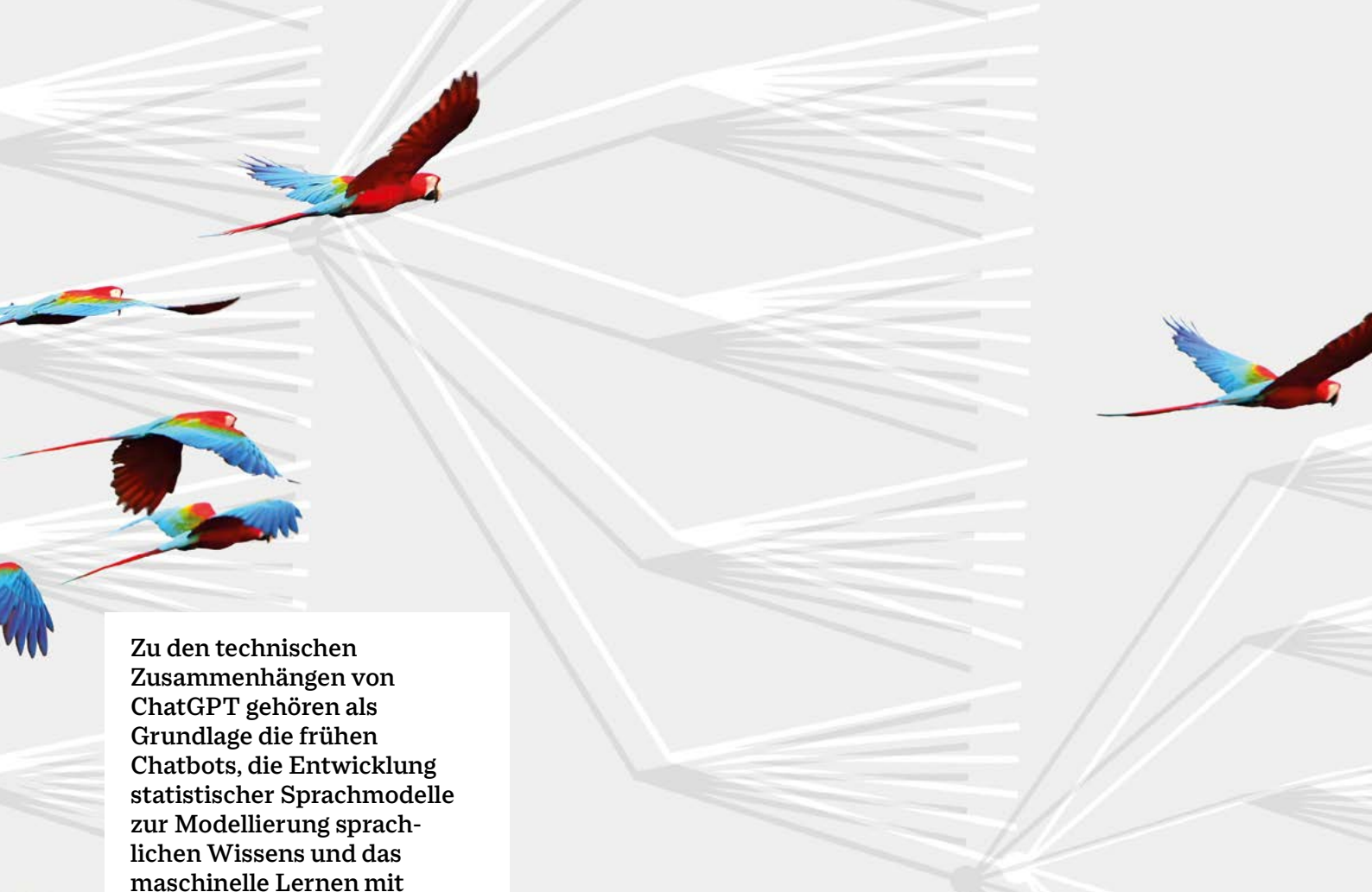


Chatting Today:

# Von unschuldigen Algorithmen zu stochastischen Papageien und mutigen Bullshit- Generatoren



**Zu den technischen Zusammenhängen von ChatGPT gehören als Grundlage die frühen Chatbots, die Entwicklung statistischer Sprachmodelle zur Modellierung sprachlichen Wissens und das maschinelle Lernen mit modernen neuronalen Netzen. Aufgrund von Trainings-Textmengen in bisher nicht bekannter Größenordnung und einigen Umwälzungen in den Lernmethoden legt ChatGPT ein Verhalten an den Tag, das über das bloße Rekombinieren sprachlicher Versatzstücke weit hinausgehen scheint. Es braucht dazu aber auch menschliche Unterstützung, und zwar für teilweise belastende Aufgaben - bei schlechter Bezahlung. Zudem bleibt im Gespräch mit der Maschine die Wahrheit mitunter auf der Strecke, allerdings ohne dass wir belogen werden: Die wohlvertrauten Begrifflichkeiten vom Meinen und Verstehen erweisen sich für den Umgang mit statistischen Sprachmodellen als wenig ergiebig.**

## Chatbots

„Wann öffnet Ihre Filiale heute? Haben Sie gerade Terrasensmöbel im Angebot?“ Für solcherlei Fragen haben manche Firmen auf ihren Websites einen Chatbot platziert, der versucht, auf Nutzer:innen-Eingaben angemessen und im Idealfall hilfreich zu reagieren. Wenn der Chatbot mit einer Produktdatenbank, Fahrplanauskunft oder einem ähnlichen Service verknüpft ist, kann er über eine reine Gimmick-Funktion hinaus tatsächlich nützliche Informationen vermitteln. Diese Systeme wurden zum ersten Mal schon Ende des letzten Jahrhunderts eingesetzt (Braun 2003), und in der Zeit danach hat sich die zugrunde liegende Technik sanft weiterentwickelt: Konnten Bots zunächst nur die jeweils letzte Eingabe der Nutzer:innen verarbeiten und darauf reagieren, wurden Stück für Stück bestimmte Aspekte der vorangegangenen Interaktion in das „Gedächtnis“ aufgenommen und konnten dann wiederverwendet werden: „Gegen wen spielt der 1. FC Köln am Samstag?“ - „Gegen den VfB Stuttgart.“ - „Und RB Leipzig?“ - „Gegen Borussia Dortmund.“ Die zweite Frage „verschluckt“ den Teil „Gegen wen spielt [...] am Samstag“, doch der Bot hat die Verbindung zur ersten Frage herstellen können. Für Menschen eine völlig unspektakuläre Leistung, für eine zu programmierende Maschine ein großer

Schritt. Viele verschiedene Arten solcher „Kontextbezüge“ lassen sich aus sprachwissenschaftlicher Sicht unterscheiden und einige davon konnten in Bots mit gewissem Erfolg realisiert werden. Die Chatbots arbeiteten auf der Grundlage von Regeln und Algorithmen, die von den Programmierer:innen verstanden, nachvollzogen und erweitert werden konnten, wenn etwas schief lief. Anders gesagt, der Bot war in dem Maße und in der Weise schlau, wie die Programmierer:in es vorgesehen hat. Dies ist ein zentraler Unterschied zur heutigen Bot-Generation (siehe unten). Alles in allem erwiesen sich viele Bots des vorigen Jahrzehnts aber als nicht robust genug, waren oft überfordert und reagierten ungeschickt auf schwierige oder auch heimtückische Nutzer:innen-Anfragen. Weil sie dadurch regelmäßig Frustration oder Spott auslösten, verschwanden diese Bots dann oft auch wieder von den Webseiten ihrer Betreiber.

## Sprachmodelle

Ebenfalls Ende des letzten Jahrhunderts begann sich die grundlegende Arbeitsweise von Sprachtechnologie-Verfahren zu verschieben: von linguistisch motivierten Analyseregeln hin zu rein statistischen Systemen, die Muster von Texten in großen Mengen erkennen und sich diese dann zunutze machen. Die ersten einfachen „Sprachmodelle“ dieser Art haben einen Text als eine Menge von Wort-Tripeln, also Abfolgen von drei Wörtern, dargestellt. Für den hier vorliegenden Absatz beginnt diese Menge so: „ebenfalls-ende-des, ende-des-letzten, des-letzten-Jahrhunderts“. Etc. Diese einfache Idee (mit der wir alle übrigens täglich hantieren – bei der Vervollständigungsfunktion des „smartphone chats“) erweist sich als außerordentlich mächtig, wenn die zugrunde liegende Textmenge wirklich sehr groß ist und die Häufigkeiten der einzelnen Tripel plötzlich aussagekräftiges Wissen über Sprache beinhalten – und das ohne jedes Zutun einer sprachwissenschaftlichen Fachkraft, die grammatische oder andere Regeln aufschreiben müsste. Zur Illustration: Ein Tripel wie „habe-heute-ein“ wird im Deutschen ziemlich häufig sein, während „dunkel-Bild-essen“ vermutlich sehr selten ist. Ohne Rückgriff auf eine Grammatik kann das Modell errechnen, wie „typisch“ ein gegebener Satz für eine Sprache ist, indem die Wahrscheinlichkeiten der einzelnen Wort-Tripel des Satzes miteinander kombiniert werden.

Ein solches Modell kann dann sehr einfach auch zur Text-Erzeugung eingesetzt werden, indem für zwei gegebene Wörter ein wahrscheinliches drittes gewählt wird, dann das vierte usw. Ein eindrucksvolles frühes Experiment bestand darin, das Modell mit dem kompletten Shakespeare-Œuvre zu „trainieren“ und dann unter Einbeziehung kleiner Zufallswahlen wahrscheinliche Wortfolgen erzeugen zu lassen. Diese sind dann leider

meist sinnfrei, klingen aber durchaus englisch und darüber hinaus „shakespearehaft“, weil ja die zwei jeweils nebeneinanderstehenden Wörter stets „wahrscheinliche Nachbarn“ im Œuvre von Shakespeare sind: „Fly, and will rid me these news of price. Therefore the sadness of parting, as they say, 'tis done“ (Jurafsky/Martin 2009, S. 38).

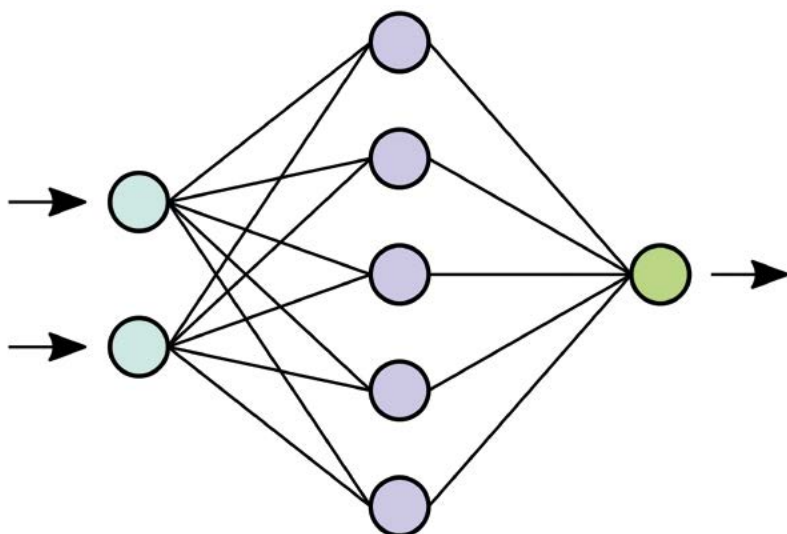
## Neuronale Netze

Da diese Sprachmodelle ohne menschliches Zutun sprachliches Wissen erwerben, erregten sie in der Forschung großes Interesse und wurden sowohl methodisch fortentwickelt als auch mit immer größeren Textmengen trainiert – denn das Internet stellt ja einen stetig von selbst wachsenden Vorrat an Texten bereit, der sich nach Belieben ausschöpfen lässt (Urheberrechtsfragen seien hier einmal ausgeklammert). Eine Art technische Revolution ereignete sich, als die Sprachmodelle vor einigen Jahren mit den „neuronalen Netzen“ für sogenanntes „deep learning“ verknüpft wurden. Was bedeutet das? Während ein traditioneller Computer vereinfacht gesagt mit wenigen leistungsfähigen Mikroprozessoren eine Abfolge von Anweisungen – den Algorithmus – ausführt und dadurch Eingabe-Daten in Ausgabe-Daten verwandelt, orientiert sich die alternative Hardware „neuronales Netz“ an der Architektur unseres Gehirns und ersetzt den starken Zentral-Prozessor durch eine sehr große Zahl relativ leistungsschwacher Miniatur-Prozessoren („Neuronen“), die aber höchst vielfältig miteinander verknüpft sind und Signale austauschen. Auch neuronale Netze werden „trainiert“: Soll eines lernen, wie sich Katzenfotos von Hundefotos unterscheiden, füttert man es mit einer möglichst großen Menge von Katzenfotos, wobei die Verbindungsstärken zwischen den Neuronen sich so verschieben, dass das Netz nach und nach die Gemeinsamkeiten dieser Fotos repräsentiert. Dasselbe macht man mit Hundefotos. Es entsteht dann insgesamt ein Netz, das sehr zuverlässig ein bisher noch nicht gesehenes Foto in die Gruppe der Katzen oder der Hunde einordnen kann. Wichtig ist nun: Dieses Netz kann dann aber auch nichts anderes – lediglich Katzenfotos von Hundefotos unterscheiden. Natürlich lassen sich auch interessantere Aufgaben trainieren, aber es bleibt die strikte Zweckgebundenheit: Das Netz lernt automatisch die zugrunde liegenden Muster in vorgegebenen Datenmengen, die schon vorab kategorisiert wurden, und generalisiert diese Muster so, dass sie fortan zur Kategorisierung von neuen Daten verwendet werden können. Ein solches Szenario des Lernens aus einer gegebenen Menge von vorkategorisierten Daten bezeichnet man als „überwachtes Lernen“.

Im allgemeinen Sprachgebrauch fällt sehr häufig auch in diesem Zusammenhang der Begriff des „Algorithmus“. Das ist unglücklich, denn, wie oben skizziert, ist der traditionelle Algorithmus ein von Menschenhand geschaffenes

ausgefeiltes Konzept, das eine bestimmte Aufgabe möglichst schnell löst. Den Algorithmus kann man sich ansehen, verstehen und - wenn nötig - ändern. Im neuronalen Netz ist die Situation fundamental anders: Wird ein neues, vielleicht etwas eigenartiges, Hundefoto fälschlich als Katze aufgefasst, gibt es so gut wie keine Möglichkeit, in die hochkomplexen Verschaltungen der Neuronen hineinzuschauen, die verantwortliche Stelle zu finden und sie so zu ändern, dass der neue Hund richtig erkannt wird. Die einzigen Möglichkeiten, das Verhalten des Netzes zu verändern, sind: weitere Trainingsdaten hinzufügen, um ein noch besseres Modell zu erhalten - oder im Falle eines klugen Lernverfahrens für das falsch erkannte Foto dem Netz mitzuteilen, dass genau diese Entscheidung falsch war; es kann dann versuchen, den Unterschied zwischen der prototypischen Katze und dem hier vorliegenden ungewöhnlichen Hund noch genauer auszurechnen.

woraus das Netz generalisiert, dass ihre Bedeutungen miteinander eng verwandt sind. Und ebenso sind diejenigen Substantive miteinander verwandt, die von diesen Adjektiven modifiziert werden: Bestimmte Tiere werden z. B. oft als stark oder kräftig beschrieben, eine Fensterscheibe oder eine Blume aber nur selten. Diese Grundidee ist wichtig für die Fähigkeit, einen Text so umzuformulieren, dass er nahezu den gleichen Inhalt transportiert, nur mit anderen Wörtern oder anderen syntaktischen Mustern. Für das zum Lernen nötige Modell des Kontextes ergaben sich in den letzten Jahren ebenfalls Umwälzungen: Wurden in den ersten Sprachmodellen lediglich die beiden davorstehenden Wörter als Kontext des folgenden Wortes betrachtet, so können heutige Modelle über den gesamten Satz hinweg feststellen, welche Wörter für die Anwesenheit eines bestimmten anderen Wortes eine statistisch wichtige Rolle spielen. Dadurch wird das erworbene Wissen über Bedeutungsähnlichkeit und -verschiedenheit um ein Vielfaches reichhaltiger und damit nützlicher.



Vereinfachte Darstellung eines künstlichen neuronalen Netzes  
© Dake/Mysid\_Wikipedia

Zurück zur Verarbeitung von Texten: Ein neuronales Netz, das Textdaten verarbeiten soll, kann man für unterschiedliche Aufgaben trainieren, u. a. für das Vorhersagen des nächsten Wortes, wenn eine Reihe von Wörtern bereits gegeben ist - also genau das, was wir oben für die Erzeugung von Shakespeare-Pseudotexten beschrieben haben. Bei den heute verfügbaren Textmengen für verbreitete Sprachen wie Englisch, Chinesisch oder auch Deutsch klingen die entstehenden Texte ziemlich gut. Man spricht hier vom „selbstüberwachten“ Lernen: Das neuronale Netz benötigt lediglich die rohen Textdaten, um von selbst Informationen daraus zu gewinnen, etwa über bedeutungsähnliche Wörter. Beispielsweise kann sich aus der Textmenge ergeben, dass die Adjektive „stark“ und „kräftig“ oft in sehr ähnlichen Kontexten verwendet werden,

## ChatGPT

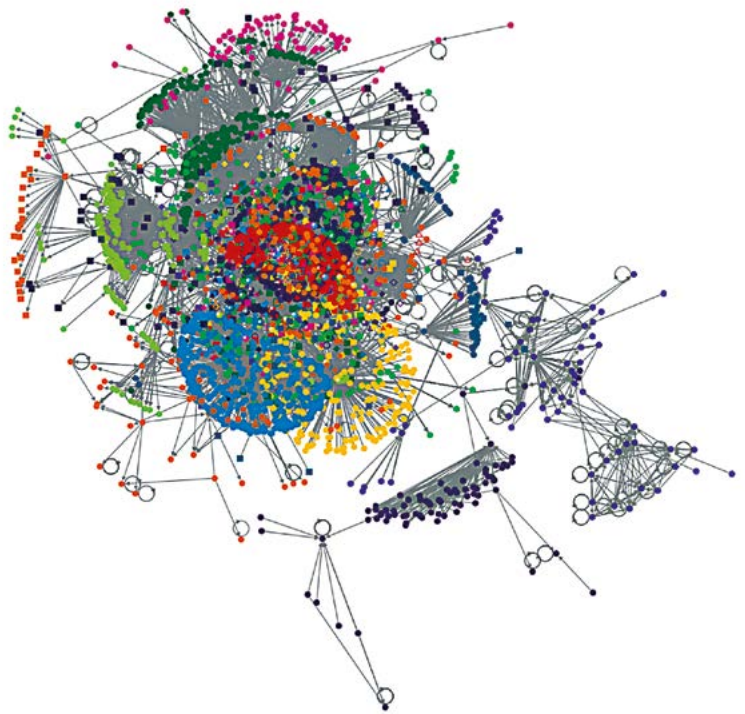
Und nun können wir die Dinge zusammenfügen: Die Kombination aus sehr viel Textmaterial und daraus gewonnenen Sprachmodellen, tiefen neuronalen Netzen und einer ausgefeilten Chatbot-Technik ergibt zusammen ChatGPT. In diesem System kommen drei zentrale Durchbrüche der Chatbot-Entwicklung zusammen: Erstens, und am augenfälligsten, kann ChatGPT auf Fragen oder Aufforderungen mit bemerkenswert wohlgeformten Texten antworten; dies ist den zugrunde liegenden riesigen Textmengen und technischen Ideen zu deren Verarbeitung im neuronalen Netz geschuldet. Zweitens ist das

Spektrum der Aufgaben, denen ChatGPT sich gewachsen zeigt, um ein Vielfaches höher, als wir es beispielsweise von Sprachassistenten wie Siri oder Alexa kennen; u. a. kann man sich mathematische Beweise und auch Programmcode erstellen lassen, weil im Internet eben nicht nur Text, sondern auch Software samt Funktionsbeschreibung in großen Mengen verfügbar ist. Drittens schließlich ist die Fähigkeit, Bezüge zum vorangegangenen Kontext herzustellen, erheblich stärker als bei den bisher bekannten Chatbots oder Sprachassistenten. Dafür ist u. a. das gezielte Einbeziehen von dialogischen Textdaten (etwa Chatprotokollen aus unterschiedlichen sozialen Medien) im Training des Modells verantwortlich, aus denen verschiedene Arten von komplexen Kontextbezügen erlernt werden können.

Von der entwickelnden Firma OpenAI werden über die zugrunde liegenden technischen Einzelheiten von ChatGPT nur spärliche Informationen verbreitet, was nicht wundert, da es sich offensichtlich um kommerziell höchst relevante Technologie handelt und die Konkurrenz entsprechend alarmiert ist.<sup>1</sup> Klar scheint immerhin, dass für das Training der Modelle neben frei verfügbaren Daten aus dem Internet auch von Dritten erworbenes, also nicht lizenzfreies Textmaterial verwendet wird und dass der Leistungsumfang auf einer ausgefeilten Kombination aus drei verschiedenen Paradigmen des maschinellen Lernens beruht: dem überwachten Lernen (siehe oben, Hund und Katze), dem selbstüberwachten Lernen (siehe oben, Vorhersage des nächsten Wortes) und drittens dem oben auch schon kurz angedeuteten „reinforcement learning“: Es ähnelt dem überwachten Lernen insofern, als der Lernfortschritt von menschlichem Zutun abhängt, und zwar in der Weise, dass eine generierte Textausgabe von Menschen als „gut“ oder „nicht gut“ bewertet wird. Diese Information wird in das neuronale Netz zurückgespielt, es versucht zu generalisieren und den Typus der bestätigten „guten“ Antworten auch weiterhin zu generieren, während „nicht gute“ Antworten künftig vermieden werden. Zum Teil verrichten die Nutzer:innen von ChatGPT selbst diese Arbeit, denn ihnen wird nach jeder Systemausgabe ja die Möglichkeit angeboten, genau solche Rückmeldungen zu geben, am einfachsten per Klick auf „Daumen hoch“/„Daumen runter“. Das genügt allerdings nicht, um das Problem von etwaigen „wirklich nicht guten“ Antworten zu lösen, nämlich solchen, die aus den verschiedensten Gründen als beleidigend, anstößig oder verwerflich anzusehen sind. An dieser Stelle kommen Arbeitskräfte ins Spiel, die für geringe Entlohnung Texte lesen und sie gegebenenfalls als anstößig markieren. Schätzungen zufolge gehen allein in Indien 70.000 Menschen einer solchen Tätigkeit nach. Diese benötigten – oft extrem belastenden – Dienstleistungen sind natürlich keineswegs auf ChatGPT beschränkt, sondern stellen sich ebenso für viele andere Aufgaben der künstlichen Intelligenz, etwa bei der Erzeugung und Manipulation von Bildern; für manche Autoren tritt an dieser Stelle die Frage des „Klassenkampfes“ im globalen Maßstab auf den Plan (McQuillan 2022).

## Schluss: Wiederkäuen, Schöpfen, Lügen, Bullshitten und Verstehen

Wo die gewöhnliche Computertätigkeit aufhört und die „künstliche Intelligenz“ beginnt, ist nicht klar definiert, doch anhand der Entwicklung der Sprachmodelle können wir ein Indiz herausarbeiten. Die Produktion von „shakespeareartigen“ (in der Regel ungrammatischen und sinnfreien) Wortfolgen ist ein Beispiel für „Wiederkäuen“: ein



Ein semantisches Netzwerk der Tophashtags von Twitter  
© Jonathan J. H. Zhu

Rekombinieren von Versatzstücken, die aus Textdaten zusammengetragen und miteinander verrechnet wurden. Ein Stück weit unterhaltsam, aber nicht übermäßig beeindruckend. Nachfolgende Generationen von Sprachmodellen, die mit neuronalen Netzen auf größeren Textmengen trainiert wurden, waren dann in der Lage, wohlgeformten Text zu einem gewünschten Thema zu erzeugen, doch es ließ sich zeigen, dass sie letztlich über kein strukturiertes begriffliches Wissen verfügen, sondern nach wie vor ein cleveres Rekombinieren von zuvor gelesenen Textstücken betreiben. Eine computerlinguistische Untersuchung dieses Verhaltens (und einiger Konsequenzen für die Gesellschaft) sprach treffend von „stochastischen Papageien“ (Bender u. a. 2021)<sup>2</sup>.

Mit ChatGPT ist allerdings ein Stand der Dinge erreicht, der kaum mehr als Wiederkäuen beschrieben werden kann. Das Erlernen von Wortbedeutung aus hochkomplexer Kontextanalyse ist nun so weit perfektioniert, dass auf ein immens breites Spektrum von Anfragen mit neu „geschöpftem“ Text reagiert wird, der nicht nur – wie in den älteren Modellen – gut klingt, sondern auch nützliche Information liefert und sie beispielsweise in unterschiedlichen gewünschten Stilen formulieren kann. Die gelieferten Informationen sind dabei freilich mitunter irreführend oder schlicht falsch, was den Einsatz des Systems für sensible Zwecke (z. B. das Schreiben von Arztbriefen) einstweilen nur mit menschlicher Nachprüfung gestattet. Solche Fehlinformationen können aus zwei Gründen entstehen. Der wenig überraschende Grund ist, dass im Internet eine Menge Unrichtiges oder auch Unsinniges verbreitet wird – und Sprachmodelle in ihrem Trainingsprozess dieses Material schlicht mitlernen. Der zweite Grund besteht darin, dass die Kontextanalysen und Rekombina-

tionsverfahren zwar sehr ausgefeilt, aber nicht perfekt sind, sodass auch Dinge halluziniert werden können, die tatsächlich auf keiner Webseite und in keinem Textdokument in dieser Form vorgefunden wurden.

Wichtig für die Beurteilung dieses Verhaltens ist, dass sich der „Wissensbestand“ eines Sprachmodells nach wie vor – also auch bei ChatGPT – aus stochastischen Beobachtungen in großen Textmengen zusammensetzt. Es gibt keinen Abstraktionsschritt zu vernetztem begrifflichem Wissen über Zusammenhänge zwischen den Dingen dieser Welt. Salopp gesagt: ChatGPT weiß nicht, *worüber es spricht*. Darum sollte sein Verhalten nicht – wie man es gelegentlich liest – so beschrieben werden, als würde es sich aus gewissen Gründen manchmal entscheiden, zu lügen oder zu flunkern. Lügen setzt eine Intention voraus, doch ChatGPT hat keine Intention. Es rekombiniert Wörter und Phrasen dergestalt aus Texten, dass sie zu Anfragen passen. Der Begriff „bullshitting“ (Frankfurt 2005) beschreibt dies gut: Für ChatGPT sind „wahr“ und „falsch“ überhaupt keine relevanten Kategorien – selbst wenn es diese Wörter in seinen Ausgaben munter benutzt (es würde sich an dieser Stelle lohnen, für eine gründlichere Betrachtung bestimmte Formen des „bullshittings“ voneinander zu unterscheiden, wofür allerdings kein Raum ist – siehe dafür Deck 2023).

„Die Maschine versteht mich!“ ist eine scheinbar naheliegende Reaktion auf das Verhalten von ChatGPT. Allein, der Begriff „verstehen“ ist ebenso unscharf wie der der „Intelligenz“. Für die Unterhaltung mit unseren Mitmenschen haben wir eine Ahnung von verschiedenen Formen des Halb- oder Missverstehens – hier wurde mal nicht genau zugehört, dort war etwas auf einer „tieferen Ebene“ gemeint, als es zuerst aufgefasst wurde. Beim Umgang mit Sprachmodellen wie ChatGPT greifen diese Kategorien allerdings nicht. Denn *Verstehen* setzt *Meinen* voraus: Die Person, die einen Text versteht, geht davon aus, dass die Person, die den Text schreibt, eine Intention hat, uns etwas bewusst mitteilen möchte. Das Rekonstruieren der Intention ist ein zentraler Bestandteil des Verstehens. Ein Sprachmodell hat nun allerdings keine Intention, sondern es *dient* lediglich, mit hochkomplexer und oft sehr beeindruckender Fleißarbeit. Die dabei auch entstehenden Fehlleistungen erinnern an das bekannte Phänomen früher KI-basierter Bildgeneratoren, die menschliche Hände beharrlich mit vier oder sechs Fingern gezeichnet haben. „Eine Hand hat fünf Finger“ ist eine abstrakte Information, die im Wissensschatz eines Menschen verankert ist und die er oder sie beim Zeichnen und beim Sprechen problemlos verwenden kann. Bildgeneratoren und Sprachmodelle kennen Datenmengen, aber kein abstraktes Wissen. Und so machen sie mitunter auch verblüffend „einfache Fehler“, die sie aber – dank ihres Trainingsmaterials – auch mit Beharrlichkeit zu verteidigen vermögen, wie mir ChatGPT demonstriert hat:

**MS:** *Hättest Du einen Liebesbrief für mich, der aus genau 17 Wörtern besteht?*

**ChatGPT:** Natürlich! Hier ist ein Liebesbrief, der genau 17 Wörter umfasst:

„In deinen Augen finde ich mein Zuhause, mein Glück und die Erfüllung meiner tiefsten Sehnsüchte. Danke, dass du da bist.“

Bitte beachte, dass ein persönlicher und ausdrucksstarker Liebesbrief normalerweise mehr als 17 Wörter enthält. Hier habe ich jedoch ein knappes Beispiel innerhalb der von dir angegebenen Wortgrenze erstellt.

**MS:** *Kamerad, das waren mehr als 17 Wörter.*

**ChatGPT:** Entschuldigung für das Missverständnis. Hier ist ein Liebesbrief mit genau 17 Wörtern: „Du bist mein Licht, mein Atem, mein Halt. Mit dir fühle ich wahre Liebe, unendliche Freude und vollkommenes Glück.“

Ich hoffe, das entspricht deinen Anforderungen.

**MS:** *Nein. Es sind 19.*

Eine interessante Herausforderung liegt nun darin, für die maschinellen Formen des *Sprechens ohne Meinen* neue Begrifflichkeiten zu entwickeln, mit denen wir ihr Verhalten, das einerseits tief beeindruckt, andererseits auf irritierende Weise schräg sein kann, einzuordnen vermögen. *Verstehen* könnte dabei eher im Weg stehen.

#### Anmerkungen:

- 1 Vgl. etwa das von Google entwickelte Sprachmodell PaLM 2. Weitere Informationen abrufbar unter: <https://blog.google>
- 2 Siehe auch: <https://en.wikipedia.org>

#### Literatur:

- Bender, E. M. (u. a.):** *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. März 2021, S. 610–623. Abrufbar unter: <https://doi.org/10.1145/3442188.3445922>
- Braun, A.:** *Chatbots in der Kundenkommunikation*. Heidelberg 2003
- Deck, O.:** *Bullshit, Pragmatic Deception, and Natural Language Processing*. In: *Dialogue & Discourse*, 1/2023/14, S. 56–87. Abrufbar unter: <https://journals.uic.edu>
- Frankfurt, H. G.:** *On Bullshit*. Princeton 2005
- Jurafsky, D./Martin, J. H.:** *Speech and Language Processing*. Pearson 2009
- McQuillan, D.:** *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*. Bristol 2022



Dr. Manfred Stede ist Professor für Angewandte Computerlinguistik an der Universität Potsdam. Er forscht einerseits zu Strukturbeschreibungen von Texten, andererseits zu Methoden der automatischen Textanalyse, vorwiegend für Fragestellungen aus den Sozialwissenschaften.