

Interview mit dem KI-Experten Pieter Buteneers

„ChatGPT ist der Wahnsinn“

Derzeit wird viel über ChatGPT von OpenAI diskutiert, und es ist durchaus beliebt, die KI zu verunglimpfen und das Ende der Menschheit heraufzubeschwören. Doch was kann ChatGPT wirklich? Und was kann es nicht? Bringt uns ChatGPT der künstlichen allgemeinen Intelligenz (AGI) näher, oder ist der ganze Hype ungerechtfertigt? Wir sprachen mit dem KI-Experten Pieter Buteneers über ChatGPT, Sprachmodelle, das Ende von Google und die Probleme mit OpenAI.

Entwickler Magazin Spezial: Vielen Dank, dass du dir die Zeit für das Interview genommen hast. Kannst du dich unseren Leser:innen kurz vorstellen?

Pieter Buteneers: Ich war früher der Director of Engineering für ML und KI bei Sinch. Kürzlich wurde ich zum Direktor von Sinch Labs befördert, dem Team, das für die Einführung neuer Out-of-the-Box-Produkte verantwortlich ist. Davor war ich CTO von Chatlayer, einem kleinen Unternehmen, das eine Plattform für technisch nicht versierte Personen zur Erstellung von Chat- und Voice-Bots anbietet. Nachdem Sinch das Unternehmen übernommen hatte, entwickelte ich AskFrank, eine Lösung, die die Erstellung von Chatbots erleichtert. AskFrank kann Inhalte in einer Wissensdatenbank durchsuchen und dort eine Antwort auf eine Frage finden. Aktuell arbeiten wir noch daran, AskFrank in die Lage zu versetzen, eine Unterhaltung fortzusetzen, so dass man Folgefragen stellen kann.

EMS: Was ist ChatGPT und wie funktioniert es?

Buteneers: ChatGPT basiert auf dem GPT-3.5-Modell, das GPT-3 sehr ähnlich ist. Es besteht aus 175 Milliar-

den Zahlen, die alle Informationen enthalten, nach denen man es fragen kann. Das ist im Grunde genommen der Großteil der Daten, die im Internet verfügbar sind. Es ist erstaunlich, wie wenig Speicherplatz für so viele Informationen benötigt wird.

GPT-3 kann jedoch nur Sätze vervollständigen. Wenn du der KI ein Stichwort gibst, fährt sie fort, je nachdem, was das Stichwort ist. Bei OpenAI dachte man sich, dass es interessant sein könnte, dieses Modell in ein Gesprächsmodell zu verwandeln, das Fragen beantworten kann, und so wurde ChatGPT entwickelt. Sie haben GPT-3 so optimiert, dass es Antworten geben kann, anstatt einfach nur Sätze zu vervollständigen.

Zur Feinabstimmung von ChatGPT ließ man eine Person ein Gespräch beginnen und beobachtete dann, was GPT-3 als Antwort lieferte. Da das Modell nicht völlig deterministisch ist, konnte man sich für eine Frage verschiedene Antworten liefern lassen. Man ließ das Modell vier verschiedene Antworten generieren, die dann durch eine Gruppe menschlicher Experten bewertet wurden. Das hat man für riesige Datenmengen gemacht und dann ein weiteres Modell darauf trainiert, die Bewertung der menschlichen Experten vorherzusagen bzw. zu imitieren, um ChatGPT auf der Grundlage dieses Bewertungsmodells automatisch feinabzustimmen.

Anschließend ließen sie die KI imaginäre Unterhaltungen über Millionen von Themen führen und nutzten die Bewertungen, um sie weiter zu verfeinern und immer bessere Ergebnisse zu erzielen. So ist ChatGPT entstanden. Es ist im Grunde eine Zusammenfassung des Internets bis Ende 2021.



Dr. Pieter Buteneers ist der ehemalige CTO von Chatlayer.ai, das 2020 von Sinch übernommen wurde. Jetzt leitet er alle Machine-Learning- und KI-Aktivitäten bei Sinch als Director of Engineering in ML&AI. Inzwischen ist er Director of Sinch Labs.

EMS: Das Training erfolgt mit Reinforcement Learning auf der Grundlage von Benutzerfeedback, aber gibt es verschiedene Ebenen zwischen dem Modell und dem tatsächlichen menschlichen Feedback? Sind hier also nur zwei Maschinen am Werk, die sich gegenseitig trainieren?

Buteneers: Ja! Man hat sich für diesen Ansatz entschieden, weil es viel einfacher ist zu verstehen, wie Menschen etwas bewerten würden, als zu verstehen, warum. ChatGPT musste das Warum herausfinden, um sich zu verbessern. Dazu eignet sich Reinforcement Learning, weil man damit einfach sagen kann: „Das ist gut“ oder „Das ist schlecht“. Man sagt dem System nicht, warum, das muss es selbst herausfinden.

Allerdings ist Reinforcement Learning sehr langsam, weil es wahnsinnig viele Daten benötigt. Da es aber so einfach ist, herauszufinden, wie Menschen die Antworten einstufen würden, ist es schneller, einfach ein Modell zu trainieren, das das kann, und es auf einer großen Menge von Daten laufen zu lassen.

EMS: Ist ChatGPT so beeindruckend, wie es behauptet wird?

Buteneers: Ich habe es schon oft gesagt, und ich sage es noch einmal: Es ist der Wahnsinn. Wenn du mich vor zwei Jahren gefragt hättest, ob sowas innerhalb der nächsten 10 Jahre möglich ist, hätte ich gesagt, dass es sehr unwahrscheinlich ist. So beeindruckend ist das.

Die vorhandene Menge an Informationen und das Verständnis für verschiedene Zusammenhänge ist unvorstellbar. Ich benutze die KI lieber als Stack Overflow und manchmal sogar lieber als die Google-Suche, weil sie so unglaublich viel Wissen über so viele verschiedene Themen bietet. Und die Antworten, die es gibt, auch wenn sie manchmal falsch oder völlig erfunden sind, sind grammatikalisch solide und überzeugend. Ich merke erst dann, ob ich hinters Licht geführt werde, wenn ich die Antwort bereits kenne. So gut ist es.

Das macht es aber auch so schwierig, der KI zu vertrauen, weil man nicht einschätzen kann, wann sie einfach falsch liegt. ChatGPTs Antworten sind grammatikalisch korrekt, und meistens inhaltlich richtig. Und wir sprechen hier nur über eine Betaversion, das ist noch nicht die offizielle Version.

Als ich ChatGPT ausprobierte, schien in der KI etwas versteckt zu sein, was auch das Internet durchsuchen kann, wodurch man Informationen aus der Zeit nach 2021 erhalten kann. Ich weiß nicht, ob das etwas Seltsames ist, was das Modell erzeugt hat, um uns zu täuschen, oder ob das echt ist. Aber es sieht so aus, als ob man bei OpenAI daran arbeiten würde, ChatGPT mit dem Internet zu verbinden. Und dann wird Google radikal ausgelöscht werden. Das Einzige, was ich dazu sagen kann, ist: Das ist der Wahnsinn.

EMS: Wenn du so etwas wie ChatGPT nicht vorhersagen konntest, was ist dann passiert, dass die Entwicklung ein solches Modell ermöglicht hat, also in Bezug auf die Technologie?

Buteneers: Das Ärgerlichste an der Sache ist, dass es gar nicht so viel ist. GPT-3 ist seit etwa zwei Jahren auf dem Markt und war damals schon sehr beeindruckend. Man konnte ihm einige Stichwörter geben, und es produzierte einen Text, der logisch ziemlich konsistent war. Allerdings waren die Antworten, die man von ihm erhalten hat, meistens falsch.

Man nahm dieses Modell, fütterte es mit neuen Daten, änderte möglicherweise ein paar Parameter und trainierte es mit menschlichem Feedback, und es wurde zu dieser wahnsinnig guten, allwissenden Person, mit der man sprechen kann. Und wenn man mit der Antwort nicht zufrieden ist, kann man genau wie bei einem echten Menschen nachfragen. Es ist absurd, wie wenig sich geändert hat, um das zu erreichen. Es brauchte nur ein wenig Feinabstimmung an der richtigen Stelle und eine Menge Rechenleistung, die wahnsinnig viel Geld gekostet haben muss. Davon abgesehen hat sich technisch nicht viel getan.

EMS: Aber ist dieses Modell nicht immer noch probabilistisch? Es „weiß“ nichts. Es kennt nur die wahrscheinlichste Antwort, entsprechend den Daten, mit denen es trainiert wurde.

Buteneers: Nun, das Gleiche gilt für dich: Weißt du etwas über ein bestimmtes Thema oder näherst du dich nur der Wirklichkeit an, aufgrund von Informationen, die in deinem Kopf gespeichert sind? Das ist genau dasselbe. Weiß ChatGPT wirklich etwas? Ja und nein. Wenn du der Meinung bist, dass du die Dinge nicht wirklich weißt, gilt das auch für ChatGPT. Aber wenn du ganz ehrlich zu dir selbst bist, weißt du, dass du Zusammenfassungen der Realität machst. Und du kannst die Realität in abstrakten Konzepten auf sehr hohem Niveau zusammenfassen. ChatGPT macht genau dasselbe. Aber ist das wichtig? Es erhält die Illusion aufrecht, dass es „weiß“, und darin ist es gut.

Viele Leute haben versucht, ChatGPT zum Programmieren oder zum Schreiben von Gedichten zu benutzen. Meine Frau aber wollte vor Kurzem wissen, wie man Joghurt macht. Also fragte ich ChatGPT: „Wie stellt man seinen eigenen Joghurt her?“ Es gab uns ein Rezept, aber es war unklar, welche Utensilien man dafür benötigt. Also habe ich gefragt: „Welche Utensilien braucht man dafür?“ Und es gab mir eine Liste von Utensilien. Davon fehlte mir aber ein zuverlässiges Thermometer. Meine nächste Frage war also: „Was ist, wenn ich kein Thermometer habe?“ Es kam ein weiteres Rezept, bei dem man Joghurt ohne Thermometer herstellen kann. Man muss sehr lange googlen, um dieselbe Antwort zu finden.

Natürlich habe ich auch versucht, ChatGPT zum Coden zu verwenden. Wenn ich mehr über eine neue Technologie oder ein neues Tool erfahren wollte, stellte ich Fragen, und ChatGPT antwortete. Wenn es eine Terminologie verwendete, die mir nicht gefiel, fragte ich noch einmal nach, und das Programm erklärte sie mir, lieferte Beispiele und so weiter. Wie lange würdest du brauchen, um die gleichen Informationen in einem Google Rabbithole zu finden? Und das ist es, was an

ChatGPT so beeindruckend ist. Es ist in der Lage, eine einigermaßen intelligente Unterhaltung zu führen.

EMS: Als wir vor einiger Zeit über Dall-E und Stable Diffusion sprachen, fragte ich, ob sie das Ende für Designer, Künstler und andere Kreative seien. Du hast damals gesagt, dass die Modelle lediglich Werkzeuge sind. Gilt das auch für ChatGPT und Schriftsteller oder Texter?

Buteneers: In seinem jetzigen Zustand kann dir ChatGPT zweifellos helfen, deine Arbeit viel schneller zu erledigen. Du kannst es zum Beispiel bitten, eine Einleitung zu diesem Interview zu schreiben. Mit den richtigen Fragen funktioniert das, dann bestimmte Abschnitte umschreiben oder bestimmte Dinge ganz streichen. Man kann das so lange machen, bis man eine schöne Einleitung hat.

Das kann ein nützliches Werkzeug sein, das dir bei solchen Dingen helfen kann. Aber es ist noch nicht so weit, dass es menschliche Kreativität ersetzen kann. Es kann sich keine völlig neuen Konzepte ausdenken. Es kann das, was es gelernt hat, auf ziemlich ähnliche Aufgaben anwenden, aber es kann keine völlig neuen Dinge hervorbringen. Es kann aber schon sein, dass einige Menschen ihre Fähigkeiten verbessern müssen.

Das ist aber immer der Fall, wenn eine neue Technologie auf den Markt kommt. Sie ist zwar noch nicht so weit, dass jeder Texter ersetzt werden kann, aber sie wird die Effizienz der Texter erhöhen und die Texte billiger machen. Mit dieser Art von Technologie braucht man vielleicht weniger Texter, oder du hast am Ende die gleiche Anzahl von Leuten, die viel mehr Arbeit erledigen können. Diejenigen, die diese Technologie zu ihrem Vorteil nutzen können, werden davon stark profitieren.

EMS: Gibt es Dinge, die ChatGPT definitiv nicht kann?

Buteneers: Nun, die Antwort ist immer ein „noch nicht“. Was es derzeit nicht kann, ist, Informationen im Internet oder in einer Wissensdatenbank nachzuschlagen, sodass man Antworten erhält, die aktuell sind. Es gibt Hinweise darauf, dass, wenn man die richtige Frage stellt, irgendwann etwas in ChatGPT aktiviert werden könnte, das die Suche im Internet ermöglicht. Das ist höchstwahrscheinlich der nächste Schritt für OpenAI. Wenn ChatGPT das Internet durchsuchen kann, werden Google und Stack Overflow Geschichte sein. Es macht Sinn, davon auszugehen, dass das passieren wird. Außerdem gibt es Modelle wie AskFrank, die bereits eine gemischte Suche durchführen können. Wir wissen also, dass das technisch möglich ist.

Was noch fehlt, ist die Fähigkeit, dir zu sagen, ob es dir die Wahrheit sagt oder dich nur täuscht, was ziemlich wichtig wäre. Vielleicht gibt es Möglichkeiten, es darauf zu trainieren, herauszufinden, wie sicher es sich mit etwas ist. Dafür ist aber vermutlich viel händische Arbeit nötig, besonders beim Labeling.

Was ChatGPT ebenfalls nicht kann, ist eine Verbindung zu einem anderen Dienst oder einem API herzustellen. Man kann fragen, „wie man X macht“, aber man kann die KI nicht bitten, X tatsächlich zu tun. Du

kannst ChatGPT darum bitten, einen Codeschnipsel zu schreiben, der X erledigt, aber du kannst nicht verlangen, dass es eine Verbindung zu einem API herstellt und Daten in einer Datenbank für dich speichert.

Es hat keine Verbindung zur Außenwelt, und ich glaube nicht, dass das Senden von Daten an ein API in absehbarer Zeit möglich sein wird. Es gibt jedoch Anzeichen dafür, dass Websuchwerkzeuge entwickelt werden, was an und für sich schon eine großartige Funktion wäre.

EMS: Und dann wird ChatGPT Google töten?

Buteneers: Es wird nicht nur Google töten, sondern auch Stack Overflow. Man sollte die KI als persönlichen Assistenten betrachten, der in der Lage sein wird, für dich eine Google-Suche durchzuführen und nur relevante Informationen anzuzeigen.

EMS: Wohin geht die Reise mit KI allgemein?

Buteneers: Forschende untersuchen wahrscheinlich gerade so ziemlich alles, von dem wir bisher ausgegangen sind, dass nur Menschen dazu in der Lage wären. Mit der Einführung von ChatGPT haben wir gesehen, dass KI das Problem der Sprache geknackt hat. Hätte jemand vor 50 Jahren, ohne Wissen von ChatGPT, den Turing-Test gemacht, wäre er oder sie zweifelsfrei zu dem Schluss gekommen: „Das ist ein Mensch.“ Alles, was mit Sprache zu tun hat und von ChatGPT noch nicht geknackt wurde, wird in Zukunft gelöst werden.

Stable Diffusion ist im Grunde die Verbindung von Sprachen und Bildern. Das war ein erster Schritt, und jetzt kann man aus einem Stück Text ein Bild erzeugen. Natürlich kann man beides auch kombinieren, und diese Mischung aus Bilderkennung und Sprache ist jetzt sozusagen völlig offen, und jeder kann sich daran versuchen, diesen Code zu knacken.

Das wirklich Interessante daran ist, dass wir uns häufig der Sprache bedienen, um unser Denken zu vermitteln. Und da wir die Sprache geknackt haben, ist unser gesamtes sprachliches Denken zweifelsohne in der Reichweite von KI. Es ist beängstigend, wie nahe wir einer AGI sind. Ich würde zwar nicht sagen, dass wir extrem nah dran sind, aber wir bewegen uns in diese Richtung.

EMS: Worauf freust du dich im Jahr 2023?

Buteneers: Ich war kein großer Fan von GPT-3. Was man mit ein wenig Input erreichen konnte, war zwar beeindruckend, aber nachdem ich gesehen habe, wie man GPT-3 in ChatGPT verwandelt hat, begann ich mich zu fragen, wie GPT-4 wohl sein wird. Ich dachte bisher, es würde vielleicht GPT-3 plus ein bisschen mehr sein. Aber jetzt bin ich überzeugt, dass es weit darüber hinausgeht und viel ausgefeilter sein wird. Es ist für mich immer noch schwierig zu fassen, was das sein könnte. Was kann da noch mehr kommen? Das weiß ich nicht, aber ich freue mich darauf, es herauszufinden.

Bei Google gibt es wahrscheinlich immer noch eine Menge Kopfzerbrechen darüber, was man mit ChatGPT machen soll. Ich glaube nicht, dass man es dort kommen

sah, auch wenn sie selbst an etwas Ähnlichem gearbeitet haben. Vor einigen Jahren hatte man auf der Google I/O eine ziemlich beeindruckende Demo mit einem Sprachbot, der über Pluto sprach. Aber an so etwas wie ChatGPT kam das nicht heran. Was sie damals hatten, war weitaus weniger fortschrittlich.

Ich weiß von meinen Kontakten innerhalb von Google Deep Mind, dass es eine kleine Krise im dortigen Team gab, als GPT-3 veröffentlicht wurde. Sie versuchten verzweifelt, so schnell wie möglich eine Lösung zu finden. Und das war nur GPT-3. Aber bei ChatGPT glaube ich, dass das gesamte Unternehmen zittert, nicht nur Google Deep Mind, sondern ganz Google oder zumindest diejenigen, die verstehen, was ChatGPT leisten kann.

Bei Google weiß man schon seit einiger Zeit über GPT-3 Bescheid. Sie haben an dieser Art Technologie gearbeitet, aber sie haben sie nicht erfunden. OpenAI hat sich das ausgedacht und bei Google wird man hart arbeiten müssen, um eine Antwort darauf zu finden. Sie könnten etwas Vergleichbares schaffen, aber ich bezweifle, dass sie OpenAI übertreffen werden.

EMS: Ist die Aufregung um ChatGPT also gerechtfertigt?

Buteneers: ChatGPT ist meiner Meinung nach das, was einer AGI am nächsten kommt. Das ist das Einzige, was ich wirklich betonen möchte. Viele Menschen auf der Welt können das, was ChatGPT kann, nicht. Natürlich haben sie dafür soziale Fähigkeiten und ein besseres Verständnis für Emotionen. Aber was diese KI kann, ist der absolut Wahnsinn. Sie kann dir die überzeugendste Lüge erzählen, mit Argumenten, die dich verblüffen werden. Wir sollten nicht unterschätzen, wie wichtig dieser Durchbruch auf dem Weg zur AGI ist.

Natürlich sind wir noch weit von der AGI entfernt, aber was ich bei ChatGPT gesehen habe, hat meine Vorhersage, wann wir sie erreichen, um 20 Jahre nähergebracht. Ich hatte immer auf 2060 getippt, während einige Forschende 2030 schätzen. Nach dem, was ich gesehen habe, glaube ich, dass 2040 wahrscheinlicher ist, was ein beunruhigender Gedanke ist.

EMS: Was ist die eine Komponente von AGI, die noch fehlt? Was fehlt, damit ChatGPT zur AGI wird?

Buteneers: Das Denken außerhalb eines bestimmten Kontexts und die Entwicklung neuer Konzepte oder Ideen. Seltsamerweise sind dazu auch nicht alle Menschen in der Lage, auch wenn jeder in der Lage ist, es in einem kleinen Kontext zu tun, zumindest allgemein gesprochen.

Was wir Menschen gut können, ist Emotionen in einem Gesichtsausdruck zu lesen oder in der Art, wie Menschen sprechen. Das sind Dinge, die ChatGPT nicht kann. Wenn man jedoch bedenkt, wie weit die Bilderkennung fortgeschritten ist, scheint es klar, dass das Lesen und Verstehen von Gesichtsausdrücken und die Bereitstellung angemessener Antworten nicht mehr weit entfernt sind.

Auch bei den Sprachmodellen haben wir enorme Fortschritte gemacht, sodass es immer weniger Dinge gibt, die diese Systeme nicht können. Schau dir Open-

Als Sprach-zu-Text-KI Whisper an, die mehr leistet als die Google Cloud APIs, Microsoft und AWS. Außerdem ist es quelloffen, sodass man es auf seinem eigenen Computer laufen lassen kann und dann über eine Sprach-zu-Text-Engine verfügt, die die Leistung der großen Anbieter übertrifft. Was das Science-Fiction-Konzept der persönlichen Roboterassistenten angeht, sind wir in der Robotik weit zurück. Aber an allem anderen, besonders auf der Ebene der Sprache, sind wir ziemlich nah dran.

EMS: Gibt es viele KI-Modelle, die quelloffen sind?

Buteneers: Gut, dass du es ansprichst, denn ChatGPT ist nicht quelloffen. Das ist frustrierend. OpenAI begann vor vielen Jahren mit der Idee, KI zu demokratisieren und zu öffnen, indem alle Forschungsergebnisse und Modelle veröffentlicht wurden. Das Ziel sollte sein, dass niemand in der Lage sein würde, die anderen auszusteichen, weil sich alle mit der gleichen Geschwindigkeit in Richtung AGI bewegen würden. Und was ist der Stand jetzt? Keiner bewegt sich so schnell in Richtung AGI wie OpenAI.

Natürlich bemüht sich die Open-Source-Gemeinschaft sehr darum, aufzuholen, und sie macht ihre Sache sehr gut. Außerdem investieren viele Leute Millionen von Dollar in diese Forschung, und ich weiß das sehr zu schätzen. Die Entscheidung von OpenAI, alles, was sie tun, als Closed Source zu behandeln und nicht zu demokratisieren, wie sie es versprochen hatten, frustriert mich jedoch.

Ich kann verstehen, dass OpenAI ihre Modelle in einer gewissen Weise feinabgestimmt hat, damit sie z. B. weniger frauenfeindlich sind, und ich verstehe, warum sie es schwer machen, ChatGPT bspw. zu fragen, wie man eine Bombe baut. Aber damit geben sie auch ein moralisches Urteil ab, ein Urteil darüber, was richtig und was falsch ist. Und sie repräsentieren nicht die gesamte Weltbevölkerung. Sie sind nur eine kleine Gruppe, die für uns entscheidet, was richtig und was falsch ist. Und sie tun das bei moralischen Fragen und auch beim Code selbst.

Von uns wird erwartet, dass wir in ihren Fähigkeiten vertrauen, moralisch einwandfreie Entscheidungen zu treffen. Wenn mich also irgendetwas an OpenAI frustriert, dann ist es, dass es den falschen Namen trägt. Es handelt sich zwar um AI, aber sie ist nicht offen. Sie veröffentlichen Papers und Blogbeiträge und geben uns Hinweise darauf, wie ihre Modelle funktionieren. Aber sie sind weit davon entfernt, offen darüber zu sprechen.

Zum Glück hat sich mit Stable Diffusion schnell eine Alternative zu DALL-E entwickelt. Die Stable-Diffusion-Community arbeitet eifrig an der Entwicklung neuer Funktionen, um die Nutzung zu vereinfachen und interessantere Aufgaben damit zu lösen. Aber es ist klar, dass OpenAI diese Technologie nicht wirklich demokratisiert. Sie behaupten, Angst davor zu haben, dass sie den falschen Leuten in die Hände fällt und sie missbraucht wird. Wir wissen es nicht, aber es ist durchaus möglich, dass sie auch bei OpenAI in den falschen Händen ist.

Die Fragen stellte Alexander Goschin.