



23.03.2023

Transkript

„Sprachmodelle, Robotik, Mensch-Maschine-Interaktion: What's next?“

Experten auf dem Podium

- ▶ **Prof. Dr. Nicole Krämer**
Leiterin des Fachgebiets Sozialpsychologie: Medien und Kommunikation, Universität Duisburg-Essen
- ▶ **Prof. Dr. Hinrich Schütze**
Inhaber des Lehrstuhls für Computerlinguistik und Direktor des Centrums für Informations- und Sprachverarbeitung, Ludwig-Maximilians-Universität München (LMU)
- ▶ **Prof. Dr. Marc Toussaint**
Leiter des Fachgebietes Intelligente Systeme, Technische Universität Berlin, und einer der Autoren des Palm-E-Papers
- ▶ **Bastian Zimmermann**
Redakteur für Digitales und Technologie, Science Media Center Germany, und Moderator dieser Veranstaltung

Mitschnitt

- ▶ Einen Videomitschnitt finden Sie unter: <https://www.sciencemediacenter.de/alle-angebote/press-briefing/details/news/sprachmodelle-robotik-mensch-maschine-interaktion-whats-next/>
- ▶ Falls Sie eine Audiodatei benötigen, können Sie sich an redaktion@sciencemediacenter.de wenden.



Transkript

Moderator [00:00:00]

Hallo, liebe Journalistinnen und Journalisten, herzlich willkommen zu unserem virtuellen Press Briefing zum Thema: Sprachmodelle, Implikationen für Robotik und Mensch-Maschine-Interaktion. Mein Name ist Bastian Zimmermann, ich bin Redakteur beim Science Media Center und mit mir habe ich heute auch noch eine Expertin und zwei Experten. Erst mal herzlich willkommen, schön, dass Sie da sind. Ich stelle Sie gleich noch alle vor.

Momentan hat man ja den Eindruck, im Bereich KI geht es gerade sehr schnell voran. Da gibt es wöchentlich neue Ankündigungen, neue Veröffentlichungen. GPT-4 vor wurde letzte Woche veröffentlicht. Kürzlich wurde dann PaLM-E veröffentlicht, zu dem Herr Toussaint noch genauer etwas sagen wird. Diese Woche wurde dann das Sprachmodell Bard von Google für eine begrenzte Anzahl von Personen freigegeben. Davon sind ja verschiedenste Themenbereiche betroffen und eben auch sehr unterschiedliche Themen aus diesen Bereichen sind dann interessant. Deswegen haben wir hier eine sehr interdisziplinäre Runde, um Fragen zu Sprachmodellen, zur Mensch-Maschine-Interaktion und zur Robotik beantworten zu können.

Bevor ich gleich zum Vorstellen komme, noch der Hinweis an Sie da draußen: Stellen Sie bitte Ihre Fragen und nehmen Sie dafür bitte die Fragefunktion von Zoom, nicht den Chat. Dann können Ihre Kolleginnen und Kollegen die Fragen auch jederzeit sehen, dann gibt es weniger Dopplungen, also bitte das Frage-und-Antwort-Modul benutzen.

Dann zu unserer Expertin und den beiden Experten. Vielen Dank, dass Sie hier sind und ich stelle Sie gerade mal der Reihe nach vor. Ich fange an mit Prof. Dr. Nicole Krämer. Sie ist Leiterin des Fachgebiets Sozialpsychologie: Medien und Kommunikation an der Universität Duisburg-Essen. Sie befassen sich ja unter anderem mit Mensch-Maschine-Interaktion einerseits in Bezug auf Roboter, aber eben auch auf Kommunikation mit Sprachmodellen und bringen heute die psychologische Perspektive ein.

Dann haben wir Prof. Dr. Hinrich Schütze. Er ist Inhaber des Lehrstuhls für Computerlinguistik und Direktor des Centrums für Informations- und Sprachverarbeitung an der Ludwig-Maximilians-Universität München. Sie sind der Experte für Sprachmodelle und Natural Language Processing.

Und dann haben wir noch Prof. Dr. Marc Toussaint, Leiter des Fachgebiets Intelligente Systeme an der Technischen Universität Berlin und auch einer der Autoren des PaLM-E-Papers. Sie befassen sich mit Robotik, aber auch mit KI-Methoden wie Reinforcement Learning. Und gerade die Kombination von Sprachmodellen mit einem visuellen Modell für Roboter, wie es bei PaLM-E-Forschung passiert ist, ist gerade jetzt auch sehr interessant, da haben Sie ja als Mitautor auch interessante Einblicke.

Zu den Eingangsfragen: Dann fangen wir mit Ihnen an, Herr Schütze und da interessiert uns: Wie ordnen Sie die aktuellen Fortschritte im Bereich der Sprachmodelle eigentlich ein und sind auch in Zukunft weiterhin große Verbesserungen durch noch größere Modelle mit mehr Parametern zu erwarten oder eher durch andere Ansätze, wie zum Beispiel die Kombination von Sprachmodellen mit Rechtersystemen?

Hinrich Schütze [00:03:15]

Zunächst mal, die Grundtechnologie ist eigentlich nicht neu, die Grundtechnologie für die Sprachmuster ist einfach immer das nächste Wort vorherzusagen, ganz stupide, immer das nächste Wort. Das ist im Grunde "AI complete", wenn man das kann, kann man eigentlich fast alles in der künstlichen Intelligenz machen und deswegen ist das so mächtig. Das gibt es aber schon seit 50 Jahren. Was sich jetzt geändert hat, ist, dass wir das auf sehr großen Textmengen trainieren können, mit



sehr, sehr großen Modellen. Und was jetzt durch OpenAI noch dazugekommen ist, ist, dass sie große Fortschritte erzielt haben darin, diese Fähigkeiten, die auf diese Weise gelernt werden, flexibel einzusetzen. Der Trainingstext enthält zum Beispiel wenig Dialoge, deswegen ist das rohe Modell für Dialoge nicht so gut einsetzbar und das wird herausgekitzelt durch spezielle Verfahren. Außerdem ist sehr wichtig, dass die groben Fehler, die passiert sind mit den ersten Generationen, dass die jetzt sehr gut unter Kontrolle sind. Rassismus und auch durchaus Halluzinationen und einige andere Kinderkrankheiten, das ist zwar nicht perfekt, aber es hat sich um Größenordnungen verbessert, wie das kontrolliert wird. Und dann natürlich die Multimodalität. Da sage ich jetzt nichts zu, weil der Kollege Toussaint sicher sehr viel besser erklären kann als ich.

Insgesamt würde ich das wirklich als eine Revolution bezeichnen, technologisch, die mit dem Internet und dem Smartphone vergleichbar ist. Es werden sich große Veränderungen darin ergeben, wie wir schreiben, wann immer wir Texte verfassen, wie wir programmieren, also Software kreieren. Das hat große Auswirkungen auf die Schule und die Universität. Da haben wir erst jetzt angefangen drüber nachzudenken. Aber genauso, wie man nicht mehr auswendig lernt, man lernt nicht mehr, Karten zu lesen, man lernt nicht mehr zu rechnen, man braucht keine Rechtschreibung mehr zu können. Jetzt wird es wirklich fundamentale Veränderungen geben, was den Lernstoff angeht, was wir überhaupt lernen wollen, und damit werden sich der Unterricht und die Inhalte sehr ändern. Das gleiche gilt für den Arbeitsalltag. Ganz viele Berufe werden wegfallen, wenn es einfach nur darum geht, Zusammenfassungen zu schreiben, Wissen zu sammeln und zu kondensieren, vielleicht auch Accounting, ganz viel im Customer Service, das hat ja schon begonnen. Das sind alle Bereiche, die von dieser Technologie gefährdet sind, aber natürlich auch von ihr profitieren können. Und genauso wie Smartphone und Internet sehr flexibel eingesetzt werden können und ganz viele verschiedene Anwendungen gefunden haben, die großen Einfluss auf unser tägliches Leben haben, so ist es auch mit dieser Grundtechnologie der Sprachmodelle. Wir können uns also noch gar nicht vorstellen, was da alles passieren wird, weil jetzt ganz viele Unternehmen anfangen, sich zu überlegen, wie können wir das kommerziell einsetzen, wie können wir damit Märkte finden und so weiter. Da wird sich also sehr, sehr viel tun.

Das Ganze hat aber auch eine Kehrseite, genau wie Smartphones und Internet eine Kehrseite hatten. Ich habe schon Arbeitslosigkeit angesprochen, Umschulung wird für viele Leute notwendig sein und es gibt auch grundlegende technologische Probleme. Es wird die Versuchung bestehen, Entscheidungen durch diese Maschinen machen zu lassen: Im Recht, in der Medizin, Steuerberatung, Vermögensverwaltung. Das Problem ist, dass die Modelle eine Entscheidung oder eine Aussage mit hoher Glaubwürdigkeit machen, sodass der Mensch denkt, das muss ja stimmen, wenn sich das Modell so sicher ist. Aber in Wirklichkeit kann das Modell seine eigene Sicherheit nicht einschätzen. Das ist eines der großen Probleme, die wir haben. Und die Menschen sind nicht daran gewöhnt, dass jemand mit hundertprozentiger Sicherheit und Selbstvertrauen spricht, aber völlig falsch liegt. Betrüger machen das vielleicht mal, aber im normalen Umgang gehen wir davon aus, wenn jemand sich sicher ist, dass da etwas dahintersteckt, aber hier ist das nicht der Fall. Das müssen wir lernen und das wird ein schwieriger Prozess sein. Wir sollten uns also nicht vorbehaltlos in diese Technologie hineinstürzen, sondern es wird sehr wichtig sein, dass wir uns selbst und die Gesellschaft darauf vorbereiten und das wird ein längerer Prozess sein.

Noch kurz zur Zukunft, wie wird die aussehen und da gibt es zwei verschiedene Meinungen zu. Die einen sagen tatsächlich: Einfach mehr Daten, einfach größere Modelle und da wird es besser werden und wir werden jedes Problem knacken, das es zu knacken gibt. Ich persönlich bin der Meinung, dass diese Grundprobleme, die ich eben angedeutet habe, die Modelle können nicht ihre eigene Sicherheit einschätzen, die Modelle können halluzinieren, denken sich einfach Sachen aus und die Modelle können auch nicht erklären, wie sie zu einer Entscheidung gekommen sind, diese Probleme werden mit mehr Daten und größeren Modellen nicht gelöst werden. Wir brauchen also grundlegend andere Ansätze, die wir vielleicht verheiraten können mit der jetzigen Technologie, um diese Probleme zu lösen und ich denke, das wird die Zukunft sein in den nächsten 10, 20, 30 Jahren, dass wir an neuen Technologien arbeiten, die das können.



Moderator [00:09:11]

Vielen Dank erst mal für die erste Einschätzung. An Herrn Toussaint die Frage: Sie haben ja an der PaLM-E-Studie mitgearbeitet. Können Sie dann vielleicht noch mal kurz zusammenfassen, was Sie da gemacht haben, was die interessanten Befunde sind und dann natürlich die Frage, welchen Stellenwert können Sprachmodelle eigentlich bei der Forschung zur Robotik haben?

Marc Toussaint [00:09:30]

Erst einmal Herr Zimmermann, vielen Dank für die Einladung und danke auch für die Einführung. Vielleicht ein Wort zu PaLM-E. Sprachmodelle sind zunächst einmal Modelle, die sich mit Sprache beschäftigen, also die textuellen Input nehmen können, auch textuellen Output geben und sehr mächtig sind und Herr Schütze hat es sogar so formuliert: Wenn sie denn alles vorhersagen könnte, wäre das "AI complete", was eine interessante Aussage ist. Die Frage stellt sich aber schon, ob solche Systeme mit mehr umgehen können als nur mit Text. Und gerade als Robotiker stellt sich die Frage – und so war die Forschungsfrage für uns –, ob man solche Systeme ähnlich wie Sprachsysteme auch an die physische Welt koppeln kann. Das bedeutet, dass sie nicht nur textuellen Input bekommen, sondern auch Bilder als Input oder Zustandsschätzungen aus der Robotik und mithilfe dieser Informationen nicht nur textuelle Probleme, sondern auch geometrische Probleme lösen können. Zum Beispiel, wenn ich ein Bild zeige und frage, ist das Objekt denn hier für den Roboter erreichbar, dann ist das eine geometrische Lösung [für ein] physikalisches Problem, also ob solche Dinge auch von solchen Systemen gelöst werden können. PaLM-E versucht genau das zu untersuchen. E steht für Embodied, wir haben das auch titulierte als ein Embodied Language Model, also ein Language Model, das in die physische Welt eingebettet sein soll und das bedeutet hier, dass es physische Sensorinputs bekommt und auch physische Dinge präzisieren muss. Und das läuft im Grunde so aus, dass der Input-Text jetzt ein Text ist, der sowohl aus Worten als auch aus Bildern, als auch aus Vektoren besteht. Zum Beispiel: Hier ist eine Szene – es kommt ein Bild. Der Roboter ist im Zustand – Vektor. Kann der Roboter mit dem Objekt – Segmentierung des Objekts – das andere Objekt – noch eine Segmentierung – erreichen? Und so ist der Input zu diesem Modell multimodal, das mischt Worte, Bilder, Zustandsvektoren und das System kann dann eben geometrische Probleme lösen. Das bedeutet es also, was ein multimodales Sprachmodell ist oder auch multimodale Foundation Models werden sie heutzutage genannt. Und das ist für die Robotik sehr spannend und ich glaube auch aus wissenschaftlicher Forschungssicht sehr spannend, weil es wirklich eine Verschränkung von Modalitäten bedeutet, eine Verschränkung von Sprache und physischer Welt. Und das wollten wir mit PaLM-E untersuchen und haben jetzt auch Erfolge damit gezeigt. GPT-4 geht auch in die Richtung, ist in kleinen Details anders.

Vielleicht auch ein Punkt zu unserer Rolle, was ich auch ganz interessant finde. Die Studie entstand in der Kooperation mit uns an der TU Berlin als auch mit Leuten von Google Brain. Die Leute von Google Brain waren sehr stark, auch mit ihrer Erfahrung, Sprachmodelle zu trainieren. Aber die Fragestellung, wie ich sie gerade skizziert habe, zu untersuchen, verlangt natürlich auch zu wissen, was sind denn eigentlich relevante Probleme, was sind relevante Problemstellungen innerhalb der Robotik. Unsere Rolle war im Grunde, weil wir eine gewisse Expertise zeigen können mit klassischen Methoden, die modellbasiert [sind], also wenn ein Modell der Szene gegeben ist, sequenzielle Handlungsplanung machen können. Wir haben klassische Problemlöser für sequenzielle Handlungsplanung oder physikalisches Denken, dass wir mithilfe unserer klassischen Methodiken Beispiele generieren können, Daten generieren können, was in dieser Szene mögliche Handlungen sind und auch wie überhaupt Kostenfunktionen oder Ziele definiert werden können. Und mithilfe dieser Expertise waren wir in der Lage, sowohl Daten zu liefern als auch Problemstellungen zu liefern, um diese großen Modelle auch multimodal zu trainieren.

Vielleicht nur abschließend: Ich sehe das auch so, dass mit diesen Entwicklungen schon etwas Fundamentales sich geändert hat, auch in Bezug auf die Robotik. Das ist einerseits etwas ziemlich



oberflächlich Klares, nämlich, dass man jetzt auch mit Sprache mit Robotern interagieren kann. Ich kann jetzt zum Beispiel auch sprachliche Ziele definieren, sprachliche Instruktionen an Roboter geben und die können interpretiert werden, mit der physischen Welt in Bezug gesetzt werden und ausgeführt werden. Das ist eine Sache, aber ich glaube, die Entwicklung geht auch dahin, dass man immer mehr und mehr versteht, dass diese Transformer mehr können, als nur mit Sprache umzugehen. Und aus meiner Sicht sind Transformer die Modelle, in denen man verstanden hat, dass man verschiedenste Modalitäten verschränken kann. Und das wird sehr mächtig sein, ich glaube weit über Sprache hinaus, um wirklich multimodale Systeme oder auch Systeme, die in der physischen Welt Entscheidungen treffen müssen, auch in der Zukunft zu trainieren. Ich belasse es mal dabei.

Moderator [00:14:26]

Wir haben ja gleich noch Zeit für die anderen Fragen, erst mal kurz alle Personen zu Wort kommen lassen. Dann haben wir noch Frau Krämer und an Sie würde mich die Frage interessieren, welche Entwicklungen jetzt eigentlich aus der Perspektive der Mensch-Maschine-Interaktion momentan bei der Interaktion mit Sprachmodellen relevant sind beziehungsweise in Zukunft und welche man jetzt schon beobachten kann.

Nicole Krämer [00:14:49]

Ich möchte mich gerne auf drei Aspekte beziehen, nämlich einerseits Transparenz, also was verstehen Menschen eigentlich von dem System, wenn sie damit interagieren, andererseits auch Vertrauen, inwieweit folgen sie zum Beispiel Empfehlungen solcher Systeme. Und der dritte Aspekt: Gibt es so etwas wie Beziehungsbildung gegenüber solchen Systemen.

Zum Transparenzbereich: Viele Studien, auch unsere eigenen Studien, haben gezeigt, dass Menschen tatsächlich noch viel zu wenig verstehen, wie solche Systeme funktionieren und das führt gegebenenfalls nicht nur zu Fehleinschätzungen, inwieweit man Vertrauen haben kann in solche Systeme und das, was sie ausgeben oder empfehlen, sondern auch zum Beispiel zu Fehleinschätzungen, was mit eigenen Daten passiert. Und da ist sicherlich noch sehr viel mehr erforderlich, dass Menschen darüber informiert werden, auch welche Daten erhoben werden, wie die eigenen Daten verarbeitet werden. Das gilt nicht nur für ChatGPT, sondern stärker für sprachliche Interaktion zum Beispiel mit Sprachassistenten, die ja mittlerweile in vielen Haushalten stehen, wie Alexa, Google [Assistant] und so weiter. Und dass teilweise ja auch Kinder mit all diesen Systemen interagieren, die eigentlich auch eine sogenannte informierte Einwilligung abgeben müssen, das ist gesetzlich vorgesehen, aber von Informiertheit kann man eigentlich nicht sprechen. Also da ist sicherlich großer Nachholbedarf. Damit Menschen sicher mit solchen Systemen umgehen können, müssen sie im Vorfeld viel besser darüber informiert werden, im besten Fall von Herstellersicht aus und das muss gesetzlich geregelt sein, was da dann auch an Informationen vorzugeben ist.

Zu dem Aspekt, wie Menschen dann mit den Empfehlungen oder dem, was sie da als Ausgabe bekommen, umgehen, möchte ich meinem Kollegen Schütze ein bisschen widersprechen, denn wir haben in der Psychologie gesehen, dass Menschen in vielerlei Hinsicht eigentlich sehr kompetent sind, wenn es um Informationssammlung geht, zum Beispiel auch, wenn im Internet Informationen gesucht werden. Da sind Menschen mittlerweile sehr gut drin und ich glaube, dass das auch bei ChatGPT zum Beispiel sehr schnell passieren wird, dass man lernt, was kann man glauben, wem soll man folgen, wem nicht. Ein Aspekt, der da sicherlich sehr schnell weiterhelfen wird, das haben wir auch in vielen Studien gesehen: Die Erwartungshaltung an solche Systeme wird natürlich auch davon geprägt, was man im Umfeld wahrnimmt, zum Beispiel auch durch Medienberichterstattung. Und da ist ja zum Beispiel auch jetzt schon sehr viel berichtet worden darüber, dass die Ergebnisse eben nicht immer zuverlässig sind. Von daher habe ich keine Sorge und auch übrigens für unseren eigenen Universitätsbetrieb keine Sorge, dass sich demnächst massenhaft Studierende darauf verlassen werden, was ChatGPT so ausgibt, weil es ja nun mannigfaltige Beispiele dafür gibt, dass man



sich eben nicht in jeder Hinsicht darauf verlassen kann. Und wir sehen in der Forschung auch, es gibt natürlich so was wie Algorithm Appreciation, also Personen, die so das Gefühl haben: "KI! Jetzt endlich sagt mir jemand, wie es wirklich ist!", aber eben auch sehr viel Algorithm Aversion, was bedeutet also das Ablehnen von Algorithmen und die grundlegende Skepsis, dass so eine KI mir doch eigentlich gar nicht sagen kann, wo es langgeht. Vor dem Hintergrund bin ich nicht der Auffassung, dass wir da in großem Maße sehen werden, dass Menschen sich darauf verlassen und das nicht mehr hinterfragen.

Der dritte Aspekt: Beziehungsbildung. Was hier eine große Rolle spielt, ist die Tatsache, dass – seien es jetzt Sprachassistenten wie Google, Alexa, aber auch ChatGPT – von Herstellerseite sehr stark die Anthropomorphisierung genutzt und eingesetzt wird. Bei den Sprachassistenten ist es natürlich die Sprache, die dann auch Menschen dazu bringt, den Eindruck zu haben, das ist wie mit einem Menschen sprechen. Und bei ChatGPT, das finde ich persönlich besonders perfide, ist es die Tatsache, dass es tippt und dadurch den Eindruck erweckt, das würde wie ein Mensch, der woanders sitzt und einem einen Ratschlag tippt und über einen Chatkanal mitgibt, auf dieser Ebene verarbeitet werden. Das ist meines Erachtens, da können die Kollegen mich korrigieren, aber technologisch überhaupt nicht notwendig, das wirklich auf "Einzeltippeebene", das hält auch manchmal an, zögert, es macht den Eindruck, als würde es weiter nachdenken oder es geht manchmal sogar zurück und korrigiert einen Buchstaben. Das ist aus meiner Sicht eine Angebotsstrategie, so zu tun, als wäre das menschlich. Wir testen jetzt gerade auch – ich habe noch leider keine Ergebnisse – inwieweit das auch das Vertrauen dem System gegenüber erhöht und das wäre natürlich dann wieder eine Sache, die es Menschen dann wiederum schwerer machen würde zu hinterfragen, ob das vielleicht alles so korrekte Informationen sind. Das erst mal von meiner Seite.

Moderator [00:20:26]

Danke schön, sehr interessant. Wir haben noch ein paar Fragen. Die ersten beiden gehen diese Service-Roboter-Geschichte an. Und auch in Bezug auf PaLM-E hätte ich die erste an Sie, Herr Toussaint: Kann man schon einschätzen, welche Kapazitäten PaLM-E für "Service-Roboter" haben wird, haben wir da einen vorbildlichen Roboter, der eigenständig lernt, was erledigt werden muss, unliebsame Aufgaben erledigt und auch aufmerksam auf Korrekturansagen reagiert?

Marc Toussaint [00:20:54]

Service- Roboter bezieht sich wahrscheinlich auf den Haushalt, das interpretiere ich jetzt mal so.

Moderator [00:20:59]

So würde ich es verstehen.

Marc Toussaint [00:21:00]

Oder vielleicht auch auf der Pflege. Um das erst mal nüchtern zu sehen: Bei PaLM-E geht es in der Forschung zunächst noch nicht darum, ein lernendes System zu haben. Das wurde direkt erwähnt, ob die Systeme autonom lernen können. Auf solchen Systemen dann noch mal ein adaptives, situatives Lernen zu realisieren, ist ganz und gar nicht so einfach, denn die Systeme sind ja offline optimiert und trainiert. Was aber, glaube ich, wirklich ermöglicht wird durch PaLM-E und solche Systeme, ist, dass man wirklich sprachlich mit dem Roboter interagieren kann und das gibt schon einen anderen Eindruck. Ich selbst hatte auch dieses Erlebnis, bisher habe ich immer nur meine Roboter programmiert und so weiter, aber wenn man mal in einem Labor ist und wirklich so ein System installiert ist und man wirklich direkt per Sprache Instruktionen geben kann, gibt es eine ganz



andere Interaktionsmöglichkeit und es ist glaube ich schon so, dass das vielleicht die Schwelle überschreitet, dass es Leute auch interessant fänden, tatsächlich so etwas für die Pflege oder für den Haushalt einzusetzen. Ich glaube jetzt nicht, dass sich durch PaLM-E fundamental ändert, was Roboter physisch können, also das sie jetzt besser Kaffee kochen können oder besser etwas lernen können. Das ist nicht der zentrale Punkt, sondern bei PaLM-E ist es wirklich der Punkt, dass jetzt plötzlich eine Verschränkung mit Sprache und auch diesem allgemeinen Sprachwissen und dem dahinter versteckten Verständnis, Common Sense, hergestellt wird, eine Verbindung dessen zur Robotik-Handlung.

Moderator [00:22:36]

Zur Akzeptanz von solchen Robotern an Sie, Frau Krämer: Haben teils selbstständig agierende Service-Roboter hierzulande eine Zukunft als geschätzte Helfer in Altenheimen, in der Gastronomie oder werden beispielsweise Europäer oder Amerikaner anders als vielleicht Menschen in Japan in solchen Robotern potenziell gefährliche, seelenlose Gestalten sehen?

Nicole Krämer [00:23:01]

Ich komme zu Beginn erst einmal noch mal zurück auf das, was ich eben sagte. Im Moment wird es sehr stark von Medien geprägt, was wir für ein Bild von Robotern haben, ob wir die akzeptieren. Und da gibt es häufiger Mediendarstellungen von irgendwelchen bösen Robotern, die die Welt-herrschaft übernehmen wollen. Deswegen mag es da noch Skepsis geben, aber in der Tat werden Roboter ja auch seit vielen Jahren schon im Alltag in Deutschland eingesetzt. Beispielsweise so kleine, wirklich einfache, nicht mit dem, was es jetzt an Entwicklung gibt vergleichbare tierartige Roboter, die in Altenheimen eingesetzt werden, damit ältere Menschen, die vielleicht nicht häufig Besuch bekommen, sich sozial ein bisschen beschäftigen können, weil diese kleine Robbe, die es da gibt, kleine soziale Interaktionen mit den Leuten macht. Und an den Stellen gibt es natürlich immer einzelne Personen, die sagen: Auf keinen Fall, so ein künstliches Ding kommt mir nicht in den Arm und andere Leute, die das lieben. Wir haben in vielen Studien gesehen, dass es da erst einmal sehr starke interindividuelle Unterschiede gibt, wie die Leute darauf reagieren. Ich würde aber erwarten, wenn erst einmal die Roboter verfügbar sind und man sie als hilfreich erlebt – Akzeptanz hängt ja auch immer sehr stark davon ab, inwieweit man erlebt, was nutzt mir etwas, sei es jetzt auf sozialer Ebene oder tatsächlich als Hilfe. Also die Krankenschwester, die erlebt, der Roboter, der jetzt auf meinen Befehl hin eine Person sanft anheben kann, ist natürlich dann eine unheimliche Hilfe im Alltag, wenn man selber nicht mehr 90 Kilo hochheben muss, um eine Person im Bett zu bewegen, dann wird sich sicherlich auch die Akzeptanz ändern. Die Tatsache, dass in Japan sehr viel mehr Akzeptanz tatsächlich auch empirisch gemessen werden kann, hängt sicherlich auch damit zusammen, dass im Alltag schon auch mehr Roboter zu finden sind im Vergleich zu bei uns.

Moderator [00:25:09]

Danke. An Herrn Schütze haben wir hier eine Frage. ChatGPT ist noch recht schlecht, was deutsche Sprache angeht. Speziellere Anforderungen wie reimen oder Silben zählen macht es schlecht oder falsch. Liegt das nur an dem unzulänglichen und unzureichenden deutschen Futter in seinem Sprachmodell? Und wo sind generell sprachliche Grenzen bei einer KI, gerade wenn es um kreative Anwendungen geht?

Hinrich Schütze [00:25:32]

Ja, ich denke, das ist wirklich der Tatsache geschuldet, dass die Trainingsdaten für das Englische sehr viel größer und umfangreicher sind und eine sehr viel bessere Abdeckung haben. Und auch



die manuell erstellten Daten, die die Fähigkeiten rauskitzeln aus dem Grundmodell, die sind ja auch Englisch vor allem. Und das führt dann auch dazu, dass das deutsche Modell, dass die deutschen Fähigkeiten des Modells nicht so gut sind. Also ich sehe da keinen Unterschied, dass jetzt Deutsch eine schwierigere Sprache wäre als Englisch. Ich denke, dass es nur eine Frage der Zeit ist, bis das gelöst ist. Es gibt natürlich weniger deutsche Daten als englische Daten, das muss man auch sehen. Aber ich denke, dass das in in relativ kurzer Zeit aufgeholt wird.

Ja, echte Kreativität, das ist eine gute Frage. Da muss man sich also überlegen, was ist das Modell, was man hat von der Funktionsweise dieser Sprachenmodelle. Und mein Modell ist eigentlich, dass sie Ähnlichkeitsmodelle sind. Das heißt, sie haben ein sehr, sehr großes Trainingsset, und wenn es im Trainingsset genug ähnliche Vorkommen gibt, die interpoliert werden müssen, die transformiert werden müssen eventuell, aber der Raum, in dem sich eine bestimmte Anfrage bewegt, ist nicht dicht genug besetzt im Trainingsset, dann brillieren sie. Aber das ist ja nicht wirkliche Kreativität. Man kann natürlich sagen, dass einen existierenden Text in einem bestimmten Stil wiederzugeben Kreativität ist, oder als Stabreim wiederzugeben, dass das Kreativität ist. Aber das ist doch eine relativ begrenzte Art von Kreativität. Man könnte sagen, es ist keine echte menschliche Kreativität. Auf der anderen Seite ist es natürlich so – ich verwende da gerne den Vergleich zur Fotografie. Bei der Fotografie könnte man auch behaupten, das ist ja nicht kreativ, man lichtet ja einfach nur ab, was in der Realität schon besteht. Aber natürlich ist bei der Fotografie doch Kreativität mit im Spiel. [Die] ist in der Auswahl, dass man die Motive auswählt. In den technischen Einstellungen des Fotoapparats und auch in der Nachbereitung und vielleicht auch Vorbereitung. Und genau dasselbe gilt für diese Modelle. Man muss sich sehr sorgfältig überlegen, was ist der Prompt? Normalerweise wird es nachbearbeitet. Also zum Beispiel dieser Vermeer, der in den Niederlanden im Museum ausgestellt wird, der ist nachbearbeitet worden. Der ist nicht einfach roh aus dem Modell da ins Museum gewandert. Und Auswahl ist auch sehr wichtig. Also wenn man 20 Texte generieren lässt von diesen Modellen, dann ist die Wahrscheinlichkeit hoch, dass einer ziemlich gut ist. Aber man braucht eben einen Menschen, der diesen einen sehr guten Text aus den 20 auswählt. Also insofern: Kreativität ja, aber ich würde immer noch denken, dass die Kombination von Mensch und Maschine notwendig ist für echte Kreativität.

Moderator [00:29:01]

Ich ziehe mal ein paar Fragen an Sie vor, Frau Krämer, weil Sie ja gleich los müssen. Hier ist die Frage von einer Journalistin. Sie sieht sehr wenig Informiertheit im Umgang mit bereits vorhandenen Systemen und sehr wenig Zurückhaltung selbst von technisch versierten Menschen, was die Verwertung der eigenen Daten angeht. Wie können wir uns psychologische Resilienz und eine neue Mündigkeit aneignen, ohne in eins der Extreme von Algorithmic Appreciation oder Aversion zu verfallen?

Nicole Krämer [00:29:28]

Vielen Dank für die Frage. Dann kann ich noch einmal etwas differenzieren. In der Tat habe ich ja eben gesagt, Menschen sind kompetent bei der Informationssuche. Aber wo Menschen, und das zeigen Studien auch immer wieder, überhaupt nicht kompetent sind, ist in der Frage, die eigenen Daten zu schützen. Und das werden wir bei all diesen Sprachmodellen haben, die ja auch Daten sammeln, zum Teil um sich selber zu verbessern. Das war bei Amazon, Alexa, Google [Assistant] immer schon so, das wird bei ChatGPT sicherlich auch nicht anders gehandhabt. Das heißt, alle Daten, die die Systeme kriegen können, sammeln die. Es gibt eine, wie ich finde, hilfreiche Theorie, der sogenannte Privacy Calculus, der davon ausgeht, dass Menschen für sich so ein bisschen berechnen: Was bringt mir das, mich jetzt hier mit Alexa zu unterhalten? Da kriege ich Informationen. Was bringt mir das bei ChatGPT, eine Anfrage einzugeben? Und da werden vielleicht meine Daten auch irgendwie gesammelt. Aber macht ja nichts, ich kriege ja dafür eine gute Antwort. Also diese



Balance zwischen: Was kriege ich dafür und was muss ich geben? Und das verstehen die aller, allermeisten Menschen nicht, dass man relativ viel gibt dafür, indem man seine eigenen Daten preisgibt, die von Unternehmen dann eben zu Geld gemacht werden. So funktioniert unsere Data-Economy-Gesellschaft. Das ist aber vielen nicht klar. Es ist auch nicht klar, wie diese Daten weiterverarbeitet werden können. Denn die Gefahr ist vielleicht auch nicht unbedingt die einzelne Info, die ich durch meine Anfrage an das System gebe, sondern die Tatsache, dass das wieder mit anderen Daten zusammengebracht werden kann. Wodurch die Möglichkeit entsteht, über mich, meine Person, meine Vorlieben, meine Religion, was auch immer, unheimlich viel zu inferieren, also zu schließen. Und das ist den Menschen nicht klar. Deswegen hatte ich eingangs gesagt, Transparenz ist ein wichtiger Aspekt. Da müssten die Systeme selber besser informieren, was mit den Daten passiert, was mit den Daten auch noch in weiteren Auswertungen gemacht werden kann.

Dann ist natürlich immer eine beliebte Forderung zu sagen: Ja, die Menschen müssen einfach besser informiert werden durch zum Beispiel Schule oder andere Bildungsanbieter. Das ist aber zum einen schwierig. Also was soll die Schule noch alles leisten? Und in manchen Bundesländern wird da schon einiges gemacht, aber das reicht sicher nicht aus. Sodass man eigentlich von gesetzgeberischer Seite auch noch mal stärker darauf hinwirken müsste, dass die Unternehmen, die Daten sammeln, eben noch einmal stärker diese informierte Einwilligung in die Nutzung der Systeme, die ohnehin schon im Gesetz steht, dass das noch weiter ausdifferenziert wird, dass klar wird, wie muss denn so eine Information erfolgen? Nämlich so, dass die Leute wirklich nachvollziehen können, was mit den Daten passiert.

Moderator [00:32:31]

Und vielleicht auch noch eine Frage, die an Sie geht, aber ansonsten können die anderen auch gerne ergänzen. KI soll ja auch jetzt noch mehr in die Suchmaschinen einziehen. Besteht dann die Gefahr, dass uns Menschen irgendwann die von der KI zusammengefasste Antwort reicht und wir gar nicht mehr das Bedürfnis haben, den Links zu folgen und die Informationen zu verifizieren? So ähnlich wie uns bei den News auf dem Smartphone auch viel kürzere Happen jetzt schon reichen?

Nicole Krämer [00:32:56]

Ja, das hat zwei Seiten. Ja, ich glaube, dass Suchmaschinen in Zukunft ersetzt werden durch ChatGPT, was damit zusammenhängt, dass ChatGPT jetzt wirklich eine neue Eigenschaft hat, die für Menschen auch ganz wichtig ist, nämlich sie agiert zum allerersten Mal wirklich dialogisch. Also man kann noch mal eine Nachfrage stellen, man kann das präzisieren lassen, man kann noch mal von einer anderen Seite her fragen, und das System kann darauf antworten. Das war für uns, und ich arbeite seit 20 Jahren in dem Bereich Dialogsystem – nicht, dass ich sie entwickle, aber dass ich die Entwicklung verfolge und wir Studien dazu machen – das konnte bisher kein System. Und das ist unheimlich hilfreich auch bei der Informationssuche. Das würde man ja auch beim Arzt zum Beispiel noch einmal machen, um zu verstehen, wie funktioniert die Behandlung. Ja, dann frage ich noch einmal nach. Aber das habe ich jetzt nicht verstanden. Wie funktioniert das? Das kann ChatGPT jetzt auf einmal tatsächlich wesentlich besser als alle vorherigen Systeme. Deswegen glaube ich, dass diese Art von System die einfache Suchmaschinensuche ablösen wird auf Dauer. Ich würde aber nicht unbedingt davon ausgehen, dass Menschen dann schlechter werden in dem Hinterfragen. Ich hatte eben schon gesagt, Menschen sind kompetent dahingehend, Dinge auch zu hinterfragen, insbesondere, wenn ihnen die Information wirklich wichtig ist. Wenn es für mich darum geht: Habe ich vielleicht eine Krebserkrankung? Muss ich zum Arzt oder nicht? Dann werde ich sorgfältig suchen. In solchen Situationen werde ich auch das, was ich da ausgespuckt bekomme, erst einmal hinterfragen. Aber es gibt gewisse Gefahren, die habe ich ja auch schon genannt. Wenn das menschlich daherkommt durch die Art und Weise, wie es gestaltet ist und wie es mit mir



spricht. Wenn ich so eine Art Vertrauensverhältnis aufbaue wie zu meinem langjährigen Arzt, dann wird es natürlich wiederum gefährlicher.

Hinrich Schütze [00:34:58]

Wenn ich eine Sache ergänzen darf. Frau Krämer hat ja mehrfach darauf hingewiesen, wie wichtig Transparenz ist. Und da sehe ich schon eine große Gefahr. Weil beim jetzigen Suchmaschinenformat ist es viel leichter, Transparenz herzustellen. Weil ich genau weiß, da gibt es eine Webseite und ich kann zum Beispiel sehen, ist das eine Organisation, die einen guten Ruf hat oder nicht? Und ich kann mir tatsächlich den Kontext ansehen, in dem [eine] bestimmte Information steht. Wenn ich mir stattdessen durchlese, was ChatGPT oder GPT-4 produziert hat, das ist dann direkt nicht mehr da. Und natürlich arbeiten die Suchmaschinenfirmen daran, dass man dann eine Ebene weiter gehen kann und sehen kann, woher kommt eigentlich die Information? Aber da sehe ich schon eine Gefahr, dass die Transparenz, die wir jetzt haben, dass die erst mal weg ist, und dass dann die Leute mehr arbeiten müssen, um die Transparenz wieder herzustellen. Also ich denke, da besteht auf jeden Fall eine Gefahr.

Moderator [00:36:07]

Nur kurz, falls Sie sich wundern, die Frau Krämer hört uns noch zu. Aber sie ist jetzt unterwegs. Und deswegen hat sie das Video aus. Eine Frage an Sie, Herr Toussaint. Beschäftigt Sie bei der Forschung an PaLM-E auch das Problem des Dual-Use und wie gehen Sie damit um?

Marc Toussaint [00:36:29]

Ja, natürlich muss einen das beschäftigen. Wir stellen ja hier Technologien her, die ganz fundamental für viele Dinge benutzt werden können, in der Robotik generell. Robotik bedeutet, dass wir Systeme haben, die in die Welt eingreifen können. Und die können ja zum Guten, zum Schlechten, also für viele Dinge eingreifen. Also uns ist das bewusst, wenn wir Robotikforschung machen. Gleichzeitig geht es uns immer wieder darum, aber auch zu versuchen, Grundlagenforschung zu machen und auch uns darauf zu fokussieren und auch immer wieder zu betonen, dass ich das ungern so sehe, dass die KI-Forschung einem einzigen Ziel hinterherrennt, nämlich dem Ziel der Schaffung einer KI oder eines Systems, das besser ist als andere. Sondern ich möchte schon eigentlich meine Forschung und auch die KI-Forschung als solche als Grundlagenforschung verstehen und insbesondere auch die Robotik. Das heißt, wir wollen möglichst viel über diese Systeme verstehen. Und was die Dual-Use-Frage angeht, versuchen wir einfach dann auch zu erklären. Ich glaube, das ist das Beste, was man als Grundlagenforscher machen kann, die Dinge, die man verstanden hat, zu erklären. Für diejenigen, die dann auch verschiedene Anwendungen damit erzielen wollen.

Ich hätte vielleicht auch noch einen Kommentar zu den Diskussionen von Frau Krämer und Herrn Schütze. Beide sind deutlich kompetenter in der wertenden Einordnung solcher Systeme. Vielleicht ein Punkt, den ich machen würde und der mir manchmal auffällt: Viele Menschen denken, wenn sie mit ChatGPT interagieren, dass sie im Grunde mit einer KI interagieren, mit einer Maschine, mit etwas Künstlichem. Das ist natürlich vollkommen korrekt. Und wenn man das so sieht, dann stellen sich alle diese Fragen, die wir auch gerade diskutiert haben. Eine andere Sichtweise ist auch, dass solche Systeme wie ChatGPT im Grunde ein Kondensat von menschlichen Daten sind, also ein Kondensat dessen, was wir in unserer menschlichen Kultur geschaffen haben, und zumindest dessen, was jetzt auch digital verfügbar ist. Das ist auch sehr biased übrigens, auch ein Punkt, dass die Daten, die über unsere menschliche Kultur vorhanden sind und digitalisiert sind, dass die sehr biased sind. Aber im Grunde sind solche Systeme wie ChatGPT, die auf massiven, allen verfügbaren digitalen Daten unserer eigenen menschlichen Kultur trainiert werden, wie ein digitaler Spiegel genau dieser Kultur. Und insofern gibt das vielleicht eine ein bisschen andere Sicht. Also, ich sehe das



manchmal gar nicht so, dass man hier wirklich mit so einem sehr künstlichen System interagiert, sondern natürlich mit einem künstlichen Kondensat, aber von Daten, die ursprünglich von Menschen kommen. Und so ist es natürlich auch so, dass viele Dinge, die jetzt in den Daten [enthalten] sind, dass die genauso unsicher und genauso unperfekt sind wie eben Dinge, die oft auch im Netz zu finden sind, die in Daten, die von Menschen erzeugt werden, zu finden sind. Das nur so als Kommentar, dass man vielleicht auch ein bisschen eine andere Perspektive darauf hat, womit man hier eigentlich interagiert.

Moderator [00:39:41]

Eine Frage haben wir noch an Sie, Herr Schütze. Sie haben ja im Eingangsstatement gesagt, dass größere Datenbasen nicht alles lösen würden, sondern dass es andere Methoden brauche. Auf welche beziehen Sie sich da? Haben Sie da vielleicht ein paar Beispiele?

Hinrich Schütze [00:39:55]

Ja, das ist in der Tat ein großes Problem. Also dass es dafür keine klare Lösung gibt. Also eine Möglichkeit sind neuere symbolische Methoden, das heißt, Methoden, die sowohl eine neuronale Komponente haben. Also die Modelle, mit denen wir jetzt arbeiten – ChatGPT, GPT-4 und so weiter –, sind ja rein neuronal. Sie haben zwar eine symbolische Ausgabe, aber die Interna sind alle neuronal und nicht wirklich verständlich für den Menschen. Und es gibt eben auch ein anderes Paradigma, das neurosymbolische Paradigma, wo auch die Interna des Modells teilweise symbolisch sind. Also das ist auf jeden Fall vielversprechend. Allerdings ist es sehr schwer, das tatsächlich umzusetzen, sodass man dann auch hochperformante Systeme hat wie GPT-4. Natürlich die Multimodalität ist ein anderer sehr wichtiger Bereich. Weil mit der Multimodalität gibt es eine Verankerung der Konzepte und der Sprache in der physischen Realität. Und das ist auch wieder etwas, was unter Umständen eine völlig andere Entwicklungsmöglichkeit ist für diese Modelle, die wegführt davon, dass man nur aus Text lernt. Es ist halt sehr schwer, einige Dinge von Text zu lernen, wenn es überhaupt möglich ist.

Moderator [00:41:26]

Ich sehe jetzt schon, dass wir mehr Fragen haben, als wir wahrscheinlich in der Zeit schaffen werden. Ich hätte hier noch eine an Sie, Herr Toussaint. Vielleicht können wir gucken, da etwas schneller durchzukommen, damit wir noch ein paar Fragen durchkriegen. Können Sie Beispiele nennen, wie PaLM-E die Interaktion mit Robotern vereinfacht? Und was leisten die neuen Algorithmen über die Sprachbefehle hinaus? Also wie genau verbessern multimodale Transformermodelle die Lösung geometrischer Probleme und letztlich die Steuerung von Robotern im Raum?

Marc Toussaint [00:41:56]

Keine so einfache Frage, die man schnell beantworten kann. Es wurde schon öfter gesagt, dass [es wirklich neu ist], dass man interaktiv, dialogisch, wie Frau Krämer das gesagt hat, mit dem System interagieren kann, insbesondere mit einem Roboter. Und dass man womöglich jetzt, weil diese Systeme multimodal sind, sich nicht nur mit Sprache auf die Welt beziehen kann, sondern auch, indem man Objekte wirklich in der Welt quasi markiert oder auf Objekte in der Welt zeigt. Dass es wirklich eine andere, eine neue Interaktionsmöglichkeiten ist mit Robotern, die zu deutlich mehr Akzeptanz führen kann, als wenn man Systeme programmieren muss oder man nur rein sprachlich mit ihnen interagieren kann. Das ist das eine. Das andere ist eben die Problemlösungsfähigkeit. Also im Netz, in unseren Daten sind natürlich auch viele Instruktionen, wie man Probleme lösen kann, also auch geometrische Probleme. Wie kann ich einen Knoten öffnen? Wie kann ich eine Flasche



press briefing

öffnen? Wie kann ich – weiß ich nicht, was man alles finden kann im Netz. Und wenn es gelänge, all diese textuellen Beschreibungen von der Manipulation der Welt, von wirklichen Interaktionen mit der Welt zu grounden, dann eröffnete sich ad hoc ein wahnsinnig großes Spektrum an Fähigkeiten dieses Roboters, die wir sonst wirklich nur sehr, sehr schwer erzeugen können als Robotiker.

Moderator [00:43:26]

Eine Frage an Sie noch, Frau Krämer. Kommt man in Zukunft noch durchs Leben, also provokativ gefragt, wenn man mit KI nichts am Hut haben will? Und werden durch KI noch mehr Leute in der Gesellschaft abgehängt?

Nicole Krämer [00:43:41]

Ja, man wird sicherlich Schwierigkeiten haben, wenn man versuchen will, um KI komplett herum zu kommen, weil KI an unglaublich vielen Stellen unseres alltäglichen Lebens vorhanden sein wird. Selbst wenn man sich entscheidet, keinen Sprachassistenten in die Wohnung zu stellen und nicht ChatGPT zu nutzen, wird man vielleicht in sozialen Medien unterwegs sein. Und auch da sorgen Algorithmen und künstliche Intelligenz dafür, dass ich bestimmte Nachrichten bekomme und andere nicht sehe et cetera. Es ist wahrscheinlich heute noch kaum vorstellbar, wie viele Bereiche des täglichen Lebens das tatsächlich betreffen wird. Ich glaube, ganz entkommen kann man der KI nicht. Der zweite Teil der Frage war?

Moderator [00:44:32]

Ob durch KI vielleicht noch mehr Leute in der Gesellschaft abgehängt werden.

Nicole Krämer [00:44:39]

Mhm, ja, das ist schwierig vorherzusagen. Und ich würde sagen, nicht zwangsläufig, weil das Abhängen von Menschen natürlich auch bisher stark dadurch passiert ist, dass man selber vielleicht einfach nicht genug technikaffin war, um sich irgendwo einzuarbeiten. Da wäre natürlich eine Hoffnung, dass die KI, so wie wir es bei ChatGPT ja eigentlich auch schon sehen, auf eine Art menschenfreundlicher wird in der Bedienung. Weil nämlich, ich hatte es eben schon angesprochen, die Bedienung dialogischer wird. Also es wird leichter, mit dem Ding einfach zu sprechen. Und das würde natürlich vielen Menschen erleichtern, auch mit KI umzugehen. Im Vergleich dazu, man müsse irgendetwas programmieren oder sich ein Gerät anschaffen, das man auch erst verstehen und zum Funktionieren bringen muss. Auf eine andere Weise gibt es natürlich andere Probleme bei KI, das wurde eben kurz angesprochen, dass vielleicht durch die Art und Weise, wie [das System] funktioniert, wie es auch auf Daten zugreift, die ja von Menschen sind, wie wir eben schon richtig gehört haben, natürlich Biases, Fehler reinkommen, die wir als Mensch halt machen – Vorurteile, Stereotype und so weiter. Da gibt es auch eine große Diskussion darüber, wie fair eigentlich solche Algorithmen und künstliche Intelligenz sind. Und die sind eben an vielen Stellen nicht fair, weil sie eben auf unfairen menschlichen Daten beruhen. Und an so einer Stelle kann man sich natürlich dann auf eine andere Art und Weise wieder vorstellen und muss in Betracht ziehen, dass eben dann wieder Menschen übervorteilt werden durch KI.

Moderator [00:46:27]

Okay, also das kann in beide Richtungen ausschlagen. Bevor wir zu den Abschlussfragen kommen, noch eine Frage an Sie, Herr Schütze. Und dann müssen wir, glaube ich, langsam zum Abschluss



kommen. Die Large Language Models laufen bisher ja in der Cloud, aber es gibt schon Beispiele von solchen Modellen, die auch lokal laufen. Wie lange dauert es noch, bis so etwas wie GPT lokal auf dem Smartphone läuft? Vielleicht nicht nur in der Ausgabe, sondern eventuell auch beim Nachtraining der Modelle.

Hinrich Schütze [00:46:54]

Also, ich bin fest davon überzeugt, dass das irgendwann passieren wird. Das ist jetzt sehr schwer, da eine Zahl zu nennen. Das hängt von vielen Faktoren ab, wie groß die Fortschritte sind bei der sogenannten Destillierung, wo wir die großen Modelle auf kleine reduzieren. Wie viel schneller die Prozessoren werden, die wir in ein Smartphone reintun können. Also ich glaube schon, dass es da in den nächsten Jahren große Fortschritte geben wird. Aber ich kann jetzt kein Datum nennen, wo wir ChatGPT auf dem Smartphone laufen lassen können.

Moderator [00:47:34]

Gut, dann komme ich zur Abschlussfrage. Wir haben ja bei den vielen Fragen – tut mir leid, dass wir es nicht geschafft haben, auf alle einzugehen – da haben wir schon gesehen, dass es in sehr viele verschiedene Bereiche reinspielt. Was denken Sie denn bei der ganzen aktuellen Entwicklung mit Sprachmodellen, was ist da Ihrer Meinung nach aus Ihrem Bereich das Wichtigste? Was würden Sie den Journalistinnen und Journalisten mitgeben, was Sie heute bei dem Thema für am wichtigsten halten? Fangen wir vielleicht mit Ihnen an, Herr Toussaint.

Marc Toussaint [00:48:07]

Aus meiner Sicht, die ist natürlich geprägt jetzt auch von mir als Robotiker, ist das Interessante, dass das, was man jetzt mit Sprachmodellen macht, über Sprache hinausgehen wird. Und das finde ich sehr spannend. Natürlich, Sprachmodelle sind sehr stark darin, Sprache zu generieren. Aber wir kennen ja auch andere generative Modelle aus der KI, die jetzt auch Bilder generieren, die Videos generieren. Und ich glaube, dass die Entwicklung eben dahin gehen kann, dass solche Systeme einfach mit ähnlichen Architekturen, wie sie jetzt für Sprache entwickelt wurden, nämlich Transformern, die sehr spannend sind, [dass die es schaffen], mehr und mehr Dinge zu verschränken und dadurch auch andere Dinge zu generieren. Das bedeutet zum Beispiel auch, dass Systeme womöglich – was mich interessieren würde, was ich faszinierend finde – auch Maschinen erfinden können, also mechanische Maschinen. Sich überlegen können, wie eine Fabrik zu designen ist. Wo also wirklich auch das physische Design als Gegenstand und vielleicht auch als Output einer KI entstehen kann. Und ich glaube, dass dahin jetzt auch Entwicklungen gehen können und dass das Spannende auch Grundlagenforschung ist.

Moderator [00:49:15]

Vielen Dank. Dann, Frau Krämer, an Sie die gleiche Frage: Was halten Sie aus den ganzen Entwicklungen jetzt in Ihrem Bereich, also Mensch-Maschine-Interaktion, Psychologie, was halten Sie da für das Wichtigste?

Nicole Krämer [00:49:27]

Ja, ich finde besonders wichtig, diesen Aspekt noch mal zu hinterfragen, wie anthropomorph sollten solche Systeme wirklich gestaltet sein? Weil es ja bestimmte Gefahren birgt, über die wir eben schon gesprochen haben. Dass es ein bisschen vielleicht verschleiert auch, was die Systeme



tatsächlich können, wie sehr man ihnen tatsächlich vertrauen kann. Dass es eventuell verschleiert, was auch mit den eigenen Daten passiert. Also kurzum mehr Aufklärung, mehr tatsächliche informierte Einwilligung zu ermöglichen, dadurch, dass Menschen tatsächlich besser informiert sind, ist aus meiner Sicht ein wichtiger Aspekt. Den kann man aus den psychologischen Forschungen schließen, weil wir da sehen, dass Menschen viel zu wenig verstehen, was da vor sich geht. Und es muss dann auch verbunden werden mit rechtswissenschaftlicher und ethischer Forschung, um eben zu schauen, wie kann man denn das einhegen beziehungsweise dafür sorgen, dass wirklich informierte Einwilligung erfolgt und dass Menschen wirklich wissen, mit [wem] sie da sprechen. Also dass es eben kein anderer Mensch ist. Oder eben ein System, das zwar auf Menschen am Ende beruht, aber auch stark fehlerbehaftet ist, um einschätzen zu können, an welcher Stelle kann ich vertrauen, an welcher Stelle sollte ich nicht vertrauen? An welcher Stelle sollte ich dann wirklich auch im Internet noch mal nachsehen? Was ist denn die eigentliche Quelle dieser Aussage? Also da werden wir noch viel mehr Forschung brauchen, um zu verstehen, unter welchen Bedingungen Menschen eben da blind vertrauen. Und wie kann man so ein blindes Vertrauen dann auch verhindern?

Moderator [00:51:25]

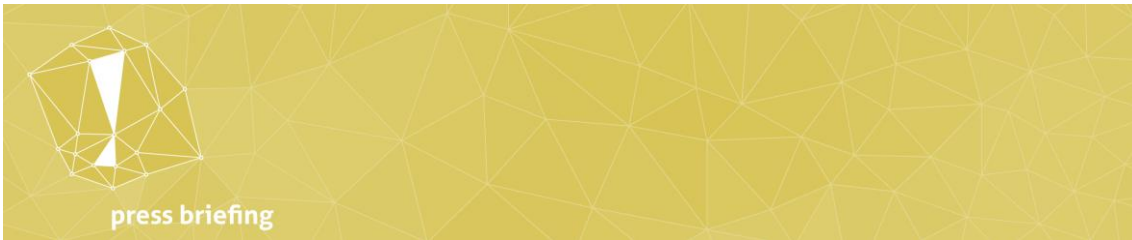
Gut. Und dann an Sie, Herr Schütze, noch mal die gleiche Frage, als jemand, der dann auch wirklich die Sprachmodelle entwickelt und in der NLP-Szene drin ist, was halten Sie da für die wichtigsten Entwicklungen momentan?

Hinrich Schütze [00:51:36]

Also ich würde da zwei nennen. Ich mag den Ansatz, dass man sich auch von Menschen inspirieren lässt. Und der Mensch hat eine sehr komplexe kognitive Architektur im Gehirn, wo zum Beispiel nicht alles mit allem verknüpft ist. Das würde ich jedenfalls so sehen. Da gibt es auch unterschiedliche Meinungen. Deswegen bin ich ein bisschen skeptisch, dass der Transformer wirklich für die nächsten 10, 20, 30 Jahre das Grundgerüst unserer Architekturen sein wird, weil das ist sozusagen eine Architektur, die überhaupt keine Annahmen macht über die Daten. Das ist völlig generalistisch, und ich kann mir nicht vorstellen, dass das richtig ist. Also bei Menschen ist es jedenfalls nicht so. Wenn das Baby geboren wird, dann ist es nicht so, dass alles mit allem gleich gut kommunizieren kann. Also ich denke, architektonisch wird es neue Entwicklungen geben, die sehr wichtig sein werden. Und das zweite, was ich sagen möchte, ist, und das geht jetzt in dieser Debatte vielleicht ganz unter, dass wir meiner Meinung nach nicht echte künstliche Intelligenz erreicht haben. Wir haben zwar enorme Fortschritte gemacht, und ich habe ja gesagt, ich halte das für eine technologische Revolution, aber echte KI heißt für mich, dass man schlussfolgern kann, dass man wirklich ein tiefes Verständnis der physischen Welt hat, der sozialen Welt hat und natürlich auch Mathematik zum Beispiel lernen kann. Und all das können diese Modelle nicht. Das heißt, da ist noch sehr, sehr viel zu tun, und ich glaube, das wird uns noch für Jahrzehnte, wenn nicht länger, beschäftigen. Da wäre ich den Journalistinnen und Journalisten dankbar, wenn sie das immer im Hinterkopf behalten, dass das jetzt nicht die Lösung aller Probleme ist.

Moderator [00:53:35]

Okay. Vielen Dank. Ja, dann ist damit die Zeit jetzt auch schon wieder vorbei. Erst einmal vielen Dank an die Journalistinnen und Journalisten für die vielen Fragen und natürlich vielen Dank vor allem an Sie, Frau Krämer, Herr Toussaint, Herr Schütze! Wir werden heute so schnell wie möglich die Aufzeichnung der Veranstaltung auf unserer Homepage online stellen. Voraussichtlich morgen Mittag oder Nachmittag kommt dann auch das Transkript dazu. Falls Sie eine Audioaufzeichnung oder die Videodatei in Sprecher- oder Galerieansicht zum Download wollen oder ein maschinell erstelltes Transkript, das von uns noch nicht überarbeitet ist, finden Sie in der Reminder-Mail von



heute Morgen einen Link. Darüber können Sie das dann herunterladen. Das laden wir alles hoch, sobald wir es haben. Dann vielen Dank für Ihre Zeit. Ich wünsche Ihnen noch einen schönen Tag und sage auf Wiedersehen.



press briefing

Ansprechpartner in der Redaktion

Bastian Zimmermann

Redakteur für Digitales und Technologie

Telefon +49 221 8888 25-0

E-Mail redaktion@sciencemediacenter.de

Impressum

Die Science Media Center Germany gGmbH (SMC) liefert Journalisten schnellen Zugang zu Stellungnahmen und Bewertungen von Experten aus der Wissenschaft – vor allem dann, wenn neuartige, ambivalente oder umstrittene Erkenntnisse aus der Wissenschaft Schlagzeilen machen oder wissenschaftliches Wissen helfen kann, aktuelle Ereignisse einzuordnen. Die Gründung geht auf eine Initiative der Wissenschafts-Pressekongress e.V. zurück und wurde möglich durch eine Förderzusage der Klaus Tschira Stiftung gGmbH.

Nähere Informationen: www.sciencemediacenter.de

Diensteanbieter im Sinne MStV/TMG

Science Media Center Germany gGmbH
Schloss-Wolfsbrunnenweg 33
69118 Heidelberg
Amtsgericht Mannheim
HRB 335493

Redaktionssitz

Science Media Center Germany gGmbH
Rosenstr. 42-44
50678 Köln

Vertretungsberechtigter Geschäftsführer

Volker Stollorz

Verantwortlich für das redaktionelle Angebot (Webmaster) im Sinne des § 18 Abs.2 MStV

Volker Stollorz

