

Wie Sprachmodelle wie ChatGPT unsere Gesellschaft verändern werden

Ein Blick ins Klassenzimmer

Martha hat eine Englischhausaufgabe bekommen. Sie soll einen Text über die drei bekanntesten Sehenswürdigkeiten von New York schreiben. New York steht im Zentrum der aktuellen Unit im Englischbuch. Martha ist nicht sehr gut in Englisch. Aber die App auf ihrem Handy ist direkt an ChatGPT angedockt. ChatGPT ist eine künstliche Intelligenz, die auf möglichst natürliche Textausgabe optimiert ist. Martha tippt in ihre App: „Write an article about the three most famous sights in New York.“, denn das ist die Aufgabenstellung, die ihre Lehrerin ihr gegeben hat. Heraus kommt folgender Text (abgedruckt ist die unveränderte Ausgabe):

New York City is one of the world's most vibrant and exciting destinations. Known for its diverse culture, bustling streets, and iconic landmarks, the city is a hub of entertainment and commerce. From the bright lights of Times Square to the grandeur of the Statue of Liberty, New York offers a wealth of sights to explore. In this article, we will take a closer look at the three most famous sights in the city.

1. The Statue of Liberty

The Statue of Liberty is one of the most iconic landmarks in New York City. Located on Liberty Island, the statue was gifted to the United States by the people of France in 1886 as a symbol of freedom and democracy. The statue is 151 feet tall and is made of copper plates, which have turned green over time due to weathering.

Visitors can take a ferry to Liberty Island and climb to the top of the pedestal or even to the crown of the statue for panoramic views of the city skyline and harbor. The statue is also visible from Battery Park in Lower Manhattan, where visitors can take in the stunning view of the statue against the backdrop of the city.

1. The Empire State Building

The Empire State Building is a towering skyscraper located in Midtown Manhattan. It was completed in 1931 and was the world's tallest building until 1971. The building stands at 1,454 feet tall and has 102 floors. Visitors can take an elevator to the observation deck on the 86th floor for breathtaking views of the city skyline. On a clear day, visitors can see up to 80 miles in every direction.

The Empire State Building has been featured in numerous movies and is a popular tourist destination. The building's art deco architecture is a sight to behold, and the nighttime lighting display is an iconic part of the city skyline.

1. Central Park

Central Park is a vast green space located in the heart of Manhattan. The park covers 843 acres and is a popular destination for both locals and tourists. Visitors can explore the park on foot, by bike, or even by horse-drawn carriage. The park features numerous walking paths, gardens, lakes, and ponds, as well as famous landmarks such as the Central Park Zoo and the Bethesda Fountain.

In the winter, visitors can go ice-skating at the Wollman Rink, and in the summer, there are numerous concerts and events held in the park. Central Park is a welcome respite from the hustle and bustle of the city and provides a tranquil oasis in the heart of Manhattan.

In conclusion, New York City is home to many famous sights, but the Statue of Liberty, Empire State Building, and Central Park are three of the most iconic. These landmarks represent the city's history, culture, and natural beauty and are must-see destinations for anyone visiting the city. Whether you're a first-time visitor or a lifelong New Yorker, these sights never cease to amaze and inspire.

Martha meldet sich und liest ihren Text vor. Ihre Lehrkraft ist verwundert. Dieser Text entspricht überhaupt nicht Marthas bisher im Unterricht gezeigten Fähigkeiten. Sie hat den Verdacht, dass dieser Text von irgendwoher kopiert sein könnte – findet im Netz aber nichts. Martha erhält eine gute Note.

Hannes und Fred präsentieren ihren Text im Unterricht. Eigentlich lässt sich zum Inhalt wenig Schlechtes sagen, aber das Geschriebene erinnert dann doch mehr an Gesprochenes: Satzzeicheninflation am Satzende, kaum Groß- und Kleinschreibung und die Zeichensetzung ist optimierbar – oder gar nicht erst vorhanden. Hannes und Fred kopieren ihren Text daraufhin in <https://www.deepl.com/write>. Das ist ein KI-basierter Webdienst, der keine Daten sammelt und bis 2000 Zeichen kostenlos ohne Account nutzbar ist. DeepL ist eigentlich ein sehr gutes Übersetzungswerkzeug, kann in einer Beta-Version jetzt aber auch Texte stilistisch überarbeiten. Hannes und Fred schauen sich die vorgeschlagenen Änderungen an. Einiges gefällt ihnen, einiges scheint den Sinn ihres Textes zu stark zu verändern. Sie entscheiden sich dazu, einige Passagen zu ändern, weil sie durch die Änderungen wirklich besser klingen.

Diese beiden Beispiele zeigen, was heute für alle bereits mit sprachbasierter künstlicher Intelligenz möglich ist. Während das erste Beispiel in vielen Kollegien in deutschen Schulen für Aufruhr sorgt, wenn das Potential von Sprachmodellen dort überhaupt ankommt, zeigt das zweite Beispiel, dass Sprachmodelle gerade unsichere Schreiber sehr gut unterstützen könnten.

Die Auseinandersetzung über ChatGPT in sozialen Netzwerken

Zwischen diesen beiden Polen bewegen sich manchmal emotional sehr aufgeladene Diskussionen in sozialen Netzwerken. Die einen sehen bestehende Strukturen und Aufgabenformate in Gefahr, die andere experimentieren mutig mit den sich ergebenden neuen Möglichkeiten. Diese Experimente bewegen sich oft primär auf einer phänomenologischen Ebene: Im Kern geht es darum, die Eingaben („Prompts“) zu finden, die eine vorher erdachte Aufgabenstellung mit Hilfe eines Sprachmodells optimal lösen.

Natürlich bleibt die Auseinandersetzung vorerst phänomenologisch. Man gibt sich Tipps, welche Eingaben zu einem gewünschten Ergebnis führen und wie der ChatBot in z.B. eine andere Rolle einnehmen kann. Das klappt auf der Oberfläche. Die technischen Abläufe in einem Sprachmodell sind gar nicht so leicht zu verstehen.

Was ist ein Sprachmodell eigentlich?

Ein Sprachmodell ist eine sehr komplexe Software, die zur Klasse der künstlichen Intelligenz gezählt wird. Schon der Begriff „Intelligenz“ ist je nach Fachdisziplin nicht einfach zu definieren. Daher sprechen einige Fachleute lieber von „maschinellern Lernen“, was das Definitionsproblem jedoch lediglich auf den Begriff des Lernens verschiebt. Als Ausweg aus diesem Dilemma haben eine Reihe

von Autor:innen nicht nur aus dem Informatikumfeld folgende Arbeitsdefinition für künstliche Intelligenz vorgeschlagen:

„Künstliche Intelligenz bezeichnet die Fähigkeit von Computersystemen, auf sie zugeschnittene Aufgaben selbsttätig zu lösen, die aufgrund ihrer Komplexität bislang menschliche Fähigkeiten erforderten“¹

Angewendet: Sprachmodelle lösen nun selbstständig das Problem der Erstellung von Texten auf Basis von Eingaben, die Nutzer:innen vorgeben. Aber wie machen sie das?

In den Anfangszeiten von künstlicher Intelligenz waren die Aufgaben und Möglichkeiten von IT-Systemen noch etwas beschränkter. Ein sehr einfaches Beispiel sind die damaligen Obstwaagen in den Frischeabteilungen der Supermärkte. Der Kunde füllte eine Ware in einen Beutel, ging zur einer Waage, legte den Beutel auf und drückte eine Taste mit einem Symbol oder einer Nummer. Die Waage druckte dann einen Kassenzettel mit Strichcode aus. Es waren aber nicht alle Kunden ehrlich, so dass man auf die Idee kam, diese Waagen durch eine Kamera zu ergänzen, mit deren Hilfe die aufgelegte Ware erkannt werden konnte, z.B. anhand von Größe oder Farbe. Im Prinzip konnte man drei Fälle unterscheiden:

- a) Wenn der Tastendruck des Kunden zur erkannten Ware passte, war die Entscheidung eindeutig.
- b) Wenn der Tastendruck eindeutig nicht zur aufgelegten Ware passte, konnte ein alternativer Kassenzettel gedruckt werden.
- c) Wenn die Erkennung selbst uneindeutig war, d.h. zwei Produkte in Frage kommen würden, aber der Kunde z.B. die Taste mit dem teureren Produkt gedrückt hat, wurde natürlich die teurere Variante gewählt – gleichzeitig ließ sich diese Situation abspeichern, so dass sie für die Zukunft mit berücksichtigt werden konnte.

Diese Waage erzeugte also auf Basis von Eingaben und der Bilder der Kamera (Eingabeschicht) Daten, die sie verarbeitete (Verarbeitungsschicht). Bei der Verarbeitung flossen die Eingabedaten ebenso ein, wie gespeicherte Fälle aus einer Datenbank. Die Daten wurden untereinander gewichtet, mit einer Datenbank voller Standardfälle abgeglichen und dann eine Ausgabe in Form eines Kassenzettels erzeugt (Ausgabeschicht). Mit einer gewissen Wahrscheinlichkeit stimmten dann Kassenzettel und Produkt tatsächlich überein. Die Wahrscheinlichkeit der Übereinstimmung war umso größer, je mehr Standardfälle in der Datenbank der Waage erfasst wurden und je mehr Eingaben die „intelligente Waage“ bereits verarbeitet hatte. Softwareentwickler:innen konnten die Standardfälle in Form einer Datenbank vorgeben.

Sprachmodelle funktionieren nach einem ähnlichen Prinzip, nur dass die Standardfälle und Ausgaben ungleich komplexer sind als bei dem Beispiel der Waage. Zudem ist verhältnismäßig wenig darüber bekannt, mit welchen Daten die zugrundeliegenden Datenbanken gefüllt wurden. Letztlich gilt aber auch hier, dass die Ausgaben lediglich mit einer gewissen Wahrscheinlichkeit korrekt sind oder gar zur Anfrage des Nutzenden passen - daher verhalten sich Hannes und Fred deutlich smarter als Martha.

Es mag auf den ersten Blick erstaunen, dass Sprachmodelle Ausgaben in der momentanen Qualität liefern. Wir halten Sprache für ein komplexes System und würden einer Maschine deren Imitation zunächst nicht so recht zutrauen. Allerdings sind Sprachmodelle nicht einfach vom Himmel gefallen.

¹Gethmann, Buxmann Distelrath, Humm, Lingner, Nitsch, Schmidt, Spiecker genannt Döhmann: „Künstliche Intelligenz in der Forschung - Neue Möglichkeiten und Herausforderungen für die Wissenschaft“, S. 10, aus der Reihe: Ethics of Science and Technology Assessment Bd. 48, bei: Springer

Die Grundlagen wurden schon früh gelegt: Sprachassistent:innen wie Siri, Alexa, Cortana usw. sind schon länger in der Lage gesprochene Sprache zu erkennen und Reaktionen auf einfache Eingaben zu erzeugen.

In der Informatik gilt der Grundsatz:

„Wenn du eine Software oder KI bauen kannst, die etwas erkennt, kannst du auch eine Software oder KI bauen, die dasselbe Ding imitiert.“²

Nebenbei bemerkt wird durch diesen schlichten Satz auch klar, dass jedes Sicherheitssystem, das auf der Erkennung von z.B. menschlicher Sprache oder der Erkennung von Gesichtern basiert, keine gute Idee zur Herstellung von Sicherheit ist – wenn es künstliche Intelligenz gibt, die menschliche Sprache und menschliche Gesichter imitieren kann und diese sich zudem weiter entwickelt.

Einen weiteren Einblick in die Komplexität von Sprachmodellen bieten übrigens auch deren Hardwareanforderungen. Vor kurzem tauchte Facebooks Sprachmodell „LLaMA“ im Netz auf. Ein handlicher Download von 268GB enthält unterschiedliche Sprachmodellvarianten. In der einfachsten 7B Variante benötigt LLaMA zwischen 12-16GB an VRAM einer Grafikkarte, möglichst viele Prozessorkerne und RAM. Die Verarbeitungszeiten selbst einfacher Prompts auf Standardhardware bewegen sich jenseits erträglicher Werte. Damit ist zumindest für den Moment klar, dass auf Basis des heutigen Wissensstandes nur sehr wenige Marktteilnehmer in der Lage sein werden, in der Breite verfügbare Sprachmodelle stabil zu betreiben.

Mögliche Auswirkungen auf das kommerzialisierte Internet

Der dazu notwendige immense Energieverbrauch ist in diesem Kontext letztlich eine Fußnote, weil Sprachmodelle das Potential besitzen, das Netz in seiner Struktur fundamental zu verändern.

Gerade das kommerzielle Internet besteht aus einer Vielzahl von Texten, die einen Gebrauchscharakter aufweisen: Nachrichten, Sportberichte, Produktbeschreibungen und Zusammenfassungen sind durch Sprachmodelle mit relativ begrenzten Anweisungen und Eingaben automatisiert erstellbar. Die dazu notwendigen Daten liegen teilweise bereits in digitalisierter Form vor. Es wird z.B. bei Handballspielen ein dezidiertes Spielprotokoll geführt, aus dem ein heutiges Sprachmodell mühelos und weitgehend automatisiert einen Spielbericht erstellen kann.

Das wird sowohl immense Auswirkungen auf jeden Marktteilnehmer haben, der sich mit Contenterstellung beschäftigt, als auch den Suchmaschinenmarkt komplett verändern: Der zur Suchanfrage passende Inhalt kann theoretisch individuell generiert werden – natürlich auch aus kommerziellen Quellen. Das werden damit spannende Zeiten für das Leistungsschutzrecht: Zitieren auch kleinerer Textpassagen ist im Prinzip nicht mehr notwendig. Die Ergebnisse werden einer signifikant großen Anwender:innengruppe vollkommen ausreichen. Die dann ggf. höflicherweise mit angegebenen Originalquellen erfahren keine Aufmerksamkeit durch Klicks und Besuche mehr – einer immens wichtigen ideellen Währung im Netz. Damit werden insbesondere die Internetgiganten es noch effektiver bewerkstelligen, Nutzer:innen in ihrem eigenen digitalen Ökosystem zu halten. Microsoft unternimmt bereits erste Schritte mit BingAI und der Integration von sprachmodellbasierter Korrekturfunktionen in seine Officeprodukte. Google holt als Reaktion seinen auf Lamda basierten Chatbot „Bard“ wieder auf die Bühne, dessen Produktivbetrieb zunächst auf Basis von ethischen Bedenken verworfen wurde: Bard konnte per Sprachausgabe täuschend echte Anrufe z.B. zur Reservierung von Restauranttischen oder Essensbestellungen auf Basis von

²<http://blog.fefe.de/?ts=9ae8942a>

einfachen Prompts realisieren. Das potentielle Missbrauchspotential dieser Funktionen stand einer Markteinführung 2021 noch im Weg.

Sprachmodelle als Unterstützung bei Anträgen und Dokumentationen

Deutschland hat ein ausgeprägtes Dokumentations- und Antragswesen in unterschiedlichen Bereichen etabliert: In der Medizin z.B. der Arztbericht, in Förderprogrammen umfangreiche Anträge, in der Pflege komplexe Dokumentationen, Protokolle von allen möglichen Sitzungen – die Liste ließe sich beliebig fortsetzen. Allen diesen „Textgattungen“ ist gemein, dass sie oft sehr schematisch sind – schematische Texte sind schon heute eine Domäne von Sprachmodellen.

Sprachmodelle werden dabei helfen, Menschen von diesen schematischen Aufgaben zu entlasten. Für spezielle Fachbereiche sind Textgeneratoren denkbar, die primär mit fachbezogenen Inhalten trainiert werden. Dadurch können sie bei Routineaufgaben unterstützen, wenngleich nach heutigem Stand immer noch eine Überprüfung der Ausgaben durch einen sachkundigen Menschen erforderlich ist.

Aber auch in diesem Feld ist es kompliziert. Nehmen wir das Antragswesen in Förderprogrammen: Die teilweise hohen Anforderungen an Förderanträge dienen dazu, den Zugriff auf eine begrenzte Ressource (z.B. Geld oder Personal) zu begrenzen. Sprachmodelle werden natürlich dabei helfen, bisher nicht ausreichend genutzte Fördermittel tatsächlich abrufen zu können. Es gibt jedoch Ressourcen, die völlig überlaufen sind - etwa die sonderpädagogischen Unterstützungsangebote in Schulen. Auch hier werden Sprachmodelle z.B. bei der Erstellung von Gutachten unterstützen, jedoch bleiben gerade personelle Ressourcen weiterhin knapp. Damit wird der Zugang zu ebendiesen Ressourcen auf anderen Wegen erschwert werden: Ein gut begründeter Antrag steht lediglich in Konkurrenz zu anderen sehr gut begründeten Anträgen.

Sprachmodelle als neutralere Entscheidungsinstanzen?

Sascha Lobo setzt sich in seiner Kolumne „KI kann menschliche Verantwortung nicht ersetzen? - Doch!“ mit der Stellungnahme des deutschen Ethikrates zu künstlicher Intelligenz auseinander. Er fordert, dass künstliche Intelligenz in bestimmten Bereichen an die Stelle der menschlichen Entscheidung tritt – genau da fällt für den deutschen Ethikrat die letzte Grenze. Als Beispiel kann die Diagnose eines Arztes dienen: Ein speziell mit medizinischem Wissen trainiertes Sprachmodell könnte dabei unterstützen. Als Eingabe dienen z.B. Labor- und Anamnesedaten. Das spezialisierte Sprachmodell könnte denkbare Erkrankungen auf Basis dieser Daten ermitteln und wäre dabei nicht auf das Wissen einer Einzelperson beschränkt.

Für Sascha Lobo ist das ein Fall, bei dem sich die Frage nach dem Einsatz von Sprachmodellen nicht stellt, sondern für ihn verbindlich sein sollte, weil das spezialisierte Sprachmodell dem Menschen überlegen sei. Diese Argumentation erscheint zunächst bestechend – allerdings ist ein Modell immer nur so gut wie dessen Eingabedaten, die hier im Rahmen einer Anamnese durch einen Menschen entstehen. Das kann bei einigen Krankheitsbildern komplex und fehlerträchtig sein – gerade in Zweifelsfällen: Die gleiche Krankheit kann bei unterschiedlichen Menschen völlig unterschiedliche Symptome auslösen – es wird eine spannende Frage in welchem ggf. lückenhaften Verhältnis das Erfahrungswissen eines Menschen und dessen Interaktionsmöglichkeiten mit anderen zu der von einer KI gelieferten Information steht.

Sprachmodelle und Bildungsprozesse

Kommen wir am Schluss noch einmal zurück zu Martha, Hannes und Fred: Sie wachsen in einer Welt auf, in der Sprachmodelle mit ihren immer menschenähnlicheren Interfaces wie ChatGPT in ganz kurzer Zeit selbstverständlich sein werden – gerade auch in Kombination mit Sprachein- und -

ausgabe. Nie wieder schulische Übungstexte? Nie wieder Hausaufgaben mit künstlichen Aufgabenstellungen? Sind Hausaufgaben gar tot? Wenn doch heute einfaches „Prompting“ ausreicht?

Entscheidend beim Schreiben ist der gedankliche Prozess bei der Erstellung eines Textes. Wenn wir an die Stelle des Schreibprozesses Prompting bei einem Sprachmodell setzen und den Schreibprozess selbst automatisieren, verschieben wir den Lernprozess auf eine andere Ebene. Diese Ebene erfordert letztlich große Beurteilungskompetenzen, die sich bei Schreiblernenden wie Martha, Hannes und Fred noch in einer sensiblen Entwicklungsphase befinden und damit in vielen Fällen eine Überforderung darstellen müssen. Gleichzeitig gehen mit einiger Wahrscheinlichkeit schreibkompetenten Menschen durch die Nutzung von Sprachmodellen Trainings- und Denkprozesse bei eigentlichen Erstellen von Texten verloren. An deren Stelle treten Mechanismen zur Entwicklung und Modifikation von geeigneten Prompts. Genau diese Vorgehensweisen und Anwendungen werden bereits in großer Breite in den sozialen Medien diskutiert. Warum den Einführungstext zu Immanuel Kant aus dem Schulbuch lesen? Lassen wie Schüler:innen doch einfach mit einem ChatBot „Immanuel Kant“ chatten, den eine Lehrkraft zuvor per „Megaprompt“ mit Hilfe eines Sprachmodells erschaffen hat und der auf die individuellen Fragen der Schüler:innen eingeht.

Sprachmodelle selbst können aufgrund ihrer Komplexität nur von großen Marktteilnehmern betrieben werden – das wirft die Frage nach neuen Abhängigkeitsebenen auf, gerade im Bildungssystem.

Chancen und neue Herausforderungen

Sprachmodelle berühren einen Bereich, der eine Gesellschaft ausmacht: Die Kommunikation und den Umgang mit Sprache. Sie sind nach Maßstäben der Digitalisierung nichts wirklich Neues, wie Googles ChatBot „Bard“ bereits vor drei Jahren zeigte. Mit ChatGPT steht erstmals ein Interface zum Sprachmodell GPT zur Verfügung, was durch viele Menschen niederschwellig nutzbar ist. Viele Anforderungen werden noch nicht zufriedenstellend gelöst. Jedoch können Sprachmodelle auf Millionen von Nutzer:inneneingaben zurückgreifen und auf deren Basis weiterentwickelt werden. Bereits auf dem Stand der heutigen Entwicklung ist immenses Potential auf unterschiedlichen Ebenen erahnbar. Gleichzeitig treten ganz neue Fragestellungen auf, die vor allem Juristen und Ethiker beschäftigen werden. Die Komplexität bei der Modifikation und dem Betrieb von Sprachmodellen wird dafür sorgen, dass sie primär in der Hand der Internetgiganten bleiben werden. Was diese ohne wirksame staatliche Regulierung daraus machen werden, bleibt eine offene Frage.

Maik Riecken
Medienpädagogischer Berater in der Region Süd-West
03/2023