



12.10.2020

Transkript

„Der Textgenerator GPT-3 von OpenAI – Hype oder Paradigmenwechsel?“

Experten auf dem Podium

▶ **Prof. Dr. Kristian Kersting**

Leiter des Fachgebiets Maschinelles Lernen, Technische Universität Darmstadt

▶ **Prof. Dr. Hinrich Schütze**

Inhaber des Lehrstuhls für Computerlinguistik
und Direktor des Centrums für Informations- und Sprachverarbeitung, Ludwig-Maximilians-Universität München (LMU), München

▶ **PD Dr. Jessica Heesen**

Leiterin des Forschungsschwerpunkts Medienethik und Informationstechnik, Internationales Zentrum für Ethik in den Wissenschaften, Eberhard Karls Universität Tübingen

▶ **Bastian Zimmermann**

Redakteur für Technik und Digitalisierung, Science Media Center Germany,
und Moderator dieser Veranstaltung

Video-Mitschnitt

- ▶ Das aufgezeichnete Webinar [finden Sie hier](#) im Webinar-Programm selbst. Um es anzuschauen, müssen Sie eine E-Mail-Adresse eingeben.



Transkript

Moderator: (01:57)

Guten Morgen, liebe Journalistinnen und Journalisten, herzlich willkommen zu unserem virtuellen Press Briefing zu GPT-3. Mein Name ist Bastian Zimmermann. Ich bin Redakteur beim Science Media Center und freue mich ganz besonders, auch unsere drei Expertinnen und Experten hier willkommen heißen zu dürfen. GPT-3: An dem Sparmodell von OpenAI scheiden sich ja ein bisschen die Geister. Auf der einen Seite wurden die vielen Anwendungsmöglichkeiten betont, die Gefahr durch massenhafte Verbreitung von Desinformation thematisiert oder auch das Ende des Journalismus herbeizitiert, da GPT-3 ja offenbar Artikel schreiben könne. Es gab sogar Diskussionen darüber, ob man das Sprachmodell als intelligent bezeichnen könnte. Auf der anderen Seite gab es dann die Kritik, viele der angeführten Beispiele wären selektiv ausgewählt, und ohne menschliches Kuratieren wären die Texte von GPT-3 oft Unsinn. Das Programm habe eben kein tieferes Verständnis von dem, worüber es eigentlich schreibt. Das Ganze wirft natürlich viele Fragen auf, aus technischer, aber eben auch aus ethischer Perspektive. Um einige davon beantworten zu können, haben wir heute zwei Experten und eine Expertin zu dem Thema versammelt. Die einzelnen Fachleute stelle ich gleich vor. An dieser Stelle noch der Hinweis an Sie da draußen: Wenn Sie Fragen haben, können Sie die gerne gleich direkt in den Chat schreiben. Aber bitte darauf achten, dass Sie unter dem Chatfenster eingestellt haben, dass die Fragen an alle Teilnehmer versendet werden. Bitte nicht an einzelne Personen oder nur Moderatoren. Ein Kollege von mir sammelt die Fragen nämlich und leitet sie an mich weiter. Und er muss die Fragen dann natürlich im Chat für alle sehen, sonst kann ihre Frage nicht gestellt werden. Jetzt aber zu den Expertinnen und Experten. Schön, dass Sie hier alle sind. Ich stelle sie jetzt nacheinander kurz vor, und jeder bzw. jede von ihnen kann ein kurzes Eingangsstatement geben. Als erstes haben wir hier Christian Kersting. Er ist Professor und Leiter des Fachgebiets für Maschinelles Lernen an der Technischen Universität Darmstadt. Herr Kersting, wie würden Sie GPT-3 denn einordnen? Ist es jetzt der große Durchbruch, vielleicht sogar in Richtung Artificial General Intelligence? Oder ist da vieles noch übertrieben und vielleicht auch zu sehr gehypt?

Kristian Kersting: (03:59)

Ich meine, Künstliche Intelligenz, da geht es um Computerprogramme, Kochrezepte, die irgendwie intelligentes Verhalten nachbilden oder zeigen sollen. Und wenn man das aus der Perspektive betrachtet, kann GPT-3 nicht AGI, also nicht allgemeine KI oder allgemeine Intelligenz darstellen. Weil, es hat keine Arme, es hat keine Beine, es kann sich von sich aus nicht fortbewegen. Aber ich glaube, man muss das auch ein bisschen aus der anderen Perspektive betrachten. Algorithmen für die Künstliche Intelligenz aufzubauen, ist nicht ganz einfach. Das können wir versuchen, händisch zu machen. Oder wir können eben dieses Maschinelle Lernen benutzen, was nur ein Teilgebiet der Künstlichen Intelligenz ist. Und dort muss ich dann eben nur noch diesen Lernalgorithmus implementieren und dann ganz viele Beispiele bringen. Dieses Lernen mit kuratierten Beispielmengen hat tolle Durchbrüche gebracht. Wir erinnern uns vielleicht an Go, an Schach und viele andere Dinge. Und was bei GPT-3 jetzt anfängt, und das ist aus fachlicher Sicht ganz spannend, ist, dass ich nicht mehr dieses Beispiel (geben muss): Und hier muss ich jetzt das machen oder das sagen oder das vorhersagen. Also ich muss nicht mehr den Lehrer spielen, sondern ich kann sehr viel unüberwacht, so nennen wir das, also ohne diese Labels plötzlich vorangehen. Und wenn man sich aus der Perspektive anschaut, was GPT-3 kann, finde ich das erstmal beeindruckend. Aber das ist sicherlich noch keine allgemeine KI, und das ist auch noch weit weg davon. Es gibt viele Beispiele: Ich finde immer am schönsten, gerade aus Aktualität, wenn man das System fragt, sagen wir mal einfach: Wer war der Präsident der Vereinigten Staaten von Amerika im 13. Jahrhundert? Dann wird es Ihnen trotzdem eine Antwort geben, auch wenn die USA da noch nicht existiert haben. Auf der anderen Seite, und damit will ich dann auch (mein) Eingang(sstatement) schließen, möchte ich



darauf hinweisen: Ich glaube, in der KI geht es nicht nur darum, immer perfekt zu sein, sondern ich finde das auch schon toll, wenn wir durchschnittlich sind. Aber selbst davon sind wir sicherlich weit, weit weg. Aber ich finde es beeindruckend, was man damit schon hibekommen kann. Aber es ist keine allgemeine Intelligenz und weit weg davon.

Moderator: (06:12)

Okay, vielen Dank. Dann haben wir als nächstes Hinrich Schütze. Er ist Professor und Inhaber des Lehrstuhls für Computerlinguistik und Direktor des Centrums für Informations- und Sprachverarbeitung an der Ludwig-Maximilians-Universität München. Herr Schütze, Sie beschäftigen sich ja mit Natural Language Processing. Wie würden Sie denn GPT-3 generell einordnen? Und ist das jetzt die eine Applikation, die alles kann, oder haben kleinere, spezialisierte Modelle weiterhin Ihre Anwendungsfelder?

Hinrich Schütze: (06:40)

Ich kann vielleicht zunächst kurz etwas zu meinem Fachgebiet sagen, dem Natural Language Processing oder der Computerlinguistik. Die Computerlinguistik befasst sich mit der automatischen Verarbeitung von Sprache. Das heißt, wir versuchen, Systeme zu bauen, die menschliche Sprache verstehen und in menschlicher Sprache kommunizieren. Insofern ist die Computerlinguistik ein Unterbereich der Künstlichen Intelligenz. Es wird auch dieselbe Methodologie verwendet wie sonst in der Künstlichen Intelligenz. Insbesondere das Maschinelle Lernen und das Deep Learning. In der Computerlinguistik sind wir aber der Ansicht, dass es große Unterschiede gibt zwischen den Bereichen Computer Vision und Robotik auf der einen Seite und der menschlichen Sprache auf der anderen. Man braucht ein detailliertes wissenschaftliches Verständnis von Sprache, um ein sprachverstehendes System bauen zu können. Dieses Verständnis kommt aus der Linguistik, das heißt aus den Sprachwissenschaften. In der Computerlinguistik sind wir also sowohl in der Informatik als auch in der Linguistik verankert.

Und damit komme ich auch gleich zu dem wichtigsten Schwachpunkt von GPT-3. Ich würde sagen, dass GPT-3 Sprache nicht wirklich gelernt hat. GPT-3 hat gelernt, wie man sich typischerweise ausdrückt. Wenn ich friere, dann sage ich im Deutschen "Mir ist kalt". Ich sage nicht wie im Englischen "I'm cold", "Ich bin kalt". GPT-3 weiß ganz genau, welche Wörter zusammenpassen. Weil GPT-3 die typischen Ausdrucksweisen auswendig gelernt hat, sieht der von ihnen generierte Text auf den ersten Blick sehr natürlich aus, wie vom Menschen geschrieben. Aber auf den zweiten Blick merkt man, dass kein echtes Sprachverständnis da ist. Und Herr Zimmermann hat das ja eben schon angedeutet. Ein bekanntes Beispiel dafür ist die automatische Generierung von Sportreportagen. Dafür dient eine tabellarische Datenstruktur als Input der Generierung. Diese Tabelle enthält alle Daten zum Spiel, also zum Beispiel eine Zeile "Wann? 49. Minute", "Wer? Goretzka", "Was? Tor". Oder "Wann? 74. Minute?", "Wer? Süle?", "Was? Foul". Wenn wir das Spiel Deutschland - Ukraine nehmen, dann sind also die zwei Tore für Deutschland und das eine Tor der Ukraine jeweils als eine Zeile in der Tabelle hinterlegt. Aufgrund dieser Tabelle können die aktuellen Methoden im Deep Learning eine ganz tolle Reportage generieren. Die liest sich gut, verwendet den einschlägigen Sportjargon und ist vielleicht sogar witzig. Das Problem ist, dass ein Modell wie GPT-3 als Titel der Reportage dann plötzlich "Deutschland verliert knapp gegen die Ukraine" erzeugen kann – verliert und nicht gewinnt. Aus der Tabelle ist klar ersichtlich, dass die deutsche Mannschaft zwei Tore geschossen hat und die ukrainische nur eins. Aber man muss die Grundzusammenhänge des Fußballs verstehen, um daraus abzuleiten, dass Deutschland gewonnen hat. Dieses Grundverständnis von Sprache und der Welt fehlt momentan völlig in unseren Deep-Learning- Modellen.

An meinem Institut arbeiten wir an einer Alternative zu GPT-3. Eine wirklich sehr schöne Einsicht von GPT-3 ist, dass Aufgabenbeschreibungen zentral wichtig sind für das Lernen, für menschliches Lernen genauso wie für maschinelles Lernen. Wenn Sie jemandem eine neue Aufgabe beibringen



wollen, dann fangen Sie mit einer Aufgabenbeschreibung an: "Wir wollen heute diese Setzlingen hier anpflanzen. Dazu graben wir zunächst ein Loch. Dann wird der Setzling hereingesetzt" und so weiter und so weiter. Das ist eine Aufgabenbeschreibung. Wie Herr Kersting eben schon angedeutet hat, werden im Maschinellen Lernen heute hauptsächlich Beispiele verwendet. Das heißt, Aufgabenbeschreibungen werden heutzutage kaum genutzt. Beispiel in diesem Fall wäre für das Garten-Szenario, dass wir dem Algorithmus, dem Lernalgorithmus 100 Videos geben, die zeigen, wie man einen Strauch anpflanzt. Das sind Beispiele, also keine Aufgabenbeschreibung, sondern nur Beispiele. Das ist der Standardansatz im Maschinellen Lernen. Und die Grundeinsicht von GPT-3 ist, da stimme ich Herrn Kersting zu, dass die Aufgabenbeschreibung zentral ist, und die sollten wir nutzen für das Maschinelle Lernen. Das ist wirklich sehr schön. Gleichzeitig wird aber das Lernen anhand von Beispielen radikal abgelehnt im GPT-3-Ansatz. Beispiele werden zwar verwendet, aber GPT-3 lernt nichts von ihnen. Das heißt, dass Modell lernt nichts dazu. Es hat wahnsinnig viel Vorwissen, aber für eine neue Aufgabe lernt es nichts dazu. Es muss mit dem Wissen auskommen, das es am Anfang gelernt hat. An meinem Institut arbeiten wir an einem Ansatz, der diese beiden Dinge verbindet. Genauer gesagt ist das mein Doktorand Timo Schick. Die Aufgabenbeschreibung ist bei unserem Ansatz zentral wie GPT-3. Aber unser Modell kann etwas dazulernen aufgrund von Beispielen. Und wenn Sie sich überlegen, wie Menschen typischerweise lernen, dann ist es bei denen genauso. Normalerweise nutzen wir eine Aufgabenbeschreibung, aber wir lernen auch von Beispielen. Und Timo Schick hat gezeigt, dass dieser Ansatz klar besser funktioniert als GPT-3.

Moderator: (11:46)

Da vielleicht noch mal direkt eine Nachfrage dazu. Sie haben von diesem Fehler bei der Sportreportage berichtet. Wie kommt es denn jetzt genau zu diesem Fehler? Das hängt ja so ein bisschen davon ab, was GPT-3 genau lernt. Und da war auch unsere Frage, die wir da noch hatten: Lernt es im Endeffekt nur ein Pattern Matching auf Trainingsdaten? Oder kann man auch davon reden, dass es Sachen irgendwie verstanden oder gelernt hat, zum Beispiel jetzt bei Arithmetik? Da war ja auch immer dieses bekannte Beispiel, dass es irgendwie multiplizieren kann, obwohl nicht alle diese Rechnungen so in den Trainingsdaten vorhanden sind. Also es schreibt nicht einfach eine Gleichung ab, die es mal gelesen hat, sondern kann so ein bisschen extrapolieren, wie das funktioniert oder kann dann eben andere Rechnungen durchführen. Das wäre im Endeffekt die Frage, ob es da was lernt oder ob es das eigentlich so gar nicht versteht.

Hinrich Schütze: (12:34)

Es lernt auf jeden Fall etwas. Aber die Frage ist, ob man das als Verständnis bezeichnen will. Es ist halt eine relativ primitive Generalisierung, wenn man aus Multiplikationsbeispielen, die vorkommen im Text, ein bisschen weiter hinausgehen kann über das, was wirklich wortwörtlich im Text vorkommt. Multiplikation allgemein wird natürlich nicht gelernt, das behauptet niemand. Und was an diesem Reportagebeispiel besonders ist, ist, dass es eben die Verknüpfung von Sprache und Arithmetik ist. Das heißt, man muss wissen, was "gewinnen" heißt. In einem Spiel "gewinnen" heißt, dass man mehr Punkte hat als der Gegner. Das ist das Grundverständnis von Spiel und von Fußball. Und dieses Grundverständnis und die Möglichkeit, das zu verknüpfen mit dem Wissen über die Zahlen, darüber, wie viel Tore geschossen worden sind. Diese Fähigkeit hat GPT-3 nicht.

Moderator: (13:32)

Okay, vielen Dank. Und dann kommen wir jetzt auch zur letzten im Bunde: Jessica Heesen. Sie ist Privatdozentin und Leiterin des Forschungsschwerpunkts Medienethik und Informationstechnik am Internationalen Zentrum für Ethik in den Wissenschaften an der Eberhard-Karls-Universität Tübingen. Frau Heesen, Sie sind ja die Ethikerin in der Runde. Wie sehen Sie denn eine mögliche



Gefahr, die von GPT-3 ausgeht? Vielleicht auch in Bezug auf den Bias durch die Trainingsdaten? Das ist ja ein mögliches Problem, das in vielen Berichten und eben auch im Paper und bei den Aussagen von OpenAI selbst immer wieder betont wurde.

Jessica Heesen: (14:09)

Da gehe ich gleich nochmal kurz darauf ein. Da können wir auch in der Fragerunde vielleicht noch ein bisschen stärker einsteigen. Ich würde gern mit meinem Statement anfangen, und zwar in dem Bereich eines Beispiels, das ich gefunden habe in der MIT Technology Review. Das Beispiel fand ich ganz witzig, und ist einfach auch sehr erhellend. Was GPT-3 kann und was eben nicht? Ich sage Ihnen mal, wie das lautet. "Sie veranstalten ein Abendessen mit Freunden. Sie möchten das Abendessen im Wohnzimmer servieren. Der Esszimmertisch ist breiter als die Türöffnung, sodass Sie, um ihn ins Wohnzimmer zu bekommen, folgendes tun müssen." Das war die Vorgabe. Das hatten die Autoren von diesem Artikel GPT-3 also gesagt. Und dann geht es weiter. Sie müssen Folgendes tun – jetzt kommt GPT-3: "Sie müssen die Tür entfernen. Sie haben eine Tischkreissäge, also schneiden Sie die Tür in zwei Hälften und entfernen Sie die obere Hälfte." Sie erkennen unmittelbar an diesem Beispiel, dass GPT-3 hier nicht so richtig mitgedacht hat. Dass wir natürlich ein ganz anderes Verständnis von der Situation haben, so ähnlich wie das Herr Schütze auch gerade schon bei dem Fußballbeispiel thematisiert hat. Man muss sich immer vor Augen halten, dass GPT-3 letztendlich ein statistisches Sprachmodell ist, das die Wahrscheinlichkeit berechnet, dass ein Wort auf das andere folgt. Und wenn Sie jetzt hier an diese Säge denken, also "table saw" heißt das auf Englisch, (denkt GPT-3) das hat irgendwas mit Tisch zu tun. Und dann hat GPT-3 halt diese Tischkreissäge ausgesucht als passendes Instrument, um dieses Problem zu lösen.

Also es kann sprachliche Kontexte gut einordnen und vor allen Dingen, das finde ich wirklich faszinierend, Sprachfarben gut erkennen. Also ein Raptext wird erkannt und wird von einem Wikipedia-Text unterschieden und wird dann in dieser Form auch weitergeschrieben. Aber natürlich kann das System nicht selber nachdenken, und es kann vor allen Dingen nicht wie die menschliche Intelligenz lebensweltliche Kontexte verstehen oder künstlerische oder sogar soziale Kontexte, also menschliche Interaktionsbeziehungen natürlich nicht verstehen. Und es kann vor allen Dingen auch nicht verstehen, wann man sich nicht anpassen sollte. Also es ist ja so, dass ein Sprachsystem letztendlich auf Mustererkennung beruht, auf dem, was eben üblich ist, was wir vorfinden in unserer Alltagswelt. Aber es kann ja gerade wichtig sein, in Texten genau das zu entdecken und das nach vorne zu stellen, was eben nicht dem üblichen Muster entspricht. Zum Beispiel, wenn Sie einen guten journalistischen Text schreiben, der auch gerade in Opposition sein soll zu der herrschenden Meinung. Oder eben genauso im Bereich der Kunst, wo Sie immer das finden möchten, was eben anders ist als das, was vorliegt. Also sich eben nicht nach dem Muster zu verhalten, das kann natürlich so ein System nicht.

Insgesamt ist ein menschliches Nachdenken oft tatsächlich auch von Mustern geleitet, aber an vielen wichtigen Punkten eben nicht. Und vor allen Dingen betrifft das das Feld der gesellschaftlichen Wertvorstellungen. Wertvorstellungen können häufig nicht abgeleitet werden aus dem, was der Fall ist, sondern man fragt dann danach, was sollte denn der Fall sein? Das ist so die klassische Frage der Ethik: Wie wollen wir leben? Was sollte der Fall sein, und was ist nicht gut an unserer Handlungspraxis? Das kann natürlich GPT-3 alles nicht abbilden, weil es immer auf das zurückgreift, was schon der Fall ist. Aber natürlich ist klar: Es ist beeindruckend. Es gibt ja wirklich viele tolle Beispiele überall auf Twitter. Man kann selber was ausprobieren. Und es ist schon ein Sprung, auf alle Fälle. Und wenn man sich vor Augen hält, dass ja auch unser Wissen eigentlich auf dem basiert, was wir lesen, was wir sehen. Wir reproduzieren ja oft auch nur das, was vorhandenes Wissen ist. Und in der Wissenschaft, aber auch zum Beispiel, wenn Sie sich Liedsongs anhören, fragt man sich oft: Was ist denn hier eigentlich das Neue? Das sind nur ganz kleine Variablen von dem, was sowieso schon vorliegt. Und das macht eben so ein Modell auch.



Noch komplizierter wird es natürlich, wenn wir ein bisschen in die Philosophie reinschauen, (in die Theorien des philosophischen Konstruktivismus, wonach der Mensch eigentlich nur das ist, was er sprachlich artikulieren kann. Und wenn es nun so einer Künstlichen Intelligenz oder durch das Maschinelle Lernen ermöglicht wird, auf alles zurückzugreifen, was Menschen kommunizieren, dann wird es wirklich sehr eng zu unterscheiden, was menschliche Kommunikation ist und was künstliche Kommunikation ist. Aber so weit sind wir ja noch nicht.

Ich will noch kurz darauf eingehen, wie das genau ist mit den Cloud-Infrastrukturen beziehungsweise mit diesen riesigen Datensätzen, auf die GPT-3 zurückgreifen kann. Es ist ja immer wieder jetzt die Frage in diesem Zusammenhang: Geht davon eine große Gefahr aus? Können das jetzt plötzlich alle machen? Kann jetzt jeder so rummanipulieren und [unverständlich] erstellen, die er eigentlich nicht könnte oder die nicht auf dem eigenen Wissen beruhen? Da muss man sagen, es gibt so enorme Datensätze, auf die GPT-3 zugreift, dass das natürlich für normale Menschen oder auch mittelständische Unternehmen nicht zu meistern ist. Da ist so eine natürliche Hürde eingeflogen, dass nur die großen Player in dem Bereich auch tatsächlich diese Technologie nutzen können, weil sie überhaupt die Ressourcen haben, um auf diese Daten zurückgreifen zu können. Deswegen ist Microsoft jetzt ja auch ganz vorneweg dabei und hat ja auch bezahlt dafür, um GPT-3 professionell nutzen zu können. Trotzdem möchte ich aus ethischer Perspektive noch einmal sagen, dass gerade deshalb, weil natürlich riesige Datenmengen genutzt werden und auch weil OpenAI sich ja letztendlich zur Aufgabe gemacht hat, diese Anwendung zu demokratisieren, in die Gesellschaft zu bringen, es sehr sehr wichtig ist, dass wir hier Regularien finden für KI insgesamt und natürlich auch für solche Systeme.

Ich möchte das kurz mal zum Abschluss an zwei Beispielen zeigen. Das ist einmal die Trainingsdatenqualität. Da hat OpenAI schon erkannt, dass es hier natürlich sehr wichtig ist, wie die Qualität der Daten ist, dass keine Voreingenommenheit da drin ist, kein Bias. Man hat ja selber auch erprobt, wie GPT-3 funktioniert, und hat bemerkt, dass es da diskriminierende Tendenzen gibt. Also Trainingsdatenqualitäts-Überprüfung ist enorm wichtig. Da hat OpenAI auch schon eine Art Filter eingebaut. Aber dieser Filter kann letztendlich nur einzelne Begriffe finden, die man nicht haben möchte. Die können dann eliminiert werden, detektiert werden. Und man kann rückverfolgen auf die Texte, die da entstehen. Was OpenAI nicht kann, ist natürlich, die gesamten Kontexte dieser Texte zu verändern. Weil diese Kontexte sind wir ja alle. Und diese Strukturen sind eben in die Gesellschaft reinverwoben. Also diese Texte werden einen Bias haben, genauso wie wir eben alle voreingenommen sind. Trotzdem ist es aber wichtig, dass diese Trainingsdaten überprüft werden können durch Dritte, zum Beispiel durch Zertifizierungen.

Ein anderer wichtiger Punkt ist noch die Kennzeichnungspflicht. Ich bin der Auffassung, dass maschinell erstellte Texte oder auch mal schnell erstellte Software unbedingt gekennzeichnet werden müssten. Das hat verschiedene Gründe. Erst einmal so eine Art erkenntnistheoretischen Grund. Es kann passieren, dass wir letztendlich KI-Systeme haben, die sich ihre Trainingsdaten aus Texten holen, die wiederum von KI-Systemen erstellt wurden. Und dann kommen wir letztendlich in so eine Art unkontrollierte Feedbackschleife, wo wir richtig falsche Artikel produzieren können. Das ist vor allen Dingen für den wissenschaftlichen Bereich eine Katastrophe, wenn wir dann gar nicht mehr nachvollziehen können, was eigentlich richtig ist, weil alles sich nur noch in so einer Feedbackschleife befindet.

Und dann haben wir natürlich das große Problem der Vertrauensbildung. Wir sind als Menschen darauf angewiesen, zu wissen, was durch eine Künstliche Intelligenz erstellt wurde und was durch einen Menschen. Weil wir ganz viele ethische Grundannahmen in unsere Kommunikation reinpacken. Ich will das jetzt hier stoppen, will nicht zu lang reden. Aber trotzdem, das Fazit wäre letztendlich: Diese Texterstellungssysteme können tolle Sachen machen und können für bestimmte Kontexte sehr sinnvoll sein. Aber es geht vor allem darum, zu verhindern, dass sich KI-Systeme aus KI-Systemen bedienen. Und deshalb brauchen wir vor allen Dingen eben auch ein kritisches Korrektiv im Journalismus. Und wir müssen vor allen Dingen eben die Kennzeichnungspflicht voranbringen.



Kristian Kersting: (22:10)

Der Feedbackloop KI aus KI – das ist ja eine Hypothese, ob es wirklich dadurch schlimmer wird oder nicht. Es wäre mal ein schönes Experiment, das wirklich aufzusetzen, zu sehen, ob es dadurch schlimmer wird oder ob sich gar nicht viel ändert. Das ist ja noch nicht ganz klar. Und dann wollte ich nur darauf hinweisen: Wir müssen ein bisschen aufpassen mit Bias. Das ist so ein überladener Begriff. Erst mal braucht man einen gewissen Bias, um zu lernen. Dann reden wir glaub ich über den menschlichen Bias. Und dann gibt es aktuelle Arbeiten bei uns in der Gruppe, bei Marc Riedel in der Gruppe, die eben auch zeigen, dass man zumindest eine sehr einfache und sicherlich nicht abschließende Dimension erkennen kann in diesen Modellen, auch so etwas wie Wertvorstellung zu extrahieren dieser Gesellschaft. Und die können wir wiederum benutzen, um auch das Textgenerieren zu steuern, sodass man versucht, Negatives nicht zu generieren. Und dann können wir natürlich auch wieder drüber reden, dass es natürlich Situationen gibt, wo ich ja Beispiele für negatives Verhalten geben möchte, damit ich es nicht mache. Also so, wie wenn wir lernen. Aber ich möchte noch einmal ein bisschen darauf hinweisen, dass ist immer dieses Problem, wenn wir über Künstliche Intelligenz und ihre Systeme reden: Wenn wir eine Schwachstelle entdeckt haben, kann ich natürlich versuchen, die anzugehen. Also das ist so ein Moving Target. Und deswegen möchte ich immer gerne, dass wir nicht, wenn wir über Regulierung sprechen – und ich unterstelle das jetzt Frau Heesen auch nicht, weil ich weiß, dass sie da ganz anders denkt – aber wir müssen eben nur aufpassen, dass wir nicht immer den Mensch schon als Vorbild nehmen. Weil wir eben schon so weit weg sind vom Menschen in diesem System, dass ich glaube, dass das nicht zielführend wäre. Sondern da würde ich mich freuen, wenn wir ein bisschen stärker aufgeklärt an die Sache herangehen, weil wir sonst immer wieder die Ängste weiter damit unterfüttern. Das ist eigentlich alles, was wir machen sollten. Die Kennzeichnung insbesondere von produzierten Texten finde ich eine sehr schöne Sache, kann man einfach realisieren. Aber wenn wir zum Beispiel über die Zertifizierung von Algorithmen sprechen, gibt so viele praktische Probleme dabei. Da müssen wir halt auch ein bisschen stärker noch diskutieren. Aber nochmal darauf hingewiesen: Da sind schon noch Entwicklungen unterwegs, die ja auch zeigen, dass man wiederum ganz andere tolle Dinge machen kann.

Moderator: (24:37)

Ja, genau. Wenn Sie Anmerkungen zu den Äußerungen von anderen haben, können Sie auch gerne einfach immer rein springen. Das macht Sinn. Eine etwas andere Frage jetzt zu einer anderen Sache vielleicht an Sie, Herr Schütze, oder vielleicht auch an Sie, Herr Kersting: Inwiefern der GPT-3-Ansatz noch weitergeführt werden kann aus Ihrer Sicht? Man kann ja jetzt mit der gleichen Methode einen GPT-4 machen, das noch mehr Parameter hat. Und wenn ich das richtig verstanden habe, sah es ja auch in den Graphen so aus, als steigt die Leistung immer noch linear weiter mit der verwendeten Datenanzahl. Die Frage ist, ist irgendwann ein Punkt erreicht, wo sich der Rechenaufwand nicht mehr für den Ertrag lohnt? Und würde man dann irgendwann auch nicht mehr besser werden? Oder kann man sagen, man muss einfach nur noch mehr Daten drauf kloppen, und dann wird das auch immer noch besser? Und dann vielleicht noch die Frage: Gibt es überhaupt noch mehr Datenquellen, zum Beispiel Common Crawl ist ja schon einbezogen. Gibt es jetzt überhaupt noch mehr Daten? Oder kann man vielleicht auch einfach bestehende Daten besser auswerten?

Hinrich Schütze: (25:40)

Um die erste Frage zuerst zu beantworten: Also ich bin persönlich der Meinung, dass das GPT-3-Modell in seiner jetzigen Form eigentlich nicht wirklich gut eingesetzt werden kann für tatsächliche Aufgaben. Es wurde ja getestet auf diesen Standard Tasks der NLP, diesen Standard-Aufgaben. Und die sind zum Teil künstlich, und zum Teil sind sie relativ einfache Aufgaben. Da gibt es auch ein paar schwere dabei, und bei denen hat GPT-3 sehr schlecht abgeschnitten. Die praktische Nützlichkeit von GPT-3 würde ich zum jetzigen Zeitpunkt anzweifeln. Also es ist ein schönes Spielzeug.



Und es ist faszinierend, und ich finde es als philosophischen Beitrag sozusagen zum Maschinellen Lernen, als eine Vision, wie es sein sollte, sehr interessant und einen sehr wichtigen Beitrag. Aber ich weiß nicht, ob es momentan eine so große praktische Bedeutung hat. Zum zweiten Teil der Frage: Sie haben die kritischen Punkte schon angesprochen. Die Frage ist, ob es überhaupt noch mehr Text gibt. Es gibt sicher noch mehr Text. Wenn wir alle Bücher nehmen, die je geschrieben worden sind, das ist sehr viel Text. Es kommt ja aber nicht, und das hat Frau Heesen ja auch schon gesagt, auf die Menge von Text an, sondern auf die Qualität. Die Wikipedia ist zum Beispiel ein hochqualitativer Text, wenn man was lernen will. Aber die Frage wäre: Wie viel bringen da diese anderen Texte noch? Da hätte ich auch Zweifel, dass man dann die Größe dieses Textkorpus noch mal verzehnfachen oder verhundertfachen kann und tatsächlich bessere Ergebnisse bekommt.

Kristian Kersting: (27:33)

Ich kann mir vorstellen, dass man noch ganz andere Daten hinzunehmen kann, die vielleicht gar nicht aus den Büchern an sich kommen, sondern wenn ich an die Industrie – ich weiß gar nicht, auf welcher Stufe wir jetzt sind, 11.0, keine Ahnung – dass man da sicherlich auch was nehmen kann, weil ich die Wordifikation-Ansätze mir überlegen könnte. Also Ich versuche, diese Maschinendaten als Texte darzustellen. Das käme zumindest auf einen Versuch an. Ich glaube aus reinem Maschinellen Lernen – man kann ein kleines Gedankenexperiment machen. Dieses Gedankenexperiment ist ein recht altes, weil es mit Simulation einhergeht. Wenn ich den perfekten Weltsimulator habe, dann bräuchte ich unter Umständen eben auch Maschinelles Lernen nicht. Dann hab ich ja schon den Simulator. Und so ein bisschen ähnlich ist das vielleicht auch bei GPT-3 zum gewissen Grad. Wobei eben schon so Flaschenhälse eingebaut werden. Also im Maschinellen Lernen versucht man eben, um wirklich Wissen zu generieren, man muss sich das (so) vorstellen: Die Daten sind das Wasser, und man drückt das durch so einen Flaschenhals durch, und dadurch kann nur das durch, was wirklich wichtig ist. So hofft man immer, das sei so ein bisschen mit eingebaut. Aber ich würde Herrn Schütze zum gewissen Grad widersprechen. Ich glaube, klar GPT-3 an sich, alleine (kann) wahrscheinlich keine großen Anwendungen, vielleicht. Aber ich glaube, als Teilkomponente in einem etwas komplexeren Systemen bietet es sehr spannende Möglichkeiten. Und jetzt kommt es darauf an, welchen Akkuratheitslevel man verlangt oder haben möchte. Ich glaube schon, dass wir Chatbots verbessern können durch GPT-3. Vielleicht auch andere Anwendungen, so wie Sie das ja auch angesprochen haben, dass wir jetzt vielleicht Knowledge Graphs noch mehr einbauen, das geht, glaube ich, relativ einfach. Dass wir vielleicht interaktiv mit dem System, mit dem Gesamtsystem weiterlernen. Nicht unmittelbar, aber mittelbar sehe ich sehr viel Spannendes, was man mit GPT-3 als Modellklasse, als Inspiration machen kann. Das Modell selber ist relativ einfach nachzubauen, würde ich sagen. Und ich sehe es eben auch als eine gewisse Chance für Europa, dass, weil das wurde eben angesprochen, wir Europäer diese Daten mal in die Hand nehmen und einmal dieses Modell auch für Europa und für die Welt bauen und dann vielleicht eben weitergeben, sodass es nicht in der Hand von Firmen liegt. Ich glaube schon, dass man das angehen könnte. Und dann ist der Invest vielleicht niedrig, wenn wir es weiter benutzen können.

Moderator: (29:59)

Vielleicht zu dem Thema davor noch ganz kurz eine schnelle Nachfrage. Wir haben auch aus dem Chat die Frage, welche konkreten Beispiele es denn geben könnte für eine Anwendung von GPT-3 und vielleicht auch von möglichen Nachfolgern, wenn die noch besser geworden sind. Sie haben jetzt gerade gesagt, dass Sie sich da schon ein paar Anwendungen vorstellen können: Verbesserung von Chatbots. Haben Sie da noch ein paar konkrete, nur mal kurz als Stichworte reingeworfen?



Kristian Kersting: (30:24)

Also von der Modellklasse her finde ich es ganz spannend. Weil ich glaube, jetzt für Deutschland und Europa diese ganzen Maschinendaten aus der Industrie 4.0, wo es vielleicht auch nicht darum geht, ein vollautomatisches System zu haben, sondern besser zu verstehen, "Was passiert gerade, was könnte in Zukunft passieren?" – da kann ich mir das sehr gut vorstellen. Ich glaube, es wird einfacher sein, gewisse Kommunikationstools wie zum Beispiel Chatbots aufzubauen, und das in verschiedenem Kontext. Und dann wird Herr Schütze ganz zu Recht und auch Frau Heesen ganz zu Recht sagen: Ja, aber was ist, wenn dort die falsche Antwort gegeben wird? Das sind jetzt die großen Fragen. Ich möchte aber auch immer ein bisschen gegenhalten. Ich hab letztens bei einem großen Provider angerufen, und ich weiß gar nicht, wie viele falsche und schlechte Antworten ich bekommen habe. Wir sind so sehr als Ingenieurland geprägt, dass dann die Antwort von der Maschine perfekt sein muss. Da möchte ich einfach widersprechen. Vielleicht müssen wir uns einmal daran gewöhnen, dass KI, wenn wir zumindest in diese menschenähnliche Richtung gehen, eben auch wie der Mensch nicht ganz so perfekt sein muss immer, sondern im Durchschnitt brauchbare Antworten geben kann.

Jessica Heesen: (31:34)

Ich möchte nochmal ergänzen, dass im letzten Jahr Springer ein Buch herausgegeben hat, das nur über Maschinelles Lernen oder – ich weiß nicht genau, wie das System war – aber die haben ein ganz Buch rausgegeben (über) Lithium-Ionen-Batterien, das also von einer maschinellen Autorschaft geprägt ist. Die haben es auch nicht nachbearbeitet. Ein ganz anderer großer Anwendungsbereich in der Form, wie es jetzt freigegeben ist, sind natürlich studentische Arbeiten. Das finde ich sehr kritisch. Proseminararbeiten zum Beispiel in den Geisteswissenschaften, die kann man sich super schreiben lassen durch so ein System. Das wird keiner merken. Das ist wie bei DeepL, diesem Übersetzungstool, das kann man sich ausdrucken lassen. Und dann kann man nochmal drüber gehen, was da alles falsch ist. Und so ähnlich kann man sich halt in Zukunft Seminararbeiten schreiben lassen. Es ist für den gesamten Bildungsbereich die Frage: Was sind jetzt eigentlich überhaupt noch eigene Leistungen? Und angeblich können ja auch diese Plagiatsdetektionssysteme das nicht erkennen, weil es ja auch eine Neuschöpfung ist, diese Systeme, die zusammengestellt sind dann durch GPT-3.

Moderator: (32:39)

Und vielleicht noch an Sie, Herr Schütze. Sehen Sie denn auch, dass es da noch einige Anwendungen geben könnte. Vielleicht auch direkt in Bezug auf die möglichen Abschlussarbeiten? Oder würden Sie sagen, da ist so viel Nachbearbeitung nötig, dass Sie skeptisch bleiben?

Hinrich Schütze: (32:55)

Ich würde sagen, beim jetzigen GPT-3 ist so viel Nachbearbeitung notwendig, dass ich da sehr skeptisch bin. Jedenfalls dann, wenn sich der Betreuer und der Prüfer tatsächlich die Zeit nehmen, die Arbeit zu lesen. Das ist ja vielleicht nicht immer der Fall. Aber ich glaube, da merkt man dann sehr schnell, dass das hinten und vorne nicht stimmt. Außer wenn die Nachbearbeitung wirklich sehr sorgfältig war. Und dann kann man es eigentlich auch gleich neu schreiben. DeepL ist auf jeden Fall eine sehr gute Anwendung. Das würde ich als etwas völlig anderes ansehen als GPT-3. Und es gibt auch viele andere Fälle, wo in der Sprachverarbeitung Systeme gebaut werden, die Texte künstlich generieren und die das sehr gut machen, Wetterberichte zum Beispiel oder Aktienberichte. Da kann man auch streiten, da gibt es auch sehr viele ethische Fragen, aber da ist die Qualität sehr gut. Das funktioniert aber nicht so wie GPT-3, sondern ganz anders. Der Grund, warum ich bei GPT-3 sehr skeptisch bin, das ist das, was ich in meinem Statement angesprochen habe. Es lernt



nicht dazu. Und ich glaube, wenn wir wirklich gute Anwendungen wollen, dann sollten wir das zulassen, dass etwas dazugelernt wird. Und natürlich können wir (es) uns meistens nicht leisten, riesige Trainingsätze aufzubauen. Aber wenn ein Problem wichtig ist, dann kann man schon zehn oder hundert Beispiele manuell erstellen. Das ist bei einem wichtigen Problem, das man wirklich gut lösen will, eine gute Investition. Und diese 100 Beispiele sollten optimal genutzt werden. Der richtige Ansatz dafür ist das Überwachte Lernen. Und GPT-3, so sehe ich es als die zugrundeliegende Philosophie, schließt das momentan aus.

Moderator: (34:49)

Vielleicht an Sie dann direkt noch mal die Nachfrage hier aus dem Chat. Sie haben eben alternative NLP-Modelle erwähnt, die ein tieferes Verständnis von Sprache als GPT-3 vielleicht entwickeln können. Da war aus dem Chat noch mal die Bitte, ob Sie das vielleicht ein bisschen ausführen könnten, wie die funktionieren. Und kommen die dann ohne Pre-Training aus? Gerne auch konkret an einem Beispiel, wo sie gerade dran arbeiten.

Hinrich Schütze: (35:12)

Ich wollte nicht sagen, dass diese Systeme Sprache verstehen. Das ist eine ganz alte Technologie, regelbasierte Systeme. Das ist, glaube ich, das, was momentan dominiert auf dem Gebiet. Die funktioniert sehr gut, aber das ist sozusagen ein Regelsystem, das wieder eine Tabelle oft nimmt als Eingabe, zum Beispiel das Wetter. "Morgen wird es regnen." "Der Regen ist so stark." "Von da bis da scheint die Sonne" und so weiter. Und diese Daten, die als Tabelle vorliegen, die irgendwo von einer Wetteragentur kommen, die werden dann versprochen. Und das kann man mit einem Regelwerk sehr gut machen.

Kristian Kersting: (35:57)

Aber da würde ich gerne eingreifen – alles zugestimmt. Aber ich finde es eben spannend, wenn ich jetzt an die Coronakrise denke, dann muss ich unter Umständen irgendeine Entscheidung treffen unter Unsicherheit und insbesondere, ohne dass ich das Universum wirklich komplett kenne. Technisch nennt man das Open-World-Szenarien. Also ich weiß gar nicht, wie viele Entitäten wirklich gerade relevant sind. Und ich glaube, daher kommt die Faszination von dem unsupervised learning. Weil rein philosophisch ist es ja erst einmal so, dass wir alle auch nicht ständig neben uns jemanden sitzen haben, der uns sagt: richtig, falsch, falsch, richtig. Was ich jetzt hier sage – ja, vielleicht, sagt mir dann Frau Heesen heute Nachmittag in einer E-Mail "Mensch, Christian, das hättest du mal nicht sagen sollen." Aber selbst das ist nicht direkte Supervision. Da muss ich eben selber ein bisschen noch nachdenken. Dann muss man andere Techniken nehmen. Und deswegen will ich nochmal dafür so ein bisschen plädieren. Es ist, glaube ich, schon toll, dass Techniken wie GPT-3 – also mich interessiert das (einzelne) Modell dort jetzt gar nicht, sondern die Vision dahinter, das zugrundeliegende Konzept, diese sogenannten Transformer, und zu sehen, wie weit man mit einer relativ einfachen Technikidee kommen kann. Das ist das, was mich fasziniert. Und wenn man das im Kontext mit anderen Systemen weiterdenkt, glaube ich, können wir schon sehr viel weiter kommen. Eben immer unter den Randbedingungen, die Frau Heesen dann schon angesprochen hat. Das finde ich wichtig. Nehmen wir den Wetterbericht. Wenn sich dann doch plötzlich noch mal was ändert oder wenn wir dann doch leider Gottes an einem anderen Ort sind und ich das nicht realisiere, dann muss ich mit solchen Unsicherheiten irgendwie umgehen können, mit diesen neuen Situationen. Und dafür brauche ich, glaube ich, ein allgemeineres Wissen. Ich weiß nicht, ob ich das gleich Weltwissen nennen will, aber so Common Sense. Und für diesen Common Sense finde ich diese Fortschritte von GPT-3 und Konsorten extrem spannend. Aber eben nur als Modellklasse. Jetzt nicht das einzelne oder konkrete Modell.



press briefing

Hinrich Schütze: (37:57)

Nur ganz kurz dazu. Ich bin auch ein riesen Fan von unüberwachtem Lernen. Ich denke, das ist die Zukunft. Und wir arbeiten an einem Ansatz, der 90 Prozent unbewachtes Lernen ist und 10 Prozent überwachtes Lernen. Die Frage ist: Sollte man überwachtes Lernen mit dazunehmen oder nicht? Und da denke ich, dass ist für die meisten Aufgaben eine gute Idee.

Moderator: (38:20)

Jetzt haben wir hier eine Frage von außen, die mehrere Bereiche anschneidet. Ich weiß nicht genau, wer von Ihnen da zuständig ist. Sagen Sie kurz, wenn Sie was dazu denken. Und zwar die Fragen, wie man generell Urheberrecht und Datenschutz regeln kann, wenn solche Systeme sich aus offen zugänglichen Quellen bedienen. Dann: Was zukünftig als gesichertes Wissen oder seriöse Quelle gelten kann und auch wie sich Personen vielleicht davor schützen können, durch solche Systeme ausgewertet zu werden.

Jessica Heesen: (38:49)

Ich kann gern ein bisschen was dazu sagen. Also KI-Urheberrecht ist gerade ein Thema in den Rechtswissenschaften, und es gibt dafür auch einige Beispiele vor allem aus dem künstlerischen Bereich. Und es ist halt generell so, dass das Urheberrecht nur für die Eigenleistung dann letztendlich steht, also die Software, die Sie erstellen. Darauf haben Sie sozusagen das Copyright, beziehungsweise das ist das, was geschützt wird. Aber nicht sämtliche Quellen fallen darunter, genauso wie Sie ja auch in der Kunst sich bedienen können aus unterschiedlichen Quellen. Und Sie machen dann aber selber ein neues Produkt. Aber wie gesagt, das ist gerade auch in den Rechtswissenschaften ein nicht umkämpftes Feld, aber da gibt es viele neue Publikationen und auch Gutachten. Das ist noch nicht richtig ausbuchstabiert für diesen Bereich. Aber generell können Sie sich als Person natürlich schon schützen gegenüber KI in Bezug auf personenbezogene Daten. Das ist hier was anderes. Aber alles, was Sie reinstellen ins Internet, kann natürlich ausgewertet werden durch solche Systeme, und das passiert ja schon die ganze Zeit. Also bei jeder Google-Suchanfrage wird das Internet mittels KI letztendlich ja über Umwege untersucht. Also alle möglichen Inhalte. Wir haben ja jetzt schon diesen Zustand. Letztendlich können Sie sich nicht dagegen schützen. Aber ein Schutzmechanismus, den Sie in Ihrer Frage ansprechen, den hab ich ja schon genannt: Es sollte eine Kennzeichnungspflicht geben für diese maschinell erstellten Inhalte, sodass sie zumindest die Möglichkeit haben als Nutzerinnen oder Nutzer sich auszusuchen, aus welchen Quellen Sie sich bedienen möchten. Das wäre unbedingt die wichtige Sache aus meiner Perspektive.

Moderator: (40:20)

Gut. Dann eine Frage vielleicht an Sie, Herr Kersting, wie Sie diese Umstellung von OpenAI als Geschäftsmodell beurteilen. OpenAI, da ist halt nur noch der Name geblieben. Die sind jetzt in Richtung kommerzielle Anwendungen gegangen. Da gab es ja auch die Kritik, dass Forscher zum Teil schwerer Zugang zu GPT-3 bekommen würden als manche Unternehmen oder Startups. Also die Frage, was das für künftige KI-Systeme heißt und wie die Transparenz des Modells selber jetzt zu beurteilen ist.

Kristian Kersting: (40:51)

Ich würde sagen, das war absehbar. Vielleicht sollten wir OpenAI zum großen Teil jetzt ClosedAI nennen, das ist ja in Ordnung. Und es war irgendwie für mich damals, als man hörte, dass Elon Musk selber investieren will, und zwar einen größeren Betrag, absehbar. Warum sollte ein Entrepreneur



das machen, wenn er nicht auch hofft, irgendwann mal den Gegenwert wieder zu gewinnen? Finde ich das gut oder schlecht? Also ich glaube, ich habe das eben schon angedeutet. Erstmal, es ist freie Marktwirtschaft, insofern brauchen wir diese Entwicklung auch. Oder wir müssen langsam ein bisschen aufpassen, weil es geht ja nicht nur um die Algorithmen, sondern auch um die Aufklärung um die Algorithmen. Google versucht jetzt auch mehr und mehr in den universitären Bereich oder sagen wir mal Ausbildungsbereich reinzugehen. Und ich glaube, da müssen wir uns frühzeitig positionieren und vielleicht auch entsprechende Partnerschaften eingehen. Und ebenso bei den Algorithmen. Ich glaube, das wäre eine spannende Frage auch in der Regulierung, sich zu fragen, welche Algorithmen dürfen denn überhaupt patentiert werden. Also jetzt bevor alle lachen, ich weiß, dass es in gewissen Teilen der Welt Patente auf Algorithmen nicht geben kann. Aber welche sollten frei zugänglich bleiben und welche nicht, vom Typus her? Da wäre ich schon dabei. Jetzt nochmal zurück auf OpenAI. Wie gesagt, ich fand es naiv, immer zu denken, dass OpenAI wirklich open ist. Hab ich nicht verstanden. Denn wenn man sich die Papiere anschaut: Eines der ersten Papiere, wo sie GPT-3 vorgestellt haben, hat ja eben gesagt, wir können nicht alle Details angeben, weil das wäre zu gefährlich. Das fand ich als Wissenschaftler nicht gut. Dann sollte das Papier nicht angenommen werden. So ist das halt in der Wissenschaft, finde ich zumindest. Aber ich glaube, wir müssen das als Weckruf sehen und dass wir Europäer, aber auch andere Bereiche und gerne auch mit anderen Kulturen jetzt anfangen, Gegenströmungen aufzubauen, die wirklich open sind. Und das eben dann über Government, über Universitäten. Und gerade GPT-3 bietet eine Chance hier, weil es ein Einmalinvest ist. Wenn wir mal annehmen, dass die Experimente und das Lernen einigermassen vernünftig funktionieren. Davon würde ich jetzt mal ausgehen. Weil wir haben solche Experten wie Hinrich Schütze in Deutschland, in Europa. Und selber, es gibt Firmen wie Aleph Alpha, die das versuchen, auch eben gemeinschaftlich zu pushen. Und ich glaube, das sollten wir eher machen. Und wir sollten es in die Hand nehmen. Und wenn ich sehe, wofür gerade Gelder in der Coronakrise auch innerhalb der KI ausgegeben werden, dann finde ich das ein sehr viel löblicheres Ziel als so manches andere. Und ich würde mich freuen, wenn da die Politik aufwacht.

Moderator: (43:24)

Und was heißt das konkret: gemeinschaftlich pushen?

Kristian Kersting: (43:28)

Dass die die Daten, der Crawl, dass auch der Code entsprechend offen gestellt wird. Weil wir sehen ja, was passiert. Also zu sagen: Wir halten den Code zurück, weil das ist so gefährlich für die Gesellschaft. Und dann sehen wir, dass zwei Jahre später direkt Geld gemacht wird. Dann habe ich lieber den Code offen, weil es unter Umständen gleiche Rechte für alle gibt und gleiche Chancen. Aber es ist wirklich offen zu machen und damit einhergehend eben auch weiter voran die Ausbildung, das Enlightenment der Bürger. Weil ich finde – klar, ich kann alles kennzeichnen, und zum großen Teil müssen wir es machen. Aber ein Auto wird auch nicht als Auto gekennzeichnet nochmal zusätzlich. Und ich glaube, wir müssen eben frühzeitig anfangen, dass Bürgerinnen und Bürger verstehen, was Künstliche Intelligenz ist. Dass das keine Magie ist, dass das schnöde Algorithmen sind, die aber ganz spannendes Verhalten zeigen können. Und da muss man halt frühzeitig mit rangehen. Aber jetzt speziell (bei) diesem Beispiel: Ich glaube, es sollte ein europäisches GPT-X geben, auf den Bürger, Bürgerinnen zugreifen können, aber eben auch alle Firmen. Das glaube ich schon. Und das wäre schön. Die Mittel sind auf jeden Fall da, und es würde auch helfen, die europäische KI voranzutreiben, weil wir auch in die Infrastruktur ein bisschen investieren würden. Weil, ja es kostet Geld. Ja, es kostet Energie, aber eben nur einmal. Danach kann ich es pre-trained benutzen.



press briefing

Moderator: (44:53)

Herr Schütze, wie sehen Sie das generell? Halten Sie das auch für sinnvoll? Und wie leicht wäre das überhaupt machbar in Europa, GPT-X zu bauen? Wie leicht ist das quasi kopierbar, wenn der Quellcode eigentlich gar nicht öffentlich einsehbar ist?

Hinrich Schütze: (45:12)

Ich denke, da stimme ich Herrn Kersting voll zu, das wär etwas, was wirklich gut wäre für Europa. Und ich denke, das können wir machen. Ich würde sogar sagen, wir können das machen, wenn der Quellcode nicht offen ist. Die Frage ist halt, wie viel technische Innovation wirklich in GPT-3 drinsteckt. Da könnte man meinen, es ist sehr wenig. Das wirklich Innovative und Interessante und Faszinierende an GPT-3 ist dieses neue Paradigma, wie man Sprache generiert, wie man Aufgaben löst. Das ist wirklich erste Klasse. Aber die Algorithmen, das rein Technische ist sehr einfach. Das ist Standardtechnologie, das können wir in Europa auch. Ein weiteres Problem, das ich sehe, ist, man braucht eine sehr gute Infrastruktur, um GPT-3 überhaupt laufen zu lassen. Also trainieren, dafür braucht man eine noch größere Infrastruktur. Aber wenn man das einmal in Europa macht, das sehe ich wie Herr Kersting, das können wir schaffen. Aber das Modell selbst ist so groß, dass man schon allein, um es laufen zu lassen, eine ziemlich große Infrastruktur braucht. Und das ist wieder für kleine und mittelständische Unternehmen und für die meisten University labs schon eine Challenge.

Kristian Kersting: (46:36)

Aber deswegen meinte ich das wirklich als einen Appell an die Politik an sich. Weil wir könnten eben so ein KI-Cern wie es CLAIRE fordert, meinetwegen auch Alternativen, wie es ELLIS fordert, das können wir schon machen. Und es würde uns glaube ich genauso, wie wir gerade mit edudip, mit Zoom, mit anderen Techniken durch die Digitalisierung die Coronakrise sehen, eben ein Quantensprung sein, für Europa diese Infrastruktur bereitzustellen. Weil die können wir ja wiederverwenden für andere Projekte. Und nochmal: Ja, die Algorithmen sind Standard. Aber das Engineering, um diese Algorithmen auf welcher Hardware auch immer dann so hinzubekommen, dass wir GPT-3 lernen können – das sind ganz spannende Software- Engineering-Probleme. Von denen wir dann auch wieder lernen können für andere Projekte. Also, ich würde mich wirklich freuen. Ich würde es, um es jetzt mal auf den Punkt zu bringen, GPT-X mit zum Beispiel Aleph Alpha, mit Herrn Schütze, mit uns, mit Frau Heesen zehnmal spannender finden als GAIA-X. Um es auf den Punkt zu bringen.

Moderator: (47:46)

Jetzt geht die Zeit auch schon wieder schnell voran. Wir hatten noch eine Frage aus dem Chat, die ich vielleicht mal mit einer Frage von uns kombiniere. Und zwar hatten sie das ja schon mit dem Energieverbrauch angeschnitten. Und da ist jetzt die Frage: Wie ist das zu beurteilen? Man hört ja oft Kritik, dass das Anlernen von solchen Modellen einfach sehr viel Energie kostet, sehr viel Geld kostet. Ich meine, bei GPT-3 von 10 Millionen Dollar mal gelesen zu haben. Die Frage ist, wie sind solche Modelle noch haltbar in Zeiten, wo Klimaschutz immer wichtiger wird. Und da gab es im Chat auch die Frage, ob es als Regulierungsvorschlag vielleicht sinnvoll wäre, Firmen, die GPT-3 kommerziellen nutzen, anhand der verarbeiteten Datenmenge zu besteuern und so einen Anreiz zur Datensparsamkeit zu setzen.



Kristian Kersting: (48:41)

(Zum Energieaufwand.) Es ist ein Einmalaufwand mit einem aktuellen Geschäftsmodell. Und dann wird hoffentlich dieses trainierte Modell wieder und wieder verwendet, sodass dann dort das Vor-Ort Trainieren im Verhältnis kostengünstig ist. Jetzt können wir uns fragen, ob dieser Einmalinvest notwendig ist. Und wenn man das wissenschaftlich betrachtet: Wir nehmen jetzt mal Frau Heesen. Es ist klar, dass Frau Heesen nicht dieselbe Datenmenge wie GPT-3 brauchte, um mit uns so zu diskutieren, wie wir es heute tun. Das heißt, es muss eine Möglichkeit geben, das dateneffizient zu betreiben. Und das ist halt die Frage, die in der Forschung gestellt wird. Herr Schütze hat das schon angesprochen, dass er sich im Auftrag (sieht), wie können wir regelbasierte und solche eher unüberwachten Deep-Learning Verfahren kombinieren? Das ist eine ganz große Frage, die man manchmal gerne unter der dritten Welle der KI auch ein bisschen zusammenfasst. Also wie man das beides zusammenbekommt. Wissenschaftlich können wir irgendwann mal hoffentlich die Einsicht bekommen, wie man das Ganze energieeffizienter gestalten kann. Es ist aber erst einmal aktuell ein Einmalinvest. Und dann möchte ich immer darauf hinweisen: Ich finde es nicht so gut, wenn immer punktuell einzelne Dinge rausgenommen werden. Genauso gut könnte ich dann den Elektronenbeschleuniger am CERN herausholen und mal fragen: Was haben wir denn davon, den Big Bang zu verstehen? Wenn man mich fragt, haben wir ganz viel dadurch. Ich finde das eine total spannende und wichtige Frage. Aber wenn ich jetzt plötzlich wirtschaftlich und klimatechnisch rangehe, müsste ich dann auch diese Frage unter Umständen nicht mehr verfolgen. Weil ich immer noch glaube, dass der Elektronenbeschleuniger sehr viel mehr Energie verbraucht als das Lernen von GPT-3. Und deswegen finde ich einfach nur: alles d'accord. Man sollte die Frage stellen: Muss man jedes dieser Modelle trainieren? Ja, das ist eine gute Frage. Aber es gibt wissenschaftlich eben die Hoffnung, dass man das mit weniger Aufwand machen kann. Und es ist im Verhältnis gar nicht so viel Aufwand, wenn die trainierten Modelle wiederverwendet werden.

Moderator: (50:45)

OK, gut. Dann ist die Zeit auch fast vorbei. Vielleicht noch als Schlussfrage an Sie, Frau Heesen. Die beiden anderen haben sich schon dazu geäußert. Jetzt die Frage, wie es jetzt aussieht, dass die Techgiganten Amazon, Google, Microsoft und OpenAI eben auch durch Zugang zu viel mehr Daten und Computing Power quasi ein Monopol haben, was solche Forschung angeht. Wir haben ja eben schon mögliche europäische Alternativen besprochen. Inwiefern halten Sie das für ein Problem? Und können da andere Forschungsinstitute dann überhaupt noch mithalten? Oder muss man das vielleicht auch regulieren?

Jessica Heesen: (51:15)

Ja, das ist ein Riesenproblem. Kristian Kersting hat das ja auch grad schon genannt, dass wir mehr Pluralität brauchen, mehr Produkte letztendlich, die aus Europa kommen. Und dazu soll es ja auch einen sogenannten dritten Weg geben. Es wird ja in den Papieren der EU-Kommission und auch der Bundesregierung immer wieder gesagt, dass wir eben auch vorgehen wollen und tatsächlich auch Werte eine Rolle spielen sollen bei der Künstlichen-Intelligenz-Entwicklung insgesamt. Pluralität ist aus meiner Perspektive tatsächlich ein Zauberwort. Dass wir versuchen, einfach auch Mittelstand zu stärken oder auch (für) Startups und so weiter entsprechende Schnittstellen aufzumachen, sodass eben diese Dinge, die entwickelt wurden, teilweise jetzt auch in den Forschungsabteilungen der Großen, eben auch für die mittlere oder untere Ebene, jetzt mal nicht hierarchisch gedacht, sondern wirklich vom Vermögen her offen gehalten wird.



press briefing

Kristian Kersting: (52:01)

Ich möchte noch mal darauf hinweisen, dass es ja wirklich auch so ist, dass durch diese Pluralität wir bei dem Trainieren bei GPT-3 nicht gleich über IPR reden müssen, also Intellectual Property Rights. Im Zweifelsfalle muss ich dieses System dann auf intelligente Art und Weise anwenden, also einspannen für eine Aufgabe. Und darauf kann ich dann plötzlich mein IPR generieren. Also auch für die Firmen ist ein originäres Interesse da, das Ganze offenzuhalten.

Moderator: (52:31)

Gut, dann ist die Zeit für das Press Briefing jetzt auch schon um. Ich glaube, das ist ein Thema, wo es so viele Fragen dazu gibt. Wir haben jetzt vielleicht die Hälfte beantwortet von dem, was nur wir uns hier aufgeschrieben haben. Es ist ein großes Thema. Aber erst mal vielen Dank an die Kolleginnen und Kollegen draußen, die dabei waren. Und natürlich vielen Dank vor allem auch an Sie, Frau Heesen, Herr Schütze und Herr Kersting für Ihre Zeit. Und in ungefähr einer Stunde können Sie die Aufzeichnung des Press Briefings über einen Link auf unserer Homepage abrufen, das jetzt eher an die Journalistinnen und Journalisten. Und morgen werden wir dort auch ein Transkript veröffentlichen. Dann hoffe ich noch, dass es für Sie alle so informativ war wie für mich und wünsche Ihnen allen noch einen schönen Tag. Auf Wiedersehen!



press briefing

Ansprechpartner in der Redaktion

Bastian Zimmermann

Redakteur für Technik und Digitalisierung

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 RStV/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

Vertretungsberechtigte Geschäftsführer

Beate Spiegel, Volker Stollorz

Verantwortlich für das redaktionelle Angebot (Webmaster) im Sinne des §55 Abs.2 RStV

Volker Stollorz

