

Droht das Ende der Experten?

Holtel

2024

ISBN 978-3-8006-7239-4

Vahlen

schnell und portofrei erhältlich bei
beck-shop.de

Die Online-Fachbuchhandlung beck-shop.de steht für Kompetenz aus Tradition. Sie gründet auf über 250 Jahre juristische Fachbuch-Erfahrung durch die Verlage C.H.BECK und Franz Vahlen.

beck-shop.de hält Fachinformationen in allen gängigen Medienformaten bereit: über 12 Millionen Bücher, eBooks, Loseblattwerke, Zeitschriften, DVDs, Online-Datenbanken und Seminare. Besonders geschätzt wird beck-shop.de für sein umfassendes Spezialsortiment im Bereich Recht, Steuern und Wirtschaft mit rund 700.000 lieferbaren Fachbuchtiteln.

KAPITEL 13

Wie funktioniert eine Sprechmaschine?

„Ich akzeptiere Chaos, bin mir aber nicht sicher, ob es mich akzeptiert.“

– Bob Dylan

Das chinesische Zimmer

Können Maschinen menschliche Intelligenz imitieren? Oder gibt es einen fundamentalen Unterschied zwischen menschlicher und Künstlicher Intelligenz? Diese Frage trieb den Philosophen John Searle um. Mit dem Gedankenexperiment *Argument des chinesischen Zimmers* wollte er sie beantworten.²⁷² Sein Fazit lautet, ein Computer besitze keine menschliche Intelligenz, weil seine Funktionen auf formalen Regeln beruhen.

Um das zu erklären, erstellte Searle eine hypothetische Versuchsanordnung: In einen Raum sitzt eine Person. Sie spricht kein Chinesisch. Vor ihr liegt ein dickes Handbuch. Es enthält die 3.000 bis 5.000 chinesischen Schriftzeichen, die typischerweise notwendig sind, um sich im Alltag zurechtzufinden.²⁷³ Zudem liegen Anweisungen vor, wie diese Schriftzeichen verarbeitet werden müssen. Außerhalb des Raumes befindet sich eine zweite Person, die allerdings fließend Chinesisch spricht. Sie stellt Fragen in Chinesisch. Die Person im Zimmer kann mithilfe des Handbuchs auf die Fragen reagieren, indem sie die Schriftzeichen laut Regeln verarbeitet und Antworten in chinesischen Zeichen liefert. Dafür muss sie die Bedeutung der Schriftzeichen nicht kennen – und die Sprache auch nicht selbst sprechen.

Searle argumentiert, obwohl das System (bestehend aus Handbuch, Person und Regeln) in der Lage sei, auf chinesische Fragen zu antworten, würde es keine tatsächliche Intelligenz zeigen. Die Person im Zimmer könne zwar eine sinnvolle Antwort in chinesischen Schriftzeichen produzieren, aber sie verstehe nicht, was sie schreibt oder welche Bedeutung in einer Frage steckt. Searles Argumentation zielt darauf ab, die Vorstellung zu widerlegen, dass Maschinen allein kraft formaler Regeln und Algorithmen menschliche Intelligenz erzeugen oder gar übertreffen könnten. Er betont, **dass menschliche Intelligenz mehr ausmache als das Befolgen formaler Regeln und das Einhalten einer Syntax**. Ein tiefes Verständnis von Bedeutung, Kontext und Empfindungsfähigkeit sei unerlässlich, um wirkliche Verständigung und Intelligenz zu zeigen.

Searles Beschreibung war ein Gedankenexperiment. Mit ChatGPT gibt es nun eine praktisch verfügbare Versuchsanordnung, die Searles Argumente untermauert. Denn tatsächlich *simuliert* ChatGPT einen Sprachdialog – und zwar derart gut, dass es Menschen schwerfällt, maschinelle von menschlichen Antworten zu unterscheiden.²⁷⁴ ChatGPT basiert auf ausgeklügelten Techniken des maschinellen Lernens,²⁷⁵ es besitzt keine Intelligenz im menschlichen Sinn – und kann dennoch komplexe Texte verstehen und Fragen ausgiebig beantworten. **Aber was ein Benutzer von ChatGPT als intelligentes Gespräch wahrnimmt, beruht tatsächlich auf schnöder Statistik.**

Ein-Wort-Geschichten

Im Jahr 1957 gelangt der Sprachforscher John Rupert Firth zu einer folgenreichen Erkenntnis: „*Die Bedeutung eines Wortes erkennt man am Kontext, in dem es verwendet wird.*“²⁷⁶ Diese Hypothese bildet die sprachtheoretische Grundlage für ChatGPT. Der Chatbot denkt nicht, sondern er ist nur sehr überzeugend darin, so zu tun als ob. Für dieses Kunststück reicht es zu wissen, welche Worte oft mit anderen im selben Zusammenhang verwendet werden. Das reicht offenbar aus, uns mutmaßen zu lassen, dass dieses Kunststück gleichbedeutend mit Denken ist. **Aber ChatGPT spielt einfach nur Improvisationstheater.**

Der Chatbot simuliert ein Spiel, das ins Standard-Repertoire des professionellen Improvisationsschauspielers gehört. ChatGPT spielt das Spiel *Ein-Wort-Geschichten*.^{277,278} In dem geht es darum, eine Geschichte zu erzählen, bei der jeder Spieler abwechselnd das nächste Wort sagen darf. Die Eingabe jedes einzelnen Spielers entsteht aus seinem Verständnis der bisherigen Geschichte und seinen eigenen Ideen, wie die Geschichte sinnvoll weitergehen könnte. Es geht so: Alle Spieler stehen in einer Reihe und das Publikum gibt den Titel der Geschichte vor. Die Spieler erzählen die Geschichte in der Reihenfolge, in der sie stehen. Jeder Spieler fügt ein neues Wort hinzu. Es geht darum, schnell und spontan zu reagieren. Die Geschichte soll einerseits den Ansprüchen an Semantik, Syntax und Pragmatik genügen (Kapitel 11), aber auch das Publikum unterhalten. Genauso wie jeder Spieler nutzt ChatGPT sein Verständnis von Sprache und Wortmustern, um auf das bisher Gesagte zu reagieren oder eine Fortsetzung zu formulieren. So wie die Schauspieler mit dem nächsten Wort auf alles reagieren, was bis zu diesem Zeitpunkt gesagt worden ist, findet auch ChatGPT ein plausibles Wort. Die Antworten der Spieler entspringen deren Sprachverständnis, ihrer Lebenserfahrung – und dem Zufall. Auch Gedanken an das übernächste Wort sind überflüssig, denn das werden andere finden müssen. Der Fokus des Schauspielers liegt darauf, ein passendes Wort für genau diese eine nächste Stelle einer sich entwickelnden Geschichte zu finden. Das läuft bei ChatGPT genauso.

Was aber unterscheidet das Schreiben eines Textes durch ChatGPT von einer Geschichte, die mehrere Menschen zusammenspinnen? Menschen sprechen wir ein Bewusstsein zu, Gefühle und die Fähigkeit zu kreativem, reflexivem Denken. Sie formulieren Absichten, Ziele und Bedeutungen, die sie in Worte kleiden. Und das alles fließt in

ihre Entscheidung für das nächste Wort ein. Im Gegensatz dazu hat ChatGPT kein Bewusstsein, keine Emotionen und auch keine Absichten. Der Algorithmus generiert Text nach statistischen Mustern, die es antrainiert hat – ohne jegliche Grundlage für Verständnis oder einer Absicht.

Auf was greift ChatGPT zurück, wenn es das nächste Wort vorhersagt? Das System analysiert die Frage und sucht nach ähnlichen Beispielen aus seinen Trainingsdaten, um eine Antwort zu generieren. Die umfassen bei ChatGPT ca. 570 Gigabyte, in denen 300 Mrd. Wörter kodiert sind. Die Wahrscheinlichkeit ist hoch, ein Wort zu finden, das im Kontext aller bisher formulierten Worte für Menschen einen Sinn ergibt – aber sicher ist es nicht. Denn wie bei der Ein-Wort-Geschichte die Qualität des nächsten Wortes vom Verständnis und der Fantasie des aktuellen Spielers abhängt, basiert die Genauigkeit der Antworten von ChatGPT auf der Qualität und Größe der Datenbanken, die es heranzieht. Einfache Regel: Je größer und vielfältiger die Datenbanken sind, desto besser kann das System Muster erkennen und findet passende Wortfolgen.

Es gibt sogar Hinweise darauf, dass auch Menschen ähnliche Mechanismen der Sprachnutzung verwenden. Der Verhaltenswissenschaftler Morten H. Christiansen beschreibt Sprache als ein komplexes adaptives System, bei dem Sprache durch einen Prozess kultureller Evolution entsteht. Und er erkennt dabei eine Art Spiel. Christiansen argumentiert, dass Sprache nicht etwa das Produkt angeborenen Wissens oder eines angeborenen Spracherwerbs sei. Vielmehr sei sie das Ergebnis der Interaktion zwischen dem, was Menschen biologisch ausmacht und der Umgebung, in der sie Sprache lernen und verwenden.²⁷⁹ Christiansen verwendet die Spielmetapher, um daran zu erläutern, wie Sprache gelernt und über Generationen weitergetragen wird. Für ihn ist Sprache ein Spiel, das eine Gemeinschaft von Sprechern spielt, wobei sich allerdings während des Spielens im Laufe der Zeit permanent die Regeln ändern. Sie sind nicht etwa von vorneherein festgelegt und vorbestimmt, sondern ergeben sich aus den gemeinsamen Interaktionen der Spieler. Der Sprachexperte betont auch die Bedeutung von Häufigkeit und Wiederholung beim Sprachenlernen: Je häufiger jemand einer sprachlichen Struktur ausgesetzt ist, desto wahrscheinlicher ist es, dass er sie lernt, imitiert und erinnert. Damit wäre menschliches Sprachenlernen nicht nur eine Frage des Erinnerns einzelner Wörter oder Grammatiken, sondern auch der statistischen Muster, die dem Verwenden von Sprache zugrunde liegen. Es schließt an das an, was wir einige Sätze vorher über die Rolle von Statistik beim maschinellen Sprachlernen durch ChatGPT gelernt haben.

Dass die Idee des Textverstehens durch das Analysieren von Worthäufigkeiten auch andersherum funktioniert, zeigten die beiden Mathematiker Frederick Mosteller und David L. Wallace bereits im Jahr 1964.²⁸⁰ Sie wollten die Frage der Autorenschaft der berühmten *Federalist Papers* klären.²⁸¹ Das sind eine Reihe von 85 Artikeln oder Essays, die um die Jahre 1787 und 1788 veröffentlicht wurden, um die Ratifizierung der US-Verfassung zu fördern. Allerdings wurden sie unter einem Pseudonym veröffentlicht. Dafür kamen drei Autoren infrage: Alexander Hamilton, James Madison und John Jay.

Die Autorenschaft einiger dieser Artikel blieb strittig. Aber die beiden Mathematiker lösten das Rätsel. Sie untersuchten die Häufigkeit des Gebrauchs von grammatikalischen Funktionswörtern wie Präpositionen, Konjunktionen oder Artikel. Ihre Annahme war, dass diese oft unbewusst und konsistent verwendet werden, und damit den persönlichen Schreibstil eines Autors widerspiegeln. Nach einer umfassenden Analyse schlossen sie, dass die strittigen Artikel wahrscheinlich von James Madison geschrieben worden waren.

Wenden wir uns der Frage zu, wie diese theoretischen Ansätze in ein technisches Werkzeug wie ChatGPT umgesetzt wurden.

Worte sind Statistik

Es bleibt wichtig zu betonen, dass es sich bei ChatGPT um eine Maschine handelt. Sie *versteht* nicht, was sie sagt. Sie *begreift* nicht im menschlichen Sinne. Sie *antwortet* nicht, sondern führt einen angefangenen Satz weiter. Alles basiert auf statistischen Formeln, die die Häufigkeit von Wortfolgen herausfinden. ChatGPT setzt seine Antworten aus einem Wortfundus zusammen, dessen innerer Aufbau in Zahlen übersetzt worden ist. Darunter findet sich kein tieferes Verständnis der verhandelten Themen oder gar eine Art von Bewusstsein. Die Wirkung von ChatGPT entsteht aus der überzeugenden *Simulation* von Verstehen. Die Benutzer des Chatbots sind beeindruckt von dessen Antworten – denn sie nehmen zwar die Resultate wahr, sind sich aber der zugrundeliegenden Technologien nicht bewusst. Denn dann würden sie akzeptieren, dass die Antworten von ChatGPT nicht ungenau sind, irreführend und irritierend oder auch mal Grund weg falsch. Es gibt einen technischen Grund dafür. Schauen wir uns genau an, was hier passiert.

Mit dem *Höhlengleichnis* erklärt der Philosoph Platon das Wesen der Realität und das Verhältnis von Wissen und Wahrnehmung.²⁸² Dafür stellt er eine Analogie her zwischen einer Höhle und dem menschlichen Verstand: In einer Höhle sind Gefangene angekettet. Sie sind am Boden fixiert und sehen nur die Schatten auf der Höhlenwand. Alle Schatten werden von Dingen geworfen, die außerhalb ihrer Sichtweite an einer Lichtquelle vorbeigeführt werden. Für die Gefangenen sind die Schatten die einzige Realität. Sie müssen annehmen, dass dies die eigentliche Welt sei.

So ähnlich blicken wir auf ChatGPT. Wenn der Chatbot antwortet, wirft er Schatten auf die Wand. Er fungiert als eine Art Mediator, der die Schatten des Weltwissens präsentiert. So wie die Schatten in der Höhle stellt ChatGPT Information und Antworten bereit, die auf den vorhandenen Texten und dem realen Wissen basieren – aber all das sind lediglich Auszüge der Realität, nicht die Realität selbst. Ein Nutzer, der Fragen an ChatGPT stellt, befindet sich metaphorisch in dieser Höhle. Er sieht die Antworten oder Vorschläge von ChatGPT. Diese sind jedoch in der begrenzten Perspektive des Modells und des Trainings gefangen. Keinesfalls repräsentieren sie die absolute Wahrheit oder auch nur das vollständige Wissen der Menschheit, sondern einen mehr oder

weniger gut dokumentierten Ausschnitt. **Das vielfache Staunen über ChatGPT sagt wenig darüber aus, was es ist, aber sehr viel darüber, was wir hineinprojizieren.**²⁸³

Der Manager Scott Rosenberg bringt diesen Umstand auf eine einfache Formel: „*Wir vermenschlichen Elemente der Natur und abstrakte Formen – und auch Gespräche mit Computern.*“²⁸⁴

ChatGPT übersetzt Buchstaben, Phoneme und Phrasen in Zahlen.²⁸⁵ Anschließend sortiert es alles nach statistischen Gesetzmäßigkeiten. Wie das funktioniert, ist nicht so schwer zu verstehen. Um Sprache als statistisches Gebilde zu erfassen, verwenden die Entwickler von ChatGPT eine Technik namens *Worteinbettung*.^{286, 287} Das Ziel dieses Vorgehens lautet, sowohl die Syntax als auch die Semantik und Pragmatik von und zwischen den Worten und Phrasen einer Sprache in eindeutige Zahlenwerte zu überführen. Dieses Zuordnen geschieht nicht einmalig, sondern verändert sich durch jeden Text, mit dem das Sprachmodell gefüttert wird. Das Training dafür kann Tage aber auch Wochen dauern und kostet sehr viel Geld.²⁸⁸ Der Aufbau von ChatGPT beispielsweise soll bis zu 25 Mio. US Dollar verschlungen haben.²⁸⁹

Durch das Training entsteht Zug um Zug ein *Large Language Model (LLM)* (dt.: Großes Sprachmodell), das am Ende die erhoffte Leistung erbringt. Diese bemisst sich in der Anzahl seiner „Parameter“. Das meint die Menge an Werten, die zur Konfiguration und zum Training des Modells verwendet wurden. Parameter gelten als Goldstandard für die Leistungsfähigkeit eines Sprachmodells, vergleichbar mit der PS-Angabe eines Autos. Als Daumenregel gilt: Je mehr Parameter ein Sprachmodell hat, desto leistungsfähiger ist es.²⁹⁰ ChatGPT z. B. basiert auf dem Sprachmodell GPT-3, das 175 Milliarden Parameter umfasst.²⁹¹ Es gibt andere, die sehr viel mehr verarbeiten (Tabelle 3). Allerdings ist die Frage noch offen, ob diese Regel auch für noch größere Modelle gilt. Erste Stimmen vermuten, dass noch mehr der Parameter keinen Leistungszuwachs garantieren werden.²⁹²

Name	Erscheinungsjahr	Entwickler	Parameter
BERT	2018	Google	340 Mio.
XLNet	2019	Google	340 Mio.
GPT-2	2019	OpenAI	1,5 Mrd.
GPT-3	2020	OpenAI	175 Mrd.
GPT-Neo	März 2021	EleutherAI	2,7 Mrd.
GPT-J	Juni 2021	EleutherAI	6 Mrd.
Megatron-Turing NLG	Oktober 2021	Microsoft und Nvidia	530 Mrd.
Ernie 3.0 Titan	Dezember 2021	Baidu	260 Mrd.
Claude	Dezember 2021	Anthropic	52 Mrd.

Name	Erscheinungsjahr	Entwickler	Parameter
GLaM	Dezember 2021	Google	1,2 Trd.
Gopher	Dezember 2021	DeepMind	280 Mrd.
LaMDA	Januar 2022	Google	137 Mrd.
GPT-NeoX	Februar 2022	EleutherAI	20 Mrd.
Chinchilla	März 2022	DeepMind	70 Mrd.
PaLM	April 2022	Google	540 Mrd.
OPT	Mai 2022	Meta	175 Mrd.
YaLM 100B	Juni 2022	Yandex	100 Mrd.
Minerva	June 2022	Google	540 Mrd.
BLOOM	Juli 2022	Hugging Face	175 Mrd.
Galactica	November 2022	Meta	120 Mrd.
AlexaTM	November 2022	Amazon	20 Mrd.
LLaMA	Februar 2023	Meta	65 Mrd.
GPT-4	März 2023	OpenAI	unbekannt
Cerebras-GPT	März 2023	Cerebras	13 Mrd.
Falcon	März 2023	Technology Innovation Institute	40 Mrd.
BloombergGPT	März 2023	Bloomberg LP	50 Mrd.
PanGu- Σ	März 2023	Huawei	1,085 Trd.
OpenAssistant	März 2023	LAION	17 Mrd.
Jurassic-2	März 2023	AI21 Labs	unbekannt
PaLM 2	Mai 2023	Google	340 Mrd.
Llama 2	Juli 2023	Meta	70 Mrd.

Tabelle 3: Sprachmodelle und Anzahl ihrer Parameter²⁹³

Um ein Sprachmodell mit Texten zu füttern, muss zunächst ein Wortschatz aus dem Textkorpus extrahiert werden. Dann erhält jedes Wort im Vokabular auf der Grundlage seiner kontextuellen Beziehungen zu anderen Wörtern im Text einen numerischen Wert. Das werden wir uns nun anschauen am Beispiel des Aufbaus eines digitalen Touristenführers, der Ihnen für einen Wochenendurlaub in Berlin zur Seite stehen könnte.

ChatBER, Ihr Touristenführer

Sie planen eine Stadtrundfahrt durch Berlin. Auf der Liste stehen das *Brandenburger Tor*, der *Reichstag*, *Checkpoint Charlie*, eine Fahrt an den *Wannsee* und ein Anstieg auf

den *Teufelsberg*.²⁹⁴ Sie sind zum ersten Mal in Berlin, kennen sich nicht aus, brauchen also einen Touristenführer. Der ist digital, soll *ChatBER* heißen und wird sie als virtueller Touristenführer-App auf der Stadtrundfahrt ortskundig begleiten. Jetzt werden wir ihn – nur ausgestattet mit Büromitteln – in mehreren Schritten zusammenbauen.

Um *ChatBER* zu entwickeln, brauchen wir Text. Sehr viel Text. Das sind besonders solche, in denen wir plausiblerweise Daten oder Information über die Sehenswürdigkeiten von Berlin vermuten dürfen: Geschichtsbücher finden wir in der Bibliothek, Stadtpläne gibt es an der Tankstelle, Prospekte liegen im Touristenbüro, Zeitungs- und Magazinartikel verkaufen sich an Kiosken oder stapeln sich in Verlagsarchiven, Reiseberichte finden wir auf *Twitter*, *Facebook*, *Instagram* und *TikTok*. Wir haben bereits gelernt: Je mehr Text wir *ChatBER* füttern, desto leistungsfähiger wird der Chatbot später sein.

Sobald alles eingesammelt ist, bearbeiten wir diesen Korpus aus zehntausenden Textseiten. Als erstes erstellen wir einen Karteikartenstapel: Auf jede Karteikarte (bitte nur Größe A6, denn es werden sehr, sehr viele) notieren wir jedes Wort, das mindestens einmal im gesamten Textkorpus auftaucht. Obwohl wir Füllwörter bereits weglassen, zählen wir Millionen Karteikarten – und füllen damit nicht nur den Schreibtisch, sondern Umzugskisten. Mit denen ziehen wir um in eine Sporthalle.

Dort setzen wir uns in den Mittelkreis des Basketballfelds und fangen an: Wir greifen nach den erstbesten Karteikarten und legen sie nebeneinander: *Tourist* und *Öffnungszeiten*. Jetzt stellen wir Fragen: Wie oft werden beide zusammen in den Texten verwendet? Stehen sie dann nahe beieinander oder weit auseinander? Welche anderen Worte erscheinen regelmäßig in ihrer Umgebung? In welchen Satzkonstruktionen spielen sie oft eine Rolle? Aus diesen und vielen weiteren Fragen errechnen wir nach einer sehr komplizierten Formel eine Zahl. Sie bemisst den Abstand, in dem wir beiden Karten auf dem Hallenboden auslegen.

Wir entnehmen dem Stapel die nächste Karte: *Kinder*. Es liegen bereits zwei Karten auf dem Hallenboden. Jetzt wollen wir die Beziehung zwischen allen Worten verstehen. Erstens, die Beziehung zwischen *Tourist* und *Kinder*, zweitens die zwischen *Öffnungszeiten* und *Kinder*, drittens die zwischen *Tourist* und *Öffnungszeiten*. Nachdem wir alle Beziehungen gewichtet haben, verteilen wir alle Karten neu. Jede Entfernung zwischen ihnen repräsentiert nun ihre Abhängigkeit voneinander, wenn sie zusammen in einem Text auftauchen. So geht es weiter mit dem Kartenstapel. Tausendfach. Hundertausendfach. Millionenfach. Schemenhaft bildet sich auf dem Linoleum eine Art Landkarte heraus. Die Beziehungen zwischen Worten im Textkorpus werden plötzlich allein über ihre Entfernung sichtbar: Je näher Wörter zueinander liegen, desto öfter tauchen sie zusammen in Texten auf; je weiter entfernt, desto weniger häufig sieht man sie zusammen in Gebrauch.

Jetzt werden größere Strukturen von *ChatBER* sichtbar: An einigen Stellen klumpen sich Karten, die eng in Beziehung stehen (*Potsdamer Platz*, *Sony Center*, *Bahntower*).

Andere liegen isoliert herum oder haben augenscheinlich gar nichts mit anderen gemein, z. B. die Tippfehler *Brantenburger Tor* oder *Wahnsee*. Die Anzahl der gelegten Karten wächst unaufhörlich, und gleichzeitig schälen sich die direkten oder indirekten Bezüge aller Begriffe zueinander heraus.

Nach sehr, sehr, sehr langer Zeit wandert Information darüber, in welcher Beziehung alle Worte und Phrasen in allen Texten zueinanderstehen, hinein in das Sprachmodell ChatBER. Es lässt sich z. B. schnell erkennen, welche Worte oft gemeinsam auftauchen (*Brandenburger Tor* und *Reichstag*, *Reichstag* und *Kuppel*), mit welchen vorhersehbaren Wörtern diese häufig verbunden sind (*Tourist*, *Foto*, *Eintrittspreis*). Es lässt sich aber auch ableiten, dass es Zusammenhänge gibt, die sich nicht auf den ersten Blick erschließen (*Wannsee* und *Konferenz*²⁹⁵, *Checkpoint Charlie* und *Jeffrey Harper*²⁹⁶, *Teufelsberg* und *Graffiti*²⁹⁷).

So entsteht das Sprachmodell ChatBER und avanciert zum fachkundigen Touristenführer. ChatBER wird uns helfen mit jeder interessanten oder abwegigen Information zu einem Touristenpunkt, die überhaupt irgendwo dokumentiert worden ist. Er kann mehr Fragen beantworten, als selbst gut vorbereitete Touristen sich vorher hätten ausdenken können – oder menschliche Touristenführer zum Besten gäben. Aber ChatBER produziert auch Probleme. Und die sind allen so erzeugten Sprachmodellen gemeinsam.

Wir haben gelernt, dass die Sprachmodelle keine exakten Kopien der Trainingsdaten sind. Bei der Übertragung gehen Details verloren, die in den Ursprungsdaten noch vorhanden sind. Denn Sprachmodelle speichern lediglich, wie Worte und Phrasen üblicherweise zusammenhängen. Dieses Abstrahieren von Bedeutung führt zu Mustern, die die Realität nur unscharf abbilden oder bei dem Details verloren gehen.

Noch ein anderer Aspekt kommt jetzt zum Tragen: Ein Sprachmodell übernimmt kommentarlos die kognitiven Verzerrungen seiner Trainingstexte – und verstärkt diese sogar! Erklären lässt sich das an einem anderen Beispiel in Berlin: Wer den Techno-Club *Berghain* betreten möchte, muss an Türstehern vorbei.^{298, 299} Der Auswahlprozess von Türstehern hat weitreichende Auswirkungen auf die demografische Zusammensetzung der Besucher des Berghain. Denn die stehen vor der schwierigen Aufgabe, in kurzer Zeit zu entscheiden, wer den Club betreten darf und wer nicht. Ein Türsteher, der bereits eine bestimmte Vorstellung von einer Person oder einer Gruppe hat, könnte dazu neigen, Entscheidungen zu treffen, die diese Vorstellung bestätigen. Stereotypen und Vorurteile können ebenfalls eine Rolle spielen. Wenn der Türsteher beispielsweise aufgrund von Stereotypen glaubt, dass Menschen einer bestimmten ethnischen oder Altersgruppe eher problematisch sind, könnte dies zu einer Unterrepräsentation dieser Gruppen im Club führen. Auch der erste Eindruck einer Person kann als Anker für die Entscheidungen über nachfolgende Besucher dienen. Wenn die erste Person, die der Türsteher bewertet, beispielsweise sehr stilvoll gekleidet ist, könnte er für die nächsten Personen, die weniger gut gekleidet sind, strengere Standards anlegen. Zusätzlich könnten soziale Beweise und Gruppenzwang den Türsteher beeinflussen. Wenn der Club bereits viele Besucher einer bestimmten demografischen Gruppe hat, könnte er