Using ChatGPT for Writing Articles for Patients' Education for Dermatological Diseases: A Pilot Study

Abstract

Background: Patients' education is a vital strategy for understanding a disease by patients and proper management of the condition. Physicians and academicians frequently make customized education materials for their patients. An artificial intelligence (AI)-based writer can help them write an article. Chat Generative Pre-Trained Transformer (ChatGPT) is a conversational language model developed by OpenAI (openai.com). The model can generate human-like responses. **Objective:** We aimed to evaluate the generated text from ChatGPT for its suitability in patients' education. **Materials and Methods:** We asked the ChatGPT to list common dermatological diseases. It provided a list of 14 diseases. We used the disease names to converse with the application with disease-specific input (e.g., write a patient education guide on acne). The text was copied for checking the number of words, readability, and text similarity by software. The text's accuracy was checked by a dermatologist following the structure of observed learning outcomes (SOLO) taxonomy. For the readability ease score, we compared the observed value with a score of 30. For the similarity index, we compared the observed value with 15% and tested it with a one-sample t-test. **Results:** The ChatGPT generated a paragraph of text of 377.43 ± 60.85 words for a patient education guide on skin diseases. The average text reading ease score was 46.94 ± 8.23 ($P < 0.0001$), and it indicates that this level of text can easily be understood by a high-school student to a newly joined college student. The text similarity index was higher ($27.07 \pm 11.46\%$, $P = 0.002$) than the expected limit of 15%. The text had a "relational" level of accuracy according to the SOLO taxonomy. **Conclusion:** In its current form, ChatGPT can generate a paragraph of text for patients' educational purposes that can be easily understood. However, the similarity index is high. Hence, doctors should be cautious when using the text generated by ChatGPT and must check for text similarity before using it.

Keywords: *Article, artificial intelligence, ChatGPT, dermatologists, software*

**Himel Mondal,
Shaikat Mondal[1],
Indrashis Podder[2]**

*Department of Physiology,
All India Institute of Medical
Sciences, Deoghar, Jharkhand,
[1]Department of Physiology,
Raiganj Government Medical
College and Hospital,
West Bengal, [2]Department
of Dermatology, College
of Medicine and Sagore Dutta
Hospital, Kolkata, West Bengal,
India*

Introduction

Patient's education is the process of providing information and instruction to patients and their families to help them understand and manage their health conditions.[1] This can include information about symptoms, treatment options, medications, and lifestyle changes that can improve their health. Patient's education can be provided in a variety of formats, such as written materials, videos, and interactive sessions with healthcare professionals. The goal of patient's education is to empower patients to take an active role in their own health care and improve their overall health outcomes.[2] Many physicians write their own article for dissemination among patients according to local needs. In contrast, many physicians find difficult to get time or write an article for the general population. Artificial intelligence (AI)-based software can help generate text for this purpose.

Chat Generative Pre-Trained Transformer (ChatGPT) is a conversational language model developed by OpenAI, an American artificial research laboratory. The software is based on GPT architecture, which is trained on a large dataset of conversational text.[3] When given a prompt, the model generates a response by predicting the next word in the sequence, one word at a time. The model uses the previously predicted words as context to inform the current prediction. Hence, this software can generate human-like text within seconds.[4]

Researchers around the world are hyped to use AI-generated text for their scientific

Address for correspondence:
*Dr. Himel Mondal,
Department of Physiology,
All India Institute of Medical
Science, Deoghar - 814 152,
Jharkhand, India.
E-mail: himelmkcg@gmail.com*

articles as it would write a paragraph of text in a minute, which is nearly impossible for a common researcher.[5] Some researchers are also including the ChatGPT as their coauthors as it is helping them to write the article as a coauthor helps. However, controversies are continuing about whether the name ChatGPT should be included as an author or gets only acknowledgment.[6]

The current version of the online software (version 9) is free for research purposes and preview. Anyone can generate text by asking a simple question or using only the keyword.[7] However, the text generated in the software may be sourced from texts that are from textbooks, magazines, scientific journals, or other Internet sources. Hence, it may be a risk to use those texts for the scientific article as we perceived.[3]

Hence, in this pilot study, we aimed to check the scientific accuracy of the texts, the number of words generated, the score of readability, and text similarity (i.e., text plagiarism) of patient education guides created by ChatGPT. The study result would help dermatologists to gain insight into the ChatGPT-generated text and its usability in a scientific article aimed at patients' education.

Materials and Methods

Ethics

This study was conducted on the Internet without involving any human research participants. Hence, according to the prevailing guidelines in the country, this study does not require any institutional ethics committee approval.

Type and settings

This was a cross-sectional observational study. This study was conducted on January 23, 2023. For accessing the Internet, we used a personal broadband connection and a personal computer (Asus VivoBook Max X541N).

Websites and software used

For generating text, we used the ChatGPT version 9 available at https://chat.openai.com/chat website.

For readability, we used the online calculator available at https://goodcalculators.com/flesch-kincaid-calculator website. This calculates the Flesch–Kincaid Grade Level (indicates the years of education required to understand the text) and the Flesch Reading Ease Score (the higher the score, the easier it is to read).[8] The calculator also provides information on the total number of words, total sentences, average words per sentence, and average syllables per word.

We have checked text similarity by Turnitin (https://www.turnitin.com) software, which is a subscription-based premium software. This uses the same algorithm as iThenticate, which is used by the majority of journals worldwide to screen the articles for text similarity of submitted articles.[9]

Data collection methods

First, we initiated a chat (a conversation) on the AI service (ChatGPT) with the input—"Common skin diseases." We then had a conversation by inputting a specific disease name. For example, we asked ChatGPT to write a patient education guide on "acne." After the completion of text generation, we copied the text on a notepad from the chat box for further analyses. The text was checked for the number of words, sentences, Flesch–Kincaid Grade Level, and Flesch Reading Ease Score. The similarity of text was also checked individually, and different sources were noted. The accuracy of the text was checked by the third author following the structure of observed learning outcomes (SOLO) taxonomy.[10] The study flow is shown in Figure 1. In the "X," we used individual disease names to generate a document for each disease.

Data analysis

Data were presented for observing the central tendencies in terms of mean, median, mode, standard deviation, and interquartile range. The data were checked for normal distribution, and statistical tests were chosen according to the distribution (normal—parametric test, nonnormal—nonparametric tests).[11] As the data were normally distributed, we used a one-sample *t*-test for comparing the observed value with a (hypothetical) reference value. For the Flesch Reading Ease Score, we used a value of 30 as calculated from a previously published article from India.[12] For overall text similarity, we used a reference value of 15%. Although there are debates and different
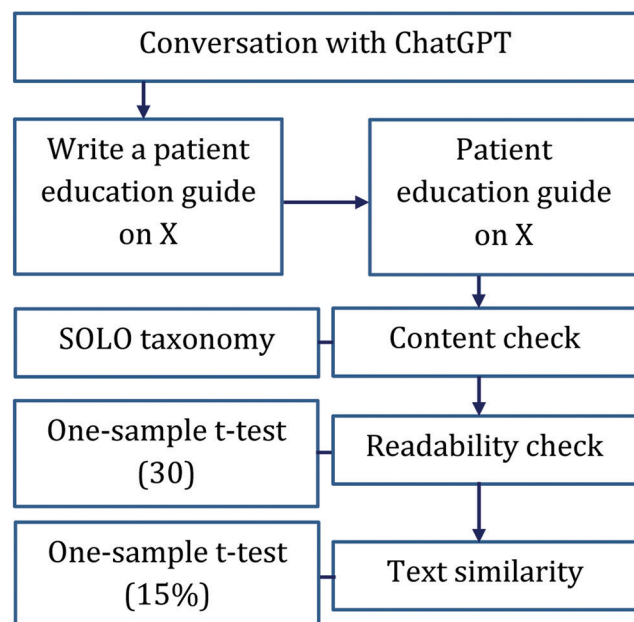

Figure 1: Data collection and analysis method (X is the individual disease name for each conversation, and the value in the bracket for *t*-test indicates the hypothetical values with which the data were compared)

journals have their right to select the percentage they tolerate,[13] a 15% text similarity is accepted by many of the journals.[14,15] We conducted Pearson's correlation coefficient to get the possible correlation between continuous variables. For all the tests, a *P* value < 0.05 was considered statistically significant. The statistical tests were conducted in GraphPad Prism 7 (GraphPad Software Inc., USA).

Results

The first conversation generated the names of 14 dermatological diseases or conditions with brief details about each, namely acne, atopic dermatitis, psoriasis, rosacea, shingles, vitiligo, warts, melanoma, basal cell carcinoma, squamous cell carcinoma, contact dermatitis, seborrheic dermatitis, ringworm, and cellulitis. A portion of the conversation is shown in Figure 2.

The ChatGPT generated patient education guide for common skin diseases with a paragraph of text of 377.43 ± 60.85 words [Table 1].

The average text reading ease score was 46.94 ± 8.23 [Table 1]. This score was significantly higher (one-sample *t*-test *P* < 0.0001) than the average reading ease score of an editorial in Indian journals (expected ease score of 30). This level of text can easily be understood by a high-school student to a newly joined college (grade level 10.53 ± 0.73) student without graduation.

The text similarity index was higher than the expected limit. There was an average of 27.07 ± 11.46% of text similarity, which is significantly (*P* = 0.002) higher than the commonly used cutoff by many of the journals (15%). We tested the data with a 20% reference value. In that case, also, there was significantly higher (one-sample *t*-test *P* = 0.038) text similarity.

There was no correlation of word count with ease score (*r* = 0.18, *P* = 0.54) and similarity index (*r* = -0.18, *P* = 0.55). The correlations are visually presented in Figure 3.

The textual content had a "relational" level of accuracy that indicates that the parts of the document are synthesized with the overall meaning of the content.

Discussion

We found that the current version of ChatGPT is capable of generating patient educational documents for common skin diseases. The text is easy to understand or interpret. A student with the tenth standard onward can easily read and understand the topic written by ChatGPT.[12] However, the text may have a high level of similarity (i.e., text plagiarism).
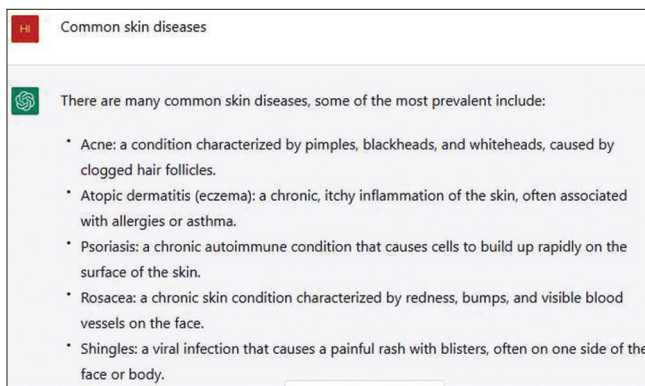
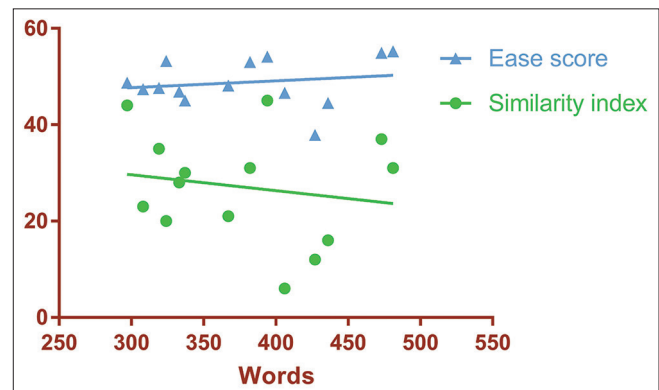Figure 2: Screenshot showing a portion of the conversation with the input—"Common skin diseases"

Figure 3: Correlation of the number of words with time, readability ease score, and text similarity index

Table 1: Characteristics of generated text (patient education guide on common skin diseases) from ChatGPT

Variable	Mean	Standard deviation	First quartile	Median	Third quartile	One-sample *t*-test
Sentences (*n*)	23.07	3.6	20.75	22	25.5	-
Words (*n*)	377.43	60.85	322.75	374.5	429.25	-
Words/sentence	16.39	1.31	15.35	16.25	17.78	-
Syllables/word	1.67	0.06	1.6	1.7	1.7	-
Grade level	10.53	0.73	10	10.4	10.95	-
Ease score	46.94	8.23	39.15	47.6	53.6	<0.0001*
Overall similarity (%)	27.07	11.46	19	29	35.5	0.002†
Internet (%)	19	11.39	6	18	31	-
Publication (%)	3.86	6.02	0	1	4.75	-
Students' paper (%)	18.79	9.22	13.75	20	23.5	-

-: Not required. *One-sample *t*-test by comparing with ease score of 30. †One-sample *t*-test by comparing with an overall similarity of 15%

ChatGPT and other AI-based systems are trained on large datasets of the text of existing written contents such as books, journal articles, and various websites. Because of this, there is a risk that the model may generate text that is similar to existing content.[16] Hence, it is important to note that the text generated by ChatGPT should be reviewed and edited by a human before publishing, to ensure that it is original and does not contain text plagiarism. In addition, when manual editing becomes cumbersome, for paraphrasing, the authors can use another tool named QuillBot (https://quillbot.com). This online software helps the author to paraphrase the text with some limitations in the free version. Furthermore, ChatGPT itself can help authors to paraphrase sentences. For that, a user needs to instruct the program to "paraphrase" with the text pasted on the chat box.

AI can help with scientific writing in a few ways. One way is through the use of natural language generation technology, which can automatically generate written text based on structured data. This could be used to generate reports, summaries, and other written materials from experimental data.[17] AI can also assist with language editing and grammar checking, helping authors improve the clarity and readability of their writing.[18] Grammarly (https://www.grammarly.com) is one such example. Additionally, AI-powered tools can help researchers organize and analyze large amounts of data, making it easier to identify patterns and draw conclusions.[19] ChatGPT has the capability to do the task of grammar correction at a glance. Users can command "correct grammar" with the text to be corrected.

AI-powered tools may help researchers to organize and analyze data, identify patterns and trends, and generate new hypotheses.[20] AI may also assist in the peer-review process of academic papers, and this could be done by providing a pre-review of the paper, flagging potential errors, or providing suggestions for improvements.[21] The Frontiers group uses such a program called artificial intelligence review assistant (AIRA) to initially screen their articles. However, whatever we get from AI should always be checked by a human for any flaw.[22]

This is a small-scale study to initially identify the issues that authors may face in using text generated from AI for their scientific article for patients' educational purposes. ChatGPT and other AI programs are evolving both for health care and academic writing. Hence, studies are needed further in due time to test newer versions and find their suitability for taking help in writing scientific articles. This study would be a reference for such studies in the future.

Conclusion

In its present form, the ChatGPT can generate an easily understandable paragraph of text for the patient education guide. The text can be read by a high-school student to a college student. Although the readability was satisfactory, the text similarity (i.e., text plagiarism) was higher than the tolerable level of text plagiarism by the majority of the scientific journals. Hence, writers who would like to take help from ChatGPT to get the text for creating an article for patients' education should use it judiciously. They should check for plagiarism and paraphrase if required.

Conflicts of interest

There are no conflicts of interest.

References

1. Paterick TE, Patel N, Tajik AJ, Chandrasekaran K. Improving health outcomes through patient education and partnerships with patients. Proc (Bayl Univ Med Cent) 2017;30:112-3.
2. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using Smartphone and Tablet Apps: Systematic review. J Med Internet Res 2020;22:e17342.
3. Gordijn B, Have HT. ChatGPT: Evolution or revolution? Med Health Care Philos 2023;26:1-2. epub ahead of print. doi: 10.1007/s11019-023-10136-0.
4. Else H. Abstracts written by ChatGPT fool scientists. Nature 2023;613:423.
5. Nazari N, Shabbir MS, Setiawan R. Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. Heliyon 2021;7:e07014.
6. Stokel-Walker C. ChatGPT listed as author on research papers: Many scientists disapprove. Nature 2023;613:620-1. doi: 10.1038/d41586-023-00107-z.
7. OpenAI's revolutionary chatbot 'ChatGPT': See what it is. Available from: https://economictimes.indiatimes.com/news/international/us/openais-revolutionary-chatbot-chatgpt-see-what-it-is/articleshow/96011256.cms?from=mdr [Last accessed on 24 January 2023].
8. Jindal P, MacDermid JC. Assessing reading levels of health information: Uses and limitations of flesch formula. Educ Health (Abingdon) 2017;30:84-8.
9. Arabyat RM, Qawasmeh BR, Al-Azzam SI, Nusair MB, Alzoubi KH. Faculty members' perceptions and attitudes towards anti-plagiarism detection tools: Applying the theory of planned behavior. J Empir Res Hum Res Ethics 2022;17:275-83.
10. Lucander H, Bondemark L, Brown G, Knutsson K. The structure of observed learning outcome (SOLO) taxonomy: A model to promote dental students' learning. Eur J Dent Educ 2010;14:145-50.
11. Mondal H, Mondal S, Majumder R, De R. Conduct common statistical tests online. Indian Dermatol Online J 2022;13:539-42.
12. Kumar KV, Aravinda K, Varadarajulu RN. The readability of

editorials in popular Indian medical journals. Indian J Endocrinol Metab 2013;17(Suppl 1):S363-6.

13. Memon AR. Similarity and plagiarism in scholarly journal submissions: Bringing clarity to the concept for authors, reviewers and editors. J Korean Med Sci 2020;35:e217.

14. Q: What is the acceptable percentage of plagiarism report?. https://www.editage.com/insights/what-is-the-acceptable-percentage-of-plagiarism-report [Last accessed on 24 January 2023].

15. Plagiarism Scores. What the Plagiarism Score Means in the Context of the Student's Academic Submission. https://www.scanmyessay.com/plagiarism/interpret-plagiarism-score.php#:~:text=Anything%20above%2015%25%20is%20indicative,a%20range%20of%20disciplinary%20measures [Last accessed on 24 January 2023].

16. Hern A. AI-assisted plagiarism? ChatGPT bot says it has an answer for that. https://www.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that [Last accessed on 24 January 2023].

17. Buchlak QD, Esmaili N, Bennett C, Farrokhi F. Natural language processing applications in the clinical neurosciences: A machine learning augmented systematic review. Acta Neurochir Suppl 2022;134:277-89.

18. Hammad M. The impact of artificial intelligence (AI) programs on writing scientific research. Ann Biomed Eng 2023;51:459-60. doi: 10.1007/s10439-023-03140-1.

19. Suarez JI. Big data/AI in neurocritical care: Maybe/Summary. Neurocrit Care 2022;37(Suppl 2):166-9.

20. Mondal S, Juhi A, Kumari A, Dhanvijay AK, Mittal S, Mondal H. Peer review in scientific publishing: Current practice, guidelines, relevancy, and way forward. CosmoDerma 2023;3:40.

21. Van Noorden R. The researchers using AI to analyse peer review. Nature 2022;609:455.

22. Artificial Intelligence to help meet global demand for high-quality, objective peer-review in publishing. https://blog.frontiersin.org/2020/07/01/artificial-intelligence-peer-review-assistant-aira/[Last accessed on 24 January 2023].