26.01.2023

Transkript

„ChatGPT und andere Sprachmodelle – zwischen Hype und Kontroverse“

Expertin und Experten auf dem Podium

- **Prof. Dr. Oliver Brock**
  Professor am Robotics and Biology Laboratory und Sprecher des Clusters "Science of Intelligence", Technische Universität Berlin
- **Dr. Thilo Hagendorff**
  Post-Doc am Exzellenzcluster "Machine Learning: New Perspectives for Science", Eberhard Karls Universität Tübingen
- **Prof. Dr. Ute Schmid**
  Leiterin der Arbeitsgruppe Kognitive Systeme, Fakultät Wirtschaftsinformatik und Angewandte Informatik, Otto-Friedrich-Universität Bamberg
- **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/chatgpt-und-andere-sprachmodelle-zwischen-hype-und-kontroverse/
- Falls Sie eine Audiodatei oder eine Sprecheransicht des Videomittschnitts benötigen, können Sie sich an redaktion@sciencemediacenter.de wenden.

Transkript

Moderator [00:00:00]

Hallo liebe Journalistinnen und Journalisten, erst mal herzlich willkommen hier zu unserem virtuellen Press Briefing zu KI-Sprachmodellen. Insbesondere interessant ist für uns da jetzt natürlich gerade ChatGPT. Mein Name ist Bastian Zimmermann. Ich bin Redakteur beim Science Media Center und habe bei mir noch eine Expertin und zwei Experten. Erst mal herzlich willkommen an Sie da draußen, und schön, dass Sie drei alle da sind. Vielen Dank erst mal, dass Sie sich die Zeit nehmen. Ich stelle Sie gleich noch im Detail vor.

Ja, den KI-Chatbot ChatGPT haben wahrscheinlich alle Anwesenden hier schon zur Genüge ausprobiert. Da brauche ich jetzt keine große Einleitung mehr. Die Medien und Twitter waren ja auch voll von Berichten über das Potenzial, aber auch Fragen zu Problemen, Fehlern. Und darüber möchten wir heute sprechen. Also so Sachen wie: Wie verlässlich sind die Aussagen von ChatGPT oder anderen Sprachmodellen? Wie sieht es mit der Skalierung solcher Sprachmodelle aus? Aber auch Probleme wie toxische Sprache oder merkwürdige Fehler könnten hier Themen sein.

Das hängt aber natürlich auch von den Fragen ab, die Sie von draußen stellen. Und dazu, bevor ich mit dem Vorstellen beginne, noch kurz der Hinweis an Sie da draußen: Wenn Sie Fragen haben, stellen Sie die bitte über das Fragemodul von Zoom, nicht über den Chat. Da sehen das manchmal nicht alle Teilnehmerinnen und Teilnehmer. Und so sehen dann eben alle die Nachrichten, dann gibt es weniger Dopplungen. Also benutzen Sie bitte das F&A-Modul und nicht den Chat.

Dann komme ich zu unserer Expertin und den Experten und stelle sie eben kurz mal alle vor. Erstens haben wir Prof. Dr. Oliver Brock, Sie sind Professor am Robotics and Biology Laboratory und Sprecher des Clusters "Science of Intelligence" an der Technischen Universität Berlin. Und gerade wegen dieses Fokusses auf Science of Intelligence ist Ihre Expertise natürlich für das Thema der Sprachmodelle sehr interessant. Man erinnere sich noch an die Debatte um Googles Lamda.

Dann haben wir noch Dr. Thilo Hagendorff. Sie sind Post-Doc am Exzellenzcluster "Machine Learning: New Perspectives for Science" an der Eberhard Karls Universität Tübingen. Sie sind von Haus aus Ethiker, beschäftigen sich da unter anderem mit KI- und Technikethik, aber seit einigen Jahren auch viel mit Sprachmodellen und auch mit den technischen Aspekten. Also haben Sie neben dem ethischen Blick auch noch das technische Verständnis zu diesem Thema.

Und zuletzt haben wir dann noch Prof. Dr. Ute Schmid. Sie ist Leiterin der Arbeitsgruppe Kognitive Systeme an der Fakultät Wirtschaftsinformatik und Angewandte Informatik an der Otto-Friedrich-Universität Bamberg. Sie beschäftigen sich ja unter anderem mit erklärbarer KI und KI und Bildung. Und gerade die Frage nach der Nachvollziehbarkeit von ChatGPTs-Aussagen und auch der Auswirkung auf Bildung und Lehre sind natürlich ebenfalls sehr interessant.

Dann kommen wir zu dem kurzen Eingangsstatements. Beginnen wir mal mit Ihnen, Herr Brock. Und die Frage, die ich an Sie habe, wäre: Inwiefern stellt ChatGPT in der KI-Forschung jetzt gerade eigentlich einen Durchbruch dar?

Oliver Brock [00:02:45]

Ja, wenn ich das jetzt knapp beantworten müsste, würde ich sagen, ChatGPT stellt in der KI-Forschung keinen Durchbruch dar. Und zwar aus zwei Gründen: weil es weder wirklich ein Durchbruch ist, noch wirklich was mit KI zu tun hat. Also ein Durchbruch ist es nicht, weil es eine serielle und inkrementelle Verbesserung von Sprachmodellen über die letzten Jahre gegeben hat, die jetzt

dazu geführt hat, dass es einfach Fähigkeiten in diesen Sprachmodellen gibt, die zu den Ergebnissen führen, die wir bei ChatGPT sehen. Also das ist kein Durchbruch in dem Sinne, sondern eher eine kontinuierliche Entwicklung.

Und dann würde ich sagen, dass die Lücke zwischen Sprachmodellen und dem generellen Feld von Künstlicher Intelligenz doch noch sehr breit ist. Also selbst wenn man ChatGPT fragt, was ist der Unterschied zwischen KI und maschinellem Lernen, gibt es einem die korrekte Antwort, dass maschinelles Lernen eine Gruppe von Methoden ist, die Anwendung finden in der Künstlichen Intelligenz, aber auch in anderen Gebieten. Also das heißt, das sind Dinge, die sich überschneiden, aber Künstliche Intelligenz ist viel breiter, und das Ziel von Künstlicher Intelligenz ist, würde ich sagen, der Versuch, das, was wir in biologischer Intelligenz verstehen, zu synthetisieren in technologischen Artefakten und dabei auch zu verstehen. Das heißt, da gibt es für mich noch einen ziemlich großen Unterschied.

Dennoch würde ich sagen, dass ChatGPT ein bedeutender Fortschritt ist und vielleicht sogar ein Durchbruch, aber auf einem anderen Gebiet. Und für mich ist das Gebiet die Schnittstelle zwischen Mensch und Internet, also diese riesigen Datenmengen, die im Internet verfügbar sind, zugänglich zu machen auf eine intuitive und natürlich sprachliche Art und Weise. Da könnte man schon durchaus davon sprechen, dass ChatGPT ein Durchbruch ist, auch weil es zum ersten Mal der Fall war, dass das mit so vielen Rechenressourcen einer breiten Öffentlichkeit zugänglich gemacht wurde, was dazu geführt hat, dass es so viel Aufmerksamkeit gab.

Trotzdem ist es natürlich total spannend zu fragen: Welche Auswirkungen können denn die Methoden, die im Kontext von diesen Sprachmodellen und spezifisch ChatGPT entwickelt wurden, auf wirkliche KI-Forschung haben? Und da bin ich aus zwei Gründen ein bisschen zurückhaltend. Erstens glaube ich, was wir momentan sehen, ist eine starke Stichprobenverzerrung. Wir sehen nur die Dinge, die funktionieren. Alle Leute sind sehr positiv motiviert, darüber zu berichten, was alles funktioniert. Wir sehen keine Berichte über das, was nicht funktioniert. Deswegen ist es so wichtig, dass man selber mal damit spielt, um zu sehen, was nicht funktioniert. Das heißt, ich glaube, momentan ist Enthusiasmus da, der die Fähigkeiten von ChatGPT überschätzt. Das ist der erste Punkt, diese Stichprobenverzerrung.

Und das Zweite ist, glaube ich, [fundamentaler]. Die Erfolgsgeschichte vom tiefen Lernen, wenn man sie sich betrachtet, wovon ja jetzt GPT und die Sprachmodelle praktisch eine Ausprägung sind, haben sich alle auf bestimmten Datentypen von bestimmter Dimensionalität abgespielt. Das ist alles niedrigdimensional. Sprache ist ja eindimensional, Bilder sind zweidimensional. Bei Videos, die dreidimensional sind mit der Zeit, wird es schon wesentlich schwerer. Das ist ein Problem, dass wenn wir jetzt – ich bin Robotiker – an Roboter denken, die in der echten Welt intelligent agieren sollen, dann haben wir viel höherdimensionale Probleme. Wir haben eine sich ständig verändernde Welt, wir haben Unsicherheit. Und das sind alles Schwierigkeiten, mit denen ChatGPT und auch viele tiefe Lern-Algorithmen noch nicht adäquat umgehen können. Das heißt, ich glaube, dass dieser Siegeszug oder diese Erfolgsgeschichte sich fortpflanzen wird in vielen Anwendungen, die sich auf diese Art von Daten beziehen, dass wir aber nicht Angst haben müssen, dass jetzt in drei Jahren es wirklich eine Künstliche Intelligenz im Sinne von der biologischen Intelligenz gibt, weil dazu gibt es noch wirklich kategorische Herausforderungen, für die wir noch keine Lösung haben und die aus meiner Sicht auch nicht durch tiefes Lernen oder Sprachmodelle gelöst werden können.

Moderator [00:06:55]

Ja, vielen Dank. Da gibt es einige Ansatzpunkte, auf die wir vielleicht auch später noch in der Diskussion zurückkommen. Aber jetzt erst mal noch die Frage an Sie, Frau Schmid: Inwiefern ist denn eigentlich erklär- und nachvollziehbar, wie Modelle wie ChatGPT jetzt zu Ihren Aussagen kommen?

Ute Schmid [00:07:11]

Das Forschungsgebiet Erklärbare Künstliche Intelligenz, Explainable AI, oft so schick mit XAI abgekürzt, das hat sich so als ein kleiner Hype in dem großen Deep Learning Hype entwickelt, ein bisschen zeitversetzt, seit etwa 2016. Und generell adressiert es das Problem, dass gerade mit den hochkomplexen neuen Netzwerkarchitekturen, wo eben auch Sprachmodelle, die auf sogenannten Transformernetzen passieren, dazugehören, dass diese natürlich aufgrund der hochkomplexen Verrechnungen nicht transparent sind. Übrigens auch nicht für diejenigen, die sie entwickeln. Man kann abstrakt natürlich beschreiben, wie solche Netze je nach Architektur funktionieren. Man kann aber im Einzelfall nicht mehr nachvollziehen, wie kommt genau diese Ausgabe zustande? Und das ist bei einem bildbasierten System, was von mir aus Hautkrebs klassifiziert, genau so der Fall wie bei ChatGPT. Und die XAI-Forschung beschäftigt sich nun mit Methoden, wie kann ich denn das, was innerhalb solcher hochkomplexen Netze, die aus vielen Daten trainiert wurden, passiert, nachvollziehbar machen? Tatsächlich gibt es bisher eigentlich kaum spezielle methodische Arbeiten, wie man das technisch machen kann. Die meisten Arbeiten aus dem XAI-Bereich haben sich bisher auf die Erklärung für bildbasierte Klassifikationsaufgaben konzentriert.

[Aber], Oliver Brock hat es gesagt, wir haben auf einmal die Möglichkeit in Dialog zu treten mit Inhalten des Internets, vermittelt eben über das Transformernetz, das Sprachmodell. Wir können natürlich probieren, nachzufragen und nachzuverfolgen, wo kommen denn bestimmte Aussagen her? Und da, denke ich, steckt vielleicht im Vergleich zu dem anderen bekannten Sprachmodell, was dann aus dem Hause Google kommt, Lamda, ein Problem, ein tieferes Problem bei den GPT-3-Modellen, dass es eben nicht möglich ist, nachzuvollziehen, welche Quellen wann wie in die jeweiligen produzierten Aussagen eingehen. Also anekdotisch, ich hatte vor Kurzem einen Vortrag halten können: Was ist KI und kann KI Gender-Biases haben. [Der Vortrag war] für ein bestimmtes Frauennetzwerk und ich habe dann eben auch vorher ChatGPT gefragt: Was wären denn Maßnahmen, um mehr Frauen für ein Informatikstudium zu begeistern? Und da kamen viele tolle Antworten, Vorschläge, die ich seit Jahren mache, aber auch durchaus neuere. Und dann dachte ich, ist ja interessant und sagte: Aus welchen Quellen stammen denn diese Vorschläge? Und dann kommt eine Antwort, die sicher jetzt alle meine Kollegen hier auch schon gehört oder gelesen haben, wenn sie mit ChatGPT-3 interagieren. Da kam dann die Aussage: Ich bin eine Künstliche Intelligenz und keine Person. Deshalb habe ich keine persönlichen Erfahrungen oder Meinungen zu diesem Thema. Die Vorschläge, die ich gemacht habe, basieren auf meinem Allgemeinwissen. Das finde ich schon mal eine große Aussage. Und meiner Fähigkeit, logisch und kreativ zu denken, da ziehen sich einem dann schon ein bisschen die Augenbrauen hoch. Und jetzt kommt es: Sie stammen nicht aus einer spezifischen Quelle, sondern sind eine Zusammenfassung verschiedener Ideen und Ansätze, die in der Vergangenheit bereits erfolgreich zur Förderung von Frauen in der Informatik eingesetzt wurden.

Okay, gut, was nehme ich daraus jetzt mit? Dass offensichtlich, wenn ich mich nicht korrekt auf Quellen beziehen kann und ich damit auch nicht deren Vertrauenswürdigkeit prüfen kann, dass es damit erst mal für Qualitätsjournalismus keine Gefahr darstellt. Bei Journalismus kommt ja auch noch dazu, gerade wenn es um aktuelle Dinge geht, dass man vielleicht manchmal auch wohin fahren muss und direkt mit Leuten reden muss. Entsprechend würde ich da die Gefahr nicht so hoch sehen. So ein Gebrauchstext: Mach doch mal ein nettes Gedicht zu Onkel Herberts 70. Geburtstag. Schön und gut. Kann man sicher machen. Ist jetzt nicht weiter schlimm.

Jetzt komme ich zu meinem Hauptthema, das Thema Bildung. Da liest man jetzt ja sehr viel als Reaktion: Oh, mein Gott, wir können keine Essays mehr schreiben lassen, Hausarbeiten gehen nicht mehr, weil das kann ChatGPT-3 ja so gut. Auch beim Programmieren, sowohl in der universitären Lehre, in der Informatik, aber auch in Schulen, sagen Lehrkräfte und Dozentinnen, Dozenten: Na ja,

ich kann bestimmte Assignments so nicht mehr stellen, weil ruck, zuck, schreibe eine Taschenrechner-Simulation in Java. Das liefert einem ChatGPT-3. Das kann also auch Programmiertexte erstellen. Ich denke eher, wir sollten uns fragen: Was für eine Chance haben wir durch solche Systeme, durch KI? Denn KI-Forschende treten doch im Allgemeinen dafür an, dass KI unsere Kompetenzen erweitert, vielleicht sogar noch fördert, aber nicht einschränkt. Das heißt, ich muss mich, denke ich, auch im Bildungsbereich fragen, wie vielleicht vor 30 Jahren zum Thema Taschenrechner: Wie kann ich denn Bildung mit KI-Systemen wie ChatGPT-3 gestalten? Und mich dann eher fragen: Was sind eigentlich Kompetenzen, die ich in der heutigen Zeit brauche? Dazu würde ich mich nachher noch auf Fragen freuen. Ich glaube, meine fünf Minuten sind um und ich höre an der Stelle mal auf.

Moderator [00:13:35]

Ja, perfekt. Wir haben auf jeden Fall noch einige Fragen, und das mit der Bildung hatte ich auch noch als explizite Frage aufgeschrieben. Aber dann machen wir die Runde noch mal komplett mit den Eingangsstatements. Und dann an Sie, Herr Hagendorff, die Frage: Wie sind Sprachmodelle wie ChatGPT jetzt eigentlich aus ethischer Perspektive zu beurteilen und was sind da wichtige Aspekte, die man bedenken muss?

Thilo Hagendorff [00:13:57]

Ja, hallo zusammen. Das Gute an der Frage ist, dass eigentlich sie auch ChatGPT beantworten könnte. Aber ich mache es trotzdem. Ich gebe einen kurzen Abriss über die einzelnen Punkte. Ich gehe nicht ins Detail. Der erste wichtige ethische Aspekt ist wahrscheinlich der offensichtlichste, nämlich dieses Problem der Diskriminierung der algorithmischen, der Reproduzierung von Stereotypen, der toxischen Sprache. Das liegt einfach daran, dass auch diese Sprachmodelle die Trainingsreize, die sie bekommen, perpetuieren oder eben reproduzieren.

Der zweite Punkt sind Informationsrisiken. Da geht es darum, dass ich Sprachmodelle möglicherweise benutzen kann, um an entweder private oder sensitive oder möglicherweise auch gefährliche Informationen zu kommen. Zum Beispiel kann ich so ein Sprachmodell fragen, wie ich bestimmte Straftaten begehen kann möglichst optimal oder wie ich sie verschleiern kann oder Ähnliches.

Der dritte große Punkt ist der der Wahrheit oder Unwahrheit. Es ist nicht sichergestellt, dass diese Modelle immer nur richtige Informationen produzieren. Sie können auch unsinnige Informationen produzieren, weil sie einfach kein Konzept davon haben, was Wahrheit und Falschheit ist, sondern sie errechnen einfach nur die Wahrscheinlichkeit für das nächste Wort. Und das kann eben zu solchen Effekten der Halluzination führen. Und da habe ich eben auch Forschung zu betreiben, kann ich gerne in einem späteren Verlauf noch mehr dazu sagen, falls es dazu Fragen gibt. Ich mache mal weiter mit dem Abriss.

Der vierte große Punkt sind natürlich Missbrauchsrisiken. Das bedeutet etwa Kosten für Desinformationskampagnen. Solange sie textbasiert sind, sinken sie natürlich massiv. Es ist aber auch möglich, das ist ein anderer Punkt, etwa Code generieren zu lassen, auch ohne Programmierkenntnisse, den man dann etwa für bestimmte mehr oder minder gefährliche Cyberangriffe einsetzen kann.

Der fünfte Punkt ist der der Mensch-am-Computer-Interaktion. Menschen können eigentlich fast nicht anders, als solche Spachmodelle zu anthropomorphisieren, also zu vermenschlichen. Und diese Anthropomorphisierung kann dann eben zu einem überhöhten Vertrauen führen, und dieses Nutzervertrauen kann man dann quasi nutzen, um auch an private Informationen zu kommen. Wenn ich jetzt ein Sprachmodell etwa frage, wie ich mit Krankheiten umgehen soll oder Ähnliches, passiert das.

Der sechste Punkt ist auch wieder ein riesiger Punkt, nämlich die sozialen Risiken. Das Offensichtliche sind Arbeitsplatzverluste oder Arbeitsplatzveränderungen in allen Branchen, die irgendwie mit Texten arbeiten. Nicht zu vergessen, wenn man über soziale Risiken spricht, ist aber auch so etwas wie Click-Arbeit von Click-Workern, die dazu eingesetzt werden, um solche Sprachmodelle zu feintunen, vor allen Dingen was Normverletzungen angeht oder eben auch diesen Punkt, dass sie möglichst bei der Wahrheit bleiben sollen. Auch dazu kann ich natürlich später gerne noch mehr sagen.

Der siebte Punkt, der letzte wichtige, ganz große Punkt, sind auch ökologische Risiken, die es gibt. Denn man darf nicht vergessen: Solche Sprachmodelle haben, je nachdem mit welchem Strommix sie trainiert und betrieben werden, auch einen gewissen CO2-Footprint, der durchaus sehr hoch sein kann.

Was ich noch unbedingt dazu sagen möchte, ist, dass ich es schade oder vielleicht auch schwierig finde, dass es so eine starke Fokussierung auf die negativen Seiten gibt. Über ChatGPT wird meiner Erfahrung nach recht häufig negativ berichtet. Aber man sollte auch die positiven Seiten sehen. Diese Sprachmodelle ermöglichen uns, in vielen Lebenslagen bessere Entscheidungen zu treffen. Sie sind durchaus auch kreativ. Sie bieten uns kreative Lösungen an, sie bereichern uns mit einem unendlichen Wissen. Die Rolle des Expertentums hat sich hier auch noch mal komplett verändert, wenn ich so einen Zugang zu so einem mächtigen Sprachmodell wie GPT habe. Und dann vielleicht noch zwei Sätze ganz zum Abschluss zur Sprache, das ist eben auch ein ethischer Aspekt: Wie sprechen wir eigentlich über solche Sprachmodelle oder GPT? Da gibt es natürlich starke Stimmen aus der Wissenschaft, die sagen, wir dürfen es nicht vermenschlichen, wir dürfen hier nicht von Intelligenz oder Wissen oder Ähnlichem reden oder das irgendwie personalisieren. Er sagt etwas, sie sagt etwas, sondern es ist ein Es. Ich persönlich denke, dass wir uns daran gewöhnen müssen, dass wir Begriffe, die bislang für Menschen oder Tiere sozusagen reserviert waren, dass die auch zunehmend auf Maschinen übertragen werden, um eben dieses maschinelle Verhalten auch erklären zu können. Und dann mache ich an der Stelle erst mal einen Punkt.

Moderator [00:19:12]

Vielen Dank erst mal, Herr Hagendorff. Jetzt sind auch schon einige Fragen von draußen reingekommen. Ich stelle die erste auch mal direkt an Sie, Herr Hagendorff, weil die beiden darauffolgenden gehen an die anderen beiden. Da können die anderen auch gerne ergänzen. Welche Möglichkeiten und Ideen gibt es denn im Moment, um KI-generierte Texte zu kennzeichnen, also zum Beispiel Wasserzeichen?

Thilo Hagendorff [00:19:34]

Das ist in der Überlegung. Das geht natürlich nur mit Texten ab einer gewissen Länge hinreichend gut. Ansonsten kann ich natürlich ein KI-System trainieren, wiederum generierte Texte zu erkennen, wobei man dann irgendwie so ein arms race hat, so ähnlich, wie sich das bei den Deep Fakes ja auch schon abgespielt hat. Gleichzeitig ist es aber auch so, dass man eigentlich auch als Mensch mit einem guten Auge zu einer gewissen Wahrscheinlichkeit einen generierten Text erkennen kann. Man kann es etwa dadurch sehen, dass nur Wörter benutzt werden, die eben auch statistisch sehr häufig vorkommen. Dass relativ einfache Sätze gebildet werden, dass diese Sätze meistens perfekt sind, was die Grammatik, was die Typologie angeht, sodass man auch ohne technische Hilfsmittel durchaus Möglichkeiten hat, Unterschiede zu sehen.

Moderator [00:20:36]

Dann an Sie, Herr Brock, die Frage: Sie hatten ja eben von dem Weg zur biologischen Intelligenz gesprochen. Welche kategorischen Herausforderungen gibt es Ihrer Ansicht denn noch bevor es große Durchbrüche beim Thema Intelligenz geben kann, also in Richtung, wie Sie sagten, biologische Intelligenz?

Oliver Brock [00:20:52]

Ja, ich denke, dass ein wirklich tiefes wissenschaftliches Verständnis von dem Konzept Intelligenz noch in so weiter Ferne ist, dass es schwer ist, über die Herausforderungen zu sprechen. Ich glaube, dass wir viele verschiedene Ansätze versuchen müssen, um uns diesem Thema zu nähern. Aber es gibt konkrete Herausforderungen. Ob die kategorisch sind, das weiß ich nicht. Aber ich glaube, eine Herausforderung ist, dass es momentan sehr viele Disziplinen gibt, die sich mit dem Thema Intelligenz beschäftigen, zum Beispiel die Psychologie oder die Neurowissenschaft oder auch Erziehungswissenschaft, und dass wir die Wissenskörper dieser Disziplinen integrieren müssen und Widersprüchlichkeiten aufdecken und diese Widersprüchlichkeiten auflösen durch neue Erkenntnisse. Also dass wir erst mal so einen wirklichen trans- oder interdisziplinären Ansatz verfolgen müssen, wo alle diese Disziplinen miteinander in Kontakt kommen. Dann, glaube ich, ist eine Herausforderung, dass wir Wissenschaftlerinnen und Wissenschaftler darin trainieren müssen, an diesen Grenzbereichen zwischen diesen Disziplinen zu forschen. Und wir müssen uns neue wissenschaftliche Methoden überlegen, weil es, glaube ich, ein neuartiges wissenschaftliches Problem ist, mit dem wir da konfrontiert sind. Und was ich jetzt sage – nicht überraschenderweise – ist genau das, was das Programm von diesem Exzellenzcluster Science of Intelligence ist. Und da haben wir natürlich ganz viele detailliertere Antworten auf diese Fragen. Aber ich möchte jetzt hier nicht zu viel Air Time in Anspruch nehmen dafür. Aber das ist genau das, womit wir uns im Konsortium von 25, 30 Forscherinnen und Forschern, also eigentlich sind wir 150 bis 170 Forscherinnen und Forscher, beschäftigen, um uns diesem Thema zu nähern.

Moderator [00:22:41]

Okay, Sie sind ja nach dem Press Briefing auch nicht aus der Welt. Vielleicht kann da noch was zustandekommen. Die nächste Frage ist, glaube ich, gut für Sie, Frau Schmid, Sie beschäftigen sich ja auch mit KI in der Medizin. Wie kann der Medizinbereich von ChatGPT profitieren? Oder kann er, und wenn ja, dann wie? Und sehen Sie da auch Gefahren durch ChatGPT für den Medizinbereich?

Ute Schmid [00:23:06]

Ich glaube, da kann man einmal ein bisschen zurück in die Geschichte gucken, als IBMs System Watson [eingeführt wurde], das ist eher ein wissensbasiertes KI-System und eben nicht wie jetzt ChatGPT ein gelerntes Transformer-Netzwerk. Aber da hat man viele Hoffnungen reingesteckt, nachdem das wirklich großartig in so etwas war wie der Quizshow Jeopardy. Und ich glaube die Uniklinik in Heidelberg hat unter anderem ja Watson dann auch eingesetzt im Medizinbereich und hat es irgendwann wieder sein lassen, aus verschiedenen guten Gründen. Und ich denke, diese Gründe sind jetzt eigentlich unabhängig davon, dass Watson ein eher wissensbasiertes KI-System und ChatGPT ein trainiertes Sprachmodell [ist]. Ich denke, es gibt wieder Chancen und Risiken.

Was ich gerade als Forschungsthema angehen will, ist tatsächlich: Kann ich ChatGPT nutzen, um so etwas, das technisch Entity Recognition heißt, in Arztbriefen zu machen. Was ja in jedem Beruf, und so auch in der Medizin, immer ein Riesenaufwand ist und wenig Vergnügen, ist eben Dokumentation. In Arztbriefen stecken jetzt jede Menge interessante Informationen, die aktuell tatsächlich gar nicht so leicht zugreifbar sind, weil diese Arztbriefe eben nicht in einer strukturierten Form vorliegen, digital, sodass man hingehen müsste und sagen müsste: Was ist die Diagnose, die da drinsteht? Von wann ist die? Handelt es sich um eine Patientin oder einen Patienten? Geht es da um Vorerkrankungen? Ist es ein Facharzt? War das eine Überweisung? Sie können es sich vorstellen. Viele Fragen, wo wir Menschen wieder ganz einfach [sagen]: Na ja, sieht man doch. Ich lese den Arztbrief. Da steht doch die Diagnose. Kommt uns gar nicht als schwieriges Problem vor. Ist aber für ein System, das erst mal auf Musterverarbeitung basiert und nicht auf dieser semantischen Verarbeitung, die wir Menschen so natürlich und unkompliziert anscheinend machen, nicht so einfach, wie man denkt. Ich denke, es kann ein nützliches Werkzeug sein in so einem Beispiel, wie ich es gerade gesagt habe, es [kann] aber natürlich – ich glaube, Thilo Hagendorff hatte es angesprochen – [...] auch Probleme machen. Ähnlich wie schon jetzt, wenn ich Dr. Google etwas frage, habe ich natürlich ähnliche Probleme. Und wie so häufig sagt der sehr kritische Begleiter der neuen Machine-Learning-Ansätze, Gary Marcus, der das Buch "Rebooting AI" geschrieben hat: "Building Artificial Intelligence We Can Trust." Der hat letztens, ich glaube in der New York Times, einen großen Beitrag gehabt, dass eben auch völlig unklar ist, wer denn haftet. Und er sagt: Wir können darauf warten, wann es die ersten Todesfälle gibt.

Jetzt stimme ich Thilo Hagendorff absolut zu, dass man auch die positiven Seiten sehen sollte, die Chancen der aktuellen Systeme. Wir sitzen hier als KI-Forscherinnen und -Forscher und sind natürlich begeistert davon und glauben auch, dass man mit KI was Gutes erreichen kann. Die Frage ist eben wirklich die soziotechnische Einbettung und viele Grundsatzfragen, die ich jetzt nicht wiederholen will, die der Thilo Hagendorff gestellt hat. Sprich: Ich glaube, es hat große Chancen für die Medizin, aber man sollte beispielsweise nicht sagen: Och ja, das baue ich jetzt in einen Chatbot ein, der berät von mir aus Teenager mit Liebeskummer. Und, um es jetzt mal drastisch zu sagen, in GPT-3, wenn man eingegeben hat – ich habe es jetzt mit ChatGPT-3 tatsächlich noch nicht probiert – und man gibt ein: Ich fühle mich so schlecht, ich möchte mich umbringen. Dann sagt GPT-3, das tut mir leid zu hören. Ich kann dir dabei helfen. [...] Da sehen Sie auch diese Mustererkennungskomponente, auf der solche Systeme basieren, weil natürlich habe ich im Internet, wenn ich Texte angucke, jede Menge Verkaufsgespräche. Und auf "I want" kommt halt sehr häufig "I can help" als Antwort. Das heißt, das ist die natürlichste Fortsetzung.

Moderator [00:28:09]

Wir haben hier praktischerweise alle Fragen, die mich auch interessiert haben, kommen ja sowieso im Chat, deswegen ist das ganz nett. Die ist jetzt vielleicht für alle. Ich stell sie mal zuerst an Sie, Herr Brock. Gerne sonst Ergänzungen. Ist es OpenAI mit GPT-3 und den Nachfolgemodellen jetzt gelungen, einen großen Vorsprung gegenüber den Konkurrenten aufzubauen? Oder gibt es da auch noch andere Modelle im Hintergrund, die genauso gut oder besser sind, nur vielleicht gerade weniger sichtbar? Und wie aufwendig ist es eigentlich, konkurrierende Modelle zum Beispiel jetzt in Deutschland oder Europa aufzubauen?

Oliver Brock [00:28:40]

Ich glaube, ein konkreter Vergleich zwischen den momentan existierenden Modellen ist schwierig ohne eine systematische Evaluation. Zu messen, wie groß der Fortschritt ist. Wo auf jeden Fall der Fortschritt ist gegenüber anderen, ist die Publicity. Das hat auf jeden Fall geklappt. Aber ich glaube, dass sich diese Modelle gegenseitig inspirieren werden und dass wir da immer wieder so ein Kopf-an-Kopf-Rennen [haben werden], und einer hat mal die Nase vorn, dann der andere. Die Frage, was wir in Deutschland tun können, ist tatsächlich, glaube ich, eine ganz fundamentale, die sich auf viele Bereiche der Künstlichen Intelligenz erstreckt, aber insbesondere auf die, wo eben sehr viel Compute-Infrastruktur notwendig ist, wo viele massive Rechenzentren notwendig sind, um die Forschung überhaupt betreiben zu können. Ich denke, dass wir da wirklich hinterher sind. [Ich weiß nicht], wer da handeln muss. Aber irgendjemand muss handeln, denn das scheint ja schon wichtig zu sein, und wir sind auf jeden Fall hinterher. Ich weiß nicht, ob die anderen Vorschläge haben, was wir da tun könnten.

Moderator [00:29:51]

Es gab ja letztens eine Machbarkeitsstudie, dass da 400 Millionen Euro reingesteckt werden müssten. Gibt es da irgendwelche Ideen in die Richtung? Gibt es da irgendwelche Entwicklungen oder irgendwelche vielversprechenden Unternehmen oder Forschungsprojekte in Deutschland, die das schon versuchen?

Oliver Brock [00:30:10]

Ich glaube, dass viele Leute jetzt damit experimentieren. Wenige Leute haben die computionalen Ressourcen, um das auf demselben Niveau zu versuchen oder auch damit nur zu konkurrieren. Und was man tun kann: Ich glaube, da gibt es ganz viele Dinge, die man tun kann. Das fängt an mit der grundsätzlichen Infrastruktur im Sinne von: Wie sind Universitäten ausgestattet, wie sind die Prozesse, Drittmittel einzuwerben bis hin zu auch tatsächlich gezielt zu investieren und zu sagen: Wir stellen an vier Standorten in Deutschland ein riesiges Rechenzentrum zur Verfügung, das nur für so eine Art von Forschung zu benutzen ist, zum Beispiel. Ich weiß nicht, ob das auf Gegenliebe stößt.

Moderator [00:30:56]

Dann eine Frage – ach so, Frau Schmid, ja.

Ute Schmid [00:30:57]

Ich wollte kurz nur ergänzen. Ich glaube, wir haben jetzt, das sollte man nicht vergessen, eine deutsche Firma im Bereich Machine Translation, maschinelle Übersetzung, die sehr erfolgreich ist, nämlich DeepL. Das ist eine deutsche Firma und da stecken natürlich auch ähnliche Modelle dahinter. Ich habe jetzt keinen Kontakt zu der Firma, aber ich würde mal vermuten, dass es sich schon lohnt, Firmen wie diese im Moment enorm zu fördern. Vielleicht ähnlich wie man zu Beginn von Covid Biontech gefördert hat. Ich kann das sagen, weil ich mit der Firma null zu tun habe und da überhaupt niemanden kenne. Es kann auch andere Firmen geben, wo das zutrifft. Ich glaube, da müsste man wirklich investieren, damit Europa da nicht das Heft aus der Hand gibt. China hat ja wohl auch sehr große Sprachmodelle, die noch deutlich größer sind und das heißt, die haben da schon die Infrastruktur und Kompetenz hört man, kann ich jetzt auch nicht beurteilen. Ich stimme Oliver Brock absolut zu, da muss für eine europäische Kompetenz unbedingt Geld in die Hand genommen werden.

Moderator [00:32:18]

Ebenfalls eine Frage, zu der Sie wahrscheinlich alle eine Meinung haben. Aber ich stelle sie erst mal an Sie, Herr Hagendorff. Wie problematisch bewerten Sie den Umstand, dass die jetzt breit genutzten Sprachmodelle und KI-Tools fast alle in der Hand von großen kommerziellen Organisationen oder Konzernen sind?

Thilo Hagendorff [00:32:34]

Grundsätzlich ist es so, dass KI-Technologien auf Netzwerkeffekten aufbauen und Zentralisierung begünstigen, Zentralisierung von Daten, große Datenberge und diese Prozesse finden dann typischerweise auch in einem organisationalen Kontext statt, der kommerzialisiert ist, also in Unternehmen. Ob ich diesen Umstand jetzt per se für problematisch halte: Nein. Ich sehe das auch ehrlich gesagt nicht so, dass wir neidisch sein müssen auf Sprachemodelle in anderen Ländern, dass Europa oder Deutschland jetzt unbedingt auch so etwas braucht. Was ich hinzufügen will, ist, dass viele dieser Modelle Open Source sind, sie können ja genutzt werden. Also es gibt Bloom, es gibt GLM, es gibt von Facebook OTP, das ist quasi wie GPT nur Open Source und sind nutzbar, ohne, dass man dafür zahlen muss. Also das GPT-3.5, was man abseits vom ChatGPT nutzt, kostet Geld. Diese Open Source-Modelle können benutzt werden in nicht kommerziellen Kontexten, sowie in kommerziellen, so wie man es eben möchte. Es ist nicht so, dass sich jetzt alles hinter den hohen Mauern von Firmen befinden würde oder dort irgendwie festgehalten wird.

Moderator [00:34:03]

Herr Brock.

Oliver Brock [00:34:03]

Ja, nur ganz kurz. Die Modelle sind natürlich frei zugänglich und manchmal werden sogar die gelernten Modelle zur Verfügung gestellt. Aber das Wichtige ist, dass das Trainieren, das Erweitern, das Fortentwickeln, obwohl man die Modelle hat, nicht möglich ist, weil man sie nicht trainieren kann. Und das ist schon ein Monopol, das mit Kapital zusammenhängt.

Moderator [00:34:27]

Und sehen Sie das als problematisch an?

Oliver Brock [00:34:32]

Das ist schwer vorherzusagen. Ich denke, dass es ein natürlicher Zyklus ist, dass Dinge aus der Forschung in die Industrie gehen. Das ist ja auch das Ziel der Forschung, wo dann bestimmte Aspekte von den erforschten Dingen besser weiterentwickelt werden können. Und die Frage ist: Sind Teile vom maschinellen Lernen jetzt an dem Punkt, wo das der Fall ist und die akademische Schule des maschinellen Lernens sollte sich auf andere Aspekte von diesem breiten Feld fokussieren, aber nicht mehr auf die Weiterentwicklung von diesen riesigen Modellen? Als Frage.

Thilo Hagendorff [00:35:06]

Vielleicht kann ich das noch ganz kurz ergänzen. Ich sehe es auch als Problem, wenn es eine Zentralisierung bei Sprachmodellen gibt, denn man darf nicht vergessen, diese Sprachmodelle kann man auch normative Fragen fragen, also nicht nur deskriptive, sondern ist etwas, ist XY schlecht oder gut. Dieser normative Standpunkt repräsentiert ja letztendlich eine Art Maschinenmoral. Wenn es eine Maschinenmoral gibt, auf die alle zugreifen, die von einer Firma determiniert wird, dann sehe ich das als sehr problematisch und deshalb glaube ich auch, dass eine Diversität an Sprachmodellen, die öffentlich verfügbar und in Benutzung sind, sehr wichtig ist.

Moderator [00:35:50]

Gut. Ich glaube, wir machen weiter mit den weiteren Fragen, eine, die ich an Sie stelle, Frau Schmid. Wie kann denn die Wissenschaft, wie können Forschende von Sprachmodellen wie ChatGPT profitieren, was müsste sich dafür ändern?

Ute Schmid [00:36:04]

Hmm, das ist jetzt eine sehr große Frage. Ich kann im Moment mit der Frage nicht viel anfangen. Ich würde mal ausweichend antworten und hoffen, dass vielleicht meine Kollegen dazu eine [Antwort] haben. Ich persönlich sehe ganz große Chancen in verschiedensten Bereichen innerhalb der KI und außerhalb, um solche Modelle einzusetzen. Und natürlich auch große Fragen, wie man solche Modelle mit anderen Technologien kombinieren kann, zum Beispiel auch datensparsame Technologien. Beispiel: Aktuell arbeiten wir in einem Verbundprojekt, was das Bundesministerium für Bildung und Forschung (BMBF) finanziert zum Thema KI in der Hochschullehre, zu Themen etwa, wie kann ich KI-Tools in der Hochschullehre einsetzen. Da kann ich etwa solche Sprachmodelle nutzen, die sehr, sehr gut darin sind, Programmcode zu produzieren, aber auch zu erklären. Ich habe sehr viel mit GPT-3 gespielt, mit so Aufgaben wie: Schreibe ein Programm, das..., habe das Programm eingegeben, erkläre mir, was dieses Programm tut, gib mir gute Testfälle, um dieses folgende Programm zu testen und so weiter. Das funktioniert wirklich sehr gut und besser, als wenn ich mit natürlicher Sprache so in Richtung Common Sense Reasoning gehe, wo das Netz schnell zu halluzinieren anfängt. Da würden wir das im Kontext von sogenannten intelligenten Tutorsystemen weiterentwickeln und dort eben diese Fähigkeiten von ChatGPT-3 nutzen, um etwa ganz gezielt bei der Vermittlung von Programmierkompetenzen individuell Feedback zu geben und Aufgaben zu generieren. Das ist jetzt ein Mini-Bereich, da hoffe ich sehr, dass meine Kollegen ergänzen, weil die Frage ja eigentlich sehr breit gestellt war und ich sie jetzt sehr eng an einem Beispiel beantwortet habe.

Moderator [00:38:15]

Herr Hagendorff.

Thilo Hagendorff [00:38:20]

Ein Aspekt, den man häufig nicht sieht, ist folgender, nämlich dass große Teile der Geisteswissenschaften, aus denen ich ursprünglich komme, sich dadurch reproduzieren, dass aus Texten Texte geschrieben werden: Dissertationen, Hausarbeiten, aber auch die Forschung kommt teilweise vollkommen ohne Empirie aus. Man studiert Bücher und setzt aus diesen Büchern Texte zusammen zu neuen Texten, die man selbst schreibt. Und dieser Prozess ist mit gewissen Ausnahmen nicht mehr nötig. Dieser Prozess kann komplett durch Sprachmodelle übernommen werden. Das heißt, ich glaube, der Impact auf diese nicht empirisch arbeitenden Fächer ist massiv, denn das sind ja auch keine innovativen Fächer in dem Sinne, dass dort viele neue Konzepte entwickelt werden, sondern hauptsächlich geht es ja um Exegese vergangener Texte, die schon Trainingsmaterial dieser großen Sprachmodelle sind. Ich glaube hier wird ein extremer Impact sein. Die Wissenschaft, die empirisch arbeitet, ist davor relativ gut gefeit.

Moderator [00:39:27]

Da kommt auch eine gute Follow up-Frage zu. Wir sieht es denn eigentlich mit Regeln aus zum Thema ChatGPT als "Studienmitautorin". Das ist ja schon ein paar Mal in Journals angenommen worden, sagt die Person, die die Frage gestellt hat. Science und Nature haben dazu zuletzt Editorials rausgegeben, wie sie damit umgehen wollen. Haben Sie da auch was mitbekommen, ist es vielleicht auch gerade im KI-Bereich interessant, gibt es da irgendwelche Richtlinien, Herr Brock, wissen Sie da was?

Oliver Brock [00:39:56]

Ich weiß nichts darüber. Ich würde Herrn Hagendorff widersprechen. Natürlich ist die Kombinatorik der existierenden Texte enorm und die Leistungen, die der Mensch da erbringt, ist, aus diesen verschiedenen Kombinatoriken eine herauszusuchen, die tatsächlich interessant ist. Und da würde ich gerne sehen, dass die Treffsicherheit von ChatGPT wirklich konsistent bleibt. Das ist noch zu beobachten, was da passiert.

Thilo Hagendorff [00:40:32]

Da will ich nicht weiter darauf eingehen, führt vielleicht zu weit. Zu der Autorenschaft: Ja, natürlich ist die Idee, dass man erst mal in die Autoren*innen- Liste reinschreibt, dass da ein Sprachemodell mit beteiligt war. Es wirkt dann so ein bisschen komisch und manche denken auch, dass man es einfach verbieten sollte oder so. Ich denke, es macht keinen Sinn, Sprachmodelle mit aufzulisten, weil früher oder später werden das alle benutzen. Man muss sich einfach vorstellen, wir sind in der zeitlichen Transformation, wo die Aufgabe des Schreibens auf ein leeres Blatt ersetzt wird durch die Aufgabe des Editierens von synthetischem Text. Wir werden einfach mehr zu Editoren und Editorinnen. Das an Texten ein Sprachmodell mitgeschrieben hat, das wird in zehn Jahren so selbstverständlich sein, dass unter jedem wissenschaftlichen Paper dann stehen müsste, dass das Sprachmodell XY mitgewirkt hat. Und dadurch ist die Information irgendwann so redundant, dass man sie auch nicht mehr braucht.

Moderator [00:41:31]

Frau Schmid.

Ute Schmid [00:41:32]

Vielleicht als Blick in die Vergangenheit: Als vor vielen Jahren Großprojekte gefördert wurden zur maschinellen Übersetzung von EU-Verträgen in alle EU-Sprachen wurde das am Ende dann erst mal auf Eis gelegt, weil für dolmetschende Menschen das Editieren von teilweise gut, teilweise schlecht übersetzten Texten natürlich viel mehr Aufwand bedeutet, als es gleich selber zu machen. Es gibt ja auch jetzt sehr schön, das kursiert im Internet, so ein Cartoon: Debugging von Programmen vor GPT-3 und jetzt mit GPT-3 und die Debugging-Zeit wird dann einfach um vielfaches größer. Das heißt, da gibt es natürlich wieder spannende Fragen für weitere Forschung, inwiefern man möglicherweise doch Unsicherheit mit berücksichtigen könnte als einen Hinweis wie: Guck dir mal diese Stellen an, Mensch, da bin ich mir unsicherer, dass man da noch mal ein bisschen Unterstützung gibt, ohne dass ich glaube, dass das vollständig lösbar wäre.

Moderator [00:42:49]

Gut. Wir nähern uns auch schon dem Ende der Zeit und haben noch sehr viele Fragen offen. Ich sehe jetzt schon, dass wir nicht durch alle kommen werden. Ich habe noch eine an Sie, Herr Brock, zu der Skalierbarkeit, also ob die Sprachmodelle proportional zu ihrer Größe – man spricht da ja immer von Parametern – besser werden oder muss man immer mehr Rechenkraft für den gleichen Fortschritt aufwenden?

Oliver Brock [00:43:11]

Prinzipiell ist es so, dass je mehr Parameter man braucht, desto mehr Daten braucht man und desto mehr Rechenzeit braucht man auch. Allerdings ist es so, dass die Trainingsprozesse von diesen Modellen jetzt viel komplizierter sind und Zwischenschritte involvieren, wo Menschen noch mal irgendwie klicken und und versuchen, dieses Lernen in die richtige Richtung zu lenken. Das heißt, dann werden ganz andere Kosten relevant, nämlich, wie viel Leute kann ich einstellen, die da mein Modell in die richtige Richtung klicken. Ich glaube, eine direkte Korrelation zwischen der Anzahl der Parameter und den Kosten gibt es nicht, aber als Approximation ist das sicherlich irgendwie passend.

Moderator [00:44:03]

Und ist immer weiteres Skalieren die Lösung oder müssen auch Kombinationen von Ansätzen in Zukunft verwendet werden oder andere Ansätze?

Oliver Brock [00:44:11]

Das wird die Zukunft zeigen. Bisher war es meiner Meinung nach so, dass bei allen Technologien die Kurve der Verbesserungen erst mal sehr stark gestiegen ist und dann irgendwie abgeflacht ist. Ich glaube, dass das auch hier der Fall sein wird. Ich sehe kein Argument, das dem widersprechen würde. Aber der Herr Hagendorff.

Thilo Hagendorff [00:44:35]

Genau, das ist ein häufiges Argument, dass man sagt: Ja, man kann diese Modelle jetzt immer größer skalieren, aber letztendlich die Schwächen bleiben die gleichen. Was man hinzufügen kann, ist, dass auch das reine Skalieren bei gleichbleibender Methode emergente Effekte zeitigen kann. Die beobachten wir überall in der Natur, die beobachten wir an Menschen selbst. Also Bewusstsein zum Beispiel aus einer chemischen Struktur ist ein emergentes Phänomen. Und genauso haben wir

solche emergenten Fähigkeiten bei Sprachmodellen. Eine dieser emergenten Fähigkeiten ist etwa, dass sie rudimentär Mathematik können. Eine andere ist, dass Sprachemodelle klassifizieren können oder auch, dass Sprachemodelle wie GPT Few-Shot-[Learning] oder sogar Zero-Shot-Learning fähig sind. Das sind Fähigkeiten, die hat man nicht bewusst eingebaut, sondern die sind emergiert einfach nur durch Skalierung. Und das ist etwas, das bei KI-Modellen sehr interessant ist, was man im Kopf behalten sollte.

Moderator [00:45:40]

Herr Brock, Sie müssen ja pünktlich weg. Deswegen würde ich Ihnen jetzt die Abschlussfrage schon mal stellen. Da wir so viele andere Fragen haben, würde ich wenn Herr Hagendorff, Frau Schmid, Sie noch fünf Minuten mehr Zeit haben vielleicht gleich noch ein, zwei Fragen stellen. Erst mal an Sie, Herr Brock die Abschlussfrage, bevor Sie los müssen. Inwiefern sind denn jetzt die neuen Sprachmodelle wie ChatGPT die große Sache oder ist momentan das Thema ein bisschen overhyped. Würden Sie sagen, das ist jetzt, nach dem was wir besprochen haben, noch ein Riesendurchbruch und das wird viel verändern oder sind die Veränderungen eher begrenzt?

Oliver Brock [00:46:14]

Es ist beides. Also es ist natürlich ein Hype in dem Sinne, dass ich glaube, der Grad der Aufregung ist durch die Sache nicht gerechtfertigt. Aber gleichzeitig passieren ganz viele Nebeneffekte durch diesen Hype, nämlich, dass sehr viel mehr Geld investiert wird, dass sehr viel mehr Aufmerksamkeit darauf kommt, dass wahrscheinlich viele junge Forscherinnen und Forscher sagen, an so was möchte ich jetzt forschen. Das heißt, dass da schon auch tatsächlich auf breiter Fläche irgendwas passiert und das kann man alles voneinander nicht unbedingt ablösen. Es wird wirklich abzuwarten sein, was da passiert.

Wenn wir uns die Geschichte der Künstlichen Intelligenz angucken, dann gab es immer wieder diese KI-Winter. Nach einer großen Aufregung war man dann enttäuscht, weil es nicht funktioniert hat. Ich glaube nicht, dass es noch mal einen KI-Winter dieser Dimension aus der Vergangenheit geben wird, aber ich glaube schon, dass es Wellen von abflachendem und aufsteigendem Hype geben wird. Insofern glaube ich, sind wir schon in einer Trajektorie in eine neue Ära in der künstlichen Intelligenz – und wenn ich künstliche Intelligenz sage, dann meine ich maschinelles Lernen –, die immer schneller und immer mehr Bereiche unseres Alltags auch berühren wird und dort Relevanz zeigen wird. Ich glaube, dass es wesentlich länger dauern wird, bis diese Entwicklung sich auf die biologische Art von Intelligenz auswirken wird. Da wird es sicherlich eine Auswirkung geben, aber ich glaube, dass da noch viel größere, fundamentalere Probleme auf uns warten, von denen wir noch nicht wissen, was da die Lösung sein könnte.

Moderator [00:47:55]

Ja, vielen Dank. Ich würde vielleicht gerne noch versuchen, zwei Fragen durchzukriegen, bevor ich bei Ihnen beiden anderen auch zu den Abschlussstatements komme. Tut mir leid, dass wir nicht alle Fragen beantworten können werden. Vielleicht an Sie, Frau Schmid, weil Sie sich ja auch mit KI und Bildung beschäftigen. Welche konkreten neuen Anforderungen ergeben sich denn jetzt durch ChatGPT für Studierende und Dozierende an den Hochschulen?

Ute Schmid [00:48:17]

Ja, also eigentlich müssen wir uns in Schulen wie Hochschulen fragen – auch schon länger, nämlich auch seit [es sehr gute Möglichkeiten gibt] Information Retrieval aus dem Internet zu betreiben, bei Wikipedia nachzugucken, was an Qualität seit Beginn wahnsinnig gewachsen ist –, wie wollen wir eigentlich in Zukunft Kompetenzen abprüfen. Was sind Kompetenzen, die wir brauchen und wie will ich sie prüfen? Ich glaube, dass es höchste Zeit ist, dass wir eben nicht mehr fragen, lernt doch bitte folgende Geschichtsdaten auswendig oder anderes Faktenwissen, weil da Schülerinnen und Schüler, Studierende schon lange sagen: Na, das kann ich ja googeln, [...] das kriege ich raus.

Vielfach ist eben Schule und zum Teil aber auch Uni ausgelegt auf effizientes Abfragen von Wissen etwa über Multiple Choice, über Lückentexte, über Abfragen von auswendig Gelerntem und es geht, glaube ich, Herr Hagendorff bis in die Geisteswissenschaften rein. Und ich finde es gerade in MINT-Fächern (Mathe, Informatik, Naturwissenschaft, Technik) extrem problematisch, dass – zumindest wie ich es aus deutschen Schulen kenne, an der Universität ist das dann zum Glück anders – auch in einem Fach wie Physik oder Mathe relativ viel auf stupides Rechnen geguckt wird. Und nicht auf Verstehen, Herleiten und Formalisieren lernen, Umgang mit Mathematik. Spätestens jetzt, wenn alle sagen: Oh mein Gott, ich kann keine Essays mehr aufgeben, ich kann kein Programming Assignment mehr machen, sind wir so weit, dass wir sagen müssen: Ja, wie prüfe ich dann Kompetenzen überhaupt noch? Und natürlich ist es so, die Art, wie ich Kompetenzen abprüfe, bedingt, wie wir lernen. Wenn ich Multiple Choice Tests als Klausur habe, dann werden unsere Studierenden auch auf diese Art von Fähigkeit Recognition lernen. Das ist in vielen Bereichen natürlich nicht sinnvoll. Wenn wir jetzt wieder zum guten alten Prüfungsgespräch gehen, zum Vorlegen von Programmen, die mit einem Übungsleiter, einer Übungsleiterin besprochen werden, macht das sehr viel Sinn, was Qualität von Bildung angeht. Das wäre aber auch enorm personalintensiv. Man kann sich jetzt wieder neuen KI-Support überlegen, wie man diese Art der Kompetenzprüfung mit KI unterstützen kann.

Es wird sich viel verändern, fasse ich zusammen und es muss sich endlich verändern. Es ist eine Riesenchance, zu überdenken, was meinen wir mit Bildung, vielleicht auch, was meinen wir in Europa mit Bildung, was meinen wir mit Bildung im Zeitalter von KI-Ansätzen.

Moderator [00:51:47]

Ja, vielen Dank. Das ist ja auch noch ein interessantes Thema, Bildung in Zeiten von KI. Aber dann hätten wir jetzt noch eine ethische Frage an Sie, Herr Hagendorff. Für umfassendes Weltwissen muss KI wie ChatGPT ja auch mit Trainingsdaten gespeist werden, die dann zum Teil auch Dinge wie schlimmste Gewalt, Mord, Missbrauch etc. umfassen. Und die Aufgabe des Klassifizierens wird teilweise an Unterfirmen in ärmeren Ländern delegiert. Wie ist das ethisch zu bewerten und wie lassen sich auch künftig KI-Modelle so trainieren, ohne dass Menschen Traumata erleben müssen?

Thilo Hagendorff [00:52:20]

Außerhalb von Sprachmodellen ist der Ansatz schon häufig der gewesen, dass man versucht, die Trainingsdaten zu reinigen von Inhalten, die man nicht wollte. Bei Sprachmodellen ist es aufgrund der riesigen Menge an Texten einfach nicht mehr möglich. Es lässt sich nicht vermeiden, dass man mit Texten trainiert, die Rassismen, Sexismen etc. beinhalten. Deshalb gibt es einen neuen Ansatz, unter anderem ein sogenanntes Reinforcement Learning from Human Feedback. Das bedeutet – genau das hat OpenAI gemacht –, dass Outputs eines Sprachmodells durch Menschen bewertet werden und die Art, wie Menschen diese Outputs bewerten, vor allen Dingen in normativen Hinsichten – also wird dort diskriminiert oder nicht –, dass diese Art des Bewertens dieser Textschnipsel wieder als Training-Size benutzt wird für ein anderes neuronales Netz, ein sogenanntes Reward Model. Und dieses Reward Model interagiert, nachdem es dann trainiert ist, mit dem eigentlichen

Sprachmodell. Das heißt, wir haben ja verschiedene KI-Systeme, die miteinander interagieren und sich gegenseitig beibringen, weniger von diesen Normen verletzenden, diskriminierenden Inhalten zu produzieren. Die lassen sich am Ende des Tages aber natürlich trotzdem nicht oder vielleicht sogar nie zu 100 Prozent vermeiden. Da tritt dann ein altes Phänomen in Kraft, das wir auch aus den sozialen Netzwerken schon kennen: Content Moderation oder eben Menschen, die in Billiglohnländern als sogenannte Klick-Arbeiter angeheuert werden, um all diese Texte, die dann gewaltverherrlichend oder ähnliches sind, zu bewerten. Diese Bewertung muss dann wieder an das KI-System zurückgespielt werden. Das ist ein extrem schwieriger Prozess, der auch mit einem großen psychologischen Preis für die betroffenen Menschen einhergeht. Aber nichtsdestotrotz schafft man es auf diese Art, Sprachemodelle zu entwickeln, die man dann wirklich einfach auch an die Öffentlichkeit geben kann, wie OpenAI das gemacht hat.

Moderator [00:54:39]

Ja, vielen Dank. Dann würde ich bei Ihnen jetzt auch zur Abschlussfrage kommen. Dann fangen wir doch mal mit Ihnen an, Frau Schmid. Inwiefern sind Sprachmodelle wie ChatGPT ein großer Durchbruch oder eine große Veränderung, oder ist das Thema ein bisschen overhyped?

Ute Schmid [00:54:56]

Ich würde sagen, es ist eine Evolution und keine Revolution. Es war natürlich schon, denke ich, ein extrem guter Schachzug von OpenAI, diese Veröffentlichung zu machen und ChatGPT-3 zur Verfügung zu stellen, weil das natürlich sehr viel Aufmerksamkeit produziert und vielleicht sogar noch mal einen Hype ausgelöst hat, wie 2014/2015, als Google Brain erstmals beim End-to-End-Learning direkt ein Bild einer Katze erkannt hat aus einem Image ohne Verarbeitung, was ja auch ein Riesendurchbruch war. Ich glaube schon, dass es sehr spannend ist, nicht ganz neu.

Ich glaube aber, es war selten so viel Bewegung in der KI-Forschung wie heute. Und Oliver Brock hat es schon gesagt, man muss interdisziplinär draufgucken, das fand ein Teil der KI-Forschenden schon immer. Es gab eine Zeit, wo KI sehr narrow war, sich sehr ausdifferenziert hat, sicher zu Recht, wo ein Teil von KI-Forschenden [sich in] Richtung – so wie mein Bereich heißt – kognitive Systeme entwickelt hat. Und da gibt es dann auch entsprechende Journals und Konferenzen und in Cognitive Science finden sie sich auch wieder. Diese interdisziplinäre Betrachtung von KI erlebt jetzt einen ganz neuen Aufschwung und bringt noch mal viele wichtige Erkenntnisse.

Was ich persönlich sehr spannend finde, ist, dass wir aktuell schon ein ganz neues, großes Interesse an den klassischen wissensbasierten Methoden und Technologien erkennen, unter dem etwas sexier klingenden Begriff neuro-symbolic AI, neuro-symbolische KI, weil doch erkannt wird, dass man eigentlich beides braucht. Das, was Menschen schon wissen, lernen sie ja auch nicht dauernd immer wieder. Also auf den Schultern von Riesen heißt ja, wir entwickeln uns weiter an unseren Kompetenzen. Je mehr ich an Wissen schon nutzen kann, desto weiter kann ich kommen. Das wird zunehmend auch erkannt im Bereich der Machine-Learning-Forschung, dass man auf Tagungen wie NeurIPS auf einmal auch Leute aus der Symbolic AI als Keynote Speakers einlädt. Das wäre vor ein paar Jahren undenkbar gewesen und das finde ich ziemlich cool.

Moderator [00:57:39]

Ja interessant. Dann die gleiche Frage noch mal an Sie, Herr Hagendorff. Also großer Durchbruch, große Veränderungen oder eher overhyped?

Thilo Hagendorff [00:57:46]

In einer gewissen Hinsicht ist es gar keine große Veränderung. Es gibt seit langem eine Vielzahl an Sprachmodellen, die man auch benutzen kann für Forschungen etc., auch Sprachmodelle, die mit diesen Human Feedback trainiert worden sind, also zum Beispiel GPT-3.5. Allerdings ist das, was jetzt neu ist und was letztendlich passiert ist, dass eine Firma es gewagt hat, so ein Sprachmodell an eine einfache Benutzerschnittstelle anzuschließen. Das ist vorher noch nicht passiert. Es gab Playgrounds, wo man diese Modelle ausprobieren konnte, aber meistens musste man eben programmieren können und unter Umständen sogar sehr große Rechner, Speicherkapazitäten haben, um sie zu nutzen. Und jetzt gibt es plötzlich für alle zugänglich eine simple Benutzerschnittstelle. Und diese Benutzerschnittstelle ist es eigentlich, was diesen wahnsinnigen Hype ausgelöst hat.

Andere Firmen werden nachziehen. Auch sie werden ihre Sprachmodelle der breiten Öffentlichkeit und Laien, Amateuren zur Verfügung stellen. Und ich glaube, das kreative Potenzial, was dann freigesetzt wird, der gesellschaftliche Impact, den es haben wird, da machen wir uns überhaupt kein Bild von. Ich halte das für massiv, was dort passiert, wenngleich meine Fantasie glaube ich, noch nicht zulässt, alles zu sehen, was da in den nächsten 10, 15, 20 Jahren passieren wird. Ich gebe nur eine Sache zu bedenken: Man überlege sich nur mal, was passiert, wenn wir jetzt KI-Systeme wie diese mächtigen Sprachmodelle mit anderen KI-Systemen kombinieren. Also wenn wir so ein Sprachmodell in einen Roboter einbauen oder wenn wir sie mit Bildgenerierungs-AIs kombinieren oder Ähnliches. Da wird sich so viel verändern, nicht nur was Medien anbelangt, sondern auch was das menschliche Zusammenleben, was die Wirtschaft angeht, dass dort einiges auf uns zukommt.

Moderator [00:59:54]

Vielen Dank. Genau. Dann ist die Zeit jetzt auch vorbei. Vielen Dank erst mal an Sie. Vielen Dank an die Journalistinnen und Journalisten. Wir sind jetzt, wie gesagt, nicht durch alle Fragen gekommen. Herr Brock hat es eben schon suggeriert. Ich hoffe, für Sie ist es auch okay, wenn vielleicht noch eine Frage offen ist, dass sich wer vielleicht noch mal per Mail oder so an Sie wendet. Es gibt noch einiges zu diskutieren bei dem Thema und gerade, wenn da in Zukunft noch was passiert. Dann werden wir heute so schnell wie möglich die Aufzeichnung auf unserer Homepage online stellen. Voraussichtlich Montagmittag werden wir dann auch das Transkript online stellen. Wenn Sie eine Audio-Aufzeichnung oder eine Video-Datei auch mit Sprecheransicht brauchen oder heute schon ein maschinell erstelltes und noch nicht redigiertes Transkript, schreiben Sie gerne an redaktion@siencemediacenter.de, dann schicken wir Ihnen das zu. Vielen Dank Ihnen erst mal noch und ich wünsche Ihnen noch einen schönen Tag. Auf Wiedersehen.

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-Pressekonferenz 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

science
media center
germany