Vortrag „So funktioniert ChatGPT“
vom 03.04.2023, Helmut Linde

Zusammenfassung der Fragerunde

Ich habe eine grundsätzliche Verständnisfrage in Richtung der Mehrsprachigkeit von ChatGPT. Ich kann meine Anfragen ja auf Deutsch, Englisch, Französisch, Spanisch und so weiter formulieren: wann wird das wo und wie im semantischen Raum und im Kontextraum berücksichtigt und umgesetzt? Ich bin kein Linguist, aber ich kann mir vorstellen, dass es bei den Sprachen in diesen Bereichen Unterschiede gibt.

Genau. Erstens wird das Modell mit Texten aus allen möglichen Sprachen trainiert – tendenziell funktionieren die Modelle dann auf Englisch besser als auf „exotischen“ Sprachen, weil es mehr Trainingsdaten gibt. Aber dadurch, dass man irgendwann auf einem sehr hohen Abstraktionslevel ist, lernen die Modelle sicherlich auch, von einer Sprache in eine andere zu übersetzen. Das heißt: Wenn ich in einer exotischen Sprache Fragen stelle, werde ich vermutlich auch gute Antworten bekommen, auch wenn es in der Sprache vielleicht gar keine Trainingsdaten zu diesem speziellen Thema gab.

Wie das ganz genau durch diese Räume läuft und an welcher Stelle die Daten aufgeteilt werden, ist schwer zu sagen, da die Modelle – wie bereits gesagt – extrem intransparent sind. Mein *best guess* wäre, dass in den untersten Ebenen die Worte schon einmal möglichst nah aneinander sortiert werden: Wenn ich das spanische Wort *casa* habe und das deutsche Wort Haus, werden die beiden Wörter im semantischen Raum relativ nahe zusammen liegen, aber vielleicht nicht unbedingt deckungsgleich sein, da sie in unterschiedlichen Situationen unterschiedliche Konnotationen haben. Meine Vermutung wäre, dass die Unterschiede zwischen diesen Wörtern, je weiter ich in das Modell gehe, von Schicht zu Schicht die sprachlichen Unterschiede immer geringer werden.
Irgendwann wird man dann ein Konzept „Haus“ haben, das von der Sprache relativ unabhängig ist.

Vielleicht ist das sogar ein Konzept „Haus“, das es in keine Sprache gibt, weil ja in jeder Sprache jedes Wort immer eine relativ schwammige Menge von Konzepten abdeckt; es ist durchaus möglich, dass das Modell Konzepte hat, die eigentlich präziser sind, als es unsere Sprache überhaupt zulässt.

Am Ende, wenn in den oberen Schichten des Modells wieder in echte Sprache zurücktransformiert wird, würde an irgendeinem Punkt die Weiche gestellt werden, ob in das Spanische oder Englische zurücktransformiert werden muss – das passiert unabhängig vom Rest des Textes.

Beruht die Basis Technologie auf neuronalen Netzen mit Backpropagation?

Sie beruht auf jeden Fall auf neuronalen Netzen, auch in diesem Attention Mechanismus. Darauf bin ich nicht eingegangen, weil es möglicherweise zu technisch ist. Auch in diesem Attention

Mechanismus ist zwischen zwei Schichten immer eine Schicht von einem klassischen neuronalen Netz dazwischengeschaltet. Zum Attention Mechanismus: Ich weiß nicht, ob man den wirklich so schön mit den Neuronen interpretieren kann, wie man es mit den klassischen neuronalen Netzen macht – also ob es hier auch eine biologische Analogie gibt. Aber das Grundkonzept der Back Propagation ist genau das Gleiche; man kann sagen, dass es eine spezielle Architektur des neuronalen Netzes ist.

Wie hoch war der Entwicklungsaufwand in Personenjahren und von welcher Gesamtkapazität der zusammengeschalteten Rechner ist auszugehen?

Beides habe ich in den Veröffentlichungen nicht gefunden. Aber der Transformer Mechanismus ist nicht wahnsinnig alt, wir reden hier vielleicht von fünf Jahren; in fünf Jahren kann man gar nicht so wahnsinnig viele Personenjahre leisten – der Aufwand ist, im Vergleich zu Unternehmenssoftware, also sicherlich beträchtlich, aber nicht unfassbar groß. Die Gesamtkapazität der Rechner kenne ich nicht und ich weiß auch nicht, ob sie veröffentlicht wird.

Ich meine, Zahlen gelesen zu haben, die besagen, dass die Trainingskosten im Millionen-Euro-Bereich liegen; daraus kann man sich vielleicht in etwa ableiten, welche hohen Stromkosten man dafür hatte – sicherlich auch Beträchtliches. Aber genaue Zahlen kennen wir nicht.

Herr Linde, vielen Dank für diesen Vortrag. Ich habe noch nie einen so guten Vortrag über die Innereien von ChatGPT gehört. Allerdings geht es mir so wie früher in der theoretischen Physik – ein Kollege hat da nach einem Vortrag den Spruch geprägt: „I'm still confused, but on a much higher level".

Ich muss sagen, dass ich etwa vier bis sechs Wochen Erfahrung mit ChatGPT Version 3.5. habe. – und am Anfang war ich wirklich verblüfft, weil ich das Gefühl hatte, dass mich jemand verstanden hat und die Antwort erstaunlich eloquent war; das habe ich noch nie gesehen.

Noch verblüffter war ich, als ich ChatGPT den Auftrag gegeben habe, mir ein Python Programm zu schreiben, das die Primzahlen von 1 bis 100 ausrechnet: In der Tat bekam ich dann einen Code heraus, der zwar nicht besonders elegant war, aber funktioniert hat.

Also war ich erst einmal begeistert. Dann liest man im Laufe der Zeit mehr über ChatGPT und darüber, wie alle möglichen Leute versuchen, das System auszutricksen. Und dann kamen die ersten Hinweise auf Halluzinationen und ich wurde etwas eigenartig berührt: Ich habe nämlich das System gebeten, meine professionelle Karriere zu erläutern.
Und das Ergebnis war ganz spektakulär – es klang, als stünde ich kurz vor dem Nobelpreis, da so viele fiktive Aspekte eingefügt wurden; das Halluzinieren scheint mir also ein großes Problem zu sein.

Dann gab es vor etwa zwei Wochen ein Interview mit Sam Altman, dem CEO von OpenAI; das war kurz nach der Ankündigung von GPT-4, und er sagte: „I am scared".
Ich finde, wenn der Erfinder schon Angst hat, sollten wir auch Angst haben.

Und warum? Er sagt, dass wir nicht genau wissen, wie das, was gerade bei ChatGPT abläuft, wirklich funktioniert und wir können das Ergebnis nicht vorhersagen. Ich glaube nicht, dass wir das so auf die Gesellschaft loslassen können.

Und er sagt auch – das war für mich völlig ungewöhnlich – dass es reguliert werden sollte. Auf Nachfrage der Interviewerin, wer denn kontrollieren solle, sagte er: „Society". Mein Einwand ist aber, dass die Gesellschaft nicht fachkundig genug zum Regulieren ist.

Daher meine Frage: Können Sie sich vorstellen, wie die Regulierung stattfinden könnte – auch in Hinblick auf den letzten Aufruf vom Future of Life Institute, dass es ein sechsmonatiges Moratorium geben sollte?

Ich glaube nicht, dass ich darauf eine gute Antwort geben kann. Ja, es gab vor einigen Tagen diesen Aufruf des Future of Life Institutes, dass vor den Gefahren dieser KI warnt. Und meines Erachtens auch aus guten Gründen.

Wie unsere Gesellschaft damit umgehen kann, ist aus meiner Sicht eine riesige, offene Frage: Man hat dabei ähnliche Aspekte wie bei anderen großen, offenen gesellschaftlichen Fragen, wie etwa beim Klimawandel oder einem Atomkrieg: Immer, wenn eine Nation etwas reguliert, hat sie auch einen Nachteil. Wenn wir die einzige Nation sind, die CO2 einspart, sind wir im strategischen Nachteil; wenn ein Land im Wettrüsten einseitig abrüstet, hat es einen strategischen Nachteil.

Das Gleiche gilt für KI: Wenn wir in Europa für ein Jahr die KI-Forschung verbieten würde (was sowieso nicht viel helfen würde), haben wir einen strategischen Nachteil und sie wird trotzdem irgendwo anders weiterentwickelt. Von daher ist meines Erachtens die einzige Art, wie eine solche Regulation letzten Endes funktionieren soll, eine globale Lösung – so wie auch bei den anderen drängenden Menschheitsfragen.
Es braucht globale Diskussionen und Richtlinien dafür, was wann gemacht werden kann und wie sich die Gesellschaft langsam darauf vorbereiten kann. Aber dass das mit einer sanften Entwicklungsgeschwindigkeit und rechtzeitig passiert, ist nicht realistisch.

Wenn man passgenaue Ergebnisse/Textbausteine von ChatGPT beim wissenschaftlichen Arbeiten nutzt, stellt sich die Frage der Quellenangabe: Was ist dabei zu berücksichtigen?

Das ist nicht wirklich meine Expertise, aber ich meine, gelesen zu haben, dass noch nicht wirklich klar ist, wer in solchen Fällen die Rechte hat. Es wurden wohl auch schon Klagen gegen OpenAI eingereicht, dass ja im Prinzip urheberrechtlich geschütztes Material zum Training verwendet wird und dann – zwar nicht in exakt gleicher Form, aber darauf aufbauend – auch weiterverwendet wird.

Andererseits müsste man sagen, dass ich auch als Mensch urheberrechtlich geschützte Bücher gelesen und dann darauf basierend eigene Artikel geschrieben habe. Es ist eine sehr schwierige Diskussion, und ich weiß nicht genau, wie der aktuelle Stand ist oder wie die Diskussion schließlich entschieden werden soll.

Mich würde interessieren, ob ChatGPT so etwas wie Begriffe bildet, die ja in den verschiedenen Sprachen durch unterschiedliche Worte belegt sind. Nicht jedes Wort in der einen Sprache hat dahinter den gleichen Begriff wie das entsprechende Wort in einer anderen Sprache. Kann man sagen, dass ChatGPT Begriffe bildet und später wieder in Sprache zurückbildet?

Ich würde sagen, dass das so sein muss. Auch hier: Wir verlassen den Bereich dessen, was wirklich wissenschaftlicher Fakt und erforscht ist, aber meine persönliche Meinung ist, dass ChatGPT Begriffe bilden muss, weil sich gewisse Antworten anders nicht erklären lassen. Zum Beispiel ist GPT in der

Lage, relativ kreativ zu sein: Wenn ich das Tool frage, wie ich ein Surfbrett einsetzen kann, um Arbeitssicherheit in einer Fabrik zu erhöhen, bekomme ich Antworten wie: Das Brett kann verwendet werden, um gefährliche Bereiche abzusperren und aus dem Segel kann ein Planschbecken genäht werden, mit welchem gefährliche Stoffe aufgefangen werden können.

Das bedeutet, dass das Tool einen Begriff davon hat, was ein Surfbrett physisch ist – denn diese Art von abstrusen Fragen kommen ja nicht in den Trainingsfragen vor. Es gibt ja keine statistischen Muster, dass Surfbretter auf diese Weise eingesetzt werden – das würde nur für Fragen gelten, die sich auf das Windsurfen beziehen, da es ja viele Texte über das Windsurfen und den Einsatz von Surfbrettern beim Windsurfen gibt.

Dass das Tool aber das Konzept „Surfbrett" nimmt und auf seine physikalischen Eigenschaften reduziert – dass es hart, lang und schmal ist – und in einen neuen Kontext (nämlich als Absperrbarriere von gefährlichen Bereichen) einsetzt, das geht aus meiner Sicht nur, wenn es sich Begriffe bildet.
Das wäre auch mein Grund, warum ich der Meinung bin, dass GPT tatsächlich so etwas wie Intelligenz ist, auch wenn das in den Medien häufig abgestritten wird. Denn es ist nicht nur ein statistisches System, sondern es ist ein emergentes System, dass tatsächlich am Ende Begriffe bildet.

Ich versuche gerade, mir das Phänomen der Halluzinationen zu erklären und habe dazu eine Frage: Wenn wir vorher KI eingesetzt haben, um beispielsweise Bilder zu kategorisieren, dann hat man in den Algorithmen ja immer einen Grad der *accuracy*, also der Genauigkeit, gesehen – so etwas wie 83% oder 95%.

Ist Halluzination im Prinzip das Problem, dass das Sprachmodell das nächste Wort sucht und dabei Worte mit 80%, was eine hohe Trefferquote ist, findet, und dann irgendwann in einer Ecke ankommt, in der nur noch Worte mit einer 5% Trefferquote auftauchen – und dass das Sprachmodel dann diese wählt? Denn dann könnte man das Modell ja dazu trainieren, dass es als Ergebnis einfach „Ich weiß es nicht" sagt.

Ich glaube das Problem liegt eher darin, dass die Information schlichtweg fehlt, weil sie in gewissem Sinne komprimiert wird.

Stellen Sie sich Folgendes vor: Sie machen ein Foto, bei welchem Sie jeden Pixel mit seinem Farbwert kennen. Jetzt können Sie fragen, welche Farbe der dritte Pixel von links hat. Dann bekommen Sie die Antwort: „Der dritte Pixel von links hat diese Farbe". Wenn Sie das Foto aber komprimieren, weil Sie nicht genügend Speicherplatz haben, dann werden die Pixel ein bisschen verändert; das Foto an sich sieht noch aus wie vorher, aber wenn Sie jetzt fragen, welche Farbe der dritte Pixel von links hat, dann bekommen Sie vielleicht eine falsche Antwort, weil er vielleicht zufällig zu einem kleinen Bereich des Fotos gehört, der sich im Vergleich zu davor verändert hat.

Sie wissen aber nicht, ob die Antwort richtig oder falsch ist, weil die Information schlichtweg nicht mehr im komprimierten Bild steckt. Und so ähnlich stelle ich mir das mit dem Wissen von GPT vor: Bei Informationen, die sehr häufig in den Trainingsdaten vorkommen, hat das Modell ein sehr hochaufgelöstes Wissen: Wenn Sie fragen, wie viele Beine ein Hund hat, wird es garantiert die richtige Antwort geben. Wenn Sie aber in exotischere Bereiche vorgehen, die im Training nicht so häufig vorkommen, dann sind diese Wissensbereiche in gewisser Weise so komprimiert wie die Pixel

des Bildes. Und wenn sie dann nach einem Pixel, also einem einzelnen, kleinen Fakt fragen („Wo ist der McDonalds im Flughafen von Seattle?“), dann werden Sie mit hoher Wahrscheinlichkeit nicht den richtigen Pixel erwischen, sondern einfach nur ein großes Viereck, und sie bekommen eine Zufallsantwort, die eben nicht stimmt.

Mich treibt im Moment besonders eine Frage in Richtung Maschinenintelligenz um: Ich habe ja bei solchen Mechanismen, oder bei solchen Maschinenintelligenzen, wie ChatGPT, im Prinzip zwei Möglichkeiten, eine Antwort zu bekommen. Ich kann eine richtige Antwort bekommen, aber diese richtige Antwort ist technokratisch betrachtet und möglicherweise auch in Bezug auf den Kontext unabhängig.

Oder: ich kann eine falsche Antwort bekommen, die dann aber im kontextualen Zusammenhang richtig ist, insbesondere wenn es um die Lösung von Problemen geht, dann würde eine einfach auf mathematischer Basis arbeitende Maschinenintelligenz natürlich die Lösung bevorzugen, die – falls ein Schaden entstehen muss – den geringsten Schaden anrichtet. Möglicherweise ist das aber auch politisch nicht gewollt.

Die Frage ist: In welche Richtung wollen wir gehen, wenn die Maschinenintelligenz ChatGPT und ihre Nachfolger uns derartige Fragen beantworten?

Noch ein Beispiel für eine solche Frage: Beim Thema Klimawandel wäre die schnellste und sicherste Möglichkeit, etwas dagegen zu tun, eine Million der reichsten Menschen der Welt in das Gefängnis zu bringen, weil sie dann wesentlich weniger CO2 produzieren würden als bisher. Das ist technokratisch die richtige Antwort auf die Frage – aber sie ist politisch und gesellschaftlich wahrscheinlich nicht durchsetzbar. Die politisch und gesellschaftlich durchsetzbare Antwort ist, wir warten, bis es brennt, und hoffen, dass wir dann noch rechtzeitig löschen können.
Welche Antwort erwarten wir, wenn wir diese Frage einer Maschinenintelligenz stellen? Welche sollte sie geben?

Ich glaube, das ist vielleicht eher eine Frage für die Runde.

(in Bezug darauf) Es ist auch eine Frage an die Leute, die diese Systeme einlernen; Gut und Böse sind ja immer auf den Zeitgeist bezogen. Und die reine Mathematik und reine Informatik beschäftigt sich ja nicht mit Befindlichkeiten, sondern eben mit Zahlen, Daten und Fakten in einem sehr komplexen System. Was wollen wir, und wer sorgt dafür, dass das in der tatsächlichen technischen Umsetzung von ChatGPT passiert?

Generell lernt das System erst einmal ähnlich wie ein Mensch, das heißt, es gibt auch relativ menschliche Antworten; es würde vermutlich nicht diese Gefängnis Antwort geben, obwohl es eine logische Antwort ist – doch das System gibt tendenziell eher menschliche als logische Antworten. Es gibt ja diesen Witz, in welchem Leute aus einem Ballon herunterrufen „Wo sind wir?“ und der Mathematiker zurückruft „Ihr seid in einem Ballon!“ – das ist zwar die richtige, aber eine nutzlose Antwort.
Und genau so etwas macht GPT eben aufgrund dieses Trainingsprozesses nicht. Aber mit welchen Informationen man GPT trainiert, ist eine sehr wichtige Frage, weil es eben Konsequenzen hat – insbesondere, wenn ein solches Modell skaliert wird. Bisher scheint zumindest die Firma OpenAI relativ verantwortungsvoll damit umzugehen, da GPT immer recht ausgewogene Antworten gibt.

Aber solche Modelle werden in Zukunft große Macht haben, ähnlich wie Google sehr große Macht allein dadurch hat, dass bestimmte Suchergebnisse höher oder niedriger platziert werden.

Eine Frage, die sich auf die Diskussion bezieht: Die Frage, was das System im Umweltschutz empfiehlt, kann das System ja so nicht beantworten. Es müsste davor erst einmal eine ideologische Position eingenommen werden. Und ich könnte mir vorstellen, dass das System auch eine Antwort geben könnte, wenn es beispielsweise die ganz konkrete ökonomische Sichtweise oder die eines Bauers in Kenia einnehmen würde – also ein Werterahmen gegeben wird.

Aber wie sollte das System ohne Werterahmen auswählen können? Wir Menschen unterliegen ja mehr oder weniger unseren persönlichen Erfahrungen und bewerten; dadurch haben wir dann in unseren Überzeugungen eine Bandbreite etwa von Kommunismus bis zu Kapitalismus. Ich kann mir schon vorstellen, dass ein solches System so etwas auch lernen kann, aber ich müsste ihm dazu den entsprechenden Prompt geben. Und dann wird es sich mit Sicherheit auch intern widersprechen, weil es einmal aus der Sicht der aktuellen Diskussion der FDP oder der CDU oder der Kirche antwortet; das System könnte dann ebenfalls aus einer bestimmten Perspektive antworten, aber dazu müssten wir ihm eben unsere persönlichen Parameter mitgeben.

Das ist auf jeden Fall möglich – ich würde sogar sagen, das ist schon passiert: Die Ideologie, die ChatGPT mitgegeben wurde, ist scheinbar die großer Neutralität und politischer Korrektheit. Das wurde im Training geübt, indem immer Texte, die sehr ausgewogen und sachlich sind, höher gewichtet wurden als andere Texte. Dadurch hat der interne Literaturkritiker des Modells gelernt, dass sachlich und ausgewogen „gut“ ist, und hat dann im intrinsischen Trainingsprozess das Modell weiter dahin trainiert, möglichst sachlich und ausgewogen zu antworten. Wenn Sie also nach Umweltschutz fragen, werden Sie die Antwort bekommen, dass es hierbei unterschiedliche Positionen gibt und welche die Argumente der einen und der anderen Seite sind.

Hätte man das Modell allerdings anders trainiert, würde es auch anders antworten – also beispielsweise aus Sicht eines bestimmten politischen Spektrums, dann würde das Modell auch so antworten.

Ich denke aber, was Sie gerade sagen, ist ja auch schon wieder ideologisch: Das System ist neutral. Befragen Sie aber einmal einen Chinesen, Russen oder Nordamerikaner hierzu – ich denke, Sie werden sicherlich ganz unterschiedliche Antworten bekommen. Insofern hilft das meines Erachtens nicht weiter, sondern bringt nur unsere Perspektive des „westlichen weißen Mannes“ ins Spiel.

Ja, das ist richtig, es gibt natürlich keine absolute Neutralität. Wenn ich mir aber die durchschnittlichen Fernsehsender und Zeitungen anschaue, habe ich schon das Gefühl, dass ChatGPT neutralere Antworten gibt. Von sozialen Medien einmal ganz abgesehen.

Die Frage bezieht sich auf diese illusionistischen Antworten von ChatGPT: Wir sind in den letzten Fragen davon ausgegangen, dass immer nur eine Frage gestellt wurde und eine Antwort kommt. Aber ChatGPT ist ja ein Chat Programm – das heißt, dass ich auf jede Antwort wieder eine Gegenfrage stelle. Und dann stellt sich eine Beziehung zwischen dem Nutzer und der Maschine her – und der Nutzer bekommt so langsam die Vorstellung, dass am anderen Ende jemand ist, der ihn versteht und mit dem er sprechen kann; entsprechend bekommt der Austausch so emotionale Züge.

In Belgien gab es kürzlich einen Fall, bei welchem jemand durch einen Chat in den Selbstmord getrieben wurde: Er ist wohl schon mit einer gewissen Vorstellung in das Gespräch hineingegangen und wurde dann im Dialog, der sich wohl über mehrere Wochen hingezogen hat, in den Selbstmord getrieben, weil ihn der Algorithmus „verstanden" hat. Das heißt, das System hört nicht auf, fortwährend plausible Antworten zu geben, sondern es passt sich dem Gegenüber an; dadurch entsteht eine große Gefahr.

Man versteht ChatGPT, wenn man einfach die Frage-Antwort-Frage-Antwort Serie nachverfolgt – das habe ich auch in meinem Fall (mit dem falschen Lebenslauf) gemacht, und nach Internetlinks gefragt. Dann hat ChatGPT mir fünf Internetlinks zur Verfügung gestellt, die dann alle bei einer Fehlermeldung geendet haben (und nicht existiert haben) – aber an sich plausibel aussahen.

Ich sehe die Gefahr also in der Anpassungsfähigkeit der Inhalte. Und jetzt komme ich noch einmal zurück zur Regulierung: Man muss darüber aufklären, was bei ChatGPT passiert, und dem System vielleicht beibringen, dass es irgendwann „Das weiß ich nicht" kommuniziert, anstatt falsche Inhalte zu erfinden.

Ja, prinzipiell sehe ich die Gefahr schon. Ich glaube, bei dem Fall in Holland war es nicht ChatGPT, sondern ein anderes Produkt – da muss man sehr stark unterscheiden, welches Sprachmodell verwendet wird. ChatGPT wurde sehr sorgfältig auf ethische und Sicherheitsaspekte trainiert – ich glaube nicht, dass GPT es „schaffen" würde, jemanden in den Selbstmord zu treiben. Davor würden schon zahlreiche Aufrufe kommen, dass man sich an eine psychologische Beratung wenden sollte.

Aber es gibt zum Beispiel aktuell ein System, das Freedom GPT heißt, und im Prinzip das Gleiche wie ChatGPT ist, aber ohne die entsprechenden Sicherheitsvorkehrungen funktioniert: Dieses Modell würde einem dann wahrscheinlich schon erklären, wie man sich das Leben nehmen kann. Dann kommt es sehr stark auf den Anwender an.

Ganz kurz und knapp: DALL-E erzeugt ja Bilder aus Texteingaben. Muss ich mir einen Transformermodell vorstellen, dass es beim Kontextraum dann auch so etwas wie einen visuellen Kontextraum gibt?

Kurioserweise ja. Ich glaube nicht, dass ich das Vorgehen besonders gut erklären kann, aber: DALL-E basiert ja auch auf einem Transformermodell – das ist ja auch erstaunlich, wie flexibel dieses Modell eingesetzt werden kann.

Ich möchte einmal über eine Anwenderebene ganz anderer Art sprechen: Ich bin, wie Herr Linde, auch im Chemiebereich tätig, und stelle mir die Frage, ob man dieses Werkzeug bereits im industriellen Alltag, zum Beispiel bei der Vertriebsunterstützung, benutzen könnte. Haben Sie sich diese Frage schon gestellt und mit welchen Bereichen würden Sie, Herr Linde, anfangen?

Wir suchen sicherlich nach Anwendungsfällen – es ist ja so, dass die Entwicklung in diesem Bereich so schnell ist, dass man kaum hinterherkommt. Und GPT-4 eröffnet meiner Meinung nach noch einmal ganz neue Anwendungsfälle, was aber auch erst seit wenigen Wochen klar ist.

Erste Fälle wären auf jeden Fall erst einmal die 1:1 Kommunikation – so dass Texte schneller verfasst werden können und die Entwicklung im Unternehmen schneller abläuft. Eine andere interessante Anwendung ist eine Suchfunktion – man kann Dateien oder Daten in diesen Raum von GPT einbetten

und sie so für das Modell zugänglicher machen. Dann könnte man zum Beispiel eine Suche auf Patentdaten laufen lassen, und die Suche ist nicht abhängig von den verwendeten Wörtern, sondern von den im Patentrecht verwendeten Konzepten. Das wäre ein klassischer Anwendungsfall.

Mir ist die Frage eingefallen, weil ich letztens eine Warnung über den Einsatz von ChatGPT in Firmen gelesen habe. Weil Sie gerade Patente erwähnten: Man müsste aufpassen, dass das, was als Frage eingegeben wird, möglicherweise bei anderen als Antwort verwendet werden könnte – wie wird denn meine Frage im Datensatz von GPT gespeichert?

Die Frage geht erst einmal gar nicht direkt in das Datenmodell ein – das ist eben das Problem der fehlenden Verknüpfung zwischen Kurzzeit- und Langzeitgedächtnis. Aber es ist natürlich anzunehmen, dass Ihre Frage gespeichert wird, und möglicherweise beim Training der Folgeversion Ihre Frage auch verwendet wird. Das ist durchaus denkbar.

Wenn Sie GPT also im Firmenkontext befragen, sollten Sie also darauf achten, dass Sie keine vertraulichen Daten verwenden oder Sie eine private Instanz – eine private Version des Programms – bekommen.

Ganz kurze Frage: Wir haben jetzt viel über OpenAI und deren System erfahren. Gibt es denn andere, ähnlich weit fortgeschrittene Systeme, oder ist OpenAI da so deutlich Marktführer, dass so schnell niemand nachkommen kann?

Was die Qualität der Antworten angeht, würde ich sagen, dass OpenAI ganz klar führend ist. Allerdings gibt es durchaus auch andere große Systeme, beispielsweise von Facebook oder Google. Nach dem, was ich gehört habe, sind diese nur in Bezug auf das Endtraining nicht so stark wie ChatGPT, aber vom Grundprinzip sehr ähnlich. Es ist anzunehmen, dass sie in den nächsten Jahren entsprechend versuchen, aufzuholen – und dann wird es quasi einen Rüstungswettbewerb zwischen den großen Konzernen um das beste Sprachmodell geben.