

Generative AI (gAI) in medical education: Chat-GPT

UbX'co

Sören Moritz¹

Bernd Romeike²

Christoph Stosch¹

Daniel Tolks³

1 University of Cologne,
Medical Faculty, Cologne,
Germany

2 University Medical Center
Rostock, Academic Dean's
Office, Division of Medical
Education, Rostock, Germany

3 Leuphana University
Lüneburg, Centre for Applied
Health Promotion, Lüneburg,
Germany

Editorial

“The use of chatbots in medical education is an emerging trend that is welcomed by many educators and medical professionals. In particular, the use of ChatGPT, a large language model of OpenAI, offers a variety of benefits for students and educators alike [...]” [1]. So far so amazing, the passage already points to the whole dilemma: will teaching at universities ever be the same after ChatGPT as it never was anyways?

We had a Cologne term paper in the “field of competence carcinogenesis” (interdisciplinary teaching in the first preclinical study semester) generated in triplicate by ChatGPT, each with identical queries, and received three different two-page texts including literature citations according to APA style. These have been examined by two detector programs (Groover, Writer) to determine whether they were written by a human or a bot. Both programs could not detect them as machine-written (cave: short texts are practically undetectable). The search for plagiarism with the software PlagAware did not reveal any conspicuous passages worthy of consideration (approx. 3-5% agreement with already published texts). The papers were forwarded unchanged to the assessing tutors with the result that two papers were assessed as “passed” and one as “failed”. The poor performance was due to certain terms used in the field of competence that was not named, as well as a non-matching literature citation. What next?

Let's ask ChatGPT: “...If students were able to access ChatGPT and ask questions during the exam, they could

theoretically receive answers from ChatGPT that could help them answer exam questions...” [2].

What is the consequence?

1. *Assessments in presence*: With adequate assessment supervision, ChatGPT does not pose a danger in terms of an increase in cheating in written, oral and practical assessment in presence.
2. *Homework and presentations*: Since most of the assessments, especially in medicine, are conducted in the form of MCQs in presence, the assessment formats “homework” or “presentation” are negligible. Nevertheless, existing, purely written, non-supervised assessment formats will have to be questioned. The assessment format “presentation” could be still used if the free presentation of scientific contexts is the focus. Here, the teacher has the opportunity to also assess the degree of mastery of the subject matter as the core aspect of the oral exam.
3. *Final theses (Bachelor, Master), scientific projects (as required by the new ÄApprO), dissertation papers*: Here a still unsolved problem of possible deception arises, since written texts and thoughts don't represent own achievements, but could be generated by AI. The oral assessment (i.e., the doctoral colloquium, etc.) has an increasingly important function.

Are the implications for the assessment culture at the university the only problem for universities, or what other implications arise from using generative AIs? To do this,

one needs to understand, at least in principle, how such a generative AI works?

To this end, ChatGPT says: “ChatGPT is a so-called “language-model-AI” or “Natural Language Processing” (NLP) AI based on deep learning. It has been trained with large amounts of human text, allowing it to generate and understand human-like text. To generate texts, ChatGPT uses a neural network that consists of many layers and has millions of parameters. It works like a very complex pattern recognition system that is able to recognize contextual relationships between words and sentences. For example, when you ask ChatGPT a question or type a sentence, the model tries to generate an appropriate answer or continuation based on what it has learned from its training data. [...] It is important to emphasize that AIs like ChatGPT are just tools and that their use is the responsibility of the people who use them. It is up to us as a society to ensure that AIs are used in ways that are both effective and ethical.” [3].

In summary, ChatGPT (GPT 3.5) models human speech better than any other computer program before it. It can simulate a real conversation partner and help answer complex facts and questions surprisingly effectively.

The latest version of OpenAI’s language models is called GPT-4. It was released on March 14, 2023 and is capable of generating more accurate and reliable statements than its predecessors GPT-3 and GPT-3.5. Another advantage is that GPT-4 is multimodal and can accept images as input. It can generate image captions, classifications, and analyses. The capabilities of GPT-4 are impressive, especially in exams: In a simulated BAR test, GPT-3.5 scored in the bottom 10% percentile, whereas GPT-4 performs in the top 10% percentile [4].

By using ChatGPT and similar AI tools, open-ended questions arise at multiple levels:

- Even if questions can be answered by generative AIs (gAI) in a predominantly factually accurate manner, who is responsible for the application and use of this knowledge? How does the accountability of decision makers change, for example, at the bedside or in teaching?
- What are the legal implications of the use of gAI on areas of law such as copyright (plagiarism), liability (malpractice), or corporate law (business models)?
- What impact do gAIs have on diversity or equity?
- In the long run, will gAI replace certain professions (dermatologists, pathologists, radiologists) or the journals (e.g., the JME), or will gAI give “superpowers” to some professions?
- Is there a risk of reinforcing educational inequity by students and institutions who have access to the tools and the skills to use them and those who don’t [5]?
- How will the authenticity of digital information be ensured in the future and how will this impact our media literacy?
- How do we design digital, university learning and assessment processes so that students become academic experts and use various gAIs in a productive way?
- How does teaching and learning change when lecturers have instructional materials created by gAI?
- What are the consequences of using gAI on communication with colleagues, patients, or on reflection?
- There is also the question, at which point the illegal use starts at all. The spell checker of Word and Grammarly® is used ubiquitously and is mostly recognized as an appropriate tool. But at what point is a paraphrase really plagiarism? The “prompt” comes from the author and mostly the generated text is further adapted if necessary.

Let’s turn the tables for once, because the AI will burn itself into the monitor, and yet the universities and consequently the teaching will continue to exist: Has the vinyl record disappeared, even though there were first CD’s and now streaming? At least there has been an amazing renaissance for vinyl records since 2010 [6]. However, in the face of climate change and resource scarcity, we will very soon be limited to purely digital formats, precisely because everyone will generate professional music themselves through “whatever-to-music” AI converters and share it via social media. Putting it positively, could gAIs be helpful in teaching?

Active use of gAIs in teaching: statements about specific, medical knowledge contexts generated interactively by students using gAIs can be analyzed by them and help to train higher cognitive functions such as “evaluation and assessment” (according to Bloom [7]) and thereby get a good overview of the topics. Students are forced to change roles in this process. They are the authors of the “prompts”, receive writing support from gAI and then, however, have to prove their expertise as editors and evaluate the generated texts or correct them as well as possible. But: How do students “climb” to the higher levels according to Bloom, if knowledge acquisition with a bot works fundamentally different?

Via the use of gAIs, students can practice asking the right questions, a core competency of Evidence Based Medicine. During the development of research questions the chatbot can help sharpen and delineate them. It can also suggest different methods and study designs. The gAI is helpful and efficient in paraphrasing texts.

In medicine, training in problem-solving skills (clinical decision making) is an important competence. Here, the so-called background knowledge probably plays a central role. This background knowledge can be specifically improved by the application of gAI, in which gAI expands the views on the problem (“chatbot PbL”?). In the end, we get some differential diagnoses added that we just didn’t think of.

gAI will establish itself as an interactive reference work. It is to be expected that the use of AI on the wards will be standardized with an evidence-based approach (GPT-4). The use of these tools will then be expected as part of physician competence and, to that extent, will also need to be trained (replaces gAI the German-language test preparation software AMBOSS?).

gAI can be used to create virtual patient cases. On the one hand, the linguistic quality can be improved and on the other hand, more exciting cases can be generated, for example with the help of storytelling elements or paraphrasing.

gAI can be used to create storyboards or stage directions for educational films.

gAI can support the enforcement of competence-based forms of assessment: Alone the use of the available information for the benefit of the patient*s (i.e. the competence in dealing with the counterpart) should increasingly be the subject of new assessment scenarios (workplace-based assessment, simulations, physical examinations, oral assessments, ...).

There are first experiences that on the lecturer side gAI can be used well for the creation of MC questions: Here, it is especially about the targeted search for distractors.

Conclusion

The emergence of generative AI tools is a game changer that some experts compare to the introduction of the smartphone. Given the remarkable advances seen in recent months, as well as those expected this year, it's fair to say that the impact on humanity will be as significant as the displacement of horses from cities with the introduction of the automobile. Unlike cars, which took a century to evolve into the sophisticated machines we have today, generative AI tools need only the year 2023 for a significant transformation, resulting in a highly notable impact. Those who embrace it will have a clear advantage over those who don't.

There needs to be a clear strategy within our education system to keep up with the rapid evolution of gAI tools and to continuously integrate them into our curricula and syllabi. The opportunities and risks of gAI on our teaching and learning need to be continuously analyzed so that didactic strategies can be adapted in a timely manner. We are likely to run somewhat behind developments in the coming months and years. Therefore, it seems all the more important to consistently conduct evidence-based teaching research in all integrations of gAIs into our teaching:

In this sense, the authors of the editorial would like to invite you to use the summer semester to engage in experiences with generative AI in teaching and learning at universities, and to bring these thoughts, preliminary studies, and experiments to a bar-camp at the upcoming annual meeting of the Society for Medical Education in Osnabrück [<https://gma2023.de/>] as a contribution. Welcome Industrialization 5.0.

A collection of links on the topic of generative AI is available at: [<https://padlet.com/danieltolks1/linksammlung-didaktik-und-chatgpt-ausschuss-digitalisierung-ihupnj9wb3y0foz3>]

More articles in this issue

In their study, Gisi et al. examined the objective and subjective effects of pandemic-related changes in the course of studies on the perception of the practical year [8]. In their study, Brütting and colleagues took a closer look at the factors conducive to working as a doctor in rural regions and the awareness of districts in the catchment area of universities [9]. In an intervention study, Hopp et al. examined the influence of teaching medical students close to patients on the stigmatization of people with mental illness [10]. In their study, Kiesewetter and colleagues were able to demonstrate a positive influence in the use of dogs in therapy with children with rheumatic diseases [11]. Kruse et al. present the results of a survey on learning strategies of dental students [12]. Mand et al. present the implementation of a multi-stage observer training for medical students to assess simulated pediatric emergency situations [13]. Jannik Osten and colleagues deal with the question of whether face-to-face lectures are still up-to-date and whether synchronous online lectures are an alternative [14]. In their questionnaire survey, Daunert and colleagues investigated the question of what motivates GPs to train medical students in their practice [15]. Dasci et al. evaluated different forms of training in knotting and suturing techniques in a controlled randomized study in dentistry [16]. The working group Quality Management in Education, Training and Continuing Education of the Society for Quality Management in Health Care (Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e.V., GQMG) has published a position paper on the topic of quality management in medical studies in the competence-based learning objectives catalogue [17]. Koch et al. describe the development of an interactive elective "modified anatomy" for students within the framework of the Z-curriculum according to NKLM 2.0 [18]. Laura Wortmann et al. present the results of a survey on the topic of gender medicine in teaching [19].

Competing interests

The authors declare that they have no competing interests.

References

1. Open AI. Antwort von ChatGPT auf die Frage: "Schreibe ein Editorial zum Gebrauch von Chat GPT in der Medizinischen Lehre." ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>
2. Open AI. Antwort von ChatGPT auf die Frage: "Ist es möglich, dass Studierende bei einer Hochschulprüfung mit Chat GPT betrügen?" ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>
3. OpenAI. Antwort von ChatGPT auf die Frage: "Wie arbeitet eine KI wie ChatGPT und welche Konsequenzen ergeben sich daraus?" ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>

4. Slowik C, Kaiser F. GPT-4 vs. GPT-3. OpenAI Models' Comparison. Neoteric.eu. 2023. Zugänglich unter/available from: <https://neoteric.eu/blog/gpt-4-vs-gpt-3-openai-models-comparison/>
5. Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Stadler M, Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education. *Lern Individ Diff.* 2023;103:102274. DOI: 10.1016/j.lindif.2023.102274
6. Statista. Absatz von Schallplatten (Vinyl-LPs) in Deutschland in den Jahren 2003 bis 2021. Hamburg: Statista; 2023. Zugänglich unter/available from: <https://de.statista.com/statistik/daten/studie/256099/umfrage/absatz-von-schallplatten-in-deutschland-zeitreihe>
7. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC. A Taxonomy for Learning, Teaching, and Assessing: Pearson New International Edition: A Revision of Bloom's Taxonomy of Educational Objectives. Harlow: Pearson Education Limited; 2013. p.336.
8. Gisi M, Ferrari V, Dubon F, Fischer MR, Angstwurm M, Berndt M. Objective and subjective consequences of pandemic-related study program changes for the perceptions of the practical year – A comparison of medical students in Germany with and without the second state examination. *GMS J Med Educ.* 2023;40(4):Doc53. DOI: 10.3205/zma001035
9. Brütting C, Hergert S, Bauch -F, Nafziger M, Klingenberg A, Deutsch T, Frese T. Factors promoting willingness to practice medicine in rural regions and awareness of rural regions in the university's catchment area – cross-sectional survey among medical students in Central Germany. *GMS J Med Educ.* 2023;40(4):Doc52. DOI: 10.3205/zma001634
10. Hopp A, Dechering S, Wilm S, Pressentin M, Müller T, Richter P, Schäfer R, Franz M, Karger F. The influence of patient-centered teaching on medical students' stigmatization of the mentally ill. *GMS J Med Educ.* 2023;40(4):Doc46. DOI: 10.3205/zma001628
11. Kiesewetter J, Herbach N, Landes I, Mayer J, Elgner V, Orle K, Grunow A, Langkau R, Gratzner C, Jansson AF. Dog assisted education in children with rheumatic diseases and adolescents with chronic pain in Germany. *GMS J Med Educ.* 2023;40(4):Doc44. DOI: 10.3205/zma001626
12. Kruse AB, Isailov-Schöchl M, Giesler M, Ratka-Krüger P. Which digital learning strategies do undergraduate dentistry students favor? A questionnaire survey at a German university. *GMS J Med Educ.* 2023;40(4):Doc49. DOI: 10.3205/zma001631
13. Mand N, Stibane T, Sitter H, Maier RF, Leonhardt A. Successful implementation of a rater training program for medical students to evaluate simulated pediatric emergencies. *GMS J Med Educ.* 2023;40(4):Doc47. DOI: 10.3205/zma001629
14. Osten J, Behrens V, Behrens S, Herrler A, Clarner T. Are live lectures a discontinued model? A survey on the influence of synchronous online lecturing on the perception of teaching and assessment outcome. *GMS J Med Educ.* 2023;40(4):Doc50. DOI: 10.3205/zma001632
15. Daunert L, Schulz S, Lehmann T, Bleidorn J, Petruschke I. What motivates GPs to train medical students in their own practice? A questionnaire survey on the motivation of medical practices to train students as an approach to acquire training practices. *GMS J Med Educ.* 2023;40(4):Doc51. DOI: 10.3205/zma001633
16. Dasci S, Schrem H, Oldhafer F, Beetz O, Kleine-Döpke D, Vondran F, Beneke J, Sarisin A, Ramackers W. Learning surgical knot tying and suturing technique – effects of different forms of training in a controlled randomized trial with dental students. *GMS J Med Educ.* 2023;40(4):Doc48. DOI: 10.3205/zma001630
17. Vogeser M, Borchers K, James J, Koch J, Kurscheid-Reich D, Kuske S, Pietsch B, Zillich S. Competence-based catalog of learning objectives for the subject area of quality management in medical studies – position paper of the working group Quality Management in Education, Training and Continuing Education of the Society for Quality Management in Health Care (GQMG). *GMS J Med Educ.* 2023;40(4):Doc42. DOI: 10.3205/zma001624
18. Koch K, Hirt B, Shiozawa-Bayer T, Königsrainer A, Fusso S, Wichmann D. Development of an interactive elective "altered anatomy" for students as part of the Z-curriculum according to the NKLM 2.0. *GMS J Med Educ.* 2023;40(4):Doc43. DOI: 10.3205/zma001625
19. Wortmann L, Haarmann L, Yeboah A, Kalbe E. Gender medicine teaching increases medical students' gender awareness: results of a quantitative survey. *GMS J Med Educ.* 2023;40(4):Doc45. DOI: 10.3205/zma001627

Corresponding author:

Daniel Tolks
 Leuphana University Lüneburg, Centre for Applied Health
 Promotion, Wilschenbrucher Weg 84A, D-21335
 Lüneburg, Germany
 Daniel.tolks@leuphana.de

Please cite as

Moritz S, Romeike B, Stosch C, Tolks D. Generative AI (gAI) in medical education: Chat-GPT and co. *GMS J Med Educ.* 2023;40(4):Doc54. DOI: 10.3205/zma001636, URN: urn:nbn:de:0183-zma0016361

This article is freely available from

<https://doi.org/10.3205/zma001636>

Received: 2023-04-28

Revised: 2023-04-28

Accepted: 2023-05-03

Published: 2023-06-15

Copyright

©2023 Moritz et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Generative KI (gKI) in der medizinischen Ausbildung: Chat-GPT und Co

Sören Moritz¹

Bernd Romeike²

Christoph Stosch¹

Daniel Tolks³

1 Universität zu Köln,
Medizinische Fakultät, Köln,
Deutschland

2 Universitätsmedizin Rostock,
Studiendekanat,
Medizindidaktik, Rostock,
Deutschland

3 Leuphana Universität
Lüneburg, Zentrum für
angewandte
Gesundheitswissenschaften,
Lüneburg, Deutschland

Leitartikel

„Die Verwendung von Chatbots in der medizinischen Ausbildung ist ein aufstrebender Trend, der von vielen Pädagogen und medizinischen Fachleuten begrüßt wird. Insbesondere der Einsatz von ChatGPT, einem *large language model* von OpenAI, bietet eine Vielzahl von Vorteilen für Studenten und Ausbilder gleichermaßen [...]“ [1]. So weit so erstaunlich, zeigt die Passage schon auf das gesamte Dilemma: Wird die Lehre an den Hochschulen nach ChatGPT jemals wieder so, wie sie nie war?

Wir haben eine Kölner Hausarbeit im „Kompetenzfeld Karzinogenese“ (Interdisziplinärer Unterricht im ersten vorklinischen Studiensemester) dreifach mit jeweils identischen Anfragen durch ChatGPT erstellen lassen und drei unterschiedliche, zweiseitige Texte inklusive Literaturzitate nach APA-Stil erhalten. Diese sind durch zwei Detektorenprogramme (Groover, Writer) auf die Frage hin untersucht worden, ob sie von einem Menschen oder von einem Bot geschrieben wurden. Beide Programme konnten diese nicht als maschinengeschrieben detektieren (cave: kurze Texte sind praktisch nicht zu erkennen). Die Suche nach Plagiaten mit der Software PlagAWARE ergab keine bedenkenswerten Auffälligkeiten (ca. 3-5% Übereinstimmung mit bereits publizierten Texten). Die Arbeiten wurden unverändert an die bewertenden Tutor*innen weitergeleitet mit dem Ergebnis, dass zwei Arbeiten als „bestanden“ und eine als „nicht bestanden“

bewertet wurden. Die Schlechtleistung zielte dabei auf bestimmte inhaltliche im Kompetenzfeld verwendete Begrifflichkeiten ab, die nicht benannt wurden, sowie ein nicht passendes Literaturzitat. Was nun?

Fragen wir doch ChatGPT: „...Wenn Studierende in der Lage wären, auf ChatGPT zuzugreifen und Fragen während der Prüfung zu stellen, könnten sie theoretisch Antworten von ChatGPT erhalten, die ihnen bei der Beantwortung von Prüfungsfragen helfen könnten...“ [2].

Was ist die Konsequenz?

1. *Prüfungen in Präsenz*: Bei einer adäquaten Prüfungsaufsicht stellt ChatGPT in schriftlichen, mündlichen und praktischen Prüfungen in Präsenz keine Gefahr dar im Sinne einer Zunahme von Täuschungen.
2. *Hausarbeiten und Referate*: Da die meisten Leistungsnachweise insbesondere im Studiengang Medizin in Form von MC-Klausuren in Präsenz durchgeführt werden, sind die Prüfungsformate „Hausarbeit“ oder „Referat“ zu vernachlässigen. Gleichwohl werden bestehende, rein schriftliche, nicht beaufsichtigte Prüfungsformate in Frage zu stellen sein. Die Prüfungsform „Referat“ ließe sich heilen, wenn der freie Vortrag wissenschaftlicher Zusammenhänge im Mittelpunkt steht. Hier hat der Lehrende die Möglichkeit, den Grad der Beherrschung der Materie als Prüfungsgegenstand ebenfalls zu bewerten.

3. *Abschlussarbeiten (Bachelor, Master), Wissenschaftliche Projekte (wie sie mit der neuen ÄApprO gefordert werden), Dissertationsschriften:* Hier ergibt sich ein noch ungelöstes Problem von möglichen Täuschungen, da formulierte Texte und Gedanken keine Eigenleistungen darstellen, sondern von KI generiert sein könnten. Der mündlichen Prüfung (also dem Promotionskolloquien etc.) kommt eine zunehmend wichtigere Funktion zu.

Sind die Auswirkungen auf die universitäre Prüfungslandschaft das einzige Problem für die Hochschulen oder welche weiteren Auswirkungen ergeben sich durch Verwendung generativer KIs? Dazu muss man zumindest im Prinzip verstehen, wie eine solche generative Künstliche Intelligenz funktioniert?

Dazu sagt ChatGPT: „ChatGPT ist eine sogenannte „Sprachmodell-KI“ oder „Natural Language Processing“ (NLP) KI, die auf einer Methode des maschinellen Lernens namens „deep learning“ basiert. Sie wurde mit großen Mengen an menschlichem Text trainiert und kann dadurch menschenähnliche Texte generieren und verstehen. Um Texte zu generieren, verwendet ChatGPT ein neuronales Netzwerk, das aus vielen Schichten besteht und Millionen von Parametern hat. Es funktioniert wie ein sehr komplexes Mustererkennungssystem, das in der Lage ist, kontextbezogene Zusammenhänge zwischen Wörtern und Sätzen zu erkennen. Wenn Sie ChatGPT beispielsweise eine Frage stellen oder einen Satz eingeben, versucht das Modell, eine passende Antwort oder Fortsetzung auf der Grundlage dessen zu generieren, was es aus seinen Trainingsdaten gelernt hat. [...] Es ist wichtig zu betonen, dass KIs wie ChatGPT nur Werkzeuge sind und dass ihre Verwendung in der Verantwortung der Menschen liegt, die sie einsetzen. Es liegt an uns als Gesellschaft, sicherzustellen, dass KIs in einer Art und Weise eingesetzt werden, die sowohl effektiv als auch ethisch vertretbar ist.“ [3].

Zusammengefasst modelliert ChatGPT (GPT 3.5) menschliche Sprache so gut wie noch kein anderes Computerprogramm zuvor. Es kann einen echten Gesprächspartner simulieren und bei der Beantwortung komplexer Sachverhalte und Fragestellungen überraschend effektiv helfen.

Die neueste Version der Sprachmodelle von OpenAI nennt sich GPT-4. Am 14. März 2023 wurde es veröffentlicht und es hat die Fähigkeit, genauere und verlässlichere Aussagen zu generieren als seine Vorgänger GPT-3 und GPT-3.5. Ein weiterer Vorteil ist, dass GPT-4 multimodal ist und somit Bilder als Eingabe akzeptieren kann. Es kann Bildunterschriften, Klassifizierungen und Analysen erstellen. Die Fähigkeiten von GPT-4 sind beeindruckend, insbesondere bei Prüfungen: In einem simulierten BAR-Test erreichte GPT-3.5 das unterste 10%-Perzentil, wohin gegen GPT-4 im obersten 10%-Perzentil abschneidet [4]. Durch die Nutzung von ChatGPT und ähnlichen KI-Tools ergeben sich offene Fragen auf vielfältigen Ebenen:

- Selbst wenn Fragen von generativen KIs (gKI) überwiegend sachlich zutreffend beantwortet werden können, wer verantwortet die An- und Verwendung dieses Wissens? Wie verändert sich die Verantwortlichkeit der Entscheidungsträger zum Beispiel am Krankenbett oder der Lehre?
- Welche juristischen Implikationen hat die Verwendung der gKI auf Rechtsgebiete wie etwa das Urheberrecht (Plagiat), Haftungsrecht (Behandlungsfehler) oder Gesellschaftsrecht (Geschäftsmodelle)?
- Welchen Einfluss haben gKI auf Diversity oder Equity?
- Ersetzen gKI langfristig bestimmte Berufe (Dermato-, Patho- und Radiolog*innen) oder die Fachzeitschriften (z.B. das JME) oder verleiht gKI einigen Berufen „Superkräfte“?
- Besteht die Gefahr der Verstärkung der Bildungsungleichheit durch Studierende und Institutionen, die Zugang zu den Tools und die Kompetenzen zur Nutzung haben und denen, die nicht [5]?
- Wie wird die Authentizität von digitalen Informationen zukünftig sichergestellt und wie wirkt sich das auf unsere Medienkompetenz aus?
- Wie gestalten wir digitale, universitäre Lern- und Prüfungsprozesse so, dass die Studierenden zum akademischen Experten werden und verschiedene gKI in einer produktiven Art und Weise nutzen?
- Wie verändert sich Lehre und Lernen, wenn Dozierende die Unterrichtsmaterialien durch gKI erstellen lassen?
- Welche Konsequenzen hat die Verwendung von gKI auf die Kommunikation mit Kolleg*innen, Patient*innen bzw. auf die Reflexion?
- Es stellt sich auch die Frage, an welcher Stelle die illegale Nutzung überhaupt beginnt. Die Rechtschreibprüfung von Word und Grammarly® wird ubiquitär genutzt und wird zumeist auch als probates Hilfsmittel anerkannt. Ab wann ist aber eine Paraphrasierung wirklich ein Plagiat? Der „Prompt“ stammt vom Autor und zumeist wird der generierte Text bei Bedarf weiter angepasst.

Drehen wir den Spieß einmal um, denn die KI wird sich im Monitor einbrennen, und dennoch werden die Hochschulen und mithin die Lehre weiter existieren: Ist die Vinylplatte verschwunden, obwohl es erst CD's und jetzt Streaming gibt? Zumindest gibt es seit 2010 eine erstaunliche Renaissance für Vinylplatten [6]. In Angesicht des Klimawandels und von Ressourcenknappheit werden wir uns aber sehr bald auf rein digitale Formate beschränken, gerade auch weil jeder durch „whatever-to-music“ KI-Konverter professionelle Musik selbst generieren und über soziale Medien teilen wird. Positiv formuliert, könnten gKIs in der Lehre hilfreich sein?

Aktiver Einsatz von gKI im Unterricht: Durch Studierende interaktiv mittels gKI erzeugten Aussagen zu spezifischen, medizinischen Wissenskontexten, können durch diese analysiert werden und helfen, höhere kognitive Funktionen wie „Evaluation und Bewertung“ (nach Bloom [7]) anzutrainieren und dadurch einen guten Überblick zu den Themen zu erhalten. Studierende werden hierbei zu ei-

nem Rollenwechsel gezwungen. Sie sind die Autoren der „Prompts“, erhalten Schreib-Unterstützung durch gKI und müssen dann aber als Editoren ihr Fachwissen unter Beweis stellen und die generierten Texte bewerten bzw. möglichst gut korrigieren. Aber: Wie „erklimmen“ die Studierenden die höheren Ebenen nach Bloom, wenn der Wissenserwerb mit einem Bot fundamental anders funktioniert?

Über die Anwendung von gKIs können die Studierenden einüben, die richtigen Fragen zu stellen, eine Kernkompetenz der Evidence Based Medicine. Bei der Entwicklung von Forschungsfragen kann der Chatbot helfen diese zu schärfen und abzugrenzen. Er kann auch unterschiedliche Methoden und Studiendesigns vorschlagen. Hilfreich und effizient ist gKI derzeit beim paraphrasieren von Texten. In der Medizin ist das Training der Problemlösekompetenz (Clinical Decision Making) eine wichtige Kompetenz. Hierbei spielt das sogenannte Hintergrundwissen wahrscheinlich eine zentrale Rolle. Dieses Hintergrundwissen kann durch die Anwendung von gKI gezielt verbessert werden, in dem gKI die Sichtweisen auf das Problem erweitert („Chatbot PbL“?). Am Ende bekommen wir noch ein paar Differenzialdiagnosen hinzu, an die wir gerade nicht gedacht haben.

gKI wird sich als interaktives Nachschlagewerk etablieren. Es ist zu erwarten, dass der Einsatz von KI auf den Stationen bei evidenzbasiertem Vorgehen standardisiert erfolgt (GPT-4). Der Umgang mit diesen Werkzeugen wird dann als Teil ärztlicher Kompetenz erwartet werden und muss insofern auch trainiert werden (ersetzt gKI die deutschsprachige Prüfungsvorbereitungssoftware AM-BOSS?).

gKI können zur Erstellung von virtuellen Patientenfällen genutzt werden. Dabei kann zum einen die sprachliche Qualität verbessert und zum anderen auch spannendere Fälle generiert werden, beispielsweise mit Hilfe von Storytelling-Elementen.

gKI kann zur Erstellung von Regieanweisungen für Lehrfilme genutzt werden.

gKI kann bei der Durchsetzung kompetenzbasierter Prüfungsformen unterstützen: Alleine der Einsatz der verfügbaren Information zum Wohle der Patient*innen (also die Kompetenz im Umgang mit dem Gegenüber) sollte vermehrt Gegenstand neuer Prüfungsszenarien sein (Workplace-based Assessment, Simulationspersonenprüfungen, mündliche Prüfungen, ...).

Es gibt erste Erfahrungen dazu, dass auf Dozierendenseite gKI gut für die Erstellung von MC-Fragen verwendet werden kann: Hier geht es insbesondere um die gezielte Suche nach Distraktoren.

Schlussformel

Das Aufkommen generativer KI-Tools ist ein Umbruch, der von einigen Experten mit der Einführung des Smartphones verglichen wird. In Anbetracht der bemerkenswerten Fortschritte, die in den letzten Monaten zu beobachten sind, sowie derjenigen, die für dieses Jahr erwartet werden, wie die Verbindung von ChatGPT und Bing im

Edge-Browser, kann man wohl behaupten, dass die Auswirkungen auf die Menschheit so bedeutend sein werden wie die Verdrängung der Pferde aus den Städten mit der Einführung des Automobils. Im Gegensatz zu Autos, die ein Jahrhundert brauchten, um sich zu den hochentwickelten Maschinen zu entwickeln, die wir heute haben, benötigen generative KI-Tools lediglich das Jahr 2023 für einen Wandel, der schon jetzt sehr einprägsam ist. Diejenigen, die sich darauf einlassen, werden einen klaren Vorteil gegenüber denjenigen haben, die dies nicht tun. Es braucht eine klare Strategie innerhalb unseres Bildungssystems, um mit der rasanten Entwicklung gKI Werkzeuge Schritt zu halten und diese fortlaufend in unsere Curricula und Lehrpläne zu integrieren. Die Chancen und Risiken von gKI auf unser Lehren und Lernen müssen fortlaufend analysiert werden, sodass didaktische Strategien zeitnah adaptiert werden können. Voraussichtlich laufen wir in den nächsten Monaten und Jahren den Entwicklungen etwas hinterher. Umso wichtiger erscheint die konsequente Durchführung evidenzbasierter Lehrforschung bei allen Integrationen von gKI in unseren Lehrbetrieb:

In diesem Sinne möchten die Autoren des Editorials einladen, das Sommersemester zu nutzen um Erfahrungen mit der generativen KI in Lehre und Lernen an den Hochschulen zu beschäftigen, und diese Gedanken, Vorstudien und Versuche mitzubringen zu einem Barcamp auf der kommenden Jahrestagung der Gesellschaft für Medizinische Ausbildung in Osnabrück [<https://gma2023.de/>] als Beitrag einzubringen. Willkommen Industrialisierung 5.0.

Eine Linksammlung zum Thema Generativer KI gibt es unter: [<https://padlet.com/danieltolks1/linksammlung-didaktik-und-chatgpt-ausschuss-digitalisierung-ihupnj9wb3y0foz3>]

Weitere Artikel in dieser Ausgabe

Neben dem großen Thema der generativen KI gibt es natürlich noch weitere Themen in dieser Ausgabe.

Gisi et al. haben in ihrer Studie die objektiven und subjektiven Auswirkungen pandemiebedingter Änderungen im Studienablauf auf die Wahrnehmung des Praktischen Jahres untersucht [8]. Brütting und KollegInnen haben in ihrer Studie die förderlichen Faktoren für eine ärztliche Tätigkeit in ländlichen Regionen und die Bekanntheit von Landkreisen im Einzugsgebiet der Universitäten näher betrachtet [9]. Hopp et al. haben in einer Interventionsstudie den Einfluss von patientennaher Lehre bei Medizinstudierenden auf die Stigmatisierung von Menschen mit psychischen Erkrankungen untersucht [10]. Kiesewetter und KollegInnen haben in ihrer Studie einen positiven Einfluss bei der Nutzung von Hunden in der Therapie mit Kindern mit rheumatischen Erkrankungen nachweisen können [11]. Kruse et al. stellen die Ergebnisse einer Befragung zu Lernstrategien von Studierenden der Zahnmedizin vor [12]. Mand et al. präsentieren die Durchführung eines mehrstufigen Beobachtertrainings für Medizinstudierende zur Beurteilung simulierter pädia-

trischer Notfallsituationen [13]. Jannik Osten und KollegInnen beschäftigen sich mit der Frage, ob Präsenzvorlesungen noch zeitgemäß sind und ob synchrone Online-Lehrveranstaltungen eine Alternative darstellen [14]. Daunert und KollegInnen sind in ihrer Fragebogenerhebung der Frage nachgegangen, was Hausärzt*innen motiviert, Medizinstudierende in ihrer Praxis auszubilden [15]. Dasci et al. haben in einer kontrollierten randomisierten Studie in der Zahnmedizin verschiedene Trainingsformen in der Knoten- und Nahttechnik evaluiert [16]. Der Arbeitskreis Qualitätsmanagement in der Aus-, Fort- und Weiterbildung der Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e.V. (GQMG) hat ein Positionspapier zum Thema Qualitätsmanagement im Medizinstudium im Kompetenzbasierten Lernzielkatalog veröffentlicht [17]. Koch et al. beschreiben die Entwicklung eines interaktiven Wahlfachs „Modifizierte Anatomie“ für Studierende im Rahmen des Z-Curriculums nach NKLM 2.0 [18]. Laura Wortmann u.a. stellen die Ergebnisse einer Befragung zum Thema Gendermedizin in der Lehre vor [19].

Interessenkonflikt

Die Autoren erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

- Open AI. Antwort von ChatGPT auf die Frage: "Schreibe ein Editorial zum Gebrauch von Chat GPT in der Medizinischen Lehre." ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>
- Open AI. Antwort von ChatGPT auf die Frage: "Ist es möglich, dass Studierende bei einer Hochschulprüfung mit Chat GPT betrügen?" ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>
- OpenAI. Antwort von ChatGPT auf die Frage: "Wie arbeitet eine KI wie ChatGPT und welche Konsequenzen ergeben sich daraus?" ChatGPT; [cited 2023 Feb 15]. Zugänglich unter/available from: <https://chat.openai.com/>
- Slowik C, Kaiser F. GPT-4 vs. GPT-3. OpenAI Models' Comparison. Neoteric.eu. 2023. Zugänglich unter/available from: <https://neoteric.eu/blog/gpt-4-vs-gpt-3-openai-models-comparison/>
- Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Stadler M, Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education. *Lern Individ Diff.* 2023;103:102274. DOI: 10.1016/j.lindif.2023.102274
- Statista. Absatz von Schallplatten (Vinyl-LPs) in Deutschland in den Jahren 2003 bis 2021. Hamburg: Statista; 2023. Zugänglich unter/available from: <https://de.statista.com/statistik/daten/studie/256099/umfrage/absatz-von-schallplatten-in-deutschland-zeitreihe>
- Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC. *A Taxonomy for Learning, Teaching, and Assessing: Pearson New International Edition: A Revision of Bloom's Taxonomy of Educational Objectives.* Harlow: Pearson Education Limited; 2013. p.336.
- Gisi M, Ferrari V, Dubon F, Fischer MR, Angstwurm M, Berndt M. Objective and subjective consequences of pandemic-related study program changes for the perceptions of the practical year – A comparison of medical students in Germany with and without the second state examination. *GMS J Med Educ.* 2023;40(4):Doc53. DOI: 10.3205/zma001035
- Brütting C, Hergert S, Bauch -F, Nafziger M, Klingenberg A, Deutsch T, Frese T. Factors promoting willingness to practice medicine in rural regions and awareness of rural regions in the university's catchment area – cross-sectional survey among medical students in Central Germany. *GMS J Med Educ.* 2023;40(4):Doc52. DOI: 10.3205/zma001634
- Hopp A, Dechering S, Wilm S, Pressentin M, Müller T, Richter P, Schäfer R, Franz M, Karger F. The influence of patient-centered teaching on medical students' stigmatization of the mentally ill. *GMS J Med Educ.* 2023;40(4):Doc46. DOI: 10.3205/zma001628
- Kiesewetter J, Herbach N, Landes I, Mayer J, Elgner V, Orle K, Grunow A, Langkau R, Gratzner C, Jansson AF. Dog assisted education in children with rheumatic diseases and adolescents with chronic pain in Germany. *GMS J Med Educ.* 2023;40(4):Doc44. DOI: 10.3205/zma001626
- Kruse AB, Isailov-Schöchlin M, Giesler M, Ratka-Krüger P. Which digital learning strategies do undergraduate dentistry students favor? A questionnaire survey at a German university. *GMS J Med Educ.* 2023;40(4):Doc49. DOI: 10.3205/zma001631
- Mand N, Stibane T, Sitter H, Maier RF, Leonhardt A. Successful implementation of a rater training program for medical students to evaluate simulated pediatric emergencies. *GMS J Med Educ.* 2023;40(4):Doc47. DOI: 10.3205/zma001629
- Osten J, Behrens V, Behrens S, Herrler A, Clarner T. Are live lectures a discontinued model? A survey on the influence of synchronous online lecturing on the perception of teaching and assessment outcome. *GMS J Med Educ.* 2023;40(4):Doc50. DOI: 10.3205/zma001632
- Daunert L, Schulz S, Lehmann T, Bleidorn J, Petruschke I. What motivates GPs to train medical students in their own practice? A questionnaire survey on the motivation of medical practices to train students as an approach to acquire training practices. *GMS J Med Educ.* 2023;40(4):Doc51. DOI: 10.3205/zma001633
- Dasci S, Schrem H, Oldhafer F, Beetz O, Kleine-Döpke D, Vondran F, Beneke J, Sarisin A, Ramackers W. Learning surgical knot tying and suturing technique – effects of different forms of training in a controlled randomized trial with dental students. *GMS J Med Educ.* 2023;40(4):Doc48. DOI: 10.3205/zma001630
- Vogeser M, Borchers K, James J, Koch J, Kurscheid-Reich D, Kuske S, Pietsch B, Zillich S. Competence-based catalog of learning objectives for the subject area of quality management in medical studies – position paper of the working group Quality Management in Education, Training and Continuing Education of the Society for Quality Management in Health Care (GQMG). *GMS J Med Educ.* 2023;40(4):Doc42. DOI: 10.3205/zma001624
- Koch K, Hirt B, Shiozawa-Bayer T, Königsrainer A, Fusso S, Wichmann D. Development of an interactive elective "altered anatomy" for students as part of the Z-curriculum according to the NKLM 2.0. *GMS J Med Educ.* 2023;40(4):Doc43. DOI: 10.3205/zma001625
- Wortmann L, Haarmann L, Yeboah A, Kalbe E. Gender medicine teaching increases medical students' gender awareness: results of a quantitative survey. *GMS J Med Educ.* 2023;40(4):Doc45. DOI: 10.3205/zma001627

Korrespondenzadresse:

Daniel Tolks
Leuphana Universität Lüneburg, Zentrum für angewandte
Gesundheitswissenschaften, Wilschenbrucher Weg 84A,
21335 Lüneburg, Deutschland
Daniel.tolks@leuphana.de

Bitte zitieren als

Moritz S, Romeike B, Stosch C, Tolks D. Generative AI (gAI) in medical
education: Chat-GPT and co. *GMS J Med Educ.* 2023;40(4):Doc54.
DOI: 10.3205/zma001636, URN: urn:nbn:de:0183-zma0016361

Artikel online frei zugänglich unter
<https://doi.org/10.3205/zma001636>

Eingereicht: 28.04.2023
Überarbeitet: 28.04.2023
Angenommen: 03.05.2023
Veröffentlicht: 15.06.2023

Copyright

©2023 Moritz et al. Dieser Artikel ist ein Open-Access-Artikel und steht
unter den Lizenzbedingungen der Creative Commons Attribution 4.0
License (Namensnennung). Lizenz-Angaben siehe
<http://creativecommons.org/licenses/by/4.0/>.