

Was ist und wie funktioniert ChatGPT?

Versuch einer mathematik- und technikkfreien Rekonstruktion
anlässlich der 16. Thüringer Medienrechtlichen Gespräche

Prof. Dr. Clemens Beckstein
Institut für Informatik
Friedrich-Schiller-Universität Jena

Jena, den 9.11.2023

Was ist ChatGPT?

Ein **Software-Werkzeug, das Texte assoziiert:**

- gegeben einen triggernden Eingabetext, den sog. **Prompt**,
- produziert das einen dazu möglichst gut „**passenden**“ Ausgabertext.

Macht das nicht schon die **Google-Suche**?

Nur so was ähnliches, denn...

Texte sind bei ChatGPT

- sowohl auf der Eingabe- wie auf der Ausgabeseite
- **beliebige, „wohlgeformte“ sprachliche Konstrukte,**

z.B.:

- frei formulierte natürlichsprachliche Äusserungen,
- programmiersprachliche Fragmente,
- Musterbriefe, Tabellen und Formulare oder
- sprachliche Bildbeschreibungen.

Im Gegensatz zu Google liefert ChatGPT als Antworten **nicht einfach wörtliche Zitate** aus dem Netz.

Die Bibliothek von Babel

ChatGPT stellt den Anspruch, ein **für alle Anwendungen** passender Textassoziator zu sein.

Wie macht es das, wenn das doch nicht mal ein Mensch zuverlässig kann?

Der Mensch nutzt dafür u.a. **sprachlich kodiertes „Wissen“**, wie man es in Form von Büchern, Artikeln und anderen Veröffentlichungen in Bibliotheken findet...

Aber wie findet und erkennt er es da?

Jorge Luis Borges hat dazu mit seiner phantastischen Erzählung „**Die Bibliothek von Babel**“ (Borges 1992) schon lange vor dem Aufkommen erster programmierbarer Maschinen ein verblüffendes Gedankenexperiment angestellt:

Die Bibliothek in dieser Geschichte

- enthält **alle** ($10^{1.845.281}$) **Bücher** genau einmal, die man mit den üblichen Sprachzeichen „schreiben“ kann,
- also insbes. alle Bücher, die jemals
 - in irgendeiner Sprache geschrieben wurden oder
 - geschrieben werden könnten,

mit **allen denkbaren** Theorien, Geschichten, Gedichten und natürlich auch alle Informationen über sich selbst. aller **erdenklicher** Bücher?

Die Bibliothek birgt in sich damit im Prinzip **alle möglichen Antworten auf alle denkbaren Fragen.**

Suche nach Sinn

Trotzdem werden die meisten ihrer Bücher als **sinnlose Aneinanderreihungen von Zeichen** erscheinen:

- **Sinnvolle** (evtl. in einem, erst in der Zukunft erdachten Geheimcode geschriebene) Bücher sicher **von unsinnigen zu unterscheiden, ist ein prinzipiell unlösbares Problem.**
- Angesichts der gewaltigen Größe der Bibliothek ist es auch **praktisch nahezu unmöglich, in ihr Antworten auf spezifische Fragen zu finden.**

Und selbst wenn man ein Buch findet, das eine Antwort auf eine brennende Frage zu enthalten scheint, gibt es **keine Garantie, dass diese Antwort wahr oder zuverlässig ist:**

Für jede „richtige“ Antwort gibt es unzählige andere Antworten mit kleineren oder grösseren semantischen Abweichungen.

Deshalb braucht es **Bibliothekare**, die

- zu ordnen versuchen und
- bei der Suche helfen:

Einige von ihnen suchen nach einem bestimmten Buch („**Schlüsselbuch**“), das als Katalog oder Schlüssel für alle anderen Bücher dienen kann.

Sie entwickeln über die Jahrhunderte verschiedene **Theorien und Glaubenssysteme über die Natur und den Zweck der Bibliothek.**

Ohne die Bibliothekare ist die Bibliothek nichts.

ChatGPT als Bibliothekar oder Schlüsselbuch?

Eine entsprechende technische Realisierung müsste:

1. **den gesamten Inhalt** der „unendlichen“ Bibliothek digital speichern und
2. für die Beantwortung natürlichsprachlicher Anfragen **algorithmisch zugänglich machen.**

Ein prinzipielles Speicherproblem gäbe es dabei nicht ... ABER:

Selbst mit modernster Technologie wäre es **auch für so ein technisches System unmöglich,**

- auf beliebige Fragen an diesen ungeordneten Berg von sinnvollen und sinnlosen Büchern
- jeweils die **richtigen Antworten** zu geben.

Das würde sich schlagartig ändern, könnte man sich **beim Speichern auf den erwarteten „sinnvollen“ Teil der Bibliothek beschränken.**

Dann müsste man die Bibliothek „nur“ als eine riesige **Assoziations-Tabelle** digitalisieren, in der man,

1. gegeben einen Prompt, die ihn **thematisierenden Bücher nachschlägt** und dann
2. **deren Beiträge zum Thema** für die Generierung einer passenden Antwort verwenden.

Wie identifiziert man aber diesen Teil,

- wenn man doch in der Bibliothek gerade nach dem **Sinn** sucht, der benötigt wird,
- **um eine effiziente Suche danach erst zu ermöglichen?**

Sinn im Unsinn

Man begrenzt den **Sinnhorizont** auf das (vermutlich weitgehend sinnvolle), was die Menschheit

- **bis zu einem bestimmten Stichtag**
- in irgendeiner sprachlichen Weise

öffentlich (digital) verfügbar gemacht hat.

Bis zur jeweils nächsten Horizontziehung gibt es damit von so einem System allerdings

- weltweit nur eine, nicht dazulernende Instanz,
- mit dem für all seine Nutzer gleichen, konstanten „**Wissens**“stand (vulgo Kommunikationsverhalten).

Dieses System repräsentiert (**immer noch**) **ein gigantisches Informationsreservoir**.

Trotzdem bleibt darin - im Gegensatz zur Bibliothek von Babel - **alles an „Wissen“ unberücksichtigt, was**

- zum Zeitpunkt dieser Horizontziehung **bekannt, aber noch nicht publiziert** war oder
- **erst irgendwann in der Zukunft ersonnen** oder auch ersponnen werden könnte!

„Maschinelles Lernen“ - Programmieren durch Justieren

Wie konstruiert man nun eine Maschine,
die passende Assoziationen durchführt?

Man versucht dazu erst gar **nicht**,

- ein ungeheuer komplexes,
- vermutlich absolut undurchschaubares

Programm für einen herkömmlichen Rechner zu ersinnen.

Stattdessen verwendet man hierfür

- ein **(Berechnungs-)Modell**,
- das über **individuell einstellbare Parameter**
auf die konkrete Aufgabe einstellbar ist.

„**Programmieren**“ heißt damit dann,

1. ein passendes **Modell** zu **identifizieren / konstruieren** und
2. dessen Parameter
über ein **mathematisches Optimierungsverfahren**
so einzustellen, dass es
 - möglichst für alle Eingaben
 - die passenden Ausgabenproduziert.

Künstliche neuronale Netze als Standard-ML-Modell

Für diesen Zweck überwiegend genutzte Modelle sind spezielle, sehr **große künstliche „Neuronale Netze“**:

- So ein Netz kodiert jeweils eine ganz bestimmte, **hoch komplexe, parallele Berechnung entlang seiner Berechnungseinheiten („Neuronen“)**,
- die **über die „Gewichte“ der zahlreichen Verbindungen dieser Einheiten parametrisiert** wird.

Für deren Optimierung braucht man natürlich **geeignete Trainingsdaten**, also möglichst gutes **„Wissen“ darüber**,

- **wann sich das** aus der Anpassung resultierende **Modell** für eine gegebene Eingabe **passend verhält** und
- wann eben nicht.

Das Modell selbst

- spielt bei diesem Anpassungsprozess **keinerlei eigene, aktive Rolle**,
- es ist lediglich passiver Bearbeitungsgegenstand (Datum) des Optimierungsprozesses.

Zum Einsatz beim Nutzer kommt am Ende (simuliert auf einem üblichen Digitalrechner) **nur die spezifische, ausparametrisierte Instanz** des Ausgangs-Modells.

Dem Ergebnis dieses Konstruktionsprozesses ist prinzipiell nicht mehr ansehbar, ob bzw. dass es durch ML entstanden ist.

Was ist dabei ein „Lernerfolg“?

Wenn sich das Modell nach dem Training

- nicht nur auf den für seine Konstruktion verwendeten Trainingsdaten **passend** verhält,
- sondern, soweit **(statistisch) erwartbar**, auch **auf allen** anderen **möglichen Eingaben**.

Das **Trainingsergebnis** ist dann nichts anderes als eine algorithmische Repräsentation der **Tabelle der von den Trainingsdaten nahegelegten Assoziationen**.

Künstliche Neuronale Netze sind **effiziente Kompressoren** auch **riesiger Assoziations-Tabellen**.

Aber:

- je mehr man beim Training Wert auf eine starke Kompression (kleines und effizientes resultierendes Modell) legt,
- desto häufiger wird man mit **Kompressionsverlusten** rechnen müssen,

das System antwortet dann unpassend, es **konfabuliert**.

Im Falle von Textgeneratoren wie ChatGPT kommt erschwerend hinzu, dass **die generierten Antworten**

- unabhängig von ihrer inhaltlichen Angemessenheit
- **immer zumindest wohlgeformt** sein sollen,

was die Kompression zusätzlich deutlich erschwert und besonders für Verluste anfällig macht.

Schlüsselprobleme des Maschinellen Lernens

So gut wie nie hat man aber **vollständiges Wissen** über das richtige Verhalten des Zielsystems.

Für einen Teil der Eingaben, wird das System damit zwangsläufig

- nicht nur aufgrund der **hohen Kompression beim Training**,
- sondern vor allem auch für **Eingaben, die nicht Teil des Trainings waren**,

„falsche“ **Ausgaben** produzieren:

- es fehlt Information zur Ermittlung der passenden Antwort,
- das System muss trotzdem eine wohlgeformte Antwort liefern.

und dann bleibt halt nur noch **konfabulieren**.

Auch deshalb stellen fast alle modernen maschinellen „Lernverfahren“ sicher, daß am Ende zumindest

- **für ähnliche Eingaben**
- **ähnliche Ausgaben** produziert werden.

Das erhöht die Chance, dass sie sich auch für unbekannte Eingaben passend verhalten und verbessert außerdem deutlich ihren Kompressionsgrad.

Speziell **im Kontext von ChatGPT**

- ist dieser **Ähnlichkeitserhalt** sogar ein **Design-Kriterium sine qua non** und muss
- (mit einem riesigen Aufwand) über ebenfalls zu trainierende, sog. **word embeddings** sichergestellt werden.

Haupt Herausforderungen beim Training

1. Praktische Nicht-Erklärbarkeit ihres Verhaltens:

Eine **Erklärung**, wie wir Menschen sie verstehen, würde

- ein strukturiertes und
- zielgerichtetes

(„intelligentes“) **Vorgehen** des Systems bei der Generierung seiner Antworten voraussetzen,

- das es bei maschinell trainierten Antwortassoziatoren **so nicht gibt** und
- das **auch nicht** überzeugend **post hoc in die Antworten hineininterpretiert werden kann**.

Es ist damit also nicht nur a priori sondern auch a posteriori **nicht sicherstellbar**,

- daß die berechneten Antworten **korrekt**,
- geschweige denn (z.B. in irgendeinem gesellschaftlichen Konsens) **angemessen** sind.

Diese Nicht-Erklärbarkeit erschwert auch deutlich den Umgang mit dem zweiten Grundproblem.

2. „Bias“:

Systematische Verzerrungen in den Modell-Vorhersagen, u.a.

- a) bei der Auswahl der **Modellarchitektur** und des **Optimierungsverfahrens** und
- b) bei der Zusammenstellung der **Trainingsdaten**.

Nur wenn der Mensch beim Programmieren durch Justieren keine „Fehler“ macht, werden dabei

- **faire und genaue Modelle** entstehen,
- die **zuverlässig** für
 - verschiedenen Anwendungen und
 - verschiedene Gruppen von Menschen

funktionieren.

Die Konstruktion von ChatGPT

erfolgt in im wesentlichen **2 Phasen**:

Pretraining und Finetuning.

1. Pretraining:

Das Modell wird darauf trainiert, für einen gegebenen Prompt eine **sprachlich kohärente Antwort** zu generieren.

Nur orientiert an den digital vorliegenden Texten, wird hier

- für einen **beliebig vorgegebenen Diskursverlauf** und
- **jedes sprachlich theoretisch mögliche Wort**
(bzw. Token, bei GPT-4 gibt es davon etwa 50.000)

ausgerechnet, mit welcher **Wahrscheinlichkeit** dieses Wort das **nächste**, von ChatGPT geäußerte **Wort im Diskursverlauf** sein sollte.

Das zugehörige statistische Modell nennt man sein **Sprachmodell**, seinen **Generative Pretrained Transformer (GPT)**.

Autoregressive Textassoziation mit dem Sprachmodell:

Gegeben einen Prompt des Nutzers,
ergänzt ChatGPT mit Hilfe dieses Sprachmodells

- **solange** den soweit entwickelten Chatverlauf um je ein plausibles nächstes Wort,

bis die Folge dieser Wörter / Token aus Sicht des Sprachmodells eine **vollständige Antwort** darstellt.

Das (**weitgehend automatisch** ablaufende) Pretraining erfordert **riesige Textmengen**.

Die fraglichen Texte stammen insbes. aus **öffentlich zugänglichen Quellen aus dem Internet** – Bücher, Artikel, Websites, Forenbeiträge und viele andere Textformate.

Unter den sehr häufig verwendeten sind z.B. die Wikipedia, Webseiten-Archive, Diskussionsportale und andere große Textsammlungen.

Diese **Quellen werden für das Pretraining** von ChatGPT

- **technisch erfasst**,
- **bereinigt** und
- **tokenisiert** (in kleinere Texteinheiten zerlegt).

2. Finetuning:

Hierbei wird das Sprachmodell verfeinert, um ein **natürliches Dialogverhalten** zu befördern:

- Nutzung **aufgabenspezifischer Datensätze**, die passgenaue Antworten für spezifische Fragen enthalten oder aus besonders themenbezogenen Textmaterial bestehen.
- **Iteratives Verfeinern** der Modelle anstelle eines kompletten Neustrainings für die nächste Version des Textgenerators.

Die dafür notwendigen **Datensätze** müssen

- sowohl Eingabebeispiele
- als auch **Rankings (Label)** für prinzipiell passende Antworten enthalten.

Dieses Verfeinern erfordert einen **extrem zeit- und kostenintensiven**

Einsatz von Menschen (Clickworkern), die jeweils **über die Passung „entscheiden“** müssen:

Reinforcement Learning from Human Feedback (RLHF)

Mit dem Finetuning wird GPT zu ChatGPT.

Einflußnahme auf die Antwortgebung von ChatGPT durch den Nutzer

...im Prinzip auf drei Wegen:

1. durch das **Einstellen von problemunabhängigen Parametern vor Beginn des konkreten Chats,**
2. mit der konkreten, **problemspezifischen Ausgestaltung des Prompts** und
3. über ein **nutzerspezifisches Finetuning.**

Themenabhängiges Einstellen des Sprachmodells:

Temperatur:

- **höhere** Temperaturen steigern die Zufälligkeit der Antworten und führen zu weniger vorhersehbaren und (oft auch) kreativeren Antworten,
- **niedrigere** Temperaturen verringern die Zufälligkeit und erzeugen (oft) konsistentere, fokussiertere, aber auch konservativere Antworten.

Zu hohe Temperaturen (Fieber!) führen schnell zum **Halluzinieren**.

Beschränkung der Länge der vom Nutzer akzeptierten Antworten auf eine bestimmte Anzahl von

- Wörtern / Token (Max Token) oder
- (unabhängig davon, wieviele Token damit verbunden sind) Zeichen.

Festlegung von Stoppwörtern oder bestimmten Begriffen, die nicht in den generierten Antworten vorkommen sollen (nutzerspezifisches Alignment).

Vorgabe einer Anzahl von Versuchen, die ChatGPT durchführen darf, bevor es dann aus den entsprechenden Antworten die beste Antwort wählt und präsentiert.

Ein klarer, präziser und gut formulierter Prompt

kann das Modell dazu bringen, Antworten

- auf **im Dialog vorausgegangene Fragen** oder
- auf **spezifische (Dialog-)externe Wissensquellen** zu beziehen.

Auch die explizite Vorgabe

- des gewünschten **Eingabe-Sprachmodells**
(z.B. Englisch, Deutsch, Python oder Excel),
- der jeweiligen **Zielsprache**
(z.B. Französisch, Latein, SQL oder Cobol) oder
- des angestrebten **Ausgabestils**
(z.B. Politikersprech oder Juristendeutsch)
bzw. des Ausgabeformats
(z.B. Zusammenfassung, Mustertext, Rede oder Präsentation),

kann die Qualität der Antworten verbessern.

Nutzerspezifisches Finetuning

- für eine **konkrete Aufgabenstellung oder einen bestimmten Anwendungsbereich**
- unter Ausnutzung der von OpenAI zur Verfügung gestellten **Programmierschnittstelle (API)** zu ChatGPT.

Beispiele:

- medizinische Texte,
- rechtliche Dokumente oder
- Kundensupport-Interaktionen in einer bestimmten Branche.

Tradeoff von Aufwand und Nutzen,

denn dieses Finetuning erfordert (in Art und Umfang) geeignete Trainingsdaten.

Historische Entwicklung des Sprachmodells von ChatGPT

GPT-1 (Juni 2018, 117 Millionen Parameter, Tokenlimit: 500):

GPT-1 war das **erste Modell** in der GPT-Reihe und markierte den Beginn der Erforschung von Transformer-Modellen (Erstpublikation 2017) **für die Textgenerierung**.

GPT-2 (Februar 2019, 1,5 Milliarden Parameter, Tokenlimit: 1.000):

GPT-2 erregte aufgrund seiner überraschend hohen Qualität bei der Textgenerierung und angesichts der damit schon damals erkennbaren potenziellen **Missbrauchsgefahr** hohe Aufmerksamkeit.

Die erste Version von GPT-2 (mit 125 Millionen und dann 355 Millionen Parametern) wurde aufgrund von Sicherheitsbedenken zunächst zurückgehalten und die finale Version erst nach einer internen Debatte von OpenAI **verzögert freigegeben**.

GPT-3 (Juni 2020, 175 Milliarden Parameter, Tokenlimit: 2.000):

Bereits GPT-3 bietet **beeindruckende Fähigkeiten in Bezug auf Textverständnis, Übersetzung und Textgenerierung**.

Übergangsmoell GPT-3.5 (November 2022, 175 Milliarden Parameter, Wortlimit: 4.000):

GPT-3.5 kann auch mit Texten umgehen, die **Programmcode** darstellen.

GPT-4 (März 2023, Tokenlimit: 4.000, später 32.000):

GPT-4 ist eine noch mal verbesserte Version von GPT-3.5:

- nicht nur ein Text-zu-Text-Modell wie seine Vorgänger, sondern ein **Daten-zu-Text-Modell** (erlaubt auch Bilder und andere Datenarten als Eingabe),
- hat wohl auch deshalb ein rund zehn mal **grösseres Wortlimit**,
- verhält sich in seinen Antworten „**politisch korrekter**“ und
- **konfabuliert angeblich weniger** aufgrund einer deutlich grösseren Datengrundlage und eines konsequenteren Finetunings.

Soweit (inoffiziell, Geschäftsgeheimnis OpenAI) bekannt, weist es 1,8 Billionen Parameter auf, verteilt auf

- 16 Teil-"Expertensysteme" (dedizierte GPT-3s) mit jeweils 110 Milliarden Parametern und
- einen 55 Milliarden Parameter schweres Attentionssystem (Erstpublikation des Mechanismus bereits 2016) zur Fokussierung auf die relevanten Teile der Eingabe.

GPT-4V (November 2023, Tokenlimit in GPT-4 Turbo: 128.000):

Weiterentwicklung vom großen Sprachmodell (LLM) GPT-4 zum großen multimodalen Modell (LMM):

- GPT-4V kann **Text, Bilder, visuelle Zeiger, Szenentext und beliebige Kombinationen als Eingaben** verarbeiten. Es erzeugt **textliche Ausgaben**.
- Es befolgt effektiv Textanweisungen, lernt aus Beispielen, „**versteht**“ **visuelle Verweise und vollzieht visuelle „Schlussfolgerungen**“.

Quo vadis, ChatGPT?

Ingenieurmässige Herausforderungen:

- **Kostengünstigeres** und zugleich **effizienteres GPT-4** (plus zugehöriger API),
- **größere Kontextfenster** (Tokenlimit: 128k -> 1 Mio)
- Anpassung der **Finetuning-API** an die neuen, erweiterten Modelle,
- Bereitstellung einer **API mit Gesprächsgedächtnis**,
- konfigurierbare, **GPT-basierte Konfektions-ChatBOTS**.

Integration von **Drittanbieterdiensten und APIs**, um

- auf **ChatGPT-externe Datenquellen** (z.B. das Internet und Datenbanken) zuzugreifen und
- (z.B. über **Plugins**) spezielle Funktionen für Berechnungen oder Analysen zu realisieren, die über die natürliche Sprachverarbeitung hinausgehen (z.B. über das Mathematik-System Wolfram).

Zugriff auf diese Plugins auch über die API.

Der feuchte (**Alp?**)**Traum** der Jünger des Maschinellen Lernens:

ChatGPT als recursive, self improving and embedded **Artificial General Intelligence (AGI)**.

Konzeptionelle Herausforderungen:

Noch stärker **multimodale Modelle**, die

- nicht nur Text (wie GPT-4) und Vision (wie GPT-V4),
- sondern auch andere Datenmodalitäten wie **Audio und Video** verarbeiten können.

Das größte Problem dabei ist wohl,

- **wie genau die einzelnen Teile** des resultierenden (Hybrid-)Systems **interagieren müssen**,
- damit das **Gesamtsystem mehr als nur die Summe seiner Teile** wird

und wie man dann das **Gesamtverhalten erklärt**.

ChatGPT als supersmarter Assistent.

ChatGPT – Extrem nützlich, aber auch intelligent?

Wo findet man im heutigen ChatGPT oder bei seiner Entwicklung „Intelligenz“
(das, was eigentlich den Menschen ausmacht oder braucht)?

1. Im digitalisierten Teil des **verschriftlichen kulturellen Erbes** der Menschheit und in den Unmengen an Daten, die bisher und laufend von den Menschen abgeschnorchelt werden,
2. im Blut und Schweiß der Heerscharen schlecht bezahlter **Clickworker**, die diesen Daten beim Finetuning erst Sinn geben,
3. in den Köpfen der **Wissenschaftler und Ingenieure**, die die ChatGPT zugrundeliegende Technologie weiterentwickeln und
4. in den öffentlichen Institutionen, die sich durch **Vorgabe ethischer und politischer Rahmenbedingungen** ein gesellschaftskonformes **Alignment** und einen **verantwortungsvollen Einsatz** von ChatGPT erhoffen.

In ChatGPT selbst ist

- **trotz** seiner gar nicht überschätzbaren **universellen Nützlichkeit**
- nur **alles künstlich aber nichts intelligent**.

Ohne den Menschen ist (wird) ChatGPT nichts.

Zumindest **Stand heute ist ChatGPT** einfach

- **ein äußerst belesener,**
- **extrem gut trainierbarer** und
- gerade deshalb **äußerst nützlicher**

stochastischer Papagei (Bender u. a. 2021),

dem die Ingenieure zusätzlich ein

- einfühlsames und
- insbes. **politisch korrektes**

Sprechen über den Inhalt des Web beigebracht haben.

Bei **Google** antwortet ein **Mensch**,

bei **ChatGPT** die **statistisch gemainstreamte Menschheit**.

Könnte eine sprechende Maschine Mensch sein?

Descartes in „Abhandlungen über die Methode“ (1637):

„Jede Maschine ist naturgemäß beschränkt in dem, was sie tun kann. Während der menschliche Körper auch als eine Maschine betrachtet werden kann, könnte er nie durch eine künstliche Maschine nachgebildet werden. Dies liegt daran, dass eine Maschine niemals über genügend Vielfalt verfügen würde, um als Mensch zu agieren...

Wenn Maschinen die Fähigkeit hätten, in irgendeiner Weise zu sprechen, wie wir es tun, dann könnten sie einem jeden Menschen in allen anderen Dingen antworten; und sie würden nicht nur in gewissen Dingen schlechter als der gemeinste Mensch agieren..."

Frage an ChatGPT (den Descartes-Prompt ergänzend):

„Ist insbesondere der zweite Teil dieser Behauptungen von Descartes im Zeitalter von großen Sprachmodellen (LLMs) a la ChatGPT überholt? Begründe bitte Deine Antwort.“

ChatGPTs Replik

„LLMs generieren Antworten auf Grundlage ihrer Trainingsdaten und können nicht wirklich „lernen“ oder sich an Entwicklungen nach ihrem letzten Training anpassen. Zudem fehlt ihnen die Fähigkeit zur Intuition und kreativen Problemlösung.

Menschliche Kommunikation bezieht moralische und ethische Überlegungen ein, die auf individuellen Wertesystemen basieren. LLMs können Ethik und Moral lediglich auf Grundlage ihrer Trainingsdaten simulieren, besitzen jedoch kein eigenes Wertesystem.

Menschen können unvorhersehbar und abseits von etablierten Mustern reagieren, während LLMs aufgrund ihrer Konstruktion deterministisch sind.

LLMs können deshalb die Tiefen und Nuancen der menschlichen Kommunikation, die über bloße Sprache hinausgehen, noch nicht vollständig nachbilden. In dieser Hinsicht bleibt Descartes zumindest teilweise gültig."

Meine Damen und Herren,
ich bedanke mich für Ihre Aufmerksamkeit.

Dank an die kreativ am Vortragsdesign Beteiligten:

- **MSc. Johannes Mitschunas** - Doktorand in der AG KI
- **Daniel Motz** - Studentische Hilfskraft in der AG KI

- **Dr. Beate Mitschunas** - Fotografie

- OpenAI für **ChatGPT und Dall-E 3**
- Stability AI für DreamStudio (**Stable Diffusion**)

Bibliographie

Bender, Gebru, McMillan-Major, und Shmitchell. 2021. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* . New York, NY, USA: Association for Computing Machinery

Borges, Jorge Luis. 1992. „Die Bibliothek von Babel“. In *Fiktionen: Erzählungen 1939 - 1944*. Fischer Verlag