Prüfungsfragen erstellen mit ChatGPT

Bericht: Universitätsklinikum Bonn

KI-Modelle wie ChatGTP sind im Kommen und mittlerweile eine wichtige Stütze in vielen Bereichen – von Aufsätzen in Uni und Schule bis hin zum Schreiben von Programmcodes. Nun haben Forschende am Universitätsklinikum Bonn (UKB) Prüfungsfragen für Medizinstudierende mithilfe von ChatGPT entwickelt und diese dann in einer Studie mit den von Dozierenden erstellten Prüfungsfragen verglichen.

Dabei konnten die Studierenden nur 57 Prozent der Fragequellen richtig identifizieren. Den Forschenden zufolge führe der Einsatz von ChatGPT bei der Erstellung von Prüfungsfragen zu einer signifikanten Zeitersparnis. Darüber hinaus könnten die Studierenden sich gegenseitig mithilfe von ChatGPT-generierten Fragen testen. Die Daten wurden in der weltweit führenden Zeitschrift für Medizindidaktik veröffentlicht.

Für die Studie haben die Forschenden des UKB zwei Sätze mit 25 Multiple-Choice-Fragen (MCQs) mit jeweils fünf Antwortmöglichkeiten erstellt, von denen eine richtig war. Der erste Fragensatz wurde von einem erfahrenen medizinischen Dozenten verfasst, der zweite Satz wurde von ChatGPT erstellt. 161 Studierende beantworteten alle Fragen in zufälliger Reihenfolge. Bei jeder Frage gaben die Studierenden auch an, ob diese ihrer Meinung nach von einem Menschen oder von ChatGPT erstellt wurde.

Matthias Laupichler, einer der Studienautoren und Wissenschaftlicher Mitarbeiter am Institut für Medizindidaktik am UKB, erläutert: „Es hat uns erstaunt, dass die Schwierigkeit von menschenerstellten und ChatGPT-generierten Fragen quasi identisch war. Noch überraschender aber war für uns, dass die Studierenden den Fragenursprung in fast der Hälfte der Fälle nicht richtig identifizieren konnten. Obwohl die Ergebnisse natürlich noch in weiteren Studien repliziert werden müssen, scheint die automatisierte Erstellung von Prüfungsfragen mithilfe von ChatGPT und Co. ein viel versprechendes Werkzeug für das Medizinstudium zu sein. Seine Kollegin und Co-Autorin der Studie Johanna Rother ergänzt: „Dozierende können ChatGPT zur Ideengenerierung von Prüfungsfragen nutzen, die anschließend durch die Dozierenden kontrolliert und gegebenenfalls überarbeitet werden. Insbesondere profitieren unserer Meinung nach aber die Studierenden von der automatisierten Erstellung von medizinischen Übungsfragen, da schon seit längerem bekannt ist, dass das Selbsttesten des eigenen Wissens sehr förderlich für das Lernen ist.“

Tobias Raupach, Direktor des Instituts für Medizindidaktik, führt weiter aus: „Wir wussten aus früheren Studien, dass Sprachmodelle wie ChatGPT die Fragen in medizinischen Staatsexamina beantworten können. Wir konnten nun erstmals zeigen, dass mit der Software auch

neue Fragen geschrieben werden können, die sich kaum von denen erfahrener Lehrender unterscheiden.“

Tizian Kaiser, der Humanmedizin im siebten Semester studiert, kommentiert: „Beim Bearbeiten der Probeklausur war ich ziemlich überrascht, wie schwer es mir fiel, die Fragen auseinanderzuhalten. Mein Ansatz war, die Fragen aufgrund ihrer Länge, der Komplexität ihres Satzbaus und ihrer inhaltlichen Schwere zu unterscheiden. Aber um ehrlich zu sein, musste ich in manchen Situationen schlicht raten und die Auswertung hat gezeigt, dass mir die Unterscheidung kaum möglich war. Das bringt mich zu der Überzeugung, dass eine sinnvolle Wissensabfrage wie in dieser Klausur auch ausschließlich durch von der KI gestellte Fragen möglich ist.“

Er ist überzeugt, dass ChatGPT ein großes Potenzial für das Lernen von Studierenden habe. Es ermögliche dem Studierenden das gelernte Wissen in verschiedenen Weisen und immer wieder anders zu wiederholen. „Es gibt die Möglichkeit, sich durch die KI zu vorher definierten Themen abfragen zu lassen, sich Probeklausuren konzipieren zu lassen oder mündliche Prüfungen schriftlich zu simulieren. Die Wiederholung des Stoffes ist dadurch zielführend auf das Prüfungskonzept zugeschnitten und die Trainingsmöglichkeiten sind endlos“, sagt der Studienteilnehmer und grenzt gleichzeitig ein: „Allerdings würde ich Chat-GPT auch nur dafür benutzen und nicht schon vorher im Lernprozess, in dem die Studienthemen aufgearbeitet und zusammengefasst werden müssen. Denn während Chat-GPT hervorragend für die Repetition ist, befürchte ich, dass es bei der Aufbereitung von Lerninhalten zu Fehlern kommen kann. Diese Fehler würden mir ohne eine vorherige Übersicht über das Thema nicht auffallen.“

Aus weiteren Studien ist bekannt, dass regelmäßiges Prüfen – auch und gerade ohne Benotung – Studierende dabei unterstützt, sich Lerninhalte nachhaltiger zu merken. Solche Prüfungen können jetzt mit wenig Aufwand erstellt werden. Zunächst soll die aktuelle Studie aber auf andere Kontexte (d. h. andere Fächer, Semester und Länder) übertragen und untersucht werden, ob ChatGPT auch andere Fragen als die in der Medizin üblichen Multiple Choice-Fragen schreiben kann.

Zum Universitätsklinikum Bonn:
Im UKB werden pro Jahr etwa 500.000 Patient*innen betreut, es sind ca. 9.000 Mitarbeiter*innen beschäftigt und die Bilanzsumme beträgt 1,6 Mrd. Euro. Neben den 3.500 Medizin- und Zahnmedizin-Studierenden werden pro Jahr 550 Personen in zahlreichen Gesundheitsberufen ausgebildet. Das UKB steht im Wissenschafts-Ranking sowie in der Focus-Klinikliste auf Platz 1 unter den Universitätsklinika (UK) in NRW und weist den dritthöchsten Case Mix Index (Fallschweregrad) in Deutschland auf. Das F.A.Z.-Institut hat das UKB 2022 und 2023 als Deutschland begehrtesten Arbeitgeber und Ausbildungs-Champion unter den öffentlichen Krankenhäusern in Deutschland ausgezeichnet.

Originalpublikation:
https://journals.lww.com/academicmedicine/abstract/9900/large_language_models_in_medical_education_.719.aspx
DOI: 10.1097/ACM.0000000000005626

10.01.2024
Daria Siverina
Universitätsklinikum Bonn
www.ukbonn.de