

Erster Teil

Was heißt (künstliche) Intelligenz?

1. Worum geht's?

Dieser Teil des Buches besteht aus einem Abschnitt, in dem der Begriff Künstliche Intelligenz in seiner heutigen Verwendung entmystifiziert wird, um in einen zweiten Abschnitt zu münden, in dem – leider – der allgemeinere Begriff „Intelligenz“ remystifiziert werden muss. Diese Remystifizierung umfasst indes ganz unmystisch, dass begriffliche und empirische Probleme der Anwendung des Begriffs in Alltag und Psychologie benannt und diskutiert werden. Sie werden nicht nur in der Euphorie für, sondern auch in der Ablehnung von neuer Technik oft vergessen, sondern stellen den Begriff der Intelligenz *als Eigenschaft von einzelnen Organismen*, und noch mehr *als messbare Eigenschaft* überhaupt in Frage.

Im zweiten Teil des Buches erst versuche ich dann, empirische Belege für Intelligenz in einem neuen, nicht-normativen Sinn zu bringen, und belege diese mit der scheinbar einzigartigen Fähigkeit des Menschen zur *Vorstellungstätigkeit und Phantasie*, die wiederum tiefer als das Denken in unseren Motiven wurzelt.

Der allgemeine Hintergrund, warum es eine Intelligenzdebatte gibt und sie auch nützt, ist schnell erklärt. Zahlreiche Worte, die in der alltäglichen Kommunikation gut funktionieren, weil sie dem Gegenüber ermöglichen, seine Handlungen auf unsere abzustimmen, erweisen sich, wenn man sie begrifflich zu fassen, also zu definieren versucht, umgehend als äußerst problematisch.

Viele dieser Worte sind psychologischer Art. Wenn ich jemandem zum Beispiel sage „Ich sehe dort ein Schloss“, so ist für ihn, vorausgesetzt ein Schloss könnte sich irgendwo in der Nähe befinden, leicht zu verstehen, was ich meine. Unsere Erwartungen werden sich aufeinander einstellen. Fragt er aber nach „Was ist ein Schloss?“ oder, schlimmer, „Was meinst du mit sehen?“, so weiß ich keine schnelle und einfache Antwort und ein

Orkus an Problemen tut sich auf, auch hinsichtlich der gegenseitigen Erwartung. Ich könnte auf der pragmatischen Ebene auf das von mir gemeinte Schloss zeigen oder Merkmale des Schlosses oder des Sehens aufzählen – und die aktuelle „künstliche Intelligenz“ tut dies, indem sie „Sprache“ auf die statistisch regelhafte Aufeinanderfolge von Worten beschränkt. „Sehen heißt, etwas mit den Augen erfassen“, wäre eine Antwort – die jedoch das Definitionsproblem nur auf weitere Worte verlagert. Für die Theorie des Sehens ist mit der Antwort in psychologischer Sicht so wenig gewonnen wie durch die Worte für das Problem des Aussprechens von Gedanken. Und eben diese Gedanken, und nicht etwa das Sprechen, sind bei Menschen die Voraussetzung für „Intelligenz“. Auch dieses Wort ist, wie wir sehen werden, trotz der vielen Versuche, es pragmatisch zu definieren und damit für psychologische Experimente an Versuchspersonen zu operationalisieren, begrifflich wackelig geblieben.

Doch beginnen wir mit dem aktuellen Hype, der Angst und Euphorie hinsichtlich neuer Computerprogramme, die unter dem Titel „Künstliche Intelligenz“ beworben werden.

Seit Herbst 2022 ist es schwierig, der öffentlichen Debatte um so genannte *Künstliche Intelligenz* zu entgehen. So selten wissenschaftliche und technische Themen sonst Schlagzeilen in der Presse machen, so häufig erreichen sie uns nun bei diesem Thema, bei dem Hoffnungen, nämlich nicht nur die von Ingenieur*innen und Investor*innen, sondern auch die von politisch Verzweifelten, aber auch Ängste, nämlich die derjenigen, deren Jobs, Freuden und Lebensgefühl von der neuen Technik bedroht scheinen, aufeinanderprallen.

Zunehmend fällt mir auf, dass in solchen Großdebatten immer noch schneller beherzt losgeredet wird, bevor man halbwegs klar umreißen kann, *worüber* genau geredet wird. Dies freilich ist ein allgemeines

Merkmal eines Umfelds, in dem alle Informationen und alle Meinungen leicht zur Hand scheinen und sich dergestalt zu einer „Themakatastrophe“ verdichten. Eine solche ereignet sich, wenn unterschiedliche, aber sich überschneidende Probleme aus unterschiedlichen Perspektiven gleichzeitig besprochen werden und dabei von Perspektive zu Perspektive gesprungen wird. Offensichtlich überschneiden sich beim Thema „Künstliche Intelligenz“ tatsächlich Technik, Psychologie, genauer die so genannten Kognitionswissenschaften, Soziologie, Linguistik, aber auch ganz praktisch Politik und Psychopathologie.

Ich versuche also im Folgenden möglichst verständlich und möglichst ohne technische Geheimniskrämerei und gleichzeitig ohne geisteswissenschaftliche Gegenpolemik auf die Technik dessen, was *im Moment* Künstliche Intelligenz *genannt* wird, einzugehen, und dann diese Vorstellungen mit dem, was wir im Alltag und was auch die Psychologie als „Intelligenz“ zu verstehen versuchen, zu vergleichen. Ich tue dies, so gut es mir möglich ist, ohne zwischen Perspektiven und Themen hin und her zu springen.

Sie werden sehen, dass dies nicht immer ganz gelingt, denn der Computer ist ja, wie sein Name „Rechner“ schon sagt, selbst eine grobe Verallgemeinerung *eines* empirischen Aspekts dessen, was wir umgangssprachlich „Denken“ nennen, nämlich algebraisches Rechnen. Außerdem: Wenn wir ein so klares Modell eines Ausschnitts der Welt besitzen, dass wir mit ihm eine Klasse von Beobachtungen vorhersagen können, so kann dieses Modell als Folge von „Wenn-dann“-Regeln von einfachen Grundaussagen („Prädikaten“) wie in der formalen Logik abgeleitet werden. Denken scheint dann als schrittweise Ausführung dieser Regeln darstellbar zu sein.

Entscheidend für mein Argument wird also sein, dass diese Logik der oberflächlichste Aspekt des Den-

kens ist, denn die Zuordnung der Beobachtungen zu den Aussagen mag, sobald wir etwas klar verstanden und als Modell formuliert haben, „evident“ scheinen, beruht aber auf der sehr komplizierten Interaktion unseres Beobachtungsapparats mit der Umwelt (Raab 2024), in dem auch Funktionen wie „Vorstellungsbilder“ und „Phantasien“ zwingend, soll heißen biologisch, notwendig sind.

Desgleichen hat sich das aktuelle Modell der KI – künstliche neuronale Netze und ihr maschinelles Lernen – von der genannten klassischen algorithmischen Vorstellung des Denkens seit Längerem abgewandt (Brooks 1991). Dennoch verwenden diese Programme, wie alle Programme, denn das ist ihre Definition, Algorithmen. Dies ist für viele, die die Debatte verfolgen, als Erstes widersprüchlich. Lassen Sie uns also kurz mit der Funktionsweise generativer KI beginnen.

2. Wie macht der große Computer was?

Ich will kurz die relevanten technischen Themen wie Statistik, neuronale Netzwerke, maschinelles Lernen und *Large Language Models* besprechen. Zweck der Darstellung ist, dass Sie als Lesende/r aus einfachen Grundgedanken ableiten können, wie Programme, die unter dem Namen „KI“ laufen, ihren Output aus in ihnen ablaufenden Prozessen und dem gegebenen Input errechnen. Dies soll erstens das Unbehagen an technischen Geräten, die man nicht klar versteht, lindern, und zweitens einen Hintergrund für die folgenden Überlegungen, was Intelligenz sein könnte, geben.

Dabei bleibe ich weitgehend beim Beispiel des Textgenerators ChatGPT der Firma OpenAI und nicht etwa auf Text-zu-Bild-Generatoren wie Stable Diffusion, weil ich beruflich mit Sprache arbeite, als Übersetzer von meiner eigenen Obsoleszenz durch Übersetzungssoft-

ware bedroht bin – und weil man die Funktionsweise der neuen, erfolgreichen Programme, Transformer genannt, an diesem Beispiel einfacher darstellen kann.

Das Grundprinzip ist Folgendes: Das Programm liest eine Zeichenkette ein, die der Mensch als sprachlichen Ausdruck, also als Worte, versteht. Es setzt diese Zeichenkette sodann mit einem Wort, dem Beginn seiner „Antwort“, fort, indem es dessen Wahrscheinlichkeit auf Grundlage eines Korpus bereits existierender Texte berechnet. Das nächste Wort wird auf Grundlage des bereits generierten Wortes sowie der Korpusstatistik berechnet usw., bis das Modell einen „Stop-Token“ errechnet und die Ausgabe beendet.

Dies ist im Kern alles, was jeder Chatbot, also ein Computerprogramm, mit dem Benutzer*innen durch Sprache interagieren, macht. Auch alle anderen heute bekannten KI-Programme, zur Bildgenerierung etwa, basieren auf neuronalen Netzen, maschinellem Lernen, die durch statistische Berechnung über große Datenmengen den Output generieren.

Die statistischen Methoden und Möglichkeiten indes haben sich, zum großen Teil durch Kapazitätsverbesserungen der Hardware, stark verändert. Diese technischen Verbesserungen am so genannten „compute“, der Rechnerleistung, sind in der Tat eindrucksvoll. So haben sich, wie in Abbildung 1 dargestellt, die Anzahl der Gleitkommaberechnungen oder FLOPS sowie das Speichervolumen seit 2012 alle 3,4 Monate verdoppelt.

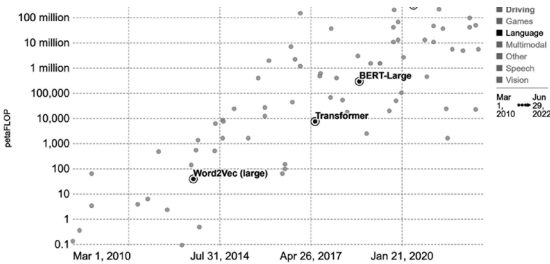


Abb. 1. Logarithmisch dargestellte Steigerung der für fünf Sprach-KI-Systeme benötigten Rechnerleistung oder „compute“ in Petaflop-Tagen 2010 bis 2022 (aus Sevilla et al. 2022). Ein Computersystem mit einer Rechenleistung von einem Petaflop/s kann eine Billion (10¹⁵) Gleitkommaoperationen pro Sekunde, also etwa 1020 pro Tag ausführen. Das Produkt aus Rechenzeit und Zeit dient als gedankliche Hilfe, ähnlich wie bei der kWh für Energie. Die Verdoppelung des „compute“ erfolgt im dargestellten Zeitraum bei bester Anpassung an die Messungen alle 3,4 Monate (OpenAI 2018).

Diese Verbesserung der rohen Rechenleistung geht mit einer Verbilligung der Kosten pro Petaflop aufgrund der Konkurrenz der Hardwarefirmen einher. Zugleich wächst allerdings der Stromverbrauch der Systeme sehr stark an.

Die Verbesserungen der Hardware sind deswegen wichtig, weil die statistischen Berechnungen in neuronalen Netzen prinzipiell anders durchgeführt werden als in der klassischen Statistik, wenngleich die Deutung des Outputs als statistische „beste Schätzung“ bleibt.

In der klassischen Statistik sind Wahrscheinlichkeiten oder deren Verteilungen Variablen zugeordnet, die man mit oder ohne Hilfsmittel beobachten, messen und mit Zahlen belegen kann. In den von künstlichen neuronalen Netzen berechneten Statistiken werden die Information solcher Variablen auf Netzwerkknoten, die Verbindungsstärke zwischen Netzwerkknoten sowie Grenzwertfunktionen, die definieren, wann Information von Knoten zu Knoten weiterverrechnet wird, „verteilt“.

Wir beginnen mit der klassischen Statistik, nicht nur weil die früheren KI-Programme, die mit dem genannten „alten“ Algorithmusbegriff und Variablen, die man außermathematisch *durch Beobachtungen* deuten konnte, operierten, diese verwendeten. In heutigen neuronalen Netzen wird die statistische Information „verteilt“, weswegen die Tatsache ihrer statistischen Funk-

tionsweise nicht mehr so leicht zu erkennen ist. Kurz, in der klassischen Statistik sind die definierten Variablen *empirisch* deutbar, in neuronalen Netzen nicht.

Die Grundlage jedes *Chatbots* sind also Worthäufigkeiten. Abbildung 2 zeigt als Beispiel den historischen Vergleich je zweier Worthäufigkeiten über etwa die letzten 220 Jahre, wie sie aus dem deutschsprachigen Google-Korpus, bestehend aus Millionen automatisch eingeleseener Bücher, errechnet wurden.

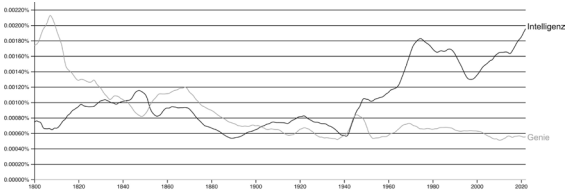


Abb. 2. Relative Häufigkeiten der Worte „Intelligenz“ und „Genie“ im deutschsprachigen Google-Korpus von 1800 bis 2019 mit Glättung über drei Jahreswerte. Deutung: Eine Wahrscheinlichkeit von 0,00100 von Intelligenz im Jahr 1800 bedeutet, dass jedes hunderttausendste Wort im Korpus dieses Jahres „Intelligenz“ ist (von <https://books.google.com/ngrams>)

Die Häufigkeit eines Wortes ist seine Anzahl geteilt durch die Anzahl aller Worte jeweils im Korpus der betrachteten Zeit. In der klassischen Statistik ordnet man also jedem Wort eine allgemeine Wahrscheinlichkeit seines Auftretens zu. Die Wahrscheinlichkeit des Folgeworts eines gegebenen Worts ist indes eine „bedingte Wahrscheinlichkeit“, die vom vorigen Wort abhängt, wie man sich problemlos klarmachen kann. Wenn das erste Wort „der“ lautet, so ist die Wahrscheinlichkeit, dass ein männliches Hauptwort folgt, höher als dass ein weibliches folgt, obwohl „der“ auch Relativsätze wie „der gestern bei mir war“ einleiten kann.

Diese bedingten Wahrscheinlichkeiten von Folgeworten ließen sich in einer Tabelle ausdrücken, in der die Folgewahrscheinlichkeiten jedes Worts auf jedes

Wort verzeichnet ist. Sie wäre nicht symmetrisch, da z.B. die Wahrscheinlichkeit, dass auf „der“ das Wort „Berg“ folgt, größer sein wird, als dass „der“ auf „Berg“ folgt usw. Angenommen, die deutsche Sprache umfasste, wenn man die Bildung und Aufnahme von neuen Worten vernachlässigt, 400.000 Worte, so wäre dies eine recht große Tabelle mit 400.000 mal 400.000 Folgewahrscheinlichkeiten. Sie stellte, könnte man sagen, das flachste „Sprachmodell des Korpus“ dar. Dies erwähne ich, um deutlich zu machen, wie das Wort „Modell“ in der Informatik beim Ausdruck „Large Language Model“ verwendet wird. Es reduziert die rein kombinatorischen Möglichkeiten, die ohne den Einbezug eines bedingten Kontexts riesig wären. Nimmt man einen Wortschatz von 400.000 und eine Satzlänge von 10 Worten an, so ergeben sich bereits mehr als 1049 verschiedene Sätze – eine riesige Zahl. Das Wort Modell hat aber nichts mit Theorie zu tun, denn diese müsste ja beschreiben, *wie* es zu den Folgewahrscheinlichkeiten überhaupt gekommen ist.

Da in dieser Tabelle alle Worte eine Folgewahrscheinlichkeit haben, lässt sich mithilfe bedingter Wahrscheinlichkeiten nun das wahrscheinlichste nächste Wort durch die Wahrscheinlichkeiten der zuvor eingelesenen oder generierten Worte errechnen. Man kann leicht einsehen, dass dies die kombinatorischen Möglichkeiten bereits enorm einschränkt.

Diese Regel lässt sich nun auf das jeweils vorige Wort erweitern. Das nächste Folgewort errechnet sich nun aus den Wahrscheinlichkeiten des Wortes zuvor, das sich wiederum auf jener des Wortes davor errechnet usw. Ich denke, Sie sehen das Problem. Der Rechenaufwand, um die Folgewahrscheinlichkeit um jeweils noch ein Wort zu errechnen, wächst exponentiell an. Der „Kontext“ des Wortes muss daher, um nicht die Rechenleistung der Hardware zu überschreiten, auf eine fix definierte Anzahl vorheriger Wörter beschränkt werden. Diese

ergeben ein Berechnungsfenster, das gleichsam „über die Wortkette“ geführt wird. Dieses Fenster heißt nur im Jargon „Kontext“, meint es doch etwas Verwandtes, aber nicht Gleiches, wie im normalen Wortgebrauch, wo der Kontext den Sinn des Wortes bestimmt. Diese Art Sinn oder Richtung oder Ziel gibt es nicht in der Informatik, worauf ich in Abschnitt 4 anhand eines Versuchs zu sprechen komme.

Diese technische Einschränkung ist im Wesentlichen der Grund, warum ältere Chatbots sehr schnell „das Thema verloren“, wenn sie nicht, wie bei der berühmten Eliza von Weizenbaum (1966), durch Regeln, die sich der Sprechgewohnheiten der Nutzer*innen annäherten, unterbrochen wurden. Die Regel „frage irgendetwas Persönliches, wenn der Satz zu lange wird“, zum Beispiel werden Sie aus Ihrem Alltag kennen: Wenn man etwas nicht versteht, ist es besser *irgendetwas* zu fragen, als weiter verwirrt zu bleiben.

Ein ähnliches Prinzip kennen Sie vermutlich, wenn Sie die automatische Textvervollständigung bei der Suche mit Google nicht ausgeschaltet haben. Die wahrscheinlichste Fortsetzung des Beginns Ihrer Eingabe wird entsprechend der bereits von anderen erfolgten Eingaben wie in Abbildung 3 angezeigt.

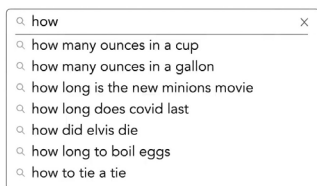


Abb. 3. Statistische Wortergänzung im Suchprogramm Google.

Mit so genannten *künstlichen neuronalen Netzen* kann man diese Rechenbeschränkung lindern, indem der „Kontext“ im Datenkorpus durch ingeniose Tricks erweitert wird. Am besten stellen Sie sich das so vor,

dass ein Wort, das 100 Worte vor dem aktuellen liegt, wiederum die Wahrscheinlichkeit eines Wortes verändert, das 1.000 Worte entfernt liegt usw. Diese rein mathematischen Abhängigkeiten im Korpus, die insofern über die klassische Statistik hinausgehen, als das neuronale Netz die Wortfolgestatistiken „verteilt“ auf Vektoren und Matrizen kodiert, die für Menschen inhaltlich nicht deutbare Regelmäßigkeiten im Korpus abbilden, sind das Fundament der so weit verkomplizierten und durch weiteren Gebrauch laufend adaptierten Wortfolgestatistik.

Die Wortfolgestatistik wird also vorderhand verkompliziert und undurchsichtiger, dafür aber zugänglich für zahlreiche mathematische Manipulationen, die im Wesentlichen auf der Addition, Subtraktion, Multiplikation, Division und Normalisierung von Vektoren und Matrizen beruhen. Als Vorstellungshilfe zeigt Abbildung 4 eine grafische Darstellung eines künstlichen neuronalen Netzes, das mit einer Eingabeschicht, drei verborgenen Schichten und einer Ausgabeschicht mathematisch wohlgeordnet ist. Die Einträge in diesen Vektoren und Matrizen simulieren Messungen der Verbindungsstärken *zwischen* und Erregungsfunktionen *in* den Knoten, die sehr grob Neuronen idealisieren.