



Can Artificial Intelligence Improve the Readability of Patient Education Materials on Aortic Stenosis? A Pilot Study

Armaun D. Rouhi · Yazid K. Ghanem · Laman Yolchieva ·
Zena Saleh · Hansa Joshi · Matthew C. Moccia · Alejandro Suarez-Pierre ·
Jason J. Han

Received: November 5, 2023 / Accepted: December 13, 2023 / Published online: January 9, 2024
© The Author(s) 2024

ABSTRACT

Introduction: The advent of generative artificial intelligence (AI) dialogue platforms and large language models (LLMs) may help facilitate ongoing efforts to improve health literacy. Additionally, recent studies have highlighted inadequate health literacy among patients with cardiac disease. The aim of the present study was to ascertain whether two freely available generative AI dialogue platforms could rewrite online aortic stenosis (AS) patient education

materials (PEMs) to meet recommended reading skill levels for the public.

Methods: Online PEMs were gathered from a professional cardiothoracic surgical society and academic institutions in the USA. PEMs were then inputted into two AI-powered LLMs, ChatGPT-3.5 and Bard, with the prompt “translate to 5th-grade reading level”. Readability of PEMs before and after AI conversion was measured using the validated Flesch Reading Ease (FRE), Flesch-Kincaid Grade Level (FKGL), Simple Measure of Gobbledygook Index (SMOGI), and Gunning-Fog Index (GFI) scores.

Results: Overall, 21 PEMs on AS were gathered. Original readability measures indicated difficult readability at the 10th–12th grade reading level. ChatGPT-3.5 successfully improved readability across all four measures ($p < 0.001$) to the approximately 6th–7th grade reading level. Bard successfully improved readability across all measures ($p < 0.001$) except for SMOGI ($p = 0.729$) to the approximately 8th–9th grade level. Neither platform generated PEMs written below the recommended 6th-grade reading level. ChatGPT-3.5 demonstrated significantly more favorable post-conversion readability scores, percentage change in readability scores, and conversion time compared to Bard (all $p < 0.001$).

Conclusion: AI dialogue platforms can enhance the readability of PEMs for patients with AS but may not fully meet recommended reading skill

A. D. Rouhi
Department of Surgery, Perelman School of
Medicine, University of Pennsylvania, Philadelphia,
PA, USA

Y. K. Ghanem · Z. Saleh · H. Joshi · M. C. Moccia
Department of Surgery, Cooper University Hospital,
Camden, NJ, USA

L. Yolchieva
College of Arts and Sciences, University of
Pennsylvania, Philadelphia, PA, USA

A. Suarez-Pierre
Department of Surgery, University of Colorado
School of Medicine, Aurora, CO, USA

J. J. Han (✉)
Division of Cardiovascular Surgery, Department of
Surgery, Perelman School of Medicine, Hospital of
the University of Pennsylvania, Philadelphia, PA,
USA
e-mail: Jason.Han@penntermedicine.upenn.edu

levels, highlighting potential tools to help strengthen cardiac health literacy in the future.

Keywords: Aortic stenosis; Heart valve disease; Readability; Health literacy; Patient education material; ChatGPT; Artificial intelligence; Large language models; Chatbots

Key Summary Points

Why carry out this study?

Since their introduction in 2022, artificial intelligence dialogue platforms have influenced many different aspects of healthcare, including efforts to improve health literacy among patients.

The present study sought to use artificial intelligence dialogue platforms to rewrite patient education materials on aortic stenosis from leading academic institutions to meet recommended reading skill levels for the public.

What was learned from the study?

Two artificial intelligence chatbots, ChatGPT and Bard, successfully improved the readability of 21 patient education materials on aortic stenosis, although neither platform reached recommended 6th-grade reading skill levels.

Virtual assistants such as recent artificial intelligence platforms demonstrate potential to improve healthcare communication, rendering existing educational resources more accessible and understandable for patients.

INTRODUCTION

The advent of generative artificial intelligence (AI) platforms has dramatically influenced many aspects of modern medicine. Large language models (LLM), or models powered by generative AI platforms to understand and

generate human language, have recently been deployed for knowledge retrieval, clinical decision support, and documentation in health systems worldwide [1, 2]. AI-powered LLMs have also been proposed as potential tools to aid ongoing efforts in improving patient education materials (PEMs) and health literacy [3]. The American Medical Association (AMA) and National Institutes of Health (NIH) recommend PEMs be written at or below a 6th-grade reading skill level [4, 5]. Previous research on the readability of online patient resources for various cardiovascular pathologies and procedures has demonstrated that most online PEMs fail to adhere to these recommendations [6–8]. In fact, most online medical literature available to the general public is written at a high school or college student level, raising serious concerns related to patient decision-making and education [9, 10].

Aortic stenosis (AS) is a prevalent valvular disease in the elderly population, with the yearly incidence of AS among patients in the USA and Europe estimated to be 4.4% [11]. Left untreated, AS can be fatal within a few years of symptom onset, which highlights the importance of improving inadequate health literacy among patients with cardiac health conditions [12]. As such, the aim of the present study was to ascertain whether two freely available, widely used AI dialogue platforms can rewrite existing PEMs on AS to adhere to reading skill level recommendations while retaining accuracy of medical content.

METHODS

Adapting previous methods by Kirchner et al., online PEMs pertaining to AS were collected on September 10, 2023 on a private browsing window on Google Chrome (version 116.0.5845.179) via web searches of the top 20 leading academic cardiology, heart, and vascular surgery institutions in the USA as per the US News and World Report (USNWR) 2023 hospital rankings [13, 14]. Additionally, PEMs on AS were gathered from available online patient resources provided by a professional cardiothoracic surgical society. Browser history and cache

were cleared prior to web searches. PEMs that were video-based or already written at or below the 6th-grade reading skill level according to at least two of four utilized readability measures were excluded from analysis. The present study was exempt from institutional review as it did not involve human subjects research. This article does not contain any new studies with human participants or animals performed by any of the authors.

Once collected, online PEMs were reviewed for patient education descriptions of AS through group discussion by multiple investigators. Once identified, these patient education descriptions were assessed for readability via an online application (<https://readable.com>, Added Bytes Ltd., Brighton, England) using four validated readability measures: Flesch Reading Ease (FRE) score, Flesch-Kincaid Grade Level (FKGL), Gunning Fog Index (GFI), and Simple Measure of Gobbledygook Index (SMOGI) [15–18]. These measures are the most commonly used readability tests in the health literacy literature, and are calculated using the following formulas [9]:

$$\text{FRE score} = 206.835 - 1.015 \left(\frac{\text{total number of words}}{\text{total number of sentences}} \right) - 84.6 \left(\frac{\text{total number of syllables}}{\text{total number of words}} \right)$$

$$\text{FKGL} = 0.39 \left(\frac{\text{total number of words}}{\text{total number of sentences}} \right) + 11.8 \left(\frac{\text{total number of syllables}}{\text{total number of words}} \right) - 15.59$$

$$\text{GFI} = 0.4 \left[\left(\frac{\text{total number of words}}{\text{total number of sentences}} \right) + 100 \left(\frac{\text{total number of complex words}}{\text{total number of words}} \right) \right]$$

$$\text{SMOGI} = 3 + \sqrt{\text{polysyllable word count}}$$

The FRE score ranges from 0 to 100 and corresponds to an American educational level, with higher scores indicating easier reading material (100–90, very easy to read/5th-grade

level; 90–80, easy to read or conversational English/6th-grade level; 80–70, fairly easy to read/7th-grade level; 70–60, plain English/8th- to 9th-grade level; 60–50, fairly difficult/10th- to 12th-grade level; 50–30, difficult/college level; 30–10, very difficult/college graduate level; 10–0, extremely difficult/professional level) [15]. FKGL, GFI, and SMOGI range from 0 to 18, 0 to 20, and 5 to 18, respectively, and each score indicates the number of years of education necessary to understand the assessed reading material [16–18]. Thus, PEMs with FKGL, GFI, and SMOGI scores of approximately 7 or lower, and FRE scores greater than 80, are generally considered in congruence with AMA and NIH readability recommendations [4, 5]. The FRE and SMOGI measures have previously been highlighted for their utility in patient education and health literacy research [9, 19].

Next, patient education descriptions of AS from each PEM were entered into the freely available LLMs ChatGPT-3.5 (<https://chat.openai.com/chat>; Version August 3, 2023, OpenAI, San Francisco, CA, USA) and Bard (<https://bard.google.com>; Version July 13, 2023, Google, Mountain View, CA, USA), preceded by the prompt “translate to 5th-grade reading level”. These AI dialogue platforms were prompted to translate text to a lower reading level than recommended by the AMA and NIH to account for potential variability in the interpretation of reading skill levels by the LLMs. The AI-generated text material from ChatGPT-3.5 and Bard was then reevaluated for readability and accuracy.

Our primary endpoint was the absolute difference in FRE, FKGL, GFI, and SMOGI scores of each PEM before and after conversion by both AI platforms. To evaluate conversion consistency, four independent conversions of PEMs by each AI dialogue platform were performed [13]. Mean readability scores were recorded and used for subsequent analysis. Percentage change in readability scores before and after AI conversion was calculated. The accuracy of medical content from converted PEMs was secondarily assessed through independent review by multiple investigators. Time (seconds) elapsed between original prompting of each AI platform

and completion of text generation was also recorded for each PEM.

Continuous variables were presented as means \pm standard deviation (SD) or medians with interquartile range if non-parametric. The normality of variables was assessed using quantile–quantile plots and Shapiro–Wilk testing. Student's *t* tests or Wilcoxon rank-sum tests were subsequently used, as appropriate, to compare text characteristics, FRE, FKGL, GFI, and SMOGI scores before and after conversion, as well as percentage change in readability scores by each AI platform. A two-sided *p* value ≤ 0.05 was considered statistically significant. Statistical analyses were performed using Stata version 18.0 (StataCorp, College Station, TX, USA).

RESULTS

Overall, 21 online PEMs on AS were gathered from USNWR-ranked institutions and a professional cardiothoracic surgical society. All 21 original PEMs were written above the 6th-grade reading level and were included for analysis. Mean (SD) FRE, FKGL, GFI, and SMOGI scores of original PEMs were 51.9 (11.6), 9.4 (2.0), 11.3 (2.1), 10.1 (1.2), respectively, indicating difficult readability at the 10th–12th grade reading level (Table 1). Each PEM was subsequently converted by two AI dialogue platforms without system error or the need for additional direction by the end user (Table 2).

Conversion of PEMs by ChatGPT-3.5 demonstrated significantly improved mean readability scores across all four validated measures: FRE (76.9 ± 1.2 vs 51.9 ± 11.6 , $p < 0.001$), FKGL (5.9 ± 0.8 vs 9.4 ± 2.0 , $p < 0.001$), GFI (7.7 ± 0.9 vs 11.3 ± 2.1 , $p < 0.001$), and SMOGI (8.8 ± 0.6 vs 10.1 ± 1.2 , $p < 0.001$) scores, indicating fairly easy readability at the approximately 6th–7th grade reading level.

Conversion of PEMs by Bard resulted in significantly improved mean readability scores among three of four validated measures: FRE (66.5 ± 5.2 vs 51.9 ± 11.6 , $p < 0.001$), FKGL (6.9 ± 0.8 vs 9.4 ± 2.0 , $p < 0.001$), and GFI (9.3 ± 1.0 vs 11.3 ± 2.1 , $p < 0.001$) scores,

indicating plain English readability at the approximately 8th–9th grade reading level. However, conversion by Bard did not show significant improvement in mean SMOGI (10.0 ± 0.8 vs 10.1 ± 1.2 , $p = 0.729$) scores.

Compared to Bard, ChatGPT-3.5 demonstrated significantly easier readability after conversion as per FRE (76.9 ± 1.2 [ChatGPT-3.5] vs 66.5 ± 5.2 [Bard], $p < 0.001$), FKGL (5.9 ± 0.8 [ChatGPT-3.5] vs 6.9 ± 0.8 [Bard], $p = 0.001$), GFI (7.7 ± 0.9 [ChatGPT-3.5] vs 9.3 ± 1.0 [Bard], $p < 0.001$), and SMOGI (8.8 ± 0.6 [ChatGPT-3.5] vs 10.0 ± 0.8 [Bard], $p < 0.001$) scores (Fig. 1). Additionally, ChatGPT-3.5 showed significantly greater mean percentage change across all four readability measures compared to Bard (Table 3). ChatGPT-3.5 also had a significantly lower mean time (seconds) to generate converted PEMs than Bard ($5.38 \text{ s} \pm 0.67 \text{ s}$ vs $8.57 \text{ s} \pm 1.29 \text{ s}$, $p < 0.001$).

After an independent review by multiple investigators of each original PEM and the AI-generated material, there were no factual errors or inaccuracies identified.

DISCUSSION

Since their introduction in November 2022, LLMs such as ChatGPT-3.5 and Bard have been utilized throughout medical education, research, and practice, with over 50 recent studies describing the great potential of these emerging platforms in healthcare [20]. In cardiology and cardiothoracic surgery, others have described the potential benefits of AI as an adjunct to clinical decision-making and diagnosis [21–23]. No previous studies to our knowledge have evaluated the utility of AI dialogue platforms in improving the readability of existing PEMs on AS. In this pilot study, we assessed the ability of ChatGPT-3.5 and Bard to rewrite PEMs to meet recommended reading skill levels for patients with AS, finding varied performance between the two AI platforms. While ChatGPT-3.5 improved readability across all four validated scores, ChatGPT-3.5 was only able to reach recommended reading skill levels among FRE and FKGL measures. Moreover, Bard improved readability in all measures except for

SMOGI but was only able to reach recommended reading skill levels in the FKGL measure. Across post-conversion readability scores, overall percentage change in readability scores, and conversion time, ChatGPT-3.5 consistently demonstrated greater utility in improving the readability of PEMs compared to Bard.

Previous studies have similarly evaluated the utility of AI chatbots in improving the readability of cardiac PEMs [24]. Moons et al. prompted ChatGPT-3.5 and Bard to reduce the reading levels of cardiac PEMs from the *Journal*

of the American Medical Association, Cochrane Library, and the *European Journal of Cardiovascular Nursing* [24]. In congruence with our findings, researchers found that while both ChatGPT-3.5 and Bard significantly improved the readability of selected PEMs, neither platform was able to reach the recommended 6th-grade reading skill level [24]. It is not surprising that these chatbots encountered difficulty in reaching lower reading levels, as the platforms themselves are designed to generate outputs written at the reading level of an average high

Table 1 Readability scores of original patient education materials on aortic stenosis prior to conversion by artificial intelligence platforms

Resource #	FRE	FKGL	GFI	SMOGI
1	54.9 (fairly difficult to read)	9.1 (9th grade)	11.7 (11th grade)	11.9 (11th grade)
2	62.1 (plain English)	7.3 (7th grade)	9.1 (9th grade)	8.3 (8th grade)
3	48.5 (difficult to read)	8.7 (8th grade)	10.5 (10th grade)	11.0 (11th grade)
4	57.8 (fairly difficult to read)	7.7 (7th grade)	9.6 (9th grade)	9.7 (9th grade)
5	56.6 (fairly difficult to read)	9.0 (9th grade)	10.5 (10th grade)	11.2 (11th grade)
6	73.6 (fairly easy to read)	7.1 (7th grade)	8.8 (8th grade)	10.9 (10th grade)
7	58.8 (fairly difficult to read)	9.6 (9th grade)	12.1 (12th grade)	9.2 (9th grade)
8	42.7 (difficult to read)	9.8 (9th grade)	13.9 (college)	10.6 (10th grade)
9	46.8 (difficult to read)	10.2 (10th grade)	12.9 (12th grade)	11.6 (11th grade)
10	34.3 (difficult to read)	12.5 (12th grade)	15.4 (college)	10.2 (10th grade)
11	45.3 (difficult to read)	9.8 (9th grade)	11.8 (11th grade)	9.8 (9th grade)
12	34.9 (difficult to read)	14.5 (college)	12.4 (12th grade)	10.0 (10th grade)
13	60.8 (plain English)	8.1 (8th grade)	10.0 (10th grade)	11.8 (11th grade)
14	51.2 (fairly difficult to read)	10.3 (10th grade)	12.7 (12th grade)	11.0 (11th grade)
15	39.3 (difficult to read)	12.0 (12th grade)	14.4 (college)	10.1 (10th grade)
16	57.1 (fairly difficult to read)	7.9 (7th grade)	10.0 (10th grade)	9.1 (9th grade)
17	62.6 (plain English)	7.5 (7th grade)	10.2 (10th grade)	9.8 (9th grade)
18	59.4 (fairly difficult to read)	9.1 (9th grade)	13.1 (college)	11.0 (11th grade)
19	47.9 (difficult to read)	8.2 (8th grade)	7.8 (7th grade)	8.5 (8th grade)
20	27.7 (very difficult to read)	11.9 (11th grade)	13.0 (college)	8.0 (8th grade)
21	65.9 (plain English)	6.8 (6th grade)	8.3 (8th grade)	8.4 (8th grade)

FKGL Fleisch–Kincaid Grade Level, *FRE* Fleisch Reading Ease, *GFI* Gunning–Fog Index, *SMOGI* Simple Measure of Gobbledygook Index

Table 2 Samples of artificial intelligence-generated patient education materials on aortic stenosis

Sample #	Converted description
1	<p>Aortic stenosis is a problem with a special door in your heart called the aortic valve. This door helps blood flow from one part of your heart to another</p> <p>When this door doesn't work right, it makes your heart work extra hard to pump blood. This can make the walls of your heart become thicker and stiff, which isn't good for your heart's job</p> <p>There are a few reasons why this door can have problems. Sometimes it's just how it formed when you were born, other times it's because of an illness you had in the past, or it could be because you're getting older</p> <p>When this door isn't working well, you might feel chest pain, faint when you're active, or even have heart problems. Doctors use different tests like looking at pictures of your heart, listening to it, and checking your blood to figure out what's wrong</p> <p>To fix it, they might give you new medicine, make the door bigger with a balloon, or replace it with a new one. This helps you feel better and keeps your heart healthy. They might even have you do special exercises to help your heart get stronger</p>
2	<p>Aortic valve stenosis is when the valve in your heart that lets blood flow out gets narrow. This narrowing can happen because of problems with the valve or just as you get older</p> <p>Symptoms</p> <p>Sometimes, when this valve gets super narrow, you might feel chest pain or even pass out when you're active. The symptoms depend on what's causing the blockage</p> <p>Sometimes, a person's aortic valve gets blocked when they're grown-ups. But it's not very common. More rarely, this valve problem happens when someone's a baby, and it can really block the blood flow out of the heart and even lead to heart failure</p> <p>Most of the time, symptoms show up when you're a bit older, usually in childhood. You might have a blocked aortic valve and not enough oxygen, but you might not notice any other symptoms. Doctors might hear a strange sound in your heart or notice unusual sounds when the valve closes. Sometimes, this condition can lead to an infection in the heart</p> <p>Causes and risk factors</p> <p>The narrowing can happen in different parts of the valve, either above, below, or right in the valve itself. Sometimes, too much calcium in the blood can make blockages above the valve, along with something called peripheral pulmonary artery stenosis. Blockages below the valve can be because of a buildup of tissue fibers or muscle</p> <p>Diagnosis</p> <p>To figure out what's going on, doctors might use tests like an electrocardiogram (ECG), which can show changes in the left ventricle's size. X-rays can also help by showing a wider aorta or an enlarged left ventricle, especially if the blockage has been there for a while. If there are signs of severe blockage or symptoms like chest pain during exercise or fainting, doctors might use echocardiography, cardiac catheterization, and angiography to take a closer look</p>

Converted description refers to output provided by an artificial intelligence platform after prompted to "translate to 5th grade reading level"

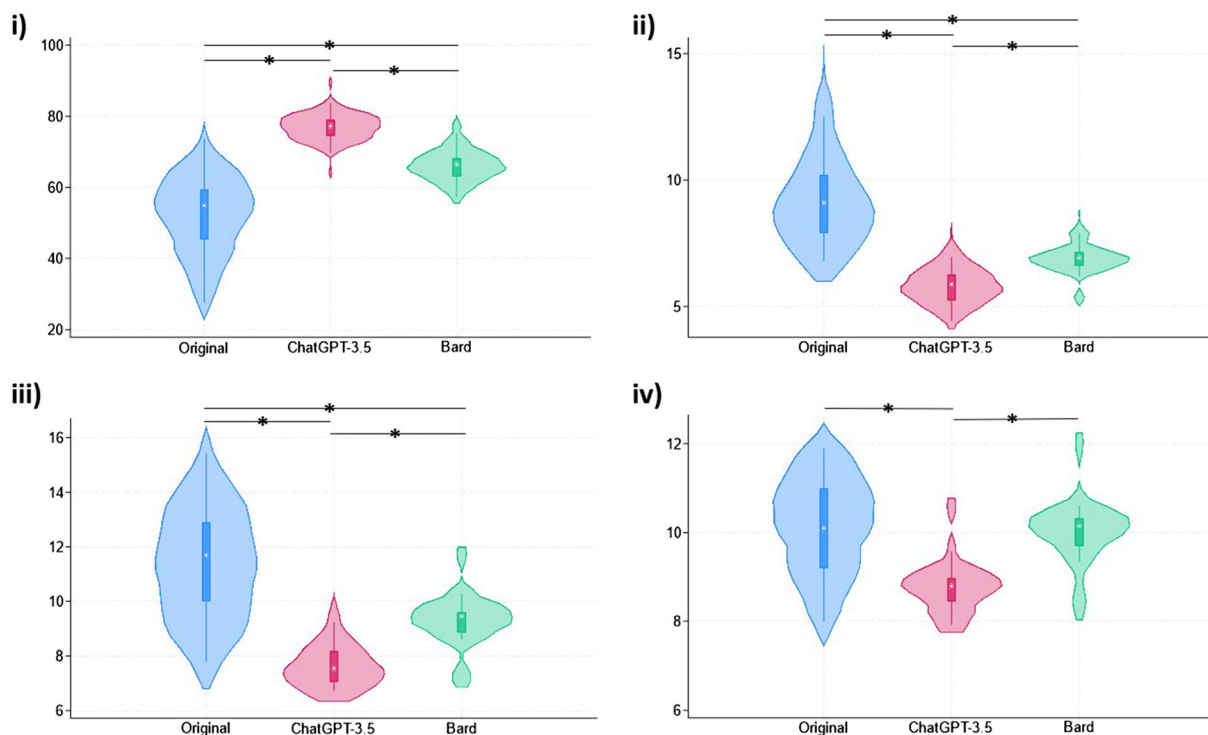


Fig. 1 Violin plots of (i) Flesch Reading Ease scores, (ii) Flesch-Kincaid Grade Level scores, (iii) Gunning-Fog Index scores, and (iv) Simple Measure of Gobbledygook

scores before and after conversion by each artificial intelligence platform. *Indicates significance ($p < 0.05$)

school student. Nevertheless, this may represent an inherent limitation of the present version of these freely available LLMs, although rapid updates and advancements in this AI technology may soon ameliorate the issue.

Despite the recommendations of the AMA and NIH, many previous studies in cardiology, cardiothoracic surgery, and vascular surgery have described most available PEMs to be widely incomprehensible to most of the public [4–8, 25, 26]. Similarly, the present study found all original PEMs on AS to be written above the recommended 6th-grade reading skill level, with most written for patients with at least a high school level of education. As low health literacy has been reported in up to 40% of patients with cardiac disease and is associated with lower rates of medication adherence and higher rates of rehospitalization, it is essential that healthcare providers prioritize the development and distribution of PEMs that are

accessible and understandable to members of the public, including those with lower health literacy [12, 27, 28].

Complex medical and surgical management of AS renders the task of accurately and comprehensibly conveying online patient literature uniquely difficult. This presents challenges across many treatment domains in twenty-first century healthcare. In a paradigm that supports shared decision-making, appropriate efforts should be taken to ensure optimal participation within this patient population as the degree of understanding and education required continues to become more sophisticated [29]. Generative AI dialogue platforms may help facilitate this process through the systematic development of comprehensible PEMs on a variety of cardiac pathologies and treatments.

While our analysis did not identify any factual errors in the AI-generated medical content, the variability in quality and accuracy of

Table 3 Percentage change in readability scores by each artificial intelligence platform

Measure	ChatGPT-3.5	Bard	<i>p</i> value
FRE	56.4 (39.8)	35.9 (39.6)	< 0.001*
FKGL	– 34.9 (17.0)	– 24.0 (14.7)	0.002*
GFI	– 29.0 (18.7)	– 15.7 (14.9)	< 0.001*
SMOGI	– 12.0 (12.5)	0.2 (10.9)	< 0.001*

Improved readability is indicated by positive percentage change in FRE scores and negative percentage change in FKGL, GFI, and SMOGI scores. All data are represented by mean (standard deviation)

*Indicates significance ($p < 0.05$)

FKGL Flesch–Kincaid Grade Level, *FRE* Flesch Reading Ease, *GFI* Gunning–Fog Index, *SMOGI* Simple Measure of Gobbledygook Index

medical information produced by ChatGPT-3.5 and other platforms has been well documented [30, 31]. A recent survey-based study presented 20 experts in congenital heart disease with AI-generated clinical vignettes and found the medical content produced by LLMs to often be incomplete and misleading [30]. Others have similarly questioned the trustworthiness and overall appropriateness of AI-generated responses to medical questions posed by patients [31, 32]. Ultimately, the potential use of AI dialogue platforms by patients to supplement medical decision-making underscores a pressing need for the involvement of regulators and healthcare professionals in establishing minimum quality standards and raising awareness among patients about the limitations of emerging AI chatbots.

The present study was not without limitations. First, the use of the USNWR to select the top 20 academic cardiology, heart, and vascular surgery institutions may limit the generalizability of our results. It is possible that PEMs from these institutions may vary in quality and readability from those at others throughout the USA. However, these major academic institutions were selected to capture a representative sample of PEMs on AS from leading centers. Second, the AI dialogue platforms only received one prompt in this study. Better phrases may exist to query LLMs regarding PEM readability improvement, although our approach was adapted from previously published literature [13]. Third, the present study utilized freely available versions of ChatGPT-3.5 and Bard,

whereas paid versions of these same AI platforms may include higher-level capabilities and functioning. Although it is possible that subscription-based versions may generate different results than those in the present study, ChatGPT-3.5 and Bard were ultimately chosen to evaluate freely accessible AI platforms that are most likely to be utilized by various stakeholders in patient engagement and health literacy [33]. Additionally, PEMs from countries outside the USA were not analyzed, which may also limit the generalizability of our results. Despite these limitations, the use of two popular AI dialogue platforms and the application of four validated readability measures to 21 PEMs from diverse sources strengthened the rigor of the present study.

CONCLUSION

AI dialogue platforms such as ChatGPT-3.5 and Bard demonstrate strong performance in improving the readability of PEMs on AS, although generated texts fall slightly short of reading skill level recommendations. On the basis of these findings, ChatGPT-3.5 may serve as a more useful tool over Bard to improve the readability of PEMs. As AI dialogue platforms rapidly advance and receive growing interest, virtual assistants may become increasingly capable of helping facilitate improvements in cardiovascular health literacy among members of the public.

Author Contributions. Conceptualization: Armaun D. Rouhi, Yazid K. Ghanem, Jason J. Han; Formal analysis and investigation: Armaun D. Rouhi, Yazid K. Ghanem, and Laman Yolchieva; Writing – original draft preparation: Armaun D. Rouhi, Yazid K. Ghanem, and Laman Yolchieva; Writing – review and editing: Armaun D. Rouhi, Yazid K. Ghanem, Laman Yolchieva, Zena Saleh, Hansa Joshi, Matthew C. Moccia, Alejandro Suarez-Pierre, and Jason J. Han. Supervision: Jason J. Han.

Funding. No funding or sponsorship was received for this study or publication of this article.

Data Availability. Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest. Armaun D. Rouhi, Yazid K. Ghanem, Laman Yolchieva, Zena Saleh, Hansa Joshi, Matthew C. Moccia, Alejandro Suarez-Pierre, and Jason J. Han declare that they have no competing interests.

Ethical Approval. The present study was exempt from institutional review as it did not involve human subjects research. This article does not contain any new studies with human participants or animals performed by any of the authors.

Open Access. This article is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative

Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. *JAMA*. 2023;330(9):866–9. <https://doi.org/10.1001/jama.2023.14217>.
2. Li R, Kumar A, Chen JH. How chatbots and large language model artificial intelligence systems will reshape modern medicine: fountain of creativity or Pandora's box? *JAMA Intern Med*. 2023;183(6):596–7. <https://doi.org/10.1001/jamainternmed.2023.1835>.
3. Goodman RS, Patrinely JR Jr, Osterman T, Wheless L, Johnson DB. On the cusp: considering the impact of artificial intelligence language models in healthcare. *Med*. 2023;4(3):139–40. <https://doi.org/10.1016/j.medj.2023.02.008>.
4. Weiss BD. Health literacy: a manual for clinicians. Chicago: American Medical Association Foundation and American Medical Association; 2003.
5. National Institutes of Health. How to write easy to read health materials. 2021. <http://www.nlm.nih.gov/medlineplus/etr.html>. Accessed 1 Nov 2023.
6. Brennan Z, Sharaf OM, Treffalls JA, et al. Quality analysis of publicly available information about hypoplastic left heart syndrome. *Cardiol Young*. 2023;33(7):1079–85. <https://doi.org/10.1017/S1047951123001294>.
7. Cajita MI, Rodney T, Xu J, Hladek M, Han HR. Quality and health literacy demand of online heart failure information. *J Cardiovasc Nurs*. 2017;32(2):156–64. <https://doi.org/10.1097/JCN.0000000000000324>.
8. Rouhi AD, Han JJ, Ghanem YK, et al. Quality and readability of online patient information on the left ventricular assist device. *Artif Organs*. 2023;47(6):1029–37. <https://doi.org/10.1111/aor.14479>.
9. Daraz L, Morrow AS, Ponce OJ, et al. Readability of online health information: a meta-narrative systematic review. *Am J Med Qual*. 2018;33(5):487–92. <https://doi.org/10.1177/1062860617751639>.

10. Diviani N, van den Putte B, Meppelink CS, van Weert JC. Exploring the role of health literacy in the evaluation of online health information: insights from a mixed-methods study. *Patient Educ Couns*. 2016;99(6):1017–25. <https://doi.org/10.1016/j.pec.2016.01.007>.
11. Durko AP, Osnabrugge RL, Van Mieghem NM, et al. Annual number of candidates for transcatheter aortic valve implantation per country: current estimates and future projections. *Eur Heart J*. 2018;39(28):2635–42. <https://doi.org/10.1093/eurheartj/ehy107>.
12. Magnani JW, Mujahid MS, Aronow HD, et al. Health literacy and cardiovascular disease: fundamental relevance to primary and secondary prevention: a scientific statement from the American Heart Association. *Circulation*. 2018;138(2):e48–74. <https://doi.org/10.1161/CIR.0000000000000579>.
13. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? *Clin Orthop Relat Res*. 2023. <https://doi.org/10.1097/CORR.0000000000002668>.
14. U.S. News & World Report. Best Hospitals: cardiology and heart surgery rankings. <https://health.usnews.com/best-hospitals/rankings/cardiology-and-heart-surgery>. Accessed Sept 2, 2023.
15. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221–33. <https://doi.org/10.1037/h0057532>.
16. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) for Navy Enlisted Personnel (1975). Institute for Simulation and Training. 56. <https://stars.library.ucf.edu/istlibrary/56>. Accessed 1 Nov 2023.
17. Gunning R. The technique of clear writing. New York: McGraw-Hill; 1952. p. 36–7.
18. McLaughlin GH. SMOG grading—a new readability formula. *J Read*. 1969;12(8):639–646. <http://www.jstor.org/stable/40011226>. Accessed 1 Nov 2023.
19. Grabeel KL, Russomanno J, Oelschlegel S, Tester E, Heidel RE. Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. *J Med Libr Assoc*. 2018;106(1):38–45. <https://doi.org/10.5195/jmla.2018.262>.
20. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. <https://doi.org/10.3390/healthcare11060887>.
21. Cardaioli F, Fovino LN, Tarantini G. Prime-time for Heart(ificial)-Team? *Am J Cardiol*. 2023;202:208–9. <https://doi.org/10.1016/j.amjcard.2023.06.085>.
22. Bart NK, Pepe S, Gregory AT, Denniss AR. Emerging roles of artificial intelligence (AI) in cardiology: benefits and barriers in a “Brave New World.” *Heart Lung Circ*. 2023;32(8):883–8. <https://doi.org/10.1016/j.hlc.2023.07.005>.
23. Seetharam K, Balla S, Bianco C, et al. Applications of machine learning in cardiology. *Cardiol Ther*. 2022;11(3):355–68. <https://doi.org/10.1007/s40119-022-00273-7>.
24. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof-of-concept. *Eur J Cardiovasc Nurs*. 2023. <https://doi.org/10.1093/eurjcn/zvad087>.
25. Rodriguez F, Ngo S, Baird G, Balla S, Miles R, Garg M. Readability of online patient educational materials for coronary artery calcium scans and implications for health disparities. *J Am Heart Assoc*. 2020;9(18): e017372. <https://doi.org/10.1161/JAHA.120.017372>.
26. Treffalls JA, Treffalls RN, Zachary H, et al. Quality analysis of online resources for patients with peripheral artery disease. *Ann Vasc Surg*. 2022;83: 1–9. <https://doi.org/10.1016/j.avsg.2021.12.079>.
27. Oscalices MIL, Okuno MFP, Lopes MCBT, Batista REA, Campanharo CRV. Health literacy and adherence to treatment of patients with heart failure. *Rev Esc Enferm USP*. 2019;53:e03447. <https://doi.org/10.1590/S1980-220X2017039803447>.
28. Cajita MI, Cajita TR, Han HR. Health literacy and heart failure: a systematic review. *J Cardiovasc Nurs*. 2016;31(2):121–30. <https://doi.org/10.1097/JCN.0000000000000229>.
29. Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, et al. Application of artificial intelligence in shared decision making: scoping review. *JMIR Med Inform*. 2022;10(8):e36199. <https://doi.org/10.2196/36199>.
30. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value and danger of ChatGPT-generated responses to health questions. *Eur J Cardiovasc Nurs*. 2023. <https://doi.org/10.1093/eurjcn/zvad038>.
31. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations

-
- obtained from a popular online chat-based artificial intelligence model. *JAMA*. 2023;329(10):842–4. <https://doi.org/10.1001/jama.2023.1044>.
32. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. *JMIR Hum Factors*. 2023;10: e47564. <https://doi.org/10.2196/47564>.
33. Scheschenja M, Viniol S, Bastian MB, Wessendorf J, König AM, Mahnken AH. Feasibility of GPT-3 and GPT-4 for in-depth patient education prior to interventional radiological procedures: a comparative analysis. *Cardiovasc Intervent Radiol*. 2023. <https://doi.org/10.1007/s00270-023-03563-2>.