln oder verwenden. Dazu gehört ein risikobasierter Umgang mit den Modellen, die sich an den Vorgaben des kommenden AI Acts orientieren.
- Kontext nutzen: Anwendungsspezifische benutzerdefinierte Modelle, die ein LLM als Grundlage verwenden, können Kontext hinzufügen, um die Richtigkeit der Antworten zu erhöhen und die Rückverfolgbarkeit zu verbessern.
- Qualifiziertes Feedback sicherstellen: Unternehmen sollten darauf achten, dass ausschließlich qualifizierte Nutzerinnen und Nutzer zum Training der Modelle beitragen, da die Qualität des menschlichen Feedbacks maßgeblich zur Qualität des Sprachmodells beiträgt.
- Implementierung von Datenschutzmaßnahmen: Da generative AI-Sprachmodelle große Mengen an Daten verarbeiten und analysieren, müssen Unternehmen sicherstellen, dass Datenschutzmaßnahmen implementiert werden. Dies umfasst unter anderem den Schutz von Kundendaten und die Einhaltung geltender Datenschutzgesetze.
- Schulung und Sensibilisierung von Mitarbeitenden: Unternehmen müssen sicherstellen, dass ihre Mitarbeitenden gut geschult sind und ein grundlegendes Verständnis für generative AI-Sprachmodelle haben. Dies ist wichtig, um sicherzustellen, dass die Technologie korrekt eingesetzt wird und die Mitarbeitenden die Risiken und Potenziale verstehen.

Gut aufgestellte Unternehmen verstehen, dass Geschäft nicht an den technologischen Fortschritt verloren geht, sondern an Konkurrenten, die Technologie besser anwenden als sie.

- **Regelmäßige Überprüfung und Aktualisierung von Modellen:** Generative AI-Sprachmodelle können im Laufe der Zeit veralten oder ungenau werden. Unternehmen müssen sicherstellen, dass ihre Modelle regelmäßig überprüft und aktualisiert werden, um die Genauigkeit und Relevanz der Ergebnisse zu gewährleisten.
- **Implementierung von Kontrollmechanismen:** Unternehmen müssen Kontrollmechanismen implementieren, um sicherzustellen, dass ihre generativen AI-Sprachmodelle nicht unvorhergesehene Ergebnisse liefern oder unerwünschte Auswirkungen haben. Dies kann die Implementierung von Tests und Überprüfungen umfassen, um sicherzustellen, dass die Ergebnisse konsistent und korrekt sind.
- **Zusammenarbeit mit Personen mit Expertise und Stakeholdern:** Unternehmen sollten sicherstellen, dass sie Personen mit Expertise und Stakeholder einbeziehen, um sicherzustellen, dass ihre generativen AI-Sprachmodelle ethisch und verantwortungsvoll eingesetzt werden. Dies kann die Zusammenarbeit mit Expertinnen und Experten aus dem Bereich der künstlichen Intelligenz, der Datenschutz- und Ethikkommissionen sowie anderen Interessengruppen umfassen.

Kontrolle über KI im eigenen Geschäftsumfeld übernehmen

- Unternehmen sollten maßgeschneiderte Modelle auf Basis von LLMs verwenden. Ein Unternehmen hat die volle Kontrolle über diese Modelle und die Nutzung der eigenen Daten der Unternehmen über die eigenen Mitarbeitenden stellt sicher, dass diese genau, rückverfolgbar und frei von Voreingenommenheit sind.
- Unternehmen sollten in die Verbesserung der Genauigkeit, Effizienz und Sicherheit großer Sprachmodelle investieren. Dazu könnte zählen, mit der KI-Forschungsgemeinschaft, der Wissenschaft und anderen Partnern der Industrie zusammenzuarbeiten und aktiv daran zu arbeiten, Voreingenommenheiten in großen Sprachmodellen zu identifizieren und abzumildern.
- Unternehmen sollten proprietäre Daten, die bei der Erledigung von Aufgaben durch Mitarbeitende entstehen, sammeln und für zukünftige Kontextualisierung nutzen. Unternehmensinterne Daten können hier eine zentrale Rolle für die Schaffung eines Wettbewerbsvorteils einnehmen.
- Unternehmen sollten Anleitungen und Schulungen für Nutzerinnen und Nutzer anbieten, die erklären, wie sie große Sprachmodelle wirksam und verantwortungsvoll nutzen können und die das Verständnis ihrer Einschränkungen und potenzielle Voreingenommenheiten einschließen.
- Unternehmen sollten ein Ökosystem mit eigenen Modellen etablieren, die von Fachleuten genutzt und trainiert werden, um das gesamte Potenzial der Technologie bei voller Kontrolle auszuschöpfen.

Die Anwendung dieser Strategien kann Unternehmen dabei helfen, große Sprachmodelle verantwortungsvoll (weiter-) zu entwickeln und zu nutzen.

Referenzen

- [1] Vaswani, A, Shazeer, N, et al. ↗ »Attention Is All You Need«, 6. Dezember 2017.
- [2] Wei, Jason, et al. ↗ »Emerging Abilities of Large Language Models«, August 2022.

Bitkom vertritt mehr als 2.000 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

Bitkom e.V.

Albrechtstraße 10
10117 Berlin
T 030 27576-0
bitkom@bitkom.org

[bitkom.org](https://www.bitkom.org)

bitkom