Einfach, weil’s wichtig ist.

ERGO

ChatGPT und Sprachmodelle: Eine Einführung mit Blick auf die Versicherungsbranche

„The reason to develop AI at all, in terms of impact on our lives and improving our lives and upside, this will be the greatest technology humanity has yet developed."

Sam Altman, CEO von OpenAI

Seitdem das Startup OpenAI im November 2022 ChatGPT veröffentlicht hat, überschlagen sich die Ereignisse und fast täglich gibt es neue Informationen über die innovative Anwendung. Die aktuellen Entwicklungen rund um generative Künstliche Intelligenz im Allgemeinen und große Sprachmodelle im Speziellen lassen vermuten, dass uns dieses spannende Thema noch länger beschäftigen wird.

Auch in der Versicherungsbranche schlägt ChatGPT hohe Wellen. Mitarbeiterinnen und Mitarbeiter verschiedenster Fachabteilungen kommen mit Anfragen oder konkreten Projektvorschlägen auf die Data-Analytics- und Innovations-Abteilungen zu. Es werden unternehmensinterne Informationsveranstaltungen abgehalten, um die notwendige Aufklärungsarbeit zu leisten. PR-Abteilungen beantworten Presseanfragen, in denen Versicherer gebeten werden, Stellungnahmen bezüglich der Chancen, Risiken und konkreten Einsatzbereiche von ChatGPT abzugeben.

Das alles passiert in einem hochdynamischen Umfeld, in dem in erstaunlicher Geschwindigkeit eine technologische Weiterentwicklung von der nächsten abgelöst wird. Und immer schwingt die Frage mit, ob wir hier Zeitzeuge einer nie dagewesenen disruptiven Entwicklung sind.

All das ist Grund genug, sich eindringlich mit dem Thema zu beschäftigen. Das haben auch wir im ERGO Innovation Lab getan und tun es immer noch. In unseren Gesprächen mit Mitarbeiterinnen und Mitarbeitern, externen Partnern und Machine-Learning-Experten begegnen uns immer wieder zwei Extreme: Entweder sind ChatGPT und große Sprachmodelle die Lösung vieler Probleme und führen uns in eine goldene Zukunft oder sie sind gehypte, seelenlose, ethisch-fragwürdige stochastische Papageien. Die Wahrheit liegt bekanntlich irgendwo in der Mitte.

Dieses Whitepaper zeigt deshalb, welche Potentiale große Sprachmodelle und Anwendungen wie ChatGPT haben, gibt Einblick in die Funktionsweise, nennt Grenzen und Risiken und erläutert, welche Voraussetzungen für den Einsatz in der Versicherungsbranche geschaffen werden müssen, um die Potentiale nutzbar zu machen und die Risiken eindämmen zu können.

Inhaltsverzeichnis

Wie OpenAI mit ChatGPT einen Hype auslöste

Die Spezialisten von OpenAI forschen an Künstlicher Intelligenz

Das US-amerikanische Startup OpenAI entwickelt seit der Gründung 2015 unter anderem sogenannte Large Language Models (LLMs), die auf maschinellem Lernen und Künstlicher Intelligenz (KI) basieren[1]. Die großen Sprachmodelle (LLMs) von OpenAI werden GPT genannt, wobei GPT für „Generative Pre-Trained Transformer" steht. Das ist eine bestimmte Modellarchitektur von Deep-Learning-Netzwerken (mehr dazu im Kapitel „Ein Blick in den Maschinenraum"). Die erste Modellvariante GPT-1 wurde bereits 2018 von OpenAI veröffentlicht. Bis zum Live-Gang von GPT-3 im Mai 2020 interessierte sich vor allem die Fachwelt für die immer besser werdenden Sprachmodelle[2].

Sprachmodelle drängen in die Öffentlichkeit

Seit Ende 2022 ist die Weiterentwicklung von GPT ein Thema für die breite Öffentlichkeit und auch für die Versicherungsbranche geworden. Mit verantwortlich dafür ist ein neuer Ansatz, den OpenAI bei der Veröffentlichung von ChatGPT wählte. Bis dato konnte man die Sprachmodelle nur über ein eher technisch-wirkendes Interface (Playground) oder mit Programmierkenntnissen über eine technische Schnittstelle (API) nutzen[3]. Um das Sprachmodell auch einer breiten Nutzerschaft zugänglich zu machen, entschied sich OpenAI dazu, die Bedienoberfläche wie bei einem klassischen Chatbot zu gestalten und das Sprachmodel dialogfähig zu machen – ChatGPT war geboren[4].

Interessierte Menschen können sich seitdem kostenlos ein Nutzerkonto anlegen und loschatten. Das taten nach zwei Monaten bereits 100 Millionen registrierte Nutzer. ChatGPT ist damit die am schnellsten wachsende digitale Anwendung der Menschheitsgeschichte. Der bisherige Rekordhalter Tiktok brauchte noch 9 Monate, um die 100 Mio-Marke zu knacken[5]. Die Anzahl der registrierten ChatGPT-Nutzer im März 2023 ist unbekannt, allerdings stiegen die monatlichen Visits auf über eine Milliarde an[6]. Ein beispielloser Erfolg, der sich nicht nur mit einer guten Bedienoberfläche und fehlenden Nutzungsgebühren erklären lässt.

ChatGPT bzw. das Sprachmodell hinter ChatGPT war und ist überraschend gut dabei, menschlich anmutende Sprache zu generieren. Das ist unter anderem damit zu erklären, dass das Sprachmodell auch von Menschen trainiert wurde. Ein paar kommunikative Eigenheiten fallen im „Gespräch" mit dem Chatbot auf – er antwortet gerne etwas ausgiebig und repetitiv („schwadroniert"). Auch tut er sich mit dem Zugeben von Wissenslücken in Form eines „darauf habe ich leider keine Antwort" sehr schwer. Jede Antwort wird mit großem Selbstbewusstsein ausgegeben. Der Nutzer hat aber nicht das Gefühl, mit einer künstlichen Intelligenz zu interagieren. Auch ist der Chatbot bemüht, keine radikalen Positionen zu vertreten. Das wurde dem Twitterbot Tay von Microsoft noch zum Verhängnis. Er wurde 2016 abgeschaltet, als er rassistische Antworten gab[6].

Die große Frage war und bleibt, ob die Antworten von ChatGPT auch nützlich und richtig sind. Manche Nutzer waren begeistert von der Qualität der Antworten und Andere machten sich einen Spaß daraus, die Grenzen und Irrungen von ChatGPT aufzudecken. Schnell konnte man im Internet auf nützliche und unnütze ChatGPT Ein- und/oder Ausgaben zugreifen[7,8]. Beide Gruppen verstärkten den ChatGPT-Hype, indem sie mehr Fans und Skeptiker hervorbrachten, die eines taten – sich einen Account zu beschaffen und sich selbst ein Bild von den Fähigkeiten der KI zu machen.

Erfolgsfaktoren von ChatGPT

Einfache Benutzeroberfläche
ChatGPT ermöglicht eine simple Interaktion über ein Dialogsystem wie bei einem ChatBot, mit einer einfachen Frage- & Antwortlogik.

Kostenloser Service
Für die Nutzung ist lediglich eine Registrierung erforderlich und etwas Geduld, wenn das hohe Nutzeraufkommen zu Wartezeiten führt.

Überzeugende, menschliche Sprachverarbeitung
ChatGPT versteht die Sprache der Nutzer und antwortet mit natürlich anmutender und ausgewogener Sprache. Dafür wurde das Sprachmodell mit menschlichem Feedback und Präferenzen trainiert.

Beeindruckende Fähigkeiten
Die Fähigkeiten von ChatGPT bei der Erstellung von Texten beeindrucken, auch wenn noch nicht alle Antworten perfekt sind.

Word-of-Mouth/Mouse-Effekt
Fans und Kritiker sorgten für immer mehr positive und negative Empfehlungen, eine verstärkte Berichterstattung und damit auch zu steigenden Nutzungszahlen.

OpenAI ist kein Tech-Gigant
OpenAI wurde als Startup gesehen und hatte damit eine andere Wahrnehmung in der Öffentlichkeit[9]. Nutzer waren nachsichtig bei Fehlern. Google's ChatGPT-Alternative Bard wird aktuell, auch bei kleinen Fehlern, eher kritisch betrachtet[10].

Was können ChatGPT und die zugrundeliegenden Sprachmodelle

GPT arbeitet mit Sprache jeglicher Art

Als ChatBot der auf einem GPT-Sprachmodell basiert, kann ChatGPT vor allem eins – mit Sprache umgehen. Die Künstliche Intelligenz hat aus digitalisierten Büchern und heruntergeladenen Webinhalten gelernt, Sprache zu verstehen, auf Fragen zu antworten und Dialoge möglichst menschenähnlich zu führen. GPT-Sprachmodelle sind nicht für einen bestimmten Zweck gebaut.

Basierend auf einer eingegebenen Textsequenz können sie das nächste wahrscheinliche Wort bestimmen und erstellen damit schrittweise Texte. Sie sind auf allgemeinen Daten trainiert und damit vielseitig einsetzbar. Deshalb ist ihre Aufgabenpalette groß. GPT-basierte Anwendungen wie ChatGPT können verschiedene Arten von Text erstellen, abwandeln, ergänzen, übersetzen, zusammenfassen, klassifizieren und Informationen extrahieren. Dazu können sie Programmcode erstellen und mittlerweile auch schon mit Bildern in der Eingabe arbeiten.

Die Anwendungsfälle lassen sich auf einer abstrakteren Ebene in sechs Kernfunktionen einteilen:

Nummer	Kernfunktion	Beschreibung
1	Kreieren	Die Fähigkeit, Texte zu generieren, z. B. E-Mails, Konzepte, Rechtstexte, Gedichte, Programmcode.
2	Simulieren	Die Fähigkeit, in Texten etwas zu simulieren bzw. jemanden nachzuahmen, z. B. Rollen einnehmen, im Stile einer Person schreiben.
3	Transferieren	Die Fähigkeit, Texte in andere Formen/Formate zu übertragen, z. B. Texte in andere Sprache übersetzen, einen Code (z. B. C+) in einen anderen Code (z. B. Python) umwandeln.
4	Modifizieren	Die Fähigkeit, Texte zu verändern, z. B. Individualisierung/Personalisierung von Texten, Variationen von bestehenden Texten erstellen.
5	Verbessern	Die Fähigkeit, Texte zu verbessern, z. B. Rechtschreibkorrektur, Anpassung an Sprachstile, Strukturierung von Texten und Argumenten.
6	Reduzieren	Die Fähigkeit, Texte zusammenzufassen, Inhalte zu extrahieren.

So nutzen Sie ChatGPT

Die Anwendungsmöglichkeiten sind vielfältig. Probieren Sie es einmal selbst aus oder lassen sich von unseren nachfolgenden Beispielen inspirieren.

1. Webseite besuchen: chat.openai.com
2. Registrieren
3. Chatten

Am besten chattet man in den frühen Morgenstunden, da die US-amerikanischen Nutzer dann noch schlafen und der Dienst noch nicht überlastet ist. Das funktioniert leider auch nicht immer – ein Abo-Modell namens ChatGPT Plus (monatliche Kosten: 23,80 Euro) bietet eine bessere Verfügbarkeit und Zugang zu neueren Modellen (z. B. ein schnelleres GPT-3.5 oder mit Einschränkungen schon GPT-4). Trotz Premium-Abo gibt es bisweilen Einschränkungen, z. B. beim Zugriff auf die Chat-Historie.

Achtung: **Lesen Sie sich die Nutzungsbedingungen von OpenAI sorgfältig durch, bevor Sie diese akzeptieren. Lassen Sie die notwendige Sorgfalt in Bezug auf den Datenschutz walten und beachten Sie auch die Regelungen zur Nutzung von Inhalten, die unter die Verschwiegenheitspflicht und das Geschäftsgeheimnis fallen.**

Sprachmodelle schneiden bei einer Reihe von Tests beeindruckend gut ab

Nachdem ChatGPT veröffentlicht wurde, gab es zügig eine Reihe von Nachrichten, welche die beeindruckenden Fähigkeiten des Chatbots untermauerten. Es wurde beispielsweise davon berichtet, dass ChatGPT ...

– eine Prüfung des MBA-Programms der renommierten US-amerikanischen Wharton School bestand [11].

– die Prüfung des US Medical Licensing Exam meisterte, welche Voraussetzung für die medizinische Zulassung von Studenten in den USA ist [12].

– vier Klausuren der University of Minnesota Law School bestand [13].

Mit der Veröffentlichung seiner Sprachmodelle publiziert OpenAI mittlerweile eine Reihe von Benchmarks, die herangezogen werden, um die Fähigkeiten des Sprachmodells zu verdeutlichen. Diese umfassen Themenbereiche wie die Medizin, Rechtswissenschaften, Mathematik, Literatur, Geschichte, Biologie, Chemie, Physik, Wirtschaft, Kunstgeschichte bis hin zur Programmierung und der Theorie der Weinkellnerei (Sommelier). In der untenstehenden Tabelle werden beispielhaft einige Testergebnisse von GPT-4 dargestellt. Für eine detaillierte Übersicht sei auf die Website von OpenAI verwiesen. Die Tests lassen sich größtenteils den Aufnahme- und Eignungstests für angehende Studenten an US-amerikanischen Hochschulen zuordnen. Es gibt aber auch Tests speziell für Machine-Learning-Modelle [14].

Diese Testergebnisse beeindruckten vor allem vor dem Hintergrund, dass die Sprachmodelle GPT-3.5 und GPT-4, die bei ChatGPT zum Einsatz kommen, für keine der Aufgaben gesondert trainiert wurden. Die Erfolge in den Prüfungen wurden mit dem Standardtraining der Sprachmodelle möglich, obwohl die Tests für Menschen mit gehobenen Fähigkeiten konzipiert wurden. Die Tests zeigen, dass die Sprachmodelle von OpenAI in der Lage sind, eine Reihe von kognitiv anspruchsvollen Aufgaben erfolgreich zu bearbeiten.

Erste Ergebnisse zum Effekt von ChatGPT im Arbeitseinsatz sind beeindruckend

Aufnahme- und Zulassungstests sind das eine, doch wie schneiden Sprachmodelle im Arbeitsalltag ab? Ein Forscherteam des MIT untersuchte bereits die Auswirkungen von ChatGPT auf die Produktivität von Arbeitnehmern und veröffentlichte die Ergebnisse in einem ersten Arbeitspapier [15].

In einem Online-Experiment mit 444 erfahrenen, akademisch gebildeten Teilnehmern wurden diese instruiert, berufsbezogene Schreibaufgaben innerhalb von 30 Minuten zu bearbeiten. Dabei wurden Anreize für qualitativ hochwertige Arbeit geschaffen, indem externe Reviewer die Aufgaben beurteilten und monetäre Belohnungen vergaben. Die Teilnehmer wurden zufällig einer Experimental- oder Kontrollgruppe zugewiesen. In einer ersten Aufgabenrunde durften in beiden Gruppen keinerlei Hilfsmittel benutzt werden. In einer zweiten Runde durfte die Experimentalgruppe ChatGPT nutzen und die Kontrollgruppe einen Text-Editor einsetzen.

Die Studie zeigte, dass 81% der Experimentalgruppe ChatGPT aktiv bei der Aufgabenbearbeitung nutzten. Die Verwendung von ChatGPT führte zu einer Reduzierung der benötigten Bearbeitungszeit für Aufgaben um 10 Minuten (-37%) und einer Steigerung der Noten um 0,45 Standardabweichungen.

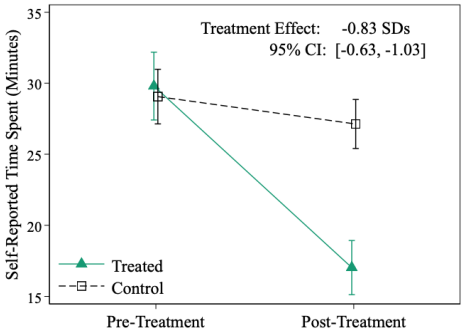
Die Ergebnisse deuten darauf hin, dass ChatGPT die Produktivität von Arbeitnehmern positiv beeinflussen kann, wobei insbesondere Arbeitnehmer, die bei der ersten Aufgabe eine niedrige Note erhielten, von einer Steigerung der Noten und einer Reduzierung der benötigten Zeit profitierten.

Test	Zweck des Tests	Testergebnis	Einordnung
Uniform Bar Exam (U-BE)	Standardisierte Anwaltsprüfung, testet Wissen für Anwaltszulassung in USA	298/400	~ Top 10 von 100
SAT Reading & Writing	Zulassungstest für Universitäten in USA mit Schwerpunkt Lesen & Schreiben	810/800	~ Top 7 von 100
SAT Math	Testet Anwendung mathematischer Konzepte und Fähigkeiten für die Universität und das Berufsleben	700/800	~ Top 11 von 100
Graduate Record Examination (GRE) – Quantitative	Testet Beherrschung der Grundrechenarten, Algebra, Geometrie, der Datenanalyse, Dateninterpretation und Problemlösungskompetenz	163/170	~ Top 20 von 100
Graduate Record Examination (GRE) – Verbal	Testet Fähigkeiten bei Satzergänzungen, Antonymen, verbale Analogien und Leseverständnis	169/170	~ Top 1 von 100
Graduate Record Examination (GRE) – Writing	Testet Fähigkeiten, Argumente zu formulieren, zielgerichtete und zusammenhängende Diskussionen zu führen	4/6	~ Top 46 von 100

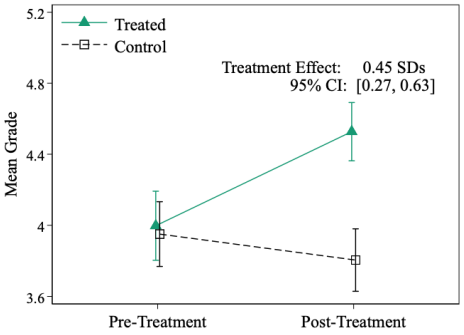
Die ChatGPT-Nutzer verbrachten im Arbeitsablauf weniger Zeit beim Brainstorming und bei der Erstellung einer Rohfassung. Die Arbeit verschob sich mehr in Richtung Bearbeitung und Feinschliff der Arbeitsergebnisse.

Die Studie zeigt eindringlich, dass durch den Einsatz generativer KI hohe Produktivitätssteigerungen zu erwarten sind und sich Arbeitsmodi und -abläufe verändern werden.

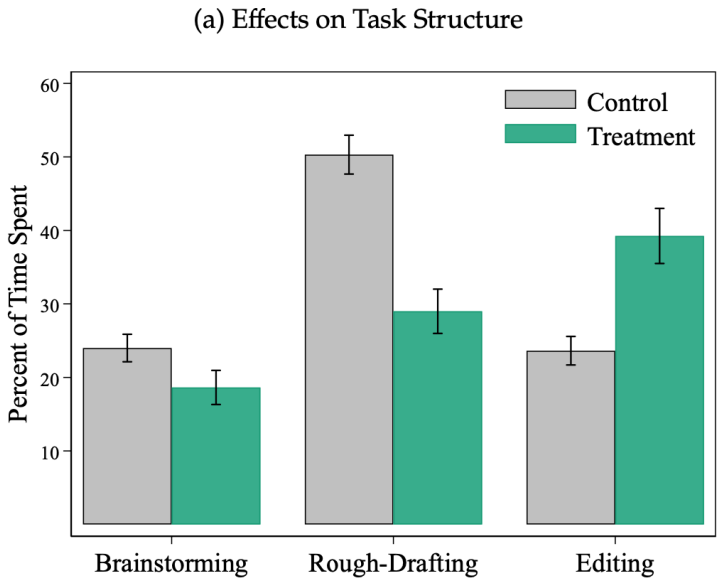
Geringere Bearbeitungszeit



Verbesserte Bewertung



Veränderte Arbeitsabläufe



Anwendungsbezogene ChatGPT-Beispiele für die Versicherungsbranche

Versicherung lebt von und durch Sprache

Die oben genannten Kernfunktionen und Anwendungsbeispiele scheinen sehr generisch und können immer dort hilfreich sein, wo mit Texten gearbeitet wird. Gerade in der Versicherung kommen sehr viele Texte zum Einsatz. Als immaterielles Gut wird eine Versicherung erst durch Schriftstücke (z. B. Versicherungsbedingungen), Prozess-Dokumentation (z. B. Schadenprotokolle) und Kommunikation (z. B. Kundenservice) zum Leben erweckt. Entlang der kompletten Wertschöpfungskette der Versicherung ergeben sich deshalb mögliche Einsatzbereiche für Sprachmodelle. Angefangen bei der Produktentwicklung, über den Vertrieb, die Kundenberatung, hin zur Schadenbearbeitung, dem Eingangsmanagement, der Abrechnung und dem Kundensupport im Allgemeinen.

Die nachfolgenden, anwendungsbezogenen Beispiele dienen rein illustrativen Zwecken und sind frei erdacht. Mit Hilfe der Beispiele soll ein erstes Bewusstsein für die Einsatzmöglichkeiten von Sprachmodellen im Versicherungskontext geschaffen werden. Alle Beispiel-Prompts wurden in ChatGPT Plus eingegeben. ChatGPT darf hier nur als ein sehr nutzerfreundliches Tool verstanden werden, welches ein einfaches Ausprobieren ermöglicht. Die eigentliche Bearbeitung der Prompts erfolgt mit dem Sprachmodell GPT-4. Um die Lesbarkeit zu gewährleisten, sind die Beispielantworten von ChatGPT eher kurzgehalten und teilweise nur als Auszüge ausgegeben. GPT-4 kann ungefähr 25.000 Wörter (genauer 32.768 Tokens) als Kontext im „Kopf“ behalten und darauf referenzieren[1].

„Garbage in, Garbage out“ – Wer gute Antworten will, muss gute Befehle geben

Damit ChatGPT weiß, was von ihm verlangt wird, muss der Nutzer über ein Dialogfenster eine Eingabe – einen sogenannten Prompt – tätigen. Dabei gilt: Je besser die Eingabe des Menschen, desto besser die Ausgabe von ChatGPT. Es lohnt also, sich Gedanken bei der Prompt-Erstellung zu machen. Das Denken nimmt ChatGPT seinen Nutzern noch nicht ab.

Mit dem Prompt-Engineering hat sich auch schon eine eigene Disziplin zur Optimierung von Prompts herausgebildet[16]. Prinzipiell gilt: Je bessere Vorgaben gemacht werden, desto bedarfsgerechter ist die Ausgabe und desto weniger muss im Nachhinein editiert werden.

Erfolgreiches Prompten

Rollenverständnis: Gebe ChatGPT vor, aus welcher Rolle heraus er auf die Eingabe antworten soll. Das Rollenverständnis beeinflusst die Perspektive und damit die Art und Weise, wie ChatGPT antwortet. Beispielhafte Rollen sind Ernährungsberater, Medizinstudent, Autor, Komiker und für uns eher relevante Versicherungsberater, Datenschützer, Marketingberater.

Klare Aufgabenstellung: Nenne eine klare Aufgabenstellung. Was soll ChatGPT genau tun? Einen einzelnen Text oder Textvariationen erstellen, Information aus Texten extrahieren, bestehende Vorlagen umschreiben, übersetzen, Texte analysieren oder Dialoge führen?

Kriterien/Kontext/Beschränkungen definieren: Weitere Informationen helfen die Ausgabe zu verbessern. Soll ein Aspekt in der Ausgabe unbedingt genannt oder weggelassen werden? Soll die Antwort eine bestimmte Perspektive einnehmen oder zwischen Positionen abwägen? Wird ein Text für eine unerfahrene Leserschaft oder Experten geschrieben? Ein Text kann Vieles sein – ein Gedicht, Blogbeitrag, eine wissenschaftliche Abhandlung, usw.

Zielsetzung kommunizieren: Mit welcher Intention wird ein Text geschrieben? Gilt es zu informieren, inspirieren, beraten, überzeugen, entertainen? Eine Zielsetzung hilft ChatGPT, bessere Ausgaben zu erzeugen.

Formatvorgaben machen: Welche Form soll die Ausgabe haben? Wird ein Fließtext oder werden Stichpunkte benötigt, wie viele Wörter dürfen es sein? Soll ein bestimmter Ton getroffen werden (z. B. freundlich vs. streng)? Diese Informationen sorgen für eine bedarfsgerechte Ausgabe.

Feedback geben: Der gewünschte Output stellt sich nicht immer beim ersten Versuch ein. Ein iteratives Vorgehen mit einer wiederholten Anpassung des Prompts ist erfolgsversprechend. Dazu muss ChatGPT mitgeteilt werden, wie die Ausgabe angepasst werden soll, welche Aspekte fehlen oder stören.

Allgemeines
Es muss nicht immer jeder Aspekt berücksichtigt sein und die Reihenfolge scheint nebensächlich. Manchmal reicht auch eine kurze knackige Eingabe, um das Ziel zu erreichen, z. B. muss man beim Übersetzen nicht unbedingt spezifizieren, dass ChatGPT wie ein Übersetzer agieren soll. Wichtig sind vor allem eine klare, einfache Sprache und gute Struktur des Prompts. Fachbegriffe sollten nur verwendet werden, wenn es die Aufgabe bzw. das Thema erfordert.

Texte erstellen

Versicherungsbeschreibungen für das Internet

Agiere wie ein Website-Texter, der für ein Versicherungsunternehmen arbeitet.

Erstelle einen Artikel für eine Website.

Dein Ziel ist es, auf der Unternehmenswebsite Interesse für eine Wohngebäudeversicherung bei Eigenheimbesitzern zu wecken. Erkläre, wie die Versicherung funktioniert, welche Schäden abgesichert sind, gebe zwei kurze aber konkrete Beispiele an.

Gib eine Überschrift und einen abschließenden Call-to-Action aus.

Der Artikel muss zwischen 150 und 180 Wörter beinhalten.

Überschrift: Schützen Sie Ihr Eigenheim mit unserer Wohngebäudeversicherung!

Ihr Eigenheim ist nicht nur ein Ort zum Leben, sondern auch eine wertvolle Investition. Umso wichtiger ist es, dieses Vermögen bestmöglich zu schützen. Unsere Wohngebäudeversicherung bietet genau den Schutz, den Sie benötigen, um ruhig schlafen zu können.

Die Wohngebäudeversicherung deckt eine Vielzahl von Schäden ab, darunter Feuer, Sturm, Hagel, Leitungswasser, Frost und Elementarschäden wie Überschwemmungen, Erdrutsche und Erdbeben. Unser Versicherungsschutz stellt sicher, dass Ihr Zuhause im Falle eines Schadens schnell und effizient repariert oder wiederhergestellt wird.

Ein konkretes Beispiel: Stellen Sie sich vor, ein Sturm entwurzelt einen Baum in Ihrem Garten und dieser beschädigt das Dach Ihres Hauses. Unsere Wohngebäudeversicherung kommt für die Reparaturkosten und gegebenenfalls für den Baumfällservice auf, sodass Sie sich um nichts kümmern müssen.

Ein weiteres Beispiel: Ein Wasserrohr in Ihrem Haus bricht und verursacht einen massiven Wasserschaden. Unsere Wohngebäudeversicherung deckt die Kosten für die Reparatur des Rohrs sowie die Beseitigung des Wasserschadens.

Schützen Sie Ihr Zuhause und Ihre finanzielle Sicherheit, indem Sie sich für unsere Wohngebäudeversicherung entscheiden. Unser kompetentes Team steht Ihnen jederzeit zur Verfügung, um Ihnen den bestmöglichen Schutz für Ihr Eigenheim zu bieten.

Call-to-Action: Zögern Sie nicht länger! Kontaktieren Sie uns noch heute, um ein unverbindliches Angebot für Ihre Wohngebäudeversicherung zu erhalten und Ihr Eigenheim bestmöglich abzusichern.

Unser Prompt hat 68 Wörter. Trotzdem haben kleine Änderungen eine große Wirkung. Ändern Sie beispielsweise die Rolle von ChatGPT vom Website-Texter zum Komiker, wird Ihr Eigenheim zum Palast, Naturkatastrophen zu komischen Zufällen, Stürme zu Windstößen und die Versicherung zum Schutzzauber, der dafür sorgt, dass Ihre Festung standhaft bleibt.

Texte erstellen

Datenschutzhinweise & Einwilligungen

Agiere wie ein deutscher Datenschützer, der für ein Versicherungsunternehmen arbeitet.

Dein Unternehmen hat eine Versicherungs-App gebaut, bei dem die Kunden einen Beratungsservice per Videotelefonie nutzen können. Dafür ist jedoch ein Datentransfer in ein Drittland notwendig. Erstelle eine Einwilligungserklärung.

Informiere den potenziellen Nutzer mit einfachen Worten und stelle die notwendige Transparenz her.

Kriterien: Gehe auf die Besonderheiten der DSGVO in Bezug auf die Datenübermittlung in ein Drittland ein.

Gib eine Überschrift und Text sowie den Text für einen Opt-In Button mit. Schreibe so einfach, klar und kurz wie möglich.

Überschrift: Einwilligung zur Datenübermittlung in ein Drittland für Videotelefonie-Beratung

Text: Lieber Nutzer, für unseren Beratungsservice per Videotelefonie ist es notwendig, einige Ihrer persönlichen Daten in ein Drittland außerhalb der Europäischen Union zu übermitteln. Gemäß der Datenschutz-Grundverordnung (DSGVO) möchten wir Sie über diesen Vorgang transparent informieren und Ihre Zustimmung einholen.

Die Datenübermittlung erfolgt, um Ihnen eine qualitativ hochwertige und zuverlässige Videotelefonie-Beratung zu ermöglichen. Zu diesem Zweck werden Ihre personenbezogenen Daten, wie Name, E-Mail-Adresse und Video-chat-Protokolle, an unseren Technologiepartner übermittelt. Dieser Partner gewährleistet ein angemessenes Datenschutzniveau, das mit den Standards der DSGVO vereinbar ist. Wir haben mit unserem Partner einen Vertrag zur Datenverarbeitung abgeschlossen, der Ihre Daten und Privatsphäre schützt.

Ihre Zustimmung können Sie jederzeit widerrufen, indem Sie uns kontaktieren. Nach Ihrem Widerruf wird die Datenübermittlung gestoppt und Ihre Daten werden nicht mehr in das Drittland übermittelt.

Opt-In Button: Ich stimme der Datenübermittlung in ein Drittland zu.

Die Trainingsdaten von GPT-Modellen stammen zum großen Teil aus dem Internet. Dementsprechend kennen die Sprachmodelle eine Reihe von Datenschutzhinweisen und Einwilligungen. Deshalb ist ChatGPT in der Lage, recht überzeugende Texte zu Themen zu entwerfen, die ein gewisses Maß an Spezialbildung benötigen. Die Besonderheiten eines Unternehmens (z. B. Stil, Konsistenz, Anforderungen) kennt ChatGPT jedoch nicht. Deshalb müssen solche Texte immer kontrolliert und auf die individuellen Bedürfnisse angepasst werden.

Texte variieren

Productclaims generieren

Für die Produkteinführung einer innovativen Cybersecurity-Versicherung werden Produktclaims für die Vermarkung gesucht. In einem Workshop sollen Ideen eingebracht, diskutiert und beurteilt werden.

Erstelle 10 Varianten von Produktclaims. Die Ansprache soll auf Einfachheit, Verlässlichkeit und Partnerschaft ausgerichtet sein.

Die Ausgabe soll als Tabelle erfolgen. Erstelle die Spalten Nummer, Produktclaim, Bewertung. Lasse die Spalte Bewertung leer. Jeder Claim soll aus 8 bis 12 Wörtern bestehen.

Nummer	Produktclaim	Bewertung
1	Einfach sicher, rundum geschützt: Ihr Cyber-Partner	
2	Verlässliche Cybersecurity-Versicherung, Ihr Schutzschild	
3	Gemeinsam stark: Wir sind Ihr Cybersecurity-Bollwerk	
4	Intuitive Sicherheit für Ihre digitale Welt	
5	Mit uns an Ihrer Seite: Ihr digitaler Sicherheitsgarant	
6	Einfache Lösung für komplexe Cyber-Risiken	
7	Sicher im Netz dank unserer Cybersecurity-Partnerschaft	
8	Sorgenfrei surfen: Wir schützen Ihre digitale Existenz	
9	Ihr verlässlicher Partner für digitale Sicherheit	
10	Cybersecurity leicht gemacht: Wir stehen Ihnen bei	

Hinweis: Sprachmodelle kennen Wortzusammenhänge aus ihrer enormen Masse an Trainingsdaten. Sie können deshalb sehr leicht verschiedene Varianten von Texten erstellen. Gerade bei häufig vorkommenden Themen kann es zu Wiederholungen kommen und die Texte wirken eher wenig originell. Eine Ausgabe des Modells kann auch in der identischen Form von jemandem Dritten erdacht und veröffentlich worden sein. Da hilft es auch nicht, dass OpenAI die Urheberrechte an den Eingaben und Ausgaben der Sprachmodelle den Nutzern überlässt[17]. Die Ausgaben von ChatGPT sollten vor der Verwendung immer über eine Internetsuche auf Urheberrechtsverletzungen Dritter geprüft werden.

Texte zusammenfassen

Kernaussagen eines Presseartikels

Agiere wie ein Mitarbeiter in einer PR-Abteilung einer Versicherung.

Für einen wöchentlichen Jour fixe musst du dich auf das Thema Generative KI und ChatGPT vorbereiten. Du hast selbst aber keine Zeit, den Artikel vollständig zu lesen und benötigst Unterstützung.

Erstelle eine Zusammenfassung des nachfolgenden Artikels. Gib den Titel, den Verfasser, die Quelle, die Tonalität und den Schreibstil aus.

Finde wenn möglich ein prägnantes Zitat zum Thema Versicherung.

Antworte im inhaltlichen Teil in einem Fließtext. Gib für die Tonalität und den Schreibstil jeweils eine kommagetrennte Aufzählung von Adjektiven aus.

Inhalte des Artikels:
[Auszug] Generative KI: ChatGPT & Co. erleben ihren iPhone-Moment
Generative-KI-Tools wie ChatGPT oder Bart von Google stellen derzeit alles in den Schatten und haben auch uns bei //next schon zu manchem Selbstversuch verleitet …

Titel: Generative KI: ChatGPT & Co. erleben ihren iPhone-Moment
Verfasser: Falk Hedemann

Quelle: //next

Zusammenfassung: Der Artikel beschreibt den rasanten Aufstieg und die Bedeutung von generativen KI-Tools wie ChatGPT. Dabei wird der aktuelle Stand dieser Technologie mit dem iPhone-Moment verglichen, also dem Punkt, an dem eine Technologie plötzlich einen signifikanten Fortschritt macht.
Die Funktionsweise von ChatGPT wird erläutert und es wird auf die Chancen, Risiken und rechtlichen Fragestellungen hingewiesen. Der Autor betont, dass generative KI-Tools in Zukunft als nützliche KI-Kollegen dienen könnten, die zeitaufwändige Aufgaben für Menschen übernehmen.

Tonalität: informativ, reflektierend, ausgewogen

Schreibstil: sachlich, klar, flüssig

Ein prägnantes Zitat zum Thema Versicherung wurde im Artikel nicht gefunden.

Die Zusammenfassung von Texten haben wir hier am Beispiel eines frei zugänglichen Artikels von next by ERGO erstellen lassen. Lesen Sie den Artikel gerne selbst und gleichen Sie die Zusammenfassung mit den Inhalten ab[18].

Ein weiterer wichtiger Anwendungsfall für Zusammenfassungen sind Gesprächsnotizen bzw. sogenannte Meeting Notes. Hier muss man bei ChatGPT allerdings den Datenschutz und Verschwiegenheitspflichten beachten. Insbesondere, wenn man Daten von Gesprächspartnern (Dritten) eingibt, muss man deren Zustimmung einholen.

Dialoge führen

Fragen zu einer Versicherung beantworten (I)

Agiere wie ein Kundenberater zur ERGO Hausratversicherung.

Sei ein ERGO-ChatBot, der auf Fragen antwortet. Tritt in den Dialog ein, indem du deinen Gesprächspartner begrüßt und ihn fragst, was du für ihn tun kannst.

Beziehe deine Antworten auf das hinterlegte Bedingungswerk. Versuche das Gespräch am Laufen zu halten. Wenn du die Information nicht aus dem Bedingungswerk entnehmen kannst, so antworte mit „Das weiß ich leider nicht. Kann ich noch etwas für Sie tun?“

Antworte freundlich, aber in kurzen Sätzen.

Bedingungswerk:

[Auszug] Meine ERGO Hausratversicherung Versicherungsbedingungen KT2013HR 1...

Hallo! Willkommen bei der ERGO Hausratversicherung.
Wie kann ich Ihnen heute helfen?

Hallo, ich interessiere mich für die Hausratversicherung.
Welche Sachen sind versichert?

Versichert ist Ihr gesamter Hausrat, dazu gehören u. a.:

1. Privat genutzte Sachen inkl. Wertsachen
2. Antennenanlagen und Markisen auf dem Versicherungsgrundstück
3. Kanus, Ruderboote, Surfgeräte, Gleitschirme etc.
4. Haustiere, die üblicherweise in Wohnungen gehalten werden
5. Beruflich/gewerblich genutzte Arbeitsgeräte und Einrichtungsgegenstände
6. Handelswaren/Musterkollektionen bis 1.500 Euro je Versicherungsfall
7. Kraftfahrzeugzubehör bis 1.500 Euro je Versicherungsfall
8. Als Mieter oder Wohnungseigentümer eingefügte Sachen ins Gebäude

Wieviel kostet die Versicherung

Das weiß ich leider nicht. Kann ich noch etwas für Sie tun?

Dialoge führen

Fragen zu einer Versicherung beantworten (II)

Anreichern mit spezifischen Daten und Training

Das generelle Sprachverständnis hat GPT-4 über den Trainingsprozess erhalten. Spezifische Informationen entnimmt das Sprachmodell beigefügten Inhalten. Das Anreichern durch unternehmensinterne Daten macht Sprachmodelle zu einem mächtigen Werkzeug. Allerdings kann nicht jede Information über Copy & Paste eingefügt werden – insbesondere dann nicht, wenn die Datenmenge groß wird. In dem Beispiel ließen sich nur ca. 500 Wörter der Versicherungsbedingungen in den Prompt kopieren. Damit wäre der BeispielBot nur eingeschränkt auskunftsfähig.

Größere Datenmengen lassen sich allerdings durch sogenannte **Embeddings** hinzufügen – das sind mathematische Repräsentationen von in Text vorhandenem Wissen [19]. Des Weiteren kann ein Sprachmodell mit spezifischen Daten antrainiert werden. Dieser Prozess nennt sich **Fine-Tuning** [20]. Ein Einsatz von Embeddings und Fine-Tuning lässt sich nicht mehr über ChatGPT abbilden. Spätestens hier muss man über eine geeignete Programmierumgebung auf die Schnittstellen von OpenAI zugreifen und sich eine eigene Nutzerschnittstelle bauen [3].

Führt ChatGPT eine Versicherungsberatung durch?
Die Versicherung ist eine hochregulierte Branche. So gibt es beispielsweise eine Beratungsdokumentation, mit dem ein Versicherungsvermittler seinen Dokumentationspflichten nachkommen muss [21]. Im Falle einer Falschberatung drohen rechtliche Konsequenzen. Was passiert, wenn eine auf Sprachmodellen basierende Anwendung nicht umfassend genug aufklärt oder Informationen falsch darstellt? Es ist zu vermuten, dass für eine Beratung unter Zuhilfenahme von Sprachmodellen ebenfalls hohe Anforderungen gelten. In vielen Fällen wird eine Sprachmodell-Lösung im Kundenkontakt keine Antworten geben dürfen oder aber in Kombination mit regelbasierten Systemen in einem stärkeren Vorgaben-Gerüst antworten müssen.

Ist eine Verknüpfung mit Kunden- und Unternehmensdaten möglich?
Jede Dateneingabe über das öffentlich verfügbare ChatGPT kann zum Training weiterer Sprachmodelle verwendet werden [22]. Unter Umständen kann es also passieren, dass persönliche und vertrauliche Daten in ein neues Sprachmodell eingepflegt und später von Dritten extrahiert werden. Eine Eingabe von schützenswerten bzw. schutzbedürftigen Daten ist zu unterlassen. Sprachmodelle werden in einigen Anwendungsfällen aber nur dann hilfreiche Antworten geben und Aufgaben ausführen können, wenn sie mit Unternehmenssystemen verknüpft sind, Kundendaten entgegennehmen und verarbeiten. Hierfür ist wiederum der Bau eigener Lösungen unter Maßgabe der DSGVO notwendig. Hier gibt es schon die Möglichkeit, Sprachmodelle auf eigenen Servern von Microsoft Azure zu hosten und zu betreiben [23]. Die Daten verbleiben dann in Unternehmenskontrolle und werden gemäß den rechtlichen Vorgaben erfasst, verarbeitet und gespeichert

Klassifizieren

Kontaktinformationen aus Texten extrahieren

Agiere wie ein Mitarbeiter des Posteingangsmanagements einer Versicherung.

Du möchtest eingehende E-Mails hinsichtlich der Anliegen klassifizieren und persönliche Daten extrahieren, damit du die E-Mails entsprechend an die zuständigen Fachabteilungen weiterleiten kannst.

Erstelle basierend auf den beigefügten E-Mails eine Tabelle mit den Namen, Nachnamen, Versicherungsnummern und Anliegen der Sender.

Klassifiziere das Anliegen in eine oder mehrere Optionen:
1) Kündigung 2) Umzug 3) SEPA-Lastschriftmandat 4) Sonstiges

E-Mail 1:
Betreff: Änderungsmitteilung
Absender: max.mustermann@email.de
Sehr geehrte Damen und Herren,
hiermit kündige ich meine Reisekrankenversicherung (987654321) fristgerecht zum Ende des Quartals. Ich entziehe Ihnen gleichzeitig das Sepa-Lastschriftmandat. Bitte senden Sie mir eine Kündigungsbestätigung auf dem Postweg zu.
Mit freundlichen Grüßen
Max Mustermann

E-Mail 2:
Betreff: Wechsel der Anschrift
Absender: erika.musterfrau@email.de
Hallo liebes Versicherungsteam,
ich ziehe um und wolle Ihnen nur kurz meine neue Adresse schicken:
Musterstrasse 1, 12345 Musterstadt
Bitte schicken Sie die Post ab jetzt an die Anschrift.
Viele Grüße
Erika Musterfrau

Hier ist die erstellte Tabelle basierend auf den beigefügten E-Mails:

E-Mail 1:
Max Mustermann hat seine Reisekrankenversicherung (Versicherungsnummer: 987654321) gekündigt und das SEPA-Lastschriftmandat entzogen.

E-Mail 2:
Erika Musterfrau hat einen Umzug gemeldet und möchte, dass ihre Post an ihre neue Adresse (Musterstrasse 1, 12345 Musterstadt) geschickt wird.

E-Mail	Vorname	Nachname	Versicherungsnummer	Anliegen
1	Max	Mustermann	987654321	1) Kündigung, 3) SEPA-Lastschriftmandat
2	Erika	Musterfrau	–	2) Umzug

Die Klassifikation von Dokumenten und die Extraktion von Informationen sind für Unternehmen und im Speziellen für Versicherungen sehr relevant. Dies erfordert aber einen gewissen Grad der Integration in die technische Infrastruktur des Unternehmens. Hier gelten dann selbstverständlich die Richtlinien zur IT-Sicherheit und zum Datenschutz.

Programmcode erstellen

Den Login einer Mobilen Applikation schreiben

Agiere wie ein Entwickler.

Du möchtest eine Applikation in Python bauen.

Schreibe den Code für den Login-Screen. Frage den Nutzernamen und das Passwort ab. Das Passwort ist maskiert und wird nur bei Bedarf angezeigt. Füge ein Checkbox hinzu, die es erlaubt, dass das Passwort temporär für 14 Tage gespeichert wird.

Gerne!
Hier ist der Code für einen einfachen Login-Screen in Python:

```
import getpass

def login():
    username = input("Username: ")
    password = getpass.getpass("Password (masked): ")
    remember_password = input("Remember password for 14 days? (y/n): ")

    if remember_password.lower() == "y":
        # save password to file or database for 14 days
        pass

    # check username and password against database or API
    if check_credentials(username, password):
        print("Login successful!")
    else:
        print("Invalid username or password.")

def check_credentials(username, password):
    # check username and password against database or API
    return True  # replace with actual implementation

if __name__ == "__main__":
    login()
```

In diesem Beispiel verwenden wir die getpass-Funktion aus der Python-Standardbibliothek, um das Passwort zu maskieren und nur bei Bedarf anzuzeigen. Die Funktion check_credentials sollte durch eine tatsächliche Implementierung ersetzt werden, die die eingegebenen Anmeldeinformationen gegen eine Datenbank oder API überprüft.

Die Checkbox für das Speichern des Passworts könnte durch eine einfache if-Abfrage implementiert werden, die prüft, ob der Nutzer die Option ausgewählt hat und das Passwort entsprechend für 14 Tage speichert. Da die Speicherung von Passwörtern in der Regel jedoch aus Sicherheitsgründen vermieden wird, haben wir diese Funktionalität nicht implementiert.

Die Erstellung von Programmcode beherrscht das Sprachmodell, weil auch diese Form von Texten über die Trainingsdaten kennengelernt hat. Neben der Codeerstellung kann man mit ChatGPT auch Code kontrollieren, kommentieren, erklären lassen oder von einer Programmiersprache in eine andere Programmiersprache übersetzen.

Ein Blick in den Maschinenraum von Sprachmodellen

Sprachmodelle und viele Fragen

Aktuell sorgen die **G**enerative-**P**re-Trained-**T**ransformer-Sprachmodelle von OpenAI für Aufsehen. Aber wie kommt es zu der rasanten Entwicklung bei Sprachmodellen? Warum sind sie in der Lage, menschenähnliche Texte zu generieren? Wie funktioniert das Ganze und wo findet sich die Künstliche Intelligenz wieder?

Die Beispiele zeigen, wie mächtig Sprachmodelle geworden sind. Dennoch weisen die Sprachmodelle teilweise noch schwerwiegende Mängel auf. Warum die Entwicklung so lange gedauert hat und welche Herausforderungen von den Entwicklern zukünftig zu lösen sind, ist nur zu begreifen, wenn man ein Verständnis für die Funktionsweise der Modelle hat. Nur mit einem ausreichenden Verständnis für die Sprachmodelle ergibt sich auch das Wissen um die Schwächen und Limitationen der Technologie.

Sprache ist komplex, vielfältig und mehrdeutig

Der Wissenschaftszweig, der sich mit dem Verständnis und der Verarbeitung von menschlicher Sprache durch Maschinen beschäftigt, wird Natural Language Processing (NLP) genannt. Für Maschinen stellt die menschliche Sprache eine große Herausforderung dar. Denn Sprache unterliegt einer Vielzahl von Regeln (Syntax) und Worte haben Bedeutungen (Semantik), die sich je nach Kontext verändern. Sprache ist mehrdeutig, vielfältig und ein Ausdruck persönlicher Prägung, Bildung und regionaler Kultur. Selbst uns Menschen fällt es schwer, Sprache zu verstehen – schon ein Rechtstext, Arztbrief, Dialekt oder eine Redewendung stellen uns vor Herausforderungen beim Sprachverständnis, das Erlernen einer Fremdsprache nicht zu vergessen. Einer Maschine menschliche Sprache beizubringen, ist eine hochkomplexe Aufgabe.

Begriffserklärungen

Künstliche Intelligenz (KI)

bezeichnet die Fähigkeit von Maschinen, bestimmte menschliche Fähigkeiten wie Wahrnehmung, Entscheidungsfindung, Problemlösung und Spracherkennung zu imitieren oder zu übertreffen. Dabei kommen Algorithmen und Technologien wie maschinelles Lernen, neuronale Netze und Deep Learning zum Einsatz.

Maschinelles Lernen

ist ein Teilgebiet der künstlichen Intelligenz, bei dem Computerprogramme auf Basis von Daten autonom „lernen“ und dabei ihre Leistung verbessern.

Deep Learning

ist ein Teilbereich des maschinellen Lernens, der sich auf künstliche neuronale Netzwerke mit vielen Schichten von Neuronen konzentriert. Durch Trainieren mit großen Datenmengen kann es komplexe Muster erkennen und Vorhersagen treffen.

Neuronale Netze

sind computergestützte Systeme, welche von biologischen Nervenzellen und ihren Verbindungen im Gehirn inspiriert sind. Die Neuronen sind in Schichten organisiert und können Muster in Daten erkennen. Sie werden als Modell zur Informationsverarbeitung genutzt und finden Anwendung in verschiedenen Bereichen der künstlichen Intelligenz.

Transformer-Modelle

sind tiefe neuronale Netzwerke für maschinelles Lernen, die u. a. bei der Verarbeitung von Sprache bzw. Texten eingesetzt werden. Sie können längere Sequenzen verarbeiten und haben eine höhere Genauigkeit bei der Vorhersage. Sie sind derzeit die bevorzugte Architektur für viele Aufgaben im Bereich des Sprachverständnisses und der Sprachverarbeitung durch Maschinen.

Generative-Pre-Trained Transformer-Modelle

werden die Sprachmodelle von OpenAI genannt, die ebenfalls auf der Transformer-Architektur basieren. Sie werden auf großen Textmengen vortrainiert, um Texte zu verarbeiten und zu generieren.

Probabilistische Modelle – Große Textmengen gepaart mit viel Rechenpower

Frühe NLP-Versuche in den 1940er Jahren zielten darauf ab, menschliche Sprache für Maschinen regelbasiert, syntaxbezogen und formalisiert zu beschreiben. Damit erzielte man zunächst auch einige Erfolge, stieß aber schnell an Grenzen. Eine Gegenströmung setzte zur gleichen Zeit auf statistische und wahrscheinlichkeitsbasierte Ansätze [25]. Die statistischen und probabilistischen Ansätze verzichten weitestgehend auf die sprachlichen Regeln und gehen vereinfacht von folgender Logik aus: Die Bedeutung eines Wortes bestimmt sich aus den begleitenden Worten, die häufig nebeneinander oder in der Nähe auftreten. Man muss also nur genug Texte kennen, um zu lernen,

„You shall know a word by the company it keeps!"

J.R. Firth,
A synopsis of linguistic theory (1957)

wie Worte untereinander zusammenhängen und wie wahrscheinlich sie gemeinsam auftreten. Je mehr Informationen zum Kontext eines gesuchten Wortes bekannt sind, desto größer die Wahrscheinlichkeit, ein passendes Wort zu finden. Das Wort „Bank" erhält z. B. eine jeweils andere Bedeutung, wenn es von den Wörtern „Geld, Schalter, abheben" oder „Park, Natur, sitzen" begleitet wird.

Statistische und wahrscheinlichkeitsbasierte Ansätze wurden lange durch eine mangelnde Rechenkapazität und eine schlechte Verfügbarkeit von Text-Daten in digitaler Form zurückgehalten. Dies änderte sich spätestens durch das Aufkommen vom Personal Computer (PC) und des Internets. Seitdem erleben wir durch die Digitalisierung und Technologisierung ein Anwachsen der verfügbaren Daten, der Rechenleistung und von Ansätzen innerhalb des Natural Language Processing, die davon profitieren [25].

Für das Training von GPT-3 wurden beispielsweise umfangreiche Texte aus dem Internet genutzt. Dafür wurde das Internet gescraped, d. h. systematisch durchsucht und die Inhalte heruntergeladen. Zusätzlich wurde auf bereits digitalisierte Datenquellen (vor allem Bücher) zurückgegriffen. Die beeindruckende Menge von 570 Gigabyte an Daten aus verschiedenen Quellen wie Webseiten, Foren, Wikipedia-Artikeln und Büchern kam so zusammen. Insgesamt wurden ca. 300 Milliarden Wörter (genauer 499 Milliarden Tokens) in das Sprachmodell eingespeist [26, 27].

Welche enorme Rechenleistung für das Training eines Sprachmodells benötigt wird, zeigt der „Supercomputer" den OpenAI beim Training einsetzt. Die Zahlen klingen unglaublich: 285.000 Zentralprozessoren (CPU), 10.000 Graphikprozessoren (GPU) und 400 Gigabits pro Sekunde an verfügbarer Netzwerkverbindung für jeden einzelnen GPU. Der Supercomputer gehört zu den 5 leistungsstärksten Computern auf der Welt und kommt von Microsoft – ein Grund dafür, dass OpenAI und Microsoft eine enge strategische Kooperation pflegen [28]. OpenAI darf die technische Infrastruktur von Microsoft zu günstigen Konditionen nutzen, Microsoft erhält dafür Verwertungsrechte an den Produkten von OpenAI [29].

Word-Embeddings – semantische Strukturen in Texten codieren

Damit ein großes Sprachmodell mit Texten umgehen kann, müssen Texte zunächst in eine maschinenlesbare Form umgewandelt werden. Jedem Wort (oder auch einzelnen Wortbestandteilen), Satz- und Sonderzeichen wird dann ein numerischer Wert zugeordnet, mit dem der Computer operieren kann. Zusätzlich sollen ähnliche Worte auch ähnliche numerische Werte aufweisen. Dazu werden Wörter mit Hilfe von Vektoren in einer Art mehrdimensionalem Raum abgebildet. Ähnliche Wörter mit ähnlichen Bedeutungen werden möglichst nahe beieinander im Vektorraum positioniert. Die Ähnlichkeit bestimmt sich durch das gemeinsame Auftreten mit anderen, begleitenden Worten. Diese sogenannte Worteinbettung (Word

Zur Einordnung der schieren Datenmengen und Rechenpower

1) Wie lange bräuchte ein Mensch, um alle Wörter im Trainingsdatensatz vorzulesen?

Es würde ungefähr 2.000.000.000 Minuten, 33.333.333,33 Stunden, 198.412,70 Wochen, 46.296,30 Monate oder 3.804,57 Jahre dauern, bis eine Person 300.000.000.000 Wörter mit einer Geschwindigkeit von 150 Wörtern pro Minute liest.

2) Wie lange würde das Training von GPT-3 mit einem einzelnen Graphikprozessor dauern?

Im Jahr 2020 schätze Lambdalabs, dass das Training von GPT-3 auf einem der damals schnellsten, verfügbaren Graphikprozessoren (Tesla V100) circa 355 Jahren dauern und ca. 4,6 Millionen US-Dollar kosten würde ... angenommen, die Preise blieben über die 355 Jahre stabil [26].

Embedding) in einen multidimensionalen Raum ist die Grundlage für die Sprachverarbeitung durch große Sprachmodelle [30].

Neuronale Netze und maschinelles Lernen – Muster in den Daten finden

Um aus den Word Embeddings nun sinnvolle Texte bilden zu können, müssen Muster in den Daten identifiziert werden. Dafür wird auf neuronale Netzwerke und maschinelles Lernen zurückgegriffen. Neuronale Netzwerke sind dem menschlichen Gehirn nachempfunden und bestehen aus miteinander verbundenen Neuronen. Über die Verbindungen werden Informationen (Gewichte) von einer Schicht von Neuronen an die nächste Schicht weitergereicht und durch die Neuronen verarbeitet [31]. Jedes Neuron verarbeitet die erhaltenen Informationen über eine meist einfache mathematische Operation (z. B. Schwellenwert- oder Aktivierungsfunktionen), indem sie ein Gewicht verstärken oder dämpfen [30]. In einer ersten Trainingsphase wird das Sprachmodell auf großen Textmengen vortrainiert (pre-trained). Beim Training wird das sogenannte Masking angewandt. Das ist eine Technik im maschinellen Lernen, bei der einzelne Wörter in Texten verdeckt – also maskiert – werden. Die Modelle lernen eigenständig (unsupervised), diese maskierten Wörter basierend auf den vorherstehenden Wörtern vorherzusagen. Die Vorhersage wird mit der tatsächlichen Ausgabe, d. h. dem maskierten Wort, verglichen. Das Modell wird mit jedem weiteren Trainingsschritt besser bei der Vorhersage, indem es die Gewichte so anpasst, dass die Fehler des Modells minimiert werden. Dazu kommen Methoden wie Backpropagation und Gradient Descent zum Einsatz [31]. Am Ende entsteht durch das Training und die ausgebildeten Gewichte eine Art gelerntes Denkmuster. Damit ist das große Sprachmodell sogar in der Lage, Texte zu verarbeiten, die es im Training nie gesehen hat. Über den Trainingsprozess hat das neuronale Netz selbstständig die beste Konfiguration an Gewichten kennen, um basierend auf einer Texteingabe das nächstwahrscheinliche Wort vorherzusagen. Schrittweise bildet das Sprachmodell somit einen ganzen Text als Ausgabe.

In früheren Sprachmodellen kamen beispielsweise sogenannte Rekurrente Neuronale Netzwerke (RNN) zum Einsatz. Die bestehen aus einer Reihe hintereinander geschalteter neuronaler Netzwerke, zwischen denen eine Informationsübergabe stattfindet. Ein neuronales Netzwerk verarbeitet die Eingabe und erzeugt das nächstwahrscheinliche Wort. Die Ausgabe dieses neuronalen Netzes wird dann als zusätzliche Eingabe an ein nachgeschaltetes neuronales Netzwerk übergeben. Damit erhalten RNNs eine Art übermittelbares Gedächtnis, welches Vorhersage des nächsten Wortes im nachfolgenden Modell beeinflusst. Diese Verkettung von neuronalen Netzen sorgt vor allem bei langen Texten und weiter entfernten Bezügen im Text für Schwierigkeiten. In RNNs können Informationen über die einzelnen Schleifen verwässert bzw. verfälscht werden und die Verkettung macht sie relativ langsam. Diese Mängel beim Berücksichtigen von Kontext und der Geschwindigkeit limitieren ihren Einsatz bei der Sprachverarbeitung [32].

Transformer – Eine Revolution für Sprachmodelle

Die Vorstellung von Transformer-Modellen stellte 2017 einen Meilenstein in der Entwicklung von Sprachmodellen dar. Sie basieren auf der Transformer-Architektur, welche von Google-Forschern im Paper „Attention is all you need“ eingeführt wurde [33]. Transformer-Modelle bestehen laut dem ursprünglichen Entwurf aus Encodern und Decodern. Ein Encoder nimmt die Texteingabe entgegen und erstellt eine mathematische Repräsentation des Textes und seiner Merkmale. Der Encoder ist darauf optimiert, ein Verständnis aus der Eingabe zu erlangen. Decoder nutzen die Informationen des Encoders zusammen mit anderen Eingaben, um die Ausgabe zu generieren. Nach Einführung der Transformer-Architektur haben sich eine Reihe von großen Sprachmodellen herausgebildet, die nur Encoder, nur Decoder oder aber Encoder und Decoder nutzen. Sie werden entsprechend ihrer Fähigkeiten für unterschiedliche Zwecke eingesetzt [34:]

Reine Encoder-Modelle fokussieren auf das Verständnis des Inputs und sind für Aufgaben wie Klassifikation oder Entitätserkennung geeignet.

Reine Decoder-Modelle sind für generative Aufgaben geeignet, wie z. B. die Textgenerierung.

Encoder-Decoder-Modelle sind für generative Aufgaben wie Übersetzungen und Zusammenfassungen geeignet, die eine Eingabe erfordern.

Der Übersetzer DeepL ist beispielsweise ein Encoder-Decoder-Modell, während die GPT-Sprachmodelle von OpenAI reine Decoder-Modelle sind [32].

Aufmerksamkeit ist alles, was man braucht

Das Besondere an Transformer-Modellen ist ihr Aufmerksamkeitsmechanismus. Während andere Modelle mit RNNs die Eingabe noch nacheinander verarbeiten und damit nur auf alles vorherstehenden Bezug nehmen können, verarbeiten Transformer die gesamte Eingabe auf einmal. Dazu müssen sie aber verstehen, wie wichtig welche Wörter der Eingabe sind. Dazu wird zunächst jedes Wort der Eingabe in einen Vektor überführt. Der Aufmerksamkeitsmechanismus gleicht jedes Wort der Eingabe mit allen anderen der Eingabe ab und errechnet eine neue mathematische Darstellung. Diese berücksichtigt wie stark einzelne Wörter der Eingabe zusammenhängen und

Beispiele für Transformer-Modelle[39]

Release	Model	Architektur	Parameter	Anwendung	Entwickler
06/2018	GPT	Decoder	117 Mio	Textgenerierung, aber über Fine-Tuning anpassungsfähig für viele andere NLP-Aufgaben.	OpenAI
10/2018	BERT	Encoder	340 Mio	Allgemeines Sprachverständnis und Frage-Antworten. Viele weitere Sprachanwendungen folgten.	Google
02/2019	GPT-2	Decoder	1,5 Mrd	Wie GPT-1	OpenAI
04/2019	RoBERTa	Encoder	356 Mio	Wie BERT	Google, University of Washington
10/2019	BART	Encoder & Decoder	374 Mio	Hauptsächlich Textgenerierung, aber auch einige Textverständnisaufgaben.	Facebook
05/2020	GPT-3	Decoder	175 Mrd	Ursprünglich Textgenerierung, aber im Laufe der Zeit für eine Vielzahl von Anwendungen in Bereichen wie Code-Generierung, aber auch Bild- und Audio-Generierung verwendet worden	OpenAI
01/2022	LaMDA	Decoder	137 Mrd	Allgemeine Sprachmodellierung	Google
04/2022	PaLM	Decoder	540 Mrd	Sprachverständnis und Generierung	Google
02/2023	LLaMA	Decoder	65,2 Mrd	Common-Sense-Reasoning, geschlossene Frage-Antworten, Leseverständnis, mathematisches Denken, Code-Generierung, massives Multitask-Sprachverständnis	MetaAI
03/2023	GPT-4	Decoder	Unbekannt	Wie GTP-3, plus Multimodalität (Text-zu-Bild, Bild-zu-Text)	OpenAI

damit auch die semantische Struktur. Dabei zieht er auch Informationen zu der Position der einzelnen Wörter, über ein sogenanntes Positional Encoding, heran. Am Ende dieses Schritts sind die Eingabetexte so gewichtet, dass die Aufmerksamkeit auf relevante Wörter und deren Bezug untereinander gelenkt wird[32]. Das funktioniert auch über große Distanzen zwischen einzelnen Wörtern und in beide Richtungen, d. h. der Kontext eines Wortes kann sich auf Worte davor oder danach beziehen.

Durch die Verarbeitung der gesamten Eingabe können die Berechnungen parallelisiert werden, d. h. eine Recheneinheit wendet den Aufmerksamkeitsmechanismus auf das erste Wort an, eine weitere Recheneinheit zeitgleich auf das zweite Wort, eine weitere Recheneinheit auf das dritte Wort, usw. Damit wird die Eingabe wesentlich schneller verarbeitet, als das bei RNNs möglich ist. Die über den Aufmerksamkeitsmechanismus errechneten Vektoren werden anschließend an ein tiefes neuronales Netz übergeben[32]. Tiefe Netzwerke enthalten viele Schichten von Neuronen, die es ihnen ermöglichen, immer komplexere Merkmale und Muster in den Daten zu erkennen. In Sprachmodellen ermöglicht Deep Learning die Verarbeitung und das Verständnis von natürlicher Sprache auf einer tieferen Ebene, indem es beispielsweise Textstrukturen, Bedeutungen und sogar kulturelle Nuancen erfasst[31]. Das Netzwerk sagt nun das nächstwahrscheinliche Wort voraus. Grundlage dafür sind die im Training mit großen Textmengen gelernten Muster und die im Aufmerksamkeitsmechanismus übergebenen Informationen[30].

Ein einzelner Decoder bei GPT-Sprachmodellen besteht aus einem Aufmerksamkeitsmechanismus und einem neuronalen Netzwerk. Die Eingabe durchläuft allerdings mehrere hintereinander geschaltete Decoder, wobei in jedem einzelnen Zyklus der Fokus auf andere Merkmale der Eingabe gelegt wird und andere Gewichte in den Netzwerken existieren. Dadurch können unterschiedliche Merkmale der Eingabe berücksichtigt werden und ein tiefes Sprachverständnis erreicht werden. Am Ende liefert der Transformer eine Wahrscheinlichkeit für jedes Wort (bzw. jeden Token) in seinem Sprachschatz, welche angibt, wie wahrscheinlich es auf die Eingabe folgt. Aus den Worten mit den höchsten Wahrscheinlichkeiten wird dann ein Wort gewählt, welches an die Eingabe angefügt wird[32]. Das muss je nach Einstellung nicht immer das wahrscheinlichste Wort sein, damit die Ausgabe variiert wird und nicht roboterhaft klingt. Die jeweils neue Eingabe durchläuft dann wieder einen kompletten Decoder-Zyklus, bis derjenige Token die höchste Wahrscheinlichkeit aufzeigt, der das Ende einer Ausgabe kennzeichnet.

Menschliches Feedback – Der Weg zu besseren Texten

OpenAI nutzt nach dem unsupervised Pre-Training zusätzlich menschliches Feedback beim Training seiner Sprachmodelle durch ein Fine-Tuning und ein sogenanntes Reinforcement Learning with Human Feedback (RLHF) [35,36].

Beim Fine-Tuning werden zufällig Prompts aus einem Datensatz ausgewählt und durch einen Menschen beantwortet. Diese Eingabe-Ausgabe-Kombination wird zum Training des Sprachmodelles über ein überwachtes Lernen (supervised learning) genutzt, welches wieder zur Anpassung der Gewichte im tiefen neuronalen Netz führt.

Beim RLHF beurteilen Menschen verschiedene Ausgaben zu einem Prompt. Diese können von verschiedenen Sprachmodellvarianten stammen. Üblicherweise werden die Antworten von Labelern in eine Rangfolge gebracht, welche die menschlichen Präferenzen abbilden. Basierend auf diesem Feedback wird ein weiteres KI-Modell (Reward-Modell) trainiert, welches jeweils Belohnungen oder Bestrafungen für eine große Anzahl an Ausgaben des Sprachmodells vergibt. Hier trainiert also eine KI eine andere KI. Das Sprachmodell bzw. die neuronalen Netze passen daraufhin ihre Gewichte an, um die Belohnungen zu maximieren. Dadurch wird der Output von Sprachmodellen stärker an menschliche Präferenzen, Kommunikation und Entscheidungsfindung angepasst [36,37].

Sprachmodelle werden immer leistungsfähiger

Die Kombination von Transformer-Architekturen mit neuronalen Netzwerken, Deep Learning und Reinforcement Learning mit menschlichem Feedback hat dazu geführt, dass Sprachmodelle immer leistungsfähiger, vielseitiger und sprachgewandter werden. GPT-3, welches als Basis für ChatGPT dient, verfügt beispielsweise über 175 Milliarden Parameter – das sind die Werte, die dem Model zur Verfügung stehen, um die Vorhersage des nächsten Wortes in einem Text zu machen. Das GPT-3-Modell besteht aus 96 Schichten von neuronalen Netzen, d. h. hintereinander geschalteten Decoder-Blöcken [14,27]. Für das Training kamen ca. 499 Milliarden Tokens und 570 GB an Textdaten zum Einsatz [26]. Allein die Möglichkeit, solche großen Textmengen zu verarbeiten, hat in den letzten Jahren für eine Verbesserung der Performance von Sprachmodellen gesorgt. Mittlerweile gibt es schon Schwierigkeiten, geeignete Daten für das Training von großen Sprachmodellen zu finden [38]. OpenAI hat durch die Veröffentlichung von ChatGPT jedoch Zugang zu einem einzigartigen Datenschatz erhalten, der durch die Nutzerprompts offenbart, was Nutzer interessiert und durch Beurteilungen der Ausgabe weitere Trainingsdaten erzeugt. Die Zukunft der Sprachmodelle liegt in ihrer weiteren Verbesserung und der Minimierung der Grenzen und Risiken.

Grenzen und Risiken von Sprachmodellen

Eine wichtige Erkenntnis lautet – Sprachmodelle sind ein Abbild Ihrer zugrundeliegenden Daten und Methodik. Aus der Art und Weise wie Sprachmodelle trainiert wurden und wie sie funktionieren, ergeben sich einige Implikationen für deren Nutzung.

Sprachmodelle verstehen Sprache nicht im menschlichen Sinne

Sprachmodelle verstehen Sprache nur in der Art, dass sie über das Training Massen an Daten verarbeitet und Muster gelernt haben. Ein kausales oder temporales Verständnis haben sie nicht, wie auch keine eigenen Ziele oder Intentionen[40]. In diesem Zusammenhang wird auch öfter von stochastischen Papageien (Stochastic Parrots) gesprochen, die das Gelernte ohne ein Verständnis nachplappern[41]. Erstellen Sprachmodelle also Texte, hängen die Worte mit einer gewissen Wahrscheinlichkeit zusammen. Das heißt jedoch nicht, dass sie logisch schlüssig und richtig sein müssen. Deshalb ist die menschliche Kontrolle von Inhalten unerlässlich.

Die Daten sind nicht frei von Verzerrungen und potenziell schädlichen Inhalten

Die Daten für das Sprachmodell stammen zum großen Teil aus dem Internet, wo bekanntlich jede Meinung ihren Platz findet. Zwar wurden die Daten einer Filterung unterzogen, um qualitativ hochwertige Ausgangsdaten zu erhalten. Dies verhindert jedoch nicht, dass die Sprachmodelle mit Daten trainiert wurden, die beispielweise Falschinformationen (z. B. Fake News), implizite Stereotypen (z. B. Geschlechter- oder Arbeitsrollen) und Diskriminierungen (z. B. Rassismus) enthalten[42]. Das Sprachmodel ist und bleibt ein Spiegel der Gesellschaft und ihrer veröffentlichten Inhalte. Das bedeutet auch, dass die Trainingsdaten Beschreibungen in Bezug auf kriminelle und unethische Handlungen enthalten können. In der Presse wurde zum Beispiel berichtet, dass die Attribute männlich, weiß einen erfolgreichen Wissenschaftler ausmachen oder ChatGPT Anleitungen zur Programmierung von Schadsoftware ausgibt[43, 44].

Dem Umstand dieser möglichen Verzerrungen muss man sich bewusst sein. Sprachmodelle sind nicht objektiv und nicht neutral. Sie können – wie jedes Werkzeug – auch missbräuchlich eingesetzt werden. Im Versicherungskontext wurde bereits diskutiert, ob der Einsatz von ChatGPT zu mehr Versicherungsbetrug führen kann[45]. Dem ist entgegenzustellen, das Sprachmodelle und andere KI-Algorithmen ebenso zur Erkennung von Betrugsversuchen eingesetzt werden können.

OpenAI ist darum bemüht, mittels vorgeschalteter Filter eine Reihe von Verzerrungen zu mindern und die missbräuchliche Verwendung von ChatGPT zu verhindern[22]. Die ChatGPT-Antworten werden dann entweder verweigert oder bewusst ausgewogen dargestellt. Welche Inhalte das in welcher Form betrifft, liegt in der Hand der Entwickler selbst. Hier werden auch kritische Stimmen laut, die die Meinungsfreiheit beschnitten sehen und alternative Angebote schaffen wollen[46]. Andere sehen z. B. eine politische Neutralität als einen Schritt hin zu einer verantwortungsvollen KI[47].

Einmal Gelerntes abzutrainieren ist fast unmöglich

Sprachmodelle erhalten ihr gesammeltes Wissen durch den Trainingsprozess, auf dessen Grundlage sich die Gewichte (Verbindungen zwischen Neuronen) im tiefen neuronalen Netzwerk bestimmen. Es ist nicht (vollständig) nachvollziehbar, wie einzelne Trainingsinhalte auf die Gewichte und damit Zustände der Neuronen bei der Verarbeitung von Texten wirken. Deshalb können einmal angelernte Inhalte nicht einfach aus dem Modell entfernt werden. Dieser Umstand stellt eine große Herausforderung in Bezug auf die eben diskutierten Falschinformationen, Verzerrungen sowie potenziell schädlichen Inhalte in den Trainingsdaten dar.

Wurde ein Sprachmodell zusätzlich mit unternehmensinternen, produktspezifischen Daten trainiert, z. B. durch ein Fine-Tuning, ergibt sich eine weitere Herausforderung: Veraltete Daten. Was passiert, wenn beispielsweise Produktinformationen in Form von Versicherungsbedingun-

gen zum Training eines Sprachmodells benutzt wurden, das Produkt nun aber aus dem Produktportfolio entfernt wurde? Die alten Produktinformationen haben sich im Trainingsprozess auf die Gewichte im Sprachmodell ausgewirkt und können nicht einfach angepasst werden. Vielmehr muss das Sprachmodell dann mit angepassten Trainingsdaten erneut trainiert werden.

Will man die Anpassung von Gewichten im Modell verhindern, lohnt sich die Arbeit mit Embeddings[19]. Diese erlauben den Zugriff auf Daten, die außerhalb des Sprachmodells liegen, wobei die Gewichte jedoch nicht verändert werden. Dies funktioniert ähnlich wie in einem der gewählten Beispiele, wo die Versicherungsbedingungen als Bezugsrahmen für die Antworten des Sprachmodells dienten. Bei Embeddings kopiert man jedoch keine vollständigen Texte, sondern übergibt die modellexternen Daten als mathematische Vektor-Repräsentation. Damit lassen sich große Datenmengen auch sehr effizient an das Sprachmodell übergeben. Im Grunde genommen, hat das Sprachmodell über das Standard-Training gelernt, mit Sprache umzugehen und wendet dieses Wissen auf den Prompt und die externen Daten an.

ChatGPT ist keine Such- oder Faktenmaschinen

Sprachmodelle sind wie ein kollektives Gedächtnis, das sich alles „Gelesene“ gemerkt hat und die Informationen jederzeit abrufen kann. Das Sprachmodell kann aber keineswegs sagen, woher die ausgegebenen Informationen genau stammen. Dies verdeutlicht vor allem eine Sache: ChatGPT ist keine Suchmaschine. Sie kann (noch) nicht auf indizierte Quellen oder Referenzen verweisen.

Je häufiger und ähnlicher die Trainingsdaten einen Sachverhalt widerspiegeln, desto klarer ausgeprägte Wortwahrscheinlichkeiten wird es im Sprachmodell geben. Im Ergebnis werden die ChatGPT-Ausgaben also bei homogenen und oft beschriebenen Themen nah an den Trainingsdaten und einer Art objektiver Wahrheit liegen. Einschränkend muss jedoch festgestellt werden, dass auch zu vermeintlichen Fakten wie der Mondlandung oder dem Klimawandel sehr kontroverse Meinungen existieren. Diese Meinungen haben sehr wahrscheinlich einen Platz in den Trainingsdaten gefunden und können sich in den Ausgaben der Sprachmodelle widerspiegeln. Da ChatGPT nicht berichten kann, woher die Informationen stammen, müssen Ausgaben immer mit kritischem Blick überprüft und einem sogenannten Fact-Checking unterzogen werden.

Des Weiteren wurden für das Training nur Daten bis zu einem Stichtag berücksichtigt. Bei GPT-3.5 liegt dieser im September 2021[27]. Auf aktuelle Daten kann ChatGPT standardmäßig nicht zugreifen. Hier sind allerdings auch Neuerungen zu erwarten: Die GPT-4 Integration in die Suchmaschine Bing von Microsoft greift schon auf aktuelle Daten aus dem Internet zu[48]. Ein weiterer Schritt sind die im März 2023 vorgestellten Plug-ins von ChatGPT, die es erlauben, auf aktuelle Daten und externe Services von Dritten zuzugreifen[49].

Erfundene Inhalte werden mit großem Selbstbewusstsein kommuniziert

Wie eben beschrieben, sind Sprachmodelle keine Faktenmaschinen. Das bedeutet, dass die Inhalte, die ChatGPT erstellt, Unwahrheiten enthalten und vollständig erdacht sein können. Dieses Phänomen ist bei Sprachmodellen unter dem Begriff Halluzinieren bekannt. Das Sprachmodell basiert sein Sprachverständnis rein auf Wahrscheinlichkeiten. Das bedeutet allerdings, dass noch so kleine, eigentlich bedeutungslose oder zufallsgetriebene Zusammenhänge zwischen Wörtern existieren und für Ausgaben des Sprachmodells herangezogen werden. Problematisch ist neben dem Halluzinieren an sich auch die Art und Weise der Darbietung. Zweifel gibt es nicht. Jeder Inhalt wird plausibel, selbstbewusst und sprachlich schlüssig präsentiert. Der uninformierte Nutzer wird dadurch leicht in einer trügerischen Sicherheit gewogen. Wieder gilt: Vertrauen ist gut, Kontrolle ist besser.

Die Abminderung von Halluzinationen ist ein wichtiges Forschungsfeld im Bereich der Sprachmodelle. Mit der Veröffentlichung von GPT-4 verkündete OpenAI, dass Halluzinationen noch immer eine Herausforderung darstellen. Die Wahrscheinlichkeit der Ausgabe sachlicher bzw. faktenbasierter Antworten ist jedoch um 40% höher als bei der ursprünglichen Variante von GPT-3.5[14]. Trotzdem zeigen interne Benchmarks zur Faktenevaluation von OpenAI, dass die Antworten von ChatGPT über verschiedenen Kategorien nur zu ca. 75–85% an den von Menschen als ideal wahrgenommenen Antworten liegen[14]. Es gibt jedoch weitere Versuche, die Antworten von Sprachmodellen anhand von externen Quellen zu verifizieren. Man spricht in diesem Zusammenhang auch von Grounding oder Groundedness (Fundierung/Fundierheit). Diese gibt an, inwieweit die Antworten durch maßgebliche externe Quellen gestützt werden[50, 51].

Das Urheberrecht liegt (noch) beim Nutzer – oder doch Dritten?

Laut den Nutzungsbedingungen von OpenAI (Stand: 20.04.2023) überträgt das Unternehmen den Nutzern alle Rechte, Titel und Interessen an und in Bezug auf den Output[17]. Der Nutzer darf die Inhalte für alle Zwecke verwenden. Das schließt auch kommerzielle Zwecke wie den Verkauf und Veröffentlichungen ein. Das heißt allerdings nicht, dass der Output von ChatGPT nicht auch das Urheberrecht Dritter betreffen kann.

Da das Sprachmodell wahrscheinliche Wortkombinationen ausgibt, kann es zu folgenden Situationen kommen: Eine Ausgabe von ChatGPT kann ...

1. ... in identischer oder ähnlicher Form schon mal für einen anderen ChatGPT-Nutzer generiert worden sein. Das kann insbesondere dann passieren, wenn mehrere Nutzer gleiche oder ähnliche Anfragen mit kurzen Ausgaben stellen.

2. ... mit Texten übereinstimmen, die in den Trainingsdaten vorkommen. Da diese Texte wahrscheinlich einmal von einem Menschen erstellt wurden, sind sie nach dem Urheberrecht schützenswert.

3. ... mit aktuelleren Texten übereinstimmen, die das Sprachmodell nicht kennt, weil sie nicht Teil des Trainingsdatensatzes sind. Es ist möglich, dass ein Text schon vorhanden, aber nicht Bestandteil des Trainingsdatensatzes war oder auch erst nach dem Sammeln des Trainingsdatensatzes erstellt wurde.

Eines ist klar: ChatGPT prüft nicht, ob die Ausgabe in der Form schon einmal selbst von ChatGPT getätigt wurde, ob sie in den Trainingsdaten vorkommt oder in externen Quellen. Die Ausgabe sollte also immer auf mögliche Urheberrechtsverletzungen von Dritten geprüft werden.

Es ist zu vermuten, dass Plagiatssoftware oder sogenannte AI-Detektoren zukünftig eine breite Anwendung finden werden. Der Anbieter OriginalityAI prüft zum z. B., mit welcher Wahrscheinlichkeit ein Text von einem Menschen oder einer Künstlichen Intelligenz geschrieben wurde und ob Textbestandteile schon einmal woanders verwendet bzw. plagiiert wurden [52].

Dass OpenAI die Urheberrechte an den erstellen Inhalten so bereitwillig zur Verfügung stellt, liegt auch an der derzeitigen Rechtslage. Nach deutschem Recht kann nur ein Mensch urheberrechtlichen Schutz seiner Werke einfordern:

„Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen."

§ 2 Abs. 2 UrhG [53]

Es wird eine persönliche geistige Schöpfung eines schaffenden Menschen vorausgesetzt. Dies wird auch in der amerikanischen Rechtsprechung so vertreten. Auf der europäischen politischen Ebene gibt es bereits seit längeren Diskussionen um die Schaffung eines eigenen Rechtsubjektes (e-Person) für Künstliche Intelligenzen, um Fragen des Urheberrechts, Leistungsschutzrechtes, Patentrechtes, aber auch der Haftung zu klären [54]. Es ist zu vermuten, dass über die zunehmende Verbreitung von generativen Künstlichen Intelligenzen auch ein gewisser Handlungsdruck auf die politischen Entscheidungsträger wirkt.

In Großbritannien ist man seit geraumer Zeit schon etwas weiter:

„In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken"

Sec. 9 (3) UK Copyright Designs and Patent Act 1988 [55]

Dem britischen Ansatz folgend kann der Urheber eine Person sein, die die Generierung durch die KI veranlasst hat und erforderliche Maßnahmen dafür schafft, dass die KI das Werk überhaupt erschaffen kann. Trotzdem bleiben Fragen offen: Wer hat die erforderlichen Maßnahmen geschaffen? Die Entwickler, die den Programmcode schreiben, die Menschen, die Inhalte für die Trainingsdaten generierten oder die Nutzer, die durch ihre Kreativität bei Prompts den Output des Sprachmodells maßgeblich beeinflussen? Diese Fragen sind rechtlich noch nicht abschließend bewertet.

Mathematische Operationen sollte man überprüfen

Bittet man ChatGPT, mathematische Operationen (Addieren, Subtrahieren, Dividieren, Multiplizieren) durchzuführen, zeigt die Anwendung zum Teil große Schwächen. Während einfache Operationen wie das Einmaleins souverän gemeistert werden, passieren beim Rechnen mit größeren Zahlen häufiger Fehler. Obwohl Rechenschritte in einzelne Schritte zerlegt und richtig beschrieben werden, weichen die Ergebnisse von ChatGPT gerne auch von denen eines Taschenrechners ab. Fragt man ChatGPT nach den Gründen für die Abweichungen führt er entweder Rundungsfehler an oder behauptet stur, dass das Ergebnis richtig sei und der Nutzer falsch gerechnet hat. Beim Lösen von 1.000 mathematischen Problemstellungen zeigte ChatGPT im Vergleich zu State-Of-The-Art Algorithmen im Januar 2023 lediglich eine 60% Genauigkeit [56]. Auch in der schon genannten MBA-Prüfung machte ChatGPT überraschende Fehler bei leichten Aufgaben [11].

Auch dieses Verhalten lässt sich sehr gut mit den Eigenheiten eines Sprachmodells erklären – ChatGPT führt keine Berechnung durch. In den Trainingsdaten hat das zugrundeliegende Sprachmodells eine Vielzahl von mathematischen Operationen gesehen und die Wahrscheinlichkeiten zwischen Ziffern und Operanden gelernt. Das Einmaleins und leichtere Rechenaufgaben kommen wahrscheinlich häufig in den Testdaten vor und sind mit hoher Sicherheit replizierbar. Aufgaben, die für das Sprachmodell neu sind, werden wieder nur über Wahrscheinlichkeiten von Ziffern- und Operanden-Kombinationen bestimmt, aber eben nicht berechnet.

Es wird allerdings nur eine Frage der Zeit sein, bis das Sprachmodell erkennt, das eine mathematische Verarbeitung erforderlich ist und auf andere Rechenprogramme oder spezialisierte Mathematik-KIs zugreifen wird. Trotzdem ist es erstaunlich, dass man mit Hilfe von ChatGPT komplexe Aufgaben und Fallstudien lösen kann, aber simple mathematische Berechnungen die Grenzen des Modells aufzeigen.

Ethik – Was darf und soll KI tun?

Mit der zunehmenden Verbreitung von KI-Systemen in alle Lebensbereiche müssen auch grundsätzliche ethische Fragestellungen beleuchtet werden. Der Deutsche Ethikrat hat sich als unabhängiges Sachverständigengremium kürzlich intensiv damit beschäftigt, welche Regeln beim Einsatz von KI gelten sollen. Im März 2023 wurde ein 290 Seiten starker Bericht unter dem Titel „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ veröffentlicht, der Chancen und Risiken der KI-Entwicklung beleuchtet [57]. Der Ethikrat diskutiert das Thema ausgehend von den Kernthemen Intelligenz, Vernunft und Verantwortung. Er sieht Chancen beim Einsatz von Künstlichen Intelligenzen in den hauptsächlich untersuchten Feldern Medizin, Bildung, öffentliche Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung. Die Rolle von KI sollte aber auf die eines Erfüllungsgehilfen und Assistenten beschränkt werden. Die ethische, moralische Verantwortung und Bewertung des Handelns von KI-Systeme sollten jedoch beim Menschen selbst verbleiben. Die Maschine darf den Menschen nicht ersetzen, sondern soll die menschliche Entfaltung erweitern.

Der Ethikrat mahnt Transparenz und die Berücksichtigung sozialer Gerechtigkeit, von Machtverhältnissen, Pluralismus und freier Meinungsbildung an. Auch müssen Schwachstellen und negative Konsequenzen vom KI-Einsatz erkannt und beseitigt werden. Auch Regulierung hält der Ethikrat für denkbar, setzt aber ausdrücklich auch auf Selbstregulierung durch die Industrie, wobei freiwillig erlassene Codizes oder Leitlinien zum ethischen Umgang mit KI herangezogen werden.

Interessanterweise gibt es auch mahnende Worte aus den Kreisen der Entwickler von KI-Systemen selbst. So forderte im März 2023 eine Petition mit namenhaften Unterstützern wie Elon Musk – übrigens Mitgründer von OpenAI – und Steve Wozniak eine sechsmonatige KI-Trainingspause [58]. Hintergrund ist das aktuelle, ungezügelte Wettrennen um die beste Künstliche Intelligenz. Die Unterzeichner warnen explizit vor den Risiken einer KI, die menschliche Intelligenz erreicht oder gar übertrifft. Denn bei GPT-4 zeigen sich schon erste Anzeichen einer Artificial General Intelligence, weshalb man für die Entwicklung darüberhinausgehender Systeme Regeln schaffen sollte [59]. Die Entwicklung solcher Systeme sollte sorgfältig geplant und überwacht werden, damit die Menschheit von den Potenzialen solcher KI-Systeme profitieren kann. Die KI-Trainingspause soll dazu genutzt werden, in Zusammenarbeit mit politischen Entscheidungsträgern, die notwendigen Sicherheitsprotokolle, Prüfverfahren und robuste Steuerungssysteme zu entwickeln.

In Anbetracht der möglichen Auswirkungen der KI-Entwicklung auf eine Vielzahl von Lebens- und Arbeitsbereichen wird es zukünftig zur Ausarbeitung eines Gestaltungsrahmens für die Entwicklung und die Anwendung von KI-Systemen kommen. Es gibt bereits erste Prinzipien, die für die Entwicklung einer „responsible/trusted/ethical AI“ gelten. Dazu gehören z. B. bei Microsoft als einen der aktuellen Treiber von ChatGPT [60]: Verantwortlichkeit, Einbeziehung aller, Zuverlässigkeit und Sicherheit, Fairness, Transparenz sowie Datenschutz und Sicherheit. Auch ERGO agiert im Rahmen selbst auferlegter Leitplanken zum Umgang mit Künstlicher Intelligenz.

Der Schutz persönlicher Daten und von Geschäftsgeheimnissen ist ein hohes Gut

Wer heute Daten in ChatGPT eingibt, erlaubt OpenAI die Daten für das Training weiterer Sprachmodelle heranziehen zu dürfen [17]. Das gilt sowohl für die kostenlose als auch für die ChatGPT Plus Version. Daraus ergeben sich selbstverständlich Fragen des Datenschutzes und der Geheimhaltung. Denn die eingegebenen Daten könnten theoretisch in einer neuen Sprachmodellvariante extrahierbar werden, d. h. Nutzer zukünftiger Sprachmodelle könnten auf die personenbezogenen Daten zugreifen. Die Eingabe personenbezogene Daten der Nutzer selbst und von Dritten sollte deshalb absolut vermieden werden. Das betrifft beispielsweise die Eingabe von Namen, E-Mails, Adressen, Zahlungsdaten und besonders schützenswerter Daten wie Gesundheitsdaten, in ChatGPT. Dasselbe gilt für Daten, die der Geheimhaltung unterliegen. Dieser Aspekt ist gerade im Kontext der beruflichen Nutzung bei Geschäftsgeheimnissen zu berücksichtigen. Zwei aktuelle Beispiele verdeutlichen eindringlich, wie wichtig die Einhaltung dieser Regeln ist:

(1) Mitarbeiter von Samsung hatten ChatGPT unter anderem dafür eingesetzt, Gesprächsnotizen in Texte für Präsentation umzuwandeln und Testsequenzen für die Identifizierung von Fehlern in einem neuen Chip zu optimieren. Dabei wurden unternehmensinterne, geheime Daten auf Server von OpenAI übertragen, die diese Informationen gemäß ihrer Nutzungsbedingungen nun verwenden dürfen [61].

(2) Datenschützer haben im April 2023 ein Verbot von ChatGPT in Italien erwirkt [62]. Hintergrund ist ein Sicherheitsleck bei OpenAI, welches den Zugriff auf Nutzerprompts und Zahlungsdaten ermöglichte. Neben dem konkreten Anlass bemängelte man grundlegende Verstöße gegen die europäische Datenschutz-Grundverordnung (DSGVO, englisch: GDPR, General Data Protection Regulation) [63]. Die Datenschützer sehen keine Grundlage für die massenhafte Erhebung und Speicherung von persönlichen Daten, welche für das Training der Sprachmodell herangezogen werden. Das Einsammeln von Trainingsdaten sei für den Betrieb der ChatGPT-Plattform nicht notwendig und damit unzulässig (Stichwort: Erforderlichkeitsprinzip). Zudem wird das Alter der Nutzer bei der Registrierung nicht erfasst, welches Minderjährigen potenziell Zugriff zu ungeeigneten Inhalten ermöglicht. Dazu steht auch eine Übermittlung der erhobenen Daten in ein Drittland im Raum, da OpenAI ein US-amerikanisches Unternehmen ist.

OpenAI ist der Überzeugung, dass sie den Anforderungen der DSGVO entsprechen und hofft, dass ChatGPT bald wieder in Italien verfügbar ist [64]. Wer seine Daten heute schon besser geschützt sehen will, kann z. B. über die technische Schnittstelle (API) auf die Sprachmodelle zugreifen [3]. Hier wird die Nutzung der Daten für das Training schon ausgeschlossen. Wer einen Schritt weiter geht, kann die Sprachmodelle auch komplett selbst auf einer unternehmenseigenen bzw. den DSGVO- und IT-Sicherheits-Anforderungen entsprechenden IT-Infrastruktur betreiben [23].

Allerdings bleiben noch einige DSGVO-relevante Fragen unbeantwortet. Enthalten die Daten zum Beispiel Kundeninformationen, müssen das Auskunftsrecht der betreffenden Person, das Recht auf Berichtigung sowie das Recht auf Lösung beachtet werden [63]. Technisch lässt sich das bei Sprachmodellen aktuell nur auf Ebene der allgemeinen Trainingsdaten, Embedding-Daten oder der Fine-Tuning-Daten bewerkstelligen. Hier müssen Daten zu Auskunftszwecken bereitgehalten und gegebenenfalls notwendige Korrekturen und Löschungen vorgenommen werden, damit nur geeignete Daten von den Sprachmodellen verarbeitet werden.

Eine klare Empfehlung: Der Mensch bleibt in der Verantwortung

Noch im Dezember 2021 schrieb Sam Altman der OpenAI CEO auf Twitter [65]:
„ChatGPT ist unglaublich begrenzt, aber gut genug in einigen Dingen, um einen irreführenden Eindruck von Größe zu vermitteln. Es ist ein Fehler, sich im Moment für irgendetwas Wichtiges darauf zu verlassen. Es ist eine Vorschau auf den Fortschritt; wir müssen noch viel an der Robustheit und Wahrhaftigkeit arbeiten."

Daraus ergibt sich die Notwendigkeit über die Chancen und Risiken von Sprachmodellen und Anwendungen wie ChatGPT aufzuklären. Auch im April 2023 gilt, dass Sprachmodelle noch nicht ohne die notwendige menschliche Kontrolle verantwortungsbewusst eingesetzt werden können. Im Zuge der GPT-4 Einführung schiebt der OpenAI CEO [66]:
„[ChatGPT] ist immer noch fehlerhaft, immer noch begrenzt, und es scheint immer noch beeindruckender zu sein, wenn man es zum ersten Mal benutzt, als es ist, wenn man mehr Zeit damit verbringt"

Dementsprechend sollte man sich an folgende Anweisungen halten, die OpenAI seinen Verwendern in den Nutzungsbedingungen mitteilt [17]:
„Aufgrund des probabilistischen Charakters des maschinellen Lernens kann die Nutzung unserer Dienste in manchen Situationen zu fehlerhaften Ergebnissen führen, die reale Personen, Orte oder Fakten nicht korrekt wiedergeben. Sie sollten die Genauigkeit der Ergebnisse je nach Ihrem Anwendungsfall bewerten, auch durch eine menschliche Überprüfung der Ergebnisse."

Der Einsatz von Sprachmodellen in der Versicherungsbranche

Voraussetzungen für den Einsatz

Bei ERGO ist man sich der zahlreichen Vorteile großer Sprachmodelle wie GPT-4 als auch der allgemeinen Herausforderungen bewusst. Es gibt jedoch auch eine Reihe unternehmensspezifischer Herausforderungen, die beim Einsatz von Sprachmodellen zu beachten sind. Nachfolgend diskutieren wir die Herausforderungen im Hinblick auf den Einsatz innerhalb einer Versicherung. Diese werden in ähnlicher Form auch in anderen Branchen auftreten.

Anpassung an branchenspezifische Anforderungen

Die GPT-Sprachmodelle sind auf allgemeinen Daten trainiert und für den Einsatz bei einer Vielzahl von Aufgaben geeignet. Trotzdem müssen die Sprachmodelle an die spezifischen Anforderungen der Versicherungsbranche und die Versicherungsunternehmen angepasst werden. Hierbei sind vier Aspekte zu berücksichtigen:

(1) Sprachmodelle müssen die Eigenheiten einer Versicherungssprache lernen[67]. Die Versicherung verwendet z. B. den Begriff der Prämie anders als im üblichen Sprachgebrauch. Auch gibt es Wortschöpfungen wie die Totalentwendung und die naturbedingten Luftbewegungen, die in Zusammenhang mit den von Kunden verwendeten Begriffen (Autodiebstahl, Sturm), z. B. bei der Schadensmeldung, gebracht werden müssen.

(2) Sprachmodelle benötigen Zugriff auf unternehmensinternes Wissen. Dieses Wissen muss in geordneter Form und geeigneter Qualität vorliegen, um über das Training in Sprachmodelle einfließen zu können.

(3) Sprachmodelle sollten unternehmensinterne Anforderungen an Formulierungen und Sprachstile (Stichwort: Corporate Identity) erlernen und bei der Ausgabe berücksichtigen. Dies ist vor allem im Hinblick auf eine konsistente Kommunikation ein wichtiger Aspekt, um einen effektiven und effizienten Einsatz zu gewährleisten. Dazu müssen bei der Entwicklung und Implementierung von Modellen weitere Anstrengungen unternommen werden.

(4) Sprachmodelle müssen regulatorische Vorgaben wie die Dokumentationspflichten erfüllen, die z. B. bei einer Versicherungsberatung erforderlich sind (Stichworte: Beratungsprotokoll, Falschberatung)[21]. Hier müssen die Möglichkeiten und Grenzen des Einsatzes von Sprachmodellen eruiert werden.

Datenschutz und Datensicherheit

Die Einhaltung von gesetzlichen Vorschriften zum Datenschutz und der IT-Sicherheit sind von entscheidender Bedeutung. Große Sprachmodelle erfordern große Datenmengen zum Training, die zum Teil sensible Daten enthalten können. In der Versicherungsbranche sind diese Daten oft vertraulich und sicherheitskritisch (z. B. Geo-Lokalisation oder Gesundheitsdaten zur Risikobestimmung). Es ist zu prüfen, welche Daten aus dem Versicherungskontext überhaupt zum Training von Sprachmodellen eingesetzt werden dürfen. Hier muss beispielsweise der Zweck der Datenerhebung beachtet werden. Insgesamt muss sichergestellt werden, dass die Modelle die geltenden Datenschutzgesetze und -vorschriften strikt einhalten. Die Daten müssen gemäß der Vor- und Freigaben genutzt werden, angemessen geschützt sein und das Risiko von Datenlecks minimiert werden. Dies gilt insbesondere dann, wenn die Integration in die IT-Infrastruktur vorgenommen wird und Prozesse unter Zugriff auf Kundendaten (teil)automatisiert werden.

Skalierbarkeit und Infrastruktur

Um große Sprachmodelle wie GPT-4 betreiben zu können, sind eine erhebliche Rechenleistung und Speicherkapazität erforderlich. Die Implementierung von Sprachmodellen bei gleichzeitiger Integration in bestehende IT-Systeme kann eine Herausforderung darstellen. Gleiches gilt für die Skalierung. Dies erfordert eine sorgfältige Planung und geeignete Investitionen in die IT-Infrastruktur.

Kosten-Nutzen-Abwägung

Die Entwicklung, Anpassung und Implementierung großer Sprachmodelle kann hohe Kosten verursachen, insbeson-

dere wenn ein kontinuierliches Monitoring, eine regelmäßige Aktualisierung und die Wartung erforderlich sind. Gleiches gilt für Lizenzkosten bei der Nutzung externer Anbieter. Schwer einschätzbaren Kosten stehen schwer einschätzbare positive Effekte gegenüber. Das Kosten-Nutzenverhältnis ist ein wichtiger zu berücksichtigender Aspekt, denn mit der Implementierung von Sprachmodellen sollte auch ein positiver Return-on-Investment (ROI) einhergehen.

Bedienoberflächen und Nutzerschnittstellen

Die effektive Einbindung von Sprachmodellen in die Arbeitsabläufe von Versicherungsmitarbeitern ist unabdingbar für den Erfolg und gleichzeitig eine große Herausforderung. Die Modelle müssen die Arbeitsweise der Mitarbeiter bei der Nutzung einer Vielzahl diverser Aufgaben unter Zuhilfenahme verschiedenster fachabteilungsspezifischer Tools unterstützen.

Aufklärung und Akzeptanz

Sprachmodelle können die Arbeitsaufgaben und -abläufe von Mitarbeiterinnen und Mitarbeitern maßgeblich verändern. Es ist zu vermuten, dass durch Sprachmodelle eine Reihe von Arbeitsabläufen teilautomatisiert werden. Hier muss die notwendige Aufklärungsarbeit geleistet werden, um das Verständnis für und die Akzeptanz von neuen Tools zu fördern. Beim Einsatz von Robotic Process Automation (RPA) und Phonebots zeigt sich bereits, dass Mitarbeiterinnen und Mitarbeiter den Einsatz von Robotern und Algorithmen sehr wertschätzen[68]. Grundbedingung ist, dass sie transparent und frühzeitig in die Entwicklung eingebunden werden. Die Maschine wird dann als Unterstützerin wahrgenommen, die dabei hilft, repetitive Aufgaben zu übernehmen und Platz für das Erledigen von Aufgaben schafft, die menschliches Wissen, Kreativität und Empathie erfordern.

Ausbildung und Fachkompetenz

Sowohl die Arbeit an und mit Sprachmodellen erfordert spezielle Kenntnisse und Fähigkeiten. Es wird notwendig sein, geeignetes Fachpersonal (z. B. Data Engineers, Data Scientists, Machine-Learning-, NLP- oder KI-Experten) einzustellen und zu entwickeln, um die wachsenden Anforderungen an die Implementierung von Sprachmodellen zu bewältigen. Des Weiteren müssen Mitarbeiter ausgebildet und befähigt werden, die Sprachmodelle optimal nutzen zu können. Die erforderlichen Fach- und Anwendungskompetenzen sind kurzfristig aufzubauen. Gerade bei den Fachkompetenzen ist zukünftig, mit steigender Nachfrage über alle Branchen hinweg, ein Wettbewerb, um die besten Talente zu erwarten.

Verantwortlichkeit und Interpretierbarkeit

Die Empfehlungen von großen Sprachmodellen sind zum Teil schwer nachvollziehbar, insbesondere bei komplexen Modellen wie GPT-4. Transparenz und nachvollziehbare Entscheidungskriterien sind z. B. bei der Risikobewertung und Preisgestaltung enorm wichtig. Hier müssen Anstrengungen unternommen werden, um den Anforderungen unserer Branche gerecht zu werden. In vielen Fällen wird der Mensch die Vorschläge der Sprachmodelle kritisch beurteilen, adaptieren und final bestätigen müssen. Dies gilt insbesondere im direkten Kundendialog, wo eine Falschberatung durch ein Sprachmodell vermieden werden muss.

Verzerrungen, Fairness und ethische Fragestellungen

Wie beschrieben, spiegeln Sprachmodelle mögliche Verzerrungen aus ihren Trainingsdaten wider. In der Versicherungsbranche könnte dies beispielsweise zu diskriminierenden Entscheidungen bei der Risikobewertung, Preisgestaltung, Schadenbearbeitung, Betrugserkennung oder Kundenansprache führen. Die Modelle müssen kontinuierlich auf Verzerrungen überprüft und entsprechende Datenbereinigungen oder Korrekturen vorgenommen werden. Der Einsatz von großen Sprachmodellen kann also ethische Fragen im Zusammenhang mit Urheberrechten, geistigem Eigentum oder moralischen Verpflichtungen gegenüber Kunden aufwerfen. Es ist wichtig, diese Fragen sorgfältig zu prüfen und entsprechende Richtlinien und Verfahren zu entwickeln, um potenzielle Risiken zu minimieren.

Um die Herausforderungen erfolgreich zu bewältigen, ist es wichtig, ein umfassendes Verständnis der technologischen, organisatorischen und regulatorischen Aspekte des Einsatzes großer Sprachmodelle innerhalb von Versicherungsunternehmen zu entwickeln. Eine proaktive und vorausschauende Herangehensweise ist entscheidend, um sicherzustellen, dass die Modelle einen positiven Einfluss auf das Versicherungsunternehmen, seine Mitarbeiter und Kunden haben wird und ihre potenziellen Risiken gemindert werden. Hierbei wird es kurzfristig zu einem schrittweisen Lernprozess über einzelne Pilotprojekte kommen.

Wo Sprachmodelle bei ERGO heute schon im Einsatz sind und erprobt werden

ERGO strebt an, bis 2025 digital führend in Deutschland und den internationalen Kernmärkten zu sein. Dem Anspruch entsprechend beschäftigt sich ERGO intensiv mit neusten technologischen Entwicklungen, wie bei großen Sprachmodellen und darauf basierenden Anwendungen.

ERGO verfügt über eine große Expertise im Einsatz von KI im Allgemeinen und von Sprachtechnologien im Speziellen. So werden in der ERGO Advanced-Analytics- und KI-Einheit bereits Sprachmodelle in diversen Use Cases genutzt. Beispielhaft sei hier die Dokumentenklassifikation im Eingangsmanagement genannt.

Die ERGO Voice Unit gestaltet bereits seit 2018 Sprachassistenten auf Basis von Conversational AI für den Kundenservice. Die Voice Unit arbeitet zudem gerade an dedizierten Projekten, um GPT-Sprachmodelle in Phonebots einzubinden und den Einsatz von großen Sprachmodellen innerhalb von ERGO voranzutreiben. Bei Phonebots ergeben sich beispielsweise Potenziale bei der Generierung von Trainingsdaten oder der Generierung von alternativen Bot-Antworten. Im Bereich der regelbasierten Chatbots ändern große Sprachmodelle künftig die Art und Weise, wie Kunden generische Informationen beziehen. Auf Sprachmodellen basierende Anwendungen können hier beispielsweise aus unterschiedlichen Quellen Antworten aggregieren, die sie Nutzern schnell und unkompliziert auf einen Blick zur Verfügung stellen. Es wird kein Suchen mehr auf mehreren Webseiten nötig sein und auch die Fragen müssen dem Chatbot nicht mehr vorher mittels Baumdiagramm vordefiniert und beantwortet werden.

Neue Technologien wie ChatGPT werden des Weiteren im ERGO Innovation Lab auf ihre Potenziale und Risiken hin geprüft. Ein Arbeitsergebnis dieser Bestrebungen ist das vorliegende Whitepaper.

Vorteile durch große Sprachmodelle für die Versicherungswirtschaft werden vor allem in folgenden Anwendungsfällen gesehen:

- Beim internen Wissensmanagement für Mitarbeiterinnen und Mitarbeiter.

- Bei der Unterstützung von Mitarbeiterinnen und Mitarbeitern bei allgemeinen Schreibaufgaben wie Kommunikationsmaterialien, Meeting Notes, Reports, E-Mails.

- Bei der Erzeugung von Text-Bausteinen mit spezialisierten Anforderungen (z. B. Programmierung, Bedingungswerke, Rechtstexte).

Grundsätzlich ist zu klären, ob solche Anwendungsfälle durch Software-Programme einer standardisierten Office Suite wie z. B. Microsoft Office 365 oder Google Workspaces bearbeitbar sind oder eigene Anwendungen entwickelt und veröffentlicht werden müssen[69–71]. Auch hier werden einige Pilotprojekte notwendig sein, da die Verfügbarkeit von neuen Tools (z. B. Microsoft CoPilot) noch nicht vollumfänglich gegeben ist und auch die Preisgestaltung sowie die Fähigkeiten noch nicht bewertet werden können. Auch wird die Rolle von Startups, welche Sprachmodell-basierte Lösungen für bestimmte Use-Cases anbieten werden, noch zu bestimmen sein.

Bei einem offensichtlichen Einsatzbereich im Kundendialog ist nach Einschätzung der ERGO-Experten Vorsicht geboten. Im direkten Kundenkontakt erscheint ein autonomer Einsatz von Sprachmodellen aufgrund der beschriebenen Herausforderungen wie Halluzinationen aktuell nur begrenzt möglich[67]. Hier wären kurzfristig zwei Ansätze zu präferieren:

(1) Die Kombination von Sprachmodellen mit regelbasierten Sprachassistenten, wie er von der ERGO Voice Unit erforscht wird. Sprachmodelle erkennen z. B. die Intention von Kunden im Dialog, geben diese Information aber an einen regelbasierten Sprachassistenten/Chatbot wieder, der die Beantwortung der Antworten in einem vorgegebenen Rahmen übernimmt. Hier liegen die Grenzen jedoch weiterhin in den definierten Regeln und Antwortoptionen. Es ist zu erwarten, dass eine solche Anwendung bei fehlenden Informationen häufig keine zufriedenstellende Antwort geben können wird.

(2) Ein Vorschlagswesen durch die KI, bei der Mitarbeiterinnen und Mitarbeiter die Textvorschläge von Sprachassistenten prüfen, ändern und absenden können[67]. Die sogenannten „Human-in-the-Loop“-Ansätze kommen auch bei klassischen Chatbots bei ERGO zum Einsatz und stellen sicher, dass die Antworten immer plausibel und korrekt gewählt sind. Zudem sammelt man über diesen Ansatz potenziell Daten für das Training zukünftiger Sprachmodelle mit Versicherungsbezug.

Der Wettlauf um die besten Sprachmodelle und Anwendungen ist in vollem Gange

Ist ChatGPT der iPhone Moment für KI?

In Fachkreisen wird gemunkelt, dass OpenAI von dem starken Nutzerinteresse und Erfolg von ChatGPT überrascht wurde. Zumindest die Data-Analytics-Experten der ERGO waren über die Nutzer- und Pressereaktionen verwundert, denn die Fähigkeiten von ChatGPT kannten sie bereits seit über einem Jahr vom Sprachmodell GPT-3. Seitdem sich der Interessentenkreis erweitert hat, hört man immer wieder, dass wir mit ChatGPT den nächsten iPhone-Moment erleben [18, 72]. Damals macht das erste iPhone von Apple das Smartphone alltagstauglich und einer breiten Masse an Endnutzern zugänglich. Zugleich entstand ein Wettbewerb um die beste Hardware, die besten Applikationen und die beste Nutzerfreundlichkeit. Ähnliches erleben wir heute mit generativer KI und Sprachmodellen.

OpenAI macht große Fortschritte

Seitdem OpenAI im November 2022 ChatGPT veröffentlicht hat, kam es zu einer Reihe von Updates von ChatGPT und den verwendeten GPT-Sprachprogrammen. Diese Updates beinhalteten von Dezember 2022 bis zum Februar 2023 allgemeinen Leistungsverbesserungen der Server, Modellverbesserungen mit erhöhter Faktentreue, verbesserten mathematischen Fähigkeiten sowie die erhöhte Geschwindigkeit bei der Ausgabe [73].

Mit dem Release von GPT-4 am 14. März 2023 kam es zu verbesserten Fähigkeiten beim logischen Denken, der Verarbeitung komplexerer Anweisungen und der Kreativität. GPT-4 übertrifft sein Vorgängermodell bei einer Reihe von standardisierten Tests, welche für Menschen und Machine-Learning-Modellen erdacht worden sind. Gleichzeitig wurde das Sprachmodell im Vergleich mit GPT-3.5 sicherer: Die Wahrscheinlichkeit, auf unzulässige Anfragen zu antworten, sank um 82%, während die Wahrscheinlichkeit, sachliche Antworten zu geben, um 40% erhöht wurden. Eine wichtige Neuerung in GPT-4 ist die Möglichkeit, Bilder zu verarbeiten, d. h. neben reinem Text akzeptiert das Modell nun auch Bilder in der Eingabe. GPT-4 ist damit multimodal. Die Funktion ist bis dato eingeschränkt testbar, aber die Vorführungen zeigen, dass GPT-4 Bildinhalte beschreiben kann und Website-Code auf Basis von Zeichnungen erstellen kann. Dies erweitert die Anwendbarkeit von Sprachmodellen enorm [14]. Seit dem 23. März 2023 gibt es die experimentelle Unterstützung von KI-Plugins in ChatGPT für eine kleine Nutzer- und Entwicklergruppe [49]. Die Plugins erlauben ChatGPT auf aktuelle Daten zuzugreifen und Services von Drittanbietern zu nutzen. Erste Plugins sind zum Beispiel für Expedia, Shopify, Slack und Wolfram verfügbar.

Microsoft nutzt die Chance zum Innovieren

Microsoft kündigte als strategischer Partner und Investor von OpenAI eine Reihe von KI-Anwendungen für das Jahr 2023 an. Nachdem bereits 2019 eine Milliarde US-Dollar investiert wurde, schoss Microsoft dieses Jahr weitere zehn Mrd. US-Dollar nach und sicherte sich im Gegenzug Verwertungsrechte an den KI-Produkten [70]. GPT-Sprachmodelle findet man seitdem in der Suchmaschine Bing und in den Office-Anwendungen wie Powerpoint, Excel, Word und Teams [48]. Unter dem Label Microsoft CoPilot werden zukünftig Büroabläufe automatisiert [70, 71]. Auch das DSGVO-konforme Betreiben von Sprachmodellen auf der Microsoft Azure-Cloud ist bereits möglich [23].

Google reagiert bei KI noch zögerlich

Bei Google lösten ChatGPT und die Microsoft-Bing-Integration einen „roten Alarm“ aus [74]. Man sah die Google-Suche als Einstiegspunkt für das Internet in Gefahr und fragte sich, warum Internetnutzer noch nach Antworten googlen und sich durch die verlinken Seiten klicken sollten, wenn ChatGPT doch oft direkt eine zufriedenstellende Antwort liefert. Damit wäre das Geschäftsmodell von Google ernsthaft bedroht.

Die erste Vorstellung des ChatGPT-Gegenstücks namens Bard im Februar 2023 enttäuschte, auch weil Bard kleine Fehler unterliefen und Google eher vage bei den Ankündigungen blieb [10]. Google-Mitarbeiter kritisierten die hektisch herbeigeführte Vorstellung vom Chatbot, welcher auf einer abgespeckten Version des Sprachmodells „LaMDA“ (Language Model für Dialogue Application) basiert [75]. Google scheint insgesamt eine höhere Zurückhaltung zu zeigen und legt nach eigenen Aussagen viel Wert auf qualitativ hochwertige sowie bodenständige Antworten und

Sicherheit [76]. So wurde Bard im März 2023 einer vertrauenswürdigen Beta-Testgruppe zur Verfügung gestellt – zunächst nur in UK und USA [77]. Der Weg ist jedoch ähnlich wie bei Microsoft: In Google Workspaces mit den Anwendungen Gmail, Google Sheets und Docs erhalten Nutzer zukünftig Unterstützung durch KI-Tools, z. B. beim Schreiben von E-Mails oder der Zusammenfassung von Texten. Auch werden KI-Tools in der Google Cloud für Entwickler bereitgestellt [69].

Die Entwicklung ist noch nicht am Ende

Neben dem Wettlauf um die besten Anwendungen, entwickelt sich auch die Forschung in rasantem Tempo weiter. Seit Einführung der Transformer-Modellarchitektur 2017 gibt es ein exponentielles Wachstum des Forschungsoutputs im Bereich Machine Learning und Artificial Intelligence [78]. Der Wettlauf zwischen OpenAI/Microsoft und Google wird für eine weitere Beschleunigung sorgen. Auch Unternehmen wie Meta haben eigene generative KIs in der Pipeline oder bereits vorgestellt [79].

Sie werden beim Wettlauf um die besten KI-Anwendungen eine Rolle spielen, ebenso wie auf Branchen oder Use Cases spezialisierte Startups. Für Unternehmen ergibt sich die Notwendigkeit, die Entwicklung zu beobachten, erste Erfahrungen zu sammeln und sich zum geeigneten Zeitpunkt für externe oder interne Lösungen zu entscheiden. Es werden auf jeden Fall spannende Zeiten mit hohem Disruptionspotenzial.

Autoren

Jens Sievert
Innovation Manager, ERGO Innovation Lab,
ERGO Digital Ventures AG

Dr. Sebastian Kaiser
Head of Machine Learning,
Artificial Intelligence Advanced Analytics

Nicolas Konnerth
Head of Conversational AI,
ERGO Digital Ventures AG

Quellenverzeichnis

1 OpenAI, „OpenAI About“. https://openai.com/about (zugegriffen 20. April 2023).
2 OpenAI, „OpenAI Research“. https://openai.com/research (zugegriffen 20. April 2023).
3 OpenAI, „OpenAI Platform“. https://platform.openai.com/ (zugegriffen 20. April 2023).
4 OpenAI, „OpenAI ChatGPT“. chat.openai.com (zugegriffen 20. April 2023).
5 K. Hu, „ChatGPT sets record for fastest-growing user base – analyst note“, Reuters, 2. Februar 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (zugegriffen 20. April 2023).
6 D. F. Carr, „ChatGPT Grew Another 55.8% in March, Overtaking Bing and DuckDuckGo“, Similarweb, 3. April 2023. www.similarweb.com/blog/insights/ai-news/chatgpt-bing-duckduckgo/ (zugegriffen 20. April 2023).
7 F. K. Akın, „Awesome ChatGPT Prompts“, GitHub, 10. Dezember 2022. https://github.com/f/awesome-chatgpt-prompts/ (zugegriffen 20. April 2023).
8 G. Venuto, „LLM failure archive (ChatGPT and beyond)“, GitHub, 3. Januar 2023. https://github.com/giuven95/chatgpt-failures (zugegriffen 20. April 2023).
9 J. Zhang und S. Rai, „AI Presents a Nearly Level Playing Field for Startups and Big Tech in Asia“, Bloomberg, 1. März 2023. https://www.bloomberg.com/news/newsletters/2023-03-01/ai-chatbot-race-startups-and-big-tech-giants-compete-for-the-best (zuge-griffen 20. April 2023).
10 J. Heinrich, „Google: KI-Fehlstart, Aktie stürzt um 100 Milliarden Dollar ab“, W&V, 9. Februar 2023. https://www.wuv.de/Themen/Performance-Analytics/Google-KI-Fehlstart-Aktie-stuerzt-um-100-Milliarden-Dollar-ab (zugegriffen 24. April 2023).
11 C. Terwiesch, „Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course,“ Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania“, 2023. Zugegriffen: 20. April 2023. [Online]. Verfügbar unter: https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf
12 T. H. Kung u. a., „Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models“, PLOS Digital Health, Bd. 2, Nr. 2, S. e0000198, Feb. 2023, doi: 10.1371/JOURNAL.PDIG.0000198.
13 J. H. Choi, K. E. Hickman, A. Monahan, und D. B. Schwarcz, „ChatGPT Goes to Law School“, SSRN Electronic Journal, Jan. 2023, doi: 10.2139/SSRN.4335905.
14 OpenAI, „OpenAI GPT-4“. https://openai.com/research/gpt-4 (zugegriffen 20. April 2023).
15 S. Noy und W. Zhang, „Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence“, März 2023.
16 N. Popli, „How to Get a Six-Figure Job as an AI Prompt Engineer “, Time, 14. April 2023. https://time.com/6272103/ai-prompt-engineer-job/ (zugegriffen 20. April 2023).
17 OpenAI, „OpenAI Terms of use“. https://openai.com/policies/terms-of-use (zugegriffen 20. April 2023).
18 F. Hegemann, „Generative KI: ChatGPT & Co. erleben ihren iPhone-Moment“, //next by ERGO, 2023. https://next.ergo.com/de/KI-Robotics/2023/Was-ist-Generative-KI-AI-ChatGPT-Hype-Technologie-Trend.html (zugegriffen 20. April 2023).
19 OpenAI, „OpenAI API – Embeddings“. https://platform.openai.com/docs/guides/embeddings/what-are-embeddings (zugegriffen 20. April 2023).
20 OpenAI, „OpenAI API-Fine-tuning“. https://platform.openai.com/docs/guides/fine-tuning (zugegriffen 20. April 2023).
21 „Welche Beratungspflichten hat der Versicherer gegenüber dem Versicherungsnehmer bei Vertragsabschluss und während der Laufzeit des Versicherungsvertrags? “, BaFin, 8. September 2021. https://www.bafin.de/SharedDocs/FAQs/DE/Verbraucher/Versicherung/VertraegeAbschliessen/04_beratung_grundsatz.html (zugegriffen 20. April 2023).
22 OpenAI, „OpenAI Usage policies“, 23. März 2023. https://openai.com/policies/usage-policies (zugegriffen 20. April 2023).
23 Microsoft Learn, „Dokumentation zu Azure OpenAI Service“, 2023. https://learn.microsoft.com/de-de/azure/cognitive-services/openai/ (zugegriffen 20. April 2023).
24 J. R. Firth, „A Synopsis of Linguistic Theory 1930-55.“, Studies in Linguistic Analysis: Special Volume of the Philological Society, 1957.
25 E. Roberts, „Natural Language Processing“, 2004. https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html (zugegriffen 24. April 2023).
26 C. Li, „OpenAI's GPT-3 Language Model: A Technical Overview“, Lambda, 3. Juni 2020. https://lambdalabs.com/blog/demystifying-gpt-3 (zugegriffen 20. April 2023).
27 T. B. Brown u. a., „Language models are few-shot learners“, in Advances in Neural Information Processing Systems, 2020.
28 J. Langston, „Microsoft announces new supercomputer, lays out vision for future AI work – Source“, Microsoft News, 19. Mai 2020. https://news.microsoft.com/source/features/ai/openai-azure-supercomputer (zugegriffen 20. April 2023).
29 C. Köver, „10 Milliarden für Start-up: Wofür braucht OpenAI so viel Geld?“, Netzpolitik.org, 25. Januar 2023. https://netzpolitik.org/2023/10-milliarden-fuer-start-up-wofuer-braucht-openai-so-viel-geld/ (zugegriffen 20. April 2023).
30 „What Is ChatGPT Doing ... and Why Does It Work?“, Stephen Wolfram Writings, 14. Februar 2023. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ (zugegriffen 24. April 2023).
31 „What are Neural Networks? “, IBM. https://www.ibm.com/topics/neural-networks (zugegriffen 24. April 2023).
32 M. Bischoff, „Wie man einem Computer das Sprechen beibringt“, Spektrum.de, 9. März 2023. https://www.spektrum.de/news/wie-funktionieren-sprachmodelle-wie-chatgpt/2115924 (zugegriffen 24. April 2023).
33 A. Vaswani u. a., „Attention is all you need“, in Advances in Neural Information Processing Systems, 2017.
34 „How do Transformers work?“, Hugging Face – NLP Course. https://huggingface.co/learn/nlp-course/chapter1/4?fw=pt (zugegriffen 23. April 2023).
35 A. Tamkin und D. Ganguli, „How Large Language Models Will Transform Science, Society, and AI“, Stanford University, Human-Centered Artificial Intelligence , 5. Februar 2021. https://hai.stanford.edu/news/how-large-language-models-will-transform-science-society-and-ai (zugegriffen 20. April 2023).
36 OpenAI, „Aligning language models to follow instructions“. https://openai.com/research/instruction-following (zugegriffen 20. April 2023).
37 OpenAI, „Learning from human preferences“. https://openai.com/research/learning-from-human-preferences (zugegriffen 20. April 2023).
38 T. Xu, „KI-Modelle: Es droht ein Mangel an Trainingsdaten “, heise online, 28. November 2022. https://www.heise.de/hintergrund/KI-Modelle-Es-droht-ein-Mangel-an-Trainingsdaten-meinen-Experten-7357898.html (zugegriffen 24. April 2023).
39 J. D'Souza, „A Catalog of Transformer Models“, ORKG, März 2023. https://orkg.org/comparison/R385006/ (zugegriffen 25. April 2023).
40 A. Tamkin, M. Brundage, J. C. †3, und D. Ganguli, „Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models“, Feb. 2021, Zugegriffen: 20. April 2023. [Online]. Verfügbar unter: https://arxiv.org/abs/2102.02503v1
41 E. M. Bender, T. Gebru, A. McMillan-Major, und S. Shmitchell, „On the dangers of stochastic parrots: Can language models be too big?“, in FAccT 2021 – Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021. doi: 10.1145/3442188.3445922.
42 E. Ferrara, „Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models“, Apr. 2023, Zugegriffen: 20. April 2023. [Online]. Verfügbar unter: https://arxiv.org/abs/2304.03738v2

43 S. Grüner, „Phishing und Malware: Europol warnt vor Missbrauch von ChatGPT “, Golem.de, 28. März 2023. https://www.golem.de/news/phishing-und-malware-europol-warnt-vor-missbrauch-von-chatgpt-2303-172998.html (zugegriffen 25. April 2023).

44 D. Alba, „ChatGPT, Open AI’s Chatbot, Is Spitting Out Biased, Sexist Results“, Bloomberg, 8. Dezember 2022. https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results (zugegriffen 25. April 2023).

45 S. Schier, „ChatGPT könnte zu mehr Versicherungsbetrug führen“, Handelsblatt, 20. Februar 2023. https://www.handelsblatt.com/finanzen/banken-versicherungen/versicherer/kuenstliche-intelligenz-chatgpt-koennte-zu-mehr-versicherungsbetrug-fuehren/28987726.html (zugegriffen 20. April 2023).

46 I. Mehta, „Elon Musk wants to develop TruthGPT, ‘a maximum truth-seeking AI’ “, TechCrunch, 18. April 2023. https://techcrunch.com/2023/04/18/elon-musk-wants-to-develop-truthgpt-a-maximum-truth-seeking-ai/ (zugegriffen 20. April 2023).

47 N. Reip, „KI: Künstliche Intelligenz – Eine Gefahr für die Medienkompetenz?“, Medienkompass.de, 30. Januar 2023. https://medienkompass.de/ki-eine-gefahr-fuer-die-medienkompetenz/ (zugegriffen 20. April 2023).

48 Y. Medhi, „Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web“, Official Microsoft Blog, 7. Februar 2023. https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/ (zugegriffen 20. April 2023).

49 OpenAI, „OpenAI API – Chat Plugins“. https://platform.openai.com/docs/plugins/introduction (zugegriffen 20. April 2023).

50 H.-T. Cheng, „LaMDA: Towards Safe, Grounded, and High-Quality Dialog Models for Everything“, Google AI Blog, 21. Januar 2022. https://ai.googleblog.com/2022/01/lamda-towards-safe-grounded-and-high.html (zugegriffen 24. April 2023).

51 „Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback“, Microsoft Research, Deep Learning Group, 7. März 2023. https://www.microsoft.com/en-us/research/group/deep-learning-group/articles/check-your-facts-and-try-again-improving-large-language-models-with-external-knowledge-and-automated-feedback/ (zugegriffen 24. April 2023).

52 „Originality.AI – Most Accurate AI Content Detector and Plagiarism Checker“. https://originality.ai/ (zugegriffen 20. April 2023).

53 „§ 2 UrhG – Einzelnorm“. https://www.gesetze-im-internet.de/urhg/__2.html (zugegriffen 20. April 2023).

54 V. PAPAKONSTANTINOU und P. DE HERT, „Refusing to award legal personality to AI: Why the European Parliament got it wrong – European Law Blog“, European Law Blog, 25. November 2020. https://europeanlawblog.eu/2020/11/25/refusing-to-award-legal-personality-to-ai-why-the-european-parliament-got-it-wrong/ (zugegriffen 20. April 2023).

55 „Copyright, Designs and Patents Act 1988“. https://www.legislation.gov.uk/ukpga/1988/48/section/9 (zugegriffen 20. April 2023).

56 S. Bordow, „Do the math: ChatGPT sometimes can’t, expert says“, ASU, Arizona State University News, 21. Februar 2023. https://news.asu.edu/20230221-discoveries-do-math-chatgpt-sometimes-cant-expert-says (zugegriffen 20. April 2023).

57 „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“, März 2023. Zugegriffen: 20. April 2023. [Online]. Verfügbar unter: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

58 „Pause Giant AI Experiments: An Open Letter – Future of Life Institute“, 22. März 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (zugegriffen 20. April 2023).

59 S. Bubeck u. a., „Sparks of Artificial General Intelligence: Early experiments with GPT-4“. 22. März 2023. Zugegriffen: 20. April 2023. [Online]. Verfügbar unter: https://www.microsoft.com/en-us/research/publication/sparks-of-artificial-general-intelligence-early-experiments-with-gpt-4/

60 „Responsible and trusted AI “, Microsoft Learn, 20. April 2023. https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/best-practices/trusted-ai (zugegriffen 20. April 2023).

61 K. Nordenbrock, „Datenleck bei Samsung: Ingenieure schicken vertrauliche Daten an ChatGPT“, t3n, 8. April 2023. https://t3n.de/news/samsung-semiconductor-daten-chatgpt-datenleck-1545913/ (zugegriffen 20. April 2023).

62 S. Lohmeier, „Datenschutzrisiken: Verbot von ChatGPT in Italien“, bakertilly, 4. April 2023. https://www.bakertilly.de/news/detail/verbot-von-chatgpt-in-italien-zeigt-datenschutzrisiken-auf.html (zugegriffen 20. April 2023).

63 intersoft consulting, „Datenschutz-Grundverordnung (DSGVO) – Finaler Text der DSGVO inklusive Erwägungsgründe“. https://dsgvo-gesetz.de/ (zugegriffen 20. April 2023).

64 S. Altman, „Tweet“, Twitter, 31. März 2023. https://twitter.com/sama/status/1641897800236687360 (zugegriffen 20. April 2023).

65 S. Altman, „Sam Altman auf Twitter: „ChatGPT is incredibly limited …“, Twitter, 11. Dezember 2022. https://twitter.com/sama/status/1601731295792414720?lang=de (zugegriffen 20. April 2023).

66 S. Altman, „Sam Altman auf Twitter: „here is GPT-4, our most capable …“, Twitter. https://twitter.com/sama/status/1635687853324902401 (zugegriffen 20. April 2023).

67 C.-A. Wallaert, R. Fehling, I. Karimi, und A. Gupta, „How Can ChatGPT Be Used in Insurance?“, LinkedIn, BCG on Insurance, 20. Februar 2023. https://www.linkedin.com/pulse/how-can-chatgpt-used-insurance-bcg-on-insurance/ (zugegriffen 20. April 2023).

68 „Whitepaper: So wird Robotic Process Automation (RPA) erfolgreich“, ERGO Group AG, 1. März 2023. https://next.ergo.com/de/KI-Robotics/2022/Whitepaper-Robot-Process-Automatisation-RPA-Robotics-Expertise-Unternehmenskultur-Kulturwandel-Erfolgsfaktoren-Digitalisierung-Mark-Klein (zugegriffen 24. April 2023).

69 T. Kurian, „New AI features and tools for Google Workspace, Cloud and developers“, Google Blog, 14. März 2023. https://blog.google/technology/ai/ai-developers-google-cloud-workspace/ (zugegriffen 24. April 2023).

70 „Introducing Microsoft 365 Copilot“, Microsoft 365 Blog, 16. März 2023. https://www.microsoft.com/en-us/microsoft-365/blog/2023/03/16/introducing-microsoft-365-copilot-a-whole-new-way-to-work/ (zugegriffen 20. April 2023).

71 „Microsoft Teams Premium mit KI-basierten Features ab sofort verfügbar“, Microsoft News, 6. Februar 2023. https://news.microsoft.com/de-at/microsoft-teams-premium-mit-ki-basierten-features-ab-sofort-verfugbar/ (zugegriffen 20. April 2023).

72 S. Scheuer, „Neuer iPhone-Moment: Chat-GPT und die Tech-Revolution“, Handelsblatt, 10. Februar 2023. https://www.handelsblatt.com/meinung/kommentare/kommentar-der-neue-iphone-moment-chatgpt-und-die-tech-revolution/28969310.html (zugegriffen 20. April 2023).

73 OpenAI, „ChatGPT – Release Notes“. https://help.openai.com/en/articles/6825453-chatgpt-release-notes (zugegriffen 20. April 2023).

74 N. Grant und C. Metz, „ChatGPT and Other Chat Bots Are a ‘Code Red’ for Google Search“, The New York Times, 21. Dezember 2022. https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html (zugegriffen 24. April 2023).

75 J. von Lindern und M. Laaff, „Chatbot Bard: So will Google zurückchatten“, ZEIT ONLINE. https://www.zeit.de/digital/2023-02/bard-google-chatbot-chat-gpt?utm_referrer=https%3A%2F%2Fwww.google.com%2F (zugegriffen 24. April 2023).

76 S. Pichai, „Ein wichtiger nächster Schritt auf unserer KI-Reise“, Google Blog, 6. Februar 2023. https://blog.google/intl/de-de/unternehmen/technologie/ein-wichtiger-naechster-schritt-auf-unserer-ki-reise/ (zugegriffen 24. April 2023).

77 J. Elias, „Google opens Bard A.I. for testing by users in U.S. and UK“, CNBC, 21. März 2023. https://www.cnbc.com/2023/03/21/google-opens-bard-ai-for-testing-by-users-in-us-and-uk.html (zugegriffen 24. April 2023).

78 M. Krenn u. a., „Predicting the Future of AI with AI: High-quality link prediction in an exponentially growing knowledge network“, Sep. 2022, Zugegriffen: 24. April 2023. [Online]. Verfügbar unter: http://arxiv.org/abs/2210.00881

79 „Introducing LLaMA: A foundational, 65-billion-parameter language model“, Meta AI, 24. Februar 2023. https://ai.facebook.com/blog/large-language-model-llama-meta-ai/ (zugegriffen 24. April 2023).