Journal of Public Health
https://doi.org/10.1007/s10389-025-02643-6

ORIGINAL ARTICLE

Readability and source transparency of AI-generated health information on human metapneumovirus: A comparative evaluation of five chatbots

Jakub Brzeziński[1] · Klaudia Watros[1] · Małgorzata Mańczak[1] · Jakub Owoc[1] · Krzysztof Jeziorski[1] · Robert Olszewski[1,2]

Received: 17 July 2025 / Accepted: 15 October 2025

Abstract

Aim This study aimed to evaluate the readability and citation practices of artificial intelligence (AI)-generated responses to questions about human metapneumovirus, a respiratory virus of growing public health concern.

Subject and methods Five widely used AI chatbots—ChatGPT-4, Copilot, Gemini, Claude.ai, and Grok—were prompted with 14 standardized questions based on official guidelines from the World Health Organization, the Centers for Disease Control and Prevention, and the Australian National Health and Medical Research Council. Responses were anonymized and assessed using six established readability metrics: Flesch–Kincaid Reading Ease and Flesch–Kincaid Grade Level, Gunning Fog Index, SMOG (Simple Measure of Gobbledygook) Index, Coleman–Liau Index, and Automated Readability Index. Scores were compared to standards recommended by the American Medical Association and the National Institutes of Health. Citation frequency and credibility were also analyzed.

Results Among 70 chatbot responses, only one met the recommended readability level. Median readability scores ranged from grade 10.4 to 16.0, indicating high complexity. One chatbot generated the most readable content, while another scored lowest. Only two chatbots included source citations. One cited 68 reliable sources, primarily from health organizations and academic institutions, while the other referenced 31 sources of varying quality.

Conclusion AI-generated health content often exceeds recommended readability thresholds and lacks consistent citation practices. These issues may hinder understanding and trust. Improving default readability settings and integrating real-time citation features could enhance the accessibility and credibility of chatbot-based health communication.

Keywords Human metapneumovirus · Artificial intelligence · Chatbots · Readability · Health communication

Background

Human metapneumovirus (hMPV) is a negative-sense, single-stranded RNA virus belonging to the *Pneumoviridae* family and is closely related to avian metapneumovirus subgroup C (Jesse et al. 2022). Although it was first identified in the Netherlands in 2001, serological evidence indicates that it has been circulating among humans for over five decades (van den Hoogen et al. 2001). Clinically, hMPV is a significant cause of acute respiratory infections across

✉ Jakub Brzeziński
jakub.brzezinski@spartanska.pl

Klaudia Watros
klaudia.watros@spartanska.pl

Małgorzata Mańczak
m.manczak@op.pl

Jakub Owoc
kowoc@wp.pl

Krzysztof Jeziorski
krzysztof.jeziorski@spartanska.pl

Robert Olszewski
robert.olszewski@spartanska.pl

[1] Department of Gerontology and Public Health, National Institute of Geriatrics, Rheumatology and Rehabilitation, Spartańska 1 Street, Spartanska 1, 02-637 Warsaw, Poland

[2] Department of Ultrasound, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawińskiego 5B Street, Warsaw, Poland

Published online: 22 November 2025

all age groups, with infants, older adults, and immunocompromised individuals being particularly vulnerable (Akhras et al. 2010). Among children under 5 years, it ranks as the second most common cause of respiratory illness after respiratory syncytial virus (RSV) (Crowe and Williams 2014). Symptoms are similar to those of the common cold or influenza, including cough, fever, nasal congestion, and sore throat. In severe cases, the infection may progress to bronchitis or pneumonia (Costa-Filho et al. 2025).

The virus is primarily transmitted via respiratory droplets, direct contact, or contaminated surfaces. The incubation period is approximately 3 to 6 days (Leung 2021). Preventive measures include regular hand washing, surface disinfection, and wearing of masks in crowded areas (Jefferson et al. 2023). No specific antiviral treatment or vaccine exists, so management focuses on symptom relief (August et al. 2022).

Recent reports indicate a rise in hMPV cases in northern China in late 2024, especially among high-risk populations (European Centre for Disease Prevention and Control 2025). However, the World Health Organization (WHO) has stated that infection levels remain within seasonal expectations and have been declining since January 2025 (World Health Organization 2025). Notably, the rate of hMPV infections in northern China was reportedly declining as of January 2025 (The Washington Post 2025; The New York Times 2025).

Artificial intelligence (AI) chatbots are increasingly utilized to disseminate information across various domains, including healthcare (Labadze et al. 2023). These tools utilize machine learning algorithms to analyze and respond to user queries in real time (Lepakshi 2022). By leveraging medical literature, clinical guidelines, and epidemiological data, AI chatbots have the potential to disseminate accurate and timely information on emerging infections such as hMPV, especially in situations where access to human experts is limited.

This study aimed to evaluate the effectiveness of AI chatbots in conveying complex medical information about hMPV in an accessible and comprehensible format—an increasingly relevant issue as chatbots gain popularity in public health communication. Specifically, it evaluated the readability of chatbot-generated responses about hMPV, assessing their comprehensibility for laypersons without medical training.

Methodology

Five chatbots were selected for analysis: ChatGPT-4, Copilot, Gemini, Claude.ai, and Grok. The first four were chosen based on their established use in previous readability assessments, while Grok was included as a recent entrant with growing relevance and potential for broader adoption (Hancı et al. 2024). To minimize bias, the chatbot-generated responses were anonymized using alphabetical labels (Chatbots A–E). The researcher responsible for calculating the readability scores was blinded to the identity of the chatbot that generated each response. Each chatbot was prompted with 14 standardized questions derived from official recommendations issued by WHO, the Centers for Disease Control and Prevention (CDC), and the National Health and Medical Research Council (NHMRC) (World Health Organization 2025; Centers for Disease Control and Prevention 2025; National Health and Medical Research Council 2025). The selection of 14 questions was based on their relevance to public health guidelines and frequent appearance in patient-oriented health materials. Each question was selected from authoritative sources (WHO, CDC, NHMRC) to ensure both validity and practical relevance. The complete list of questions and their respective sources is provided in Supplemental Table 1.

Readability assessment

All chatbot responses were analyzed using the WebFX platform, a free online tool designed to assess text readability levels (https://www.webfx.com/tools/read-able/). The platform applies multiple readability formulas, including the Flesch–Kincaid Reading Ease, Flesch–Kincaid Grade Level, Gunning Fog Index, SMOG (Simple Measure of Gobbledygook) Index, Coleman–Liau Index, and Automated Readability Index. Additionally, the number of sentences in each response was recorded. These formulas assess textual

Table 1 Comparison of all chatbot responses

	Chatbot A	Chatbot B	Chatbot C	Chatbot D	Chatbot E	*p*
Number of sentences	20.5 (12–38)	8.5 (4–12)	14 (11–17)	15 (12–18)	18 (8–24)	<0.001
Flesch–Kincaid Reading Ease	30.1 (23.7–35.0)	36.4 (31.3–45.0)	42.3 (29.8–49.5)	25.8 (21.4–37.3)	34.5 (22.9–40.1)	0.006
Flesch–Kincaid Grade Level	13.0 (11.4–14.1)	12.7 (11.1–13.4)	11.5 (10.1–12.8)	15.1 (12.7–16.1)	10.9 (9.8–12.4)	0.005
Gunning Fog Index	14.2 (11.7–15.6)	14.5 (13.1–16.1)	13.3 (11.6–14.5)	16.4 (14.5–17.3)	13.2 (11.2–15.0)	0.033
SMOG Index	10.9 (9.2–11.7)	10.8 (9.4–11.6)	10.7 (9.2–10.9)	12.9 (11.2 -13.3)	9.5 (8.7–10.7)	0.001
Coleman–Liau Index	16.9 (15.8–18.5)	15.9 (15.5–17.7)	14.5 (13.5–16.5)	16.9 (15.4–18.7)	18.6 (17.5–20.0)	0.004
Automated Readability Index	13.1 (10.6–14.7)	13.6 (11.5–14.2)	10.4 (10.2–11.9)	16.0 (12.8–17.0)	12.1 (10.2–14.7)	0.001

complexity based on syntactic and lexical features. The readability scores were compared against the sixth-grade reading level recommended by the American Medical Association (AMA) and the National Institutes of Health (NIH) (Rooney 2021; Eltorai et al. 2014). According to established thresholds, a Flesch Reading Ease score (FRES) of ≥80.0 was considered acceptable, while a score of ≤7 was used as the benchmark for the remaining formulas (Kher et al. 2017). To improve comparability, we calculated the average readability scores across five scales with comparable numeric ranges (excluding FRES). These averages (Table 3) summarize each chatbot's overall readability performance.

Statistical analysis

The distribution of continuous variables was assessed using the Shapiro–Wilk test. None of the analyzed variables were normally distributed. Continuous variables are reported as medians with interquartile ranges (IQRs), and nominal variables are reported as counts and percentages. A comparison of the readability ratings of chatbots' responses was made using the Friedman analysis of variance (ANOVA) test. All analyses were performed using STATISTICA software (v. 13.1, Tulsa, OK, USA). A p-value of <0.05 was considered statistically significant.

Results

Five chatbots generated a total of 70 responses. Chatbot A generated the longest responses (median = 20.5 sentences; IQR, 12–38), whereas Chatbot B generated the shortest (median, 8.5; IQR, 4–12). According to the Flesch–Kincaid Reading Ease Scale, Chatbot E generated the least readable responses (25.8; IQR, 21.4–37.3), whereas Chatbot C produced the most readable ones (42.3; IQR, 29.8–49.5). On the Flesch–Kincaid Grade Level Scale, Chatbot D achieved the lowest score (10.9; IQR, 9.8–12.4), reflecting improved readability, whereas Chatbot E received the highest (15.1; IQR, 12.7–16.1), indicating lower readability. Similarly, the highest Gunning Fog score was assigned to Chatbot E, at 16.4 (IQR, 14.5–17.3), whereas Chatbot C received the lowest score, 13.3 (IQR, 11.6–14.5). The SMOG Index indicated that Chatbot E responses were the most difficult to understand (15.1; IQR, 12.7–16.1), while Chatbot D had the most readable ones (10.9; IQR, 9.8–12.4). In the Coleman–Liau Index, Chatbot C again ranked best (14.5; IQR, 13.5–16.5), while Chatbots A and E had the highest scores, indicating lower readability. Finally, in the Automated Readability Index, Chatbot C had the lowest score, 10.4 (10.2–11.9), whereas Chatbot E obtained the highest score at 16.0 (12.8–17.0). Table 1 presents a comparison of all chatbot responses.

Summary data are presented in Fig. 1 for comparative purposes, illustrating the performance of the five AI chatbots across six established metrics.

We also analyzed how many chatbot responses met the readability standards recommended by the AMA. Of the 70 responses analyzed, only one—generated by Chatbot D—complied with the AMA-recommended readability level of sixth to seventh grade (score 6.2). Table 2 summarizes the number of responses meeting AMA readability guidelines.

Notably, only one of the 70 responses met the AMA-recommended readability standards, underscoring a mismatch between AI-generated content and the health literacy levels of the general population. This discrepancy may hinder access to reliable health information, especially for individuals with low health literacy. Consequently, it could lead to misinformation, misinterpretation of guidelines, and reduced engagement with chatbot-delivered content. Thus, readability is not merely a stylistic concern—it has real-world implications for public health communication and equitable access to knowledge.

This suggests that chatbot responses often employ advanced, professional language structures, which may be difficult for individuals without the appropriate educational background to comprehend. The complexity of the language chatbots use often stems from their training on vast datasets, including academic and technical sources. As a result, their responses may incorporate specialized terminology, formal syntax, and nuanced phrasing that could pose challenges for users unfamiliar with the subject matter. While beneficial in professional and scientific contexts, this complex language may necessitate simplification or adaptation to ensure accessibility for a broader audience.

Since five of the six readability metrics share a similar scale and interpretation, and the differences between chatbots were relatively minor, we computed average scores to provide a clearer summary. The average scores indicate that Chatbot C achieved the best readability scores (12.22), followed by Chatbot D (13.18), Chatbot B (13.34), Chatbot A (13.50), and Chatbot E (15.13). We excluded the Flesch–Kincaid Reading Ease scale because of a different scoring range. Table 3 presents the average readability scores for each chatbot based on five comparable indices.

Only Chatbots B and C included source citations, referencing a total of 99 sources (31 from Chatbot B and 68 from Chatbot C). The references—including scientific publications, official health organizations, and academic institutions—were subsequently classified. Table 4 presents the number of sources included in chatbot responses.

After data collection, the cited sources were categorized based on their nature and credibility. A detailed analysis allowed for classification into the following groups: scientific publications, health organizations (e.g., WHO, CDC), medical society guidelines, government authorities, medical

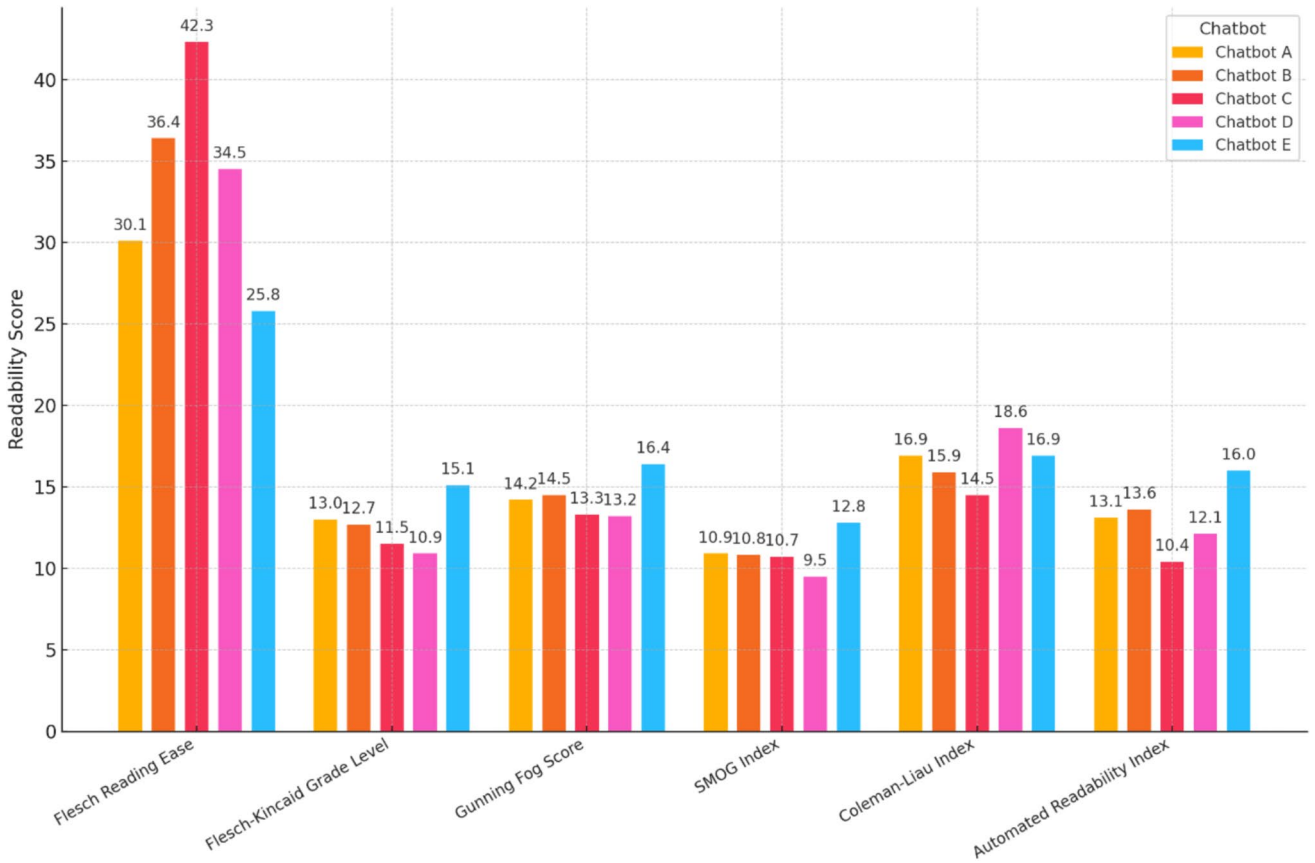

Fig. 1 Comparison of all chatbot responses

Table 2 Number of responses adhering to American Medical Association guidelines

	Chatbot A	Chatbot B	Chatbot C	Chatbot D	Chatbot E
Flesch–Kincaid Reading Ease ≥ 70	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Flesch–Kincaid Grade Level ≤ 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Gunning Fog score ≤ 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
SMOG Index ≤ 7	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)
Coleman–Liau Index ≤ 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Automated Readability Index ≤ 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 3 The average readability scores

Chatbot C	Chatbot D	Chatbot B	Chatbot A	Chatbot E
12.22	13.18	13.34	13.50	15.13

Table 4 The number of sources included in chatbot responses

Chatbot A	Chatbot B	Chatbot C	Chatbot D	Chatbot E
0	31	68	0	0

universities and hospitals, public health-related websites, sources linked to the pharmaceutical industry, and other online sources. Our classification aimed to evaluate the reliability and diversity of information provided by chatbots, with an emphasis on potential variations in source credibility and domain specificity. Table 5 categorizes the origin of cited sources in chatbot responses by type and frequency. Chatbot C most frequently referenced health organizations, with WHO and the CDC being the primary sources. The second most common sources cited by Chatbot C were medical universities and hospital websites.

In contrast, Chatbot B primarily cited websites related to public health, followed by information from health

Table 5 Origin of the sources cited in the chatbot responses

Source category	Gemini	Copilot
Scientific publications	9	1
Health organizations	31	10
Medical society guidelines	1	0
Websites related to public health	4	11
Sources associated with the pharmaceutical industry	4	1
Medical universities and hospitals	12	1
Government authorities	3	3
Other	4	4

organizations. These differences in citation patterns likely reflect varying information retrieval algorithms employed by the chatbots, which may affect both the reliability and domain relevance of their responses. Supplemental Table 2 in the Supplementary Materials categorizes the sources of information used in chatbot-generated responses, providing insights into their distribution and relative frequency. Sources used by Chatbot C included medical universities and hospital websites.

Discussion

To the best of our knowledge, this is the first study to assess how AI chatbots respond to questions concerning hMPV. While AI-driven chatbots can be valuable tools in health education, they are not substitutes for professional medical consultations, particularly in complex or personalized cases. Instead, these systems are best used as supplementary tools to enhance the public's understanding of viruses, such as hMPV, by providing structured and accessible information (ltamimi et al. 2023).

AI chatbots can help explain key aspects of hMPV, including virology, transmission, symptoms, and preventive measures. They may also summarize recent research and report regional outbreaks in accessible formats, especially during periods of increased infections. Prior studies have shown that AI-generated content often exceeds recommended readability levels, using complex language that diverges from AMA and NIH guidelines (Doğan et al. 2024; Warren et al. 2025; Chen et al. 2024; Cao et al. 2024; Nian et al. 2024; Yau et al. 2024; Olszewski et al. 2024; Meyer et al. 2024). Although some chatbots can meet these standards when prompted to simplify responses (Balta et al. 2025; Boscolo-Rizzo et al. 2025; Sharkiya 2023), their default outputs typically reflect training on technical sources. Nonetheless, these systems are capable of simplifying content when appropriately guided, offering potential for broader public engagement (Kooli 2023; Adamopoulou and Moussiades 2020). Improving readability through default simplification and user-adjustable reading levels could enhance accessibility and encourage more frequent use, especially among individuals with limited health literacy (Mennella et al. 2024; Du and Daniel 2024; Ömür Arça et al. 2024).

Chatbots such as ChatGPT, Copilot, Gemini, and Claude have been widely utilized in prior analyses. However, Grok, a recently introduced chatbot, may become increasingly relevant in future research on readability and health communication. Given its relatively recent introduction, further studies will be needed to assess how Grok compares to existing models in generating user-friendly responses. The ability of AI chatbots to provide timely and structured health-related information is valuable, particularly during outbreaks of infectious diseases. However, the results suggest that chatbot responses may require further refinement to ensure accessibility for a broader audience. Built-in readability tools—such as automatic simplification or user-adjustable reading levels—could significantly improve chatbot accessibility and usability.

This study also identified differences in information sourcing among chatbots. Only Copilot and Gemini consistently provided citations, enhancing the transparency and verifiability of their content. By contrast, the absence of citations in other chatbots limits transparency and trust. Future development should prioritize integrating verifiable sources and real-time citation retrieval systems. Another important consideration is the risk of hallucinations—instances where chatbots generate plausible but false information (Preiksaitis and Rose 2023; Bail 2024; McCoy et al. 2024; de Boer 2019). Although our focus was on readability, future evaluations should also monitor hallucination frequency, particularly in medical contexts where misinformation may compromise patient safety (Vermeir et al. 2015; Zarour et al. 2021).

Despite our emphasis on readability, we did not evaluate the clinical accuracy or currency of the responses—an essential factor in their safe use in healthcare. For end users—especially those relying on chatbots for health-related decisions—comprehensibility without accuracy can be misleading or even harmful. Misinformation, even when presented in simple language, can reinforce false beliefs or lead to inappropriate self-management. Prior studies have shown that some AI chatbots, while generally reliable, can occasionally produce outdated, incomplete, or contextually inappropriate recommendations. Future studies should incorporate dual assessment frameworks that evaluate both readability and factual correctness to ensure that chatbot-based communication meets the essential standards of clarity and clinical validity. Such research is essential to determine whether currently available models produce more straightforward text that aligns with or falls within the AMA guidelines, which recommend a readability level

equivalent to sixth to seventh grade. Such investigations will help establish whether AI-generated content is accessible to the general population and whether additional modifications are necessary to improve its comprehensibility and adherence to established readability standards. Future improvements should strive to balance clarity with clinical accuracy, ensuring essential nuance is preserved.

Strengths and limitations

While this study provides valuable insights into the readability of AI chatbot responses, several limitations should be acknowledged. First, the analysis was based on a relatively small set of 14 questions, which may limit the generalizability of the findings. Expanding the range of questions to cover additional hMPV-related topics could offer a more comprehensive evaluation. Second, the study did not assess the accuracy of chatbot responses, which is a critical factor in determining their utility in health education. Future research should investigate the accuracy and reliability of AI-generated information in conjunction with its readability and comprehensibility.

Additionally, all questions were posed in English, reflecting the predominant training language of the selected AI models. The study did not impose limits on the length of chatbot responses. Although response length was not restricted, it varied substantially across models, potentially influencing both readability scores and perceived informativeness. While our study focused on readability indices, future research could explore potential correlations between response length and readability outcomes to better understand how verbosity affects comprehension. This variability may influence both the measured readability and users' perception of informativeness, warranting more nuanced analyses in future chatbot evaluations. It may also be worth considering a brief analysis of inter-chatbot response variability, particularly whether differences in response length may have influenced the readability outcomes.

Despite these limitations, the study has several notable strengths. One is its focus on hMPV, a topic that has received increased attention in recent months. This relevance informed the decision to select hMPV as the thematic basis for chatbot queries.

Another strength lies in the inclusion of multiple AI chatbots, enabling a comprehensive comparative assessment. Notably, the study is among the first to evaluate responses generated by Grok, a newly released chatbot, adding originality and relevance to the research. A further strength is the comprehensive evaluation of readability. The study employed a full range of established readability scales, providing a robust and multidimensional assessment of textual complexity.

Finally, the reliability of the question set contributes to the study's credibility. All questions were derived from respected public health institutions known for disseminating validated information, particularly during pandemics. This foundation supports the methodological rigor of the analysis.

Conclusion

This study assessed the readability of responses generated by five widely used AI chatbots in the context of public health information related to hMPV. The findings demonstrate that none of the evaluated models consistently adhered to the readability thresholds recommended by the AMA, and only two systems provided source citations, raising concerns about accessibility and transparency.

Given the increasing integration of AI-driven tools into public health communication, developers must implement default readability-enhancing functionalities, including automated language simplification and user-adjustable complexity settings, to ensure effective communication. Such measures may improve the comprehensibility and trustworthiness of chatbot-generated outputs for nonspecialist audiences.

Future studies should assess both readability and factual accuracy to ensure that AI-generated health information is understandable and clinically reliable, particularly in contexts where misinformation may have significant implications for individual and population health outcomes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10389-025-02643-6.

Authors' contributions JB contributed to the conceptualization, methodology, investigation, resource acquisition, and drafting and revising the manuscript. RO contributed to the conceptualization, methodology, and drafting and revising the manuscript, and provided overall supervision. KW participated in the investigation and resource acquisition and contributed to the original draft. MM contributed to the methodology and formal analysis and participated in writing the original draft. JO contributed to the methodology, original drafting, and critical revision of the manuscript. KJ supervised the project, supported resource acquisition, and revised the manuscript critically. All authors read and approved the final version of the manuscript.

Funding This research received no external funding.

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Adamopoulou E, Moussiades L (2020) An overview of chatbot technology. Artif Intell Applic Innov 584:373–83. https://doi.org/10.1007/978-3-030-49186-4_31

Akhras N, Weinberg JB, Newton D (2010) Human metapneumovirus and respiratory syncytial virus: subtle differences but comparable severity. Infect Dis Rep 2(2):e12. https://doi.org/10.4081/idr.2010.e12

August A, Shaw CA, Lee H, Knightly C, Kalidindia S, Chu L, Essink BJ, Seger W, Zaks T, Smolenov I, Panther L (2022) Safety and immunogenicity of an mRNA-based human metapneumovirus and parainfluenza virus type 3 combined vaccine in healthy adults. Open Forum Infect Dis 9(7):ofac206. https://doi.org/10.1093/ofid/ofac206

Bail CA (2024) Can generative AI improve social science? Proc Natl Acad Sci U S A 121(21):e2314021121. https://doi.org/10.1073/pnas.2314021121

Balta KY, Javidan AP, Walser E, Arntfield R, Prager R (2025) Evaluating the appropriateness, consistency, and readability of ChatGPT in critical care recommendations. J Intensive Care Med 40(2):184–190. https://doi.org/10.1177/08850666241267871

Boscolo-Rizzo P, Marcuzzo AV, Lazzarin C, Giudici F, Polesel J, Stellin M, Pettorelli A, Spinato G, Ottaviano G, Ferrari M, Borsetto D, Zucchini S, Trabalzini F, Sia E, Gardenal N, Baruca R, Fortunati A, Vaira LA, Tirelli G (2025) Quality of information provided by artificial intelligence chatbots surrounding the reconstructive surgery for head and neck cancer: a comparative analysis between ChatGPT4 and Claude2. Clin Otolaryngol 50(2):330–335. https://doi.org/10.1111/coa.14261

Cao JJ, Kwon DH, Ghaziani TT, Kwo P, Tse G, Kesselman A, Kamaya A, Tse JR (2024) Large language models' responses to liver cancer surveillance, diagnosis, and management questions: accuracy, reliability, readability. Abdom Radiol (NY) 49(12):4286–4294. https://doi.org/10.1007/s00261-024-04501-7

Centers for Disease Control and Prevention (2025) About Human Metapneumovirus (hMPV) https://www.cdc.gov/human-metapneumovirus/about/index.html (Accessed: 18.01.2025)

Chen D, Parsa R, Hope A, Hannon B, Mak E, Eng L, Liu FF, Fallah-Rad N, Heesters AM, Raman S (2024) Physician and artificial intelligence chatbot responses to cancer questions from social media. JAMA Oncol 10(7):956–960. https://doi.org/10.1001/jamaoncol.2024.0836

Costa-Filho RC, Saddy F, Costa JLF, Tavares LR, Castro Faria Neto HC (2025) The silent threat of human metapneumovirus: clinical challenges and diagnostic insights from a severe pneumonia case. Microorganisms. 13(1):73. https://doi.org/10.3390/microorganisms13010073

Crowe JE Jr, Williams JV (2014) Paramyxoviruses: respiratory syncytial virus and human metapneumovirus. Viral Infect Humans:601–27. https://doi.org/10.1007/978-1-4899-7448-8_26

de Boer JN, Linszen MMJ, de Vries J, Schutte MJL, Begemann MJH, Heringa SM, Bohlken MM, Hugdahl K, Aleman A, Wijnen FNK, Sommer IEC (2019) Auditory hallucinations, top-down processing and language perception: a general population study. Psychol Med 49(16):2772–2780. https://doi.org/10.1017/S0033291719001800380X

Doğan L, Özer Özcan Z, Edhem Yılmaz I (2024) The promising role of chatbots in keratorefractive surgery patient education. J Fr Ophtalmol 48(2):104381. https://doi.org/10.1016/j.jfo.2024.104381

Du J, Daniel BK (2024) Transforming language education: a systematic review of AI-powered chatbots for English as a foreign language speaking practice. Comput Educ Art Intell 6:100230. https://doi.org/10.1016/j.caeai.2024.100230

Eltorai AE, Ghanian S, Adams CA Jr, Born CT, Daniels AH (2014) Readability of patient education materials on the American Association for Surgery of Trauma website. Arch Trauma Res 3(2):e18161. https://doi.org/10.5812/atr.18161

European Centre for Disease Prevention and Control (2025) Increase in respiratory infections in China https://www.ecdc.europa.eu/en/news-events/increase-respiratory-infections-china (Accessed: 21.01.2025)

Hancı V, Ergün B, Gül Ş, Uzun Ö, Erdemir İ, Hancı FB (2024) Assessment of readability, reliability, and quality of ChatGPT®, BARD®, Gemini®, Copilot®, Perplexity® responses on palliative care. Medicine (Baltimore) 103(33):e39305. https://doi.org/10.1097/MD.0000000000039305

https://www.who.int/westernpacific/about/how-we-work/pacific-support/news/detail/16-01-2025-who-advisory-on-trends-of-acute-respiratory-infection--including-human metapneumovirus (2025)

Itamimi I, Altamimi A, Alhumimidi A, Temsah MH (2023) Artificial intelligence (AI) chatbots in medicine: a supplement, not a substitute. Cureus 15(6):e40922. https://doi.org/10.7759/cureus.40922

Jefferson T, Dooley L, Ferroni E, Al-Ansary LA, van Driel ML, Bawazeer GA, Jones MA, Hoffmann TC, Clark J, Beller EM, Glasziou PP, Conly JM (2023) Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst Rev 1(1):CD006207. https://doi.org/10.1002/14651858.CD006207.pub6

Jesse ST, Ludlow M, Osterhaus ADME (2022) Zoonotic origins of human metapneumovirus: a journey from birds to humans. Viruses 14(4):677. https://doi.org/10.3390/v14040677

Kher A, Johnson S, Griffith R (2017) Readability assessment of online patient education material on congestive heart failure. Adv Prev Med 2017:9780317. https://doi.org/10.1155/2017/9780317

Kooli C (2023) Chatbots in education and research: a critical examination of ethical implications and solutions. Sustainability 15(7):5614. https://doi.org/10.3390/su15075614

Labadze L, Grigolia M, Machaidze L (2023) Role of AI chatbots in education: systematic literature review. Int J Educ Technol High Educ 20:56. https://doi.org/10.1186/s41239-023-00426-1

Lepakshi VA (2022) Machine learning and deep learning based AI Tools for development of diagnostic tools. computational approaches for novel therapeutic and diagnostic designing to mitigate SARS-CoV-2. Infection:399–420. https://doi.org/10.1016/B978-0-323-91172-6.00011-X.

Leung NHL (2021) Transmissibility and transmission of respiratory viruses. Nat Rev Microbiol 19(8):528–545. https://doi.org/10.1038/s41579-021-00535-6

McCoy LG, Ci Ng FY, Sauer CM, Yap Legaspi KE, Jain B, Gallifant J, McClurkin M, Hammond A, Goode D, Gichoya J, Celi LA (2024) Understanding and training for the impact of large

language models and artificial intelligence in healthcare practice: a narrative review. BMC Med Educ 24(1):1096. https://doi.org/10.1186/s12909-024-06048-z

Mennella C, Maniscalco U, De Pietro G, Esposito M (2024) Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon 10(4):e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

Meyer MKR, Kandathil CK, Davis SJ, Durairaj KK, Patel PN, Pepper JP, Spataro EA, Most SP (2024) Evaluation of rhinoplasty information from ChatGPT, Gemini, and Claude for readability and accuracy. Aesthet Plast Surg. https://doi.org/10.1007/s00266-024-04343-0

National Health and Medical Research Council (2025). Staying Healthy Guidelines – Fact sheet: Human Metapneumovirus. https://www.nhmrc.gov.au/about-us/publications/staying-healthy-guidelines/fact-sheets/human-metapneumovirus (Accessed: 18.01.2025)

Nian PP, Saleet J, Magruder M, Wellington IJ, Choueka J, Houten JK, Saleh A, Razi AE, Ng MK (2024) ChatGPT as a source of patient information for lumbar spinal fusion and laminectomy: a comparative analysis against Google web search. Clin Spine Surg 37(10):E394–E403. https://doi.org/10.1097/BSD.0000000000001582

Olszewski R, Watros K, Mańczak M, Owoc J, Jeziorski K, Brzeziński J (2024) Assessing the response quality and readability of chatbots in cardiovascular health, oncology, and psoriasis: a comparative study. Int J Med Inform 190:105562. https://doi.org/10.1016/j.ijmedinf.2024.105562

Ömür Arça D, Erdemir İ, Kara F, Shermatov N, Odacioğlu M, İbişoğlu E, Hanci FB, Sağiroğlu G, Hanci V (2024) Assessing the readability, reliability, and quality of artificial intelligence chatbot responses to the 100 most searched queries about cardiopulmonary resuscitation: an observational study. Medicine (Baltimore) 103(22):e38352. https://doi.org/10.1097/MD.0000000000038352

Preiksaitis C, Rose C (2023) Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Med Educ 9:e48785. https://doi.org/10.2196/48785

Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, Jagsi R, Golden DW (2021) Readability of patient education materials from high-impact medical journals: a 20-year analysis. J Patient Exp 8:2374373521998847. https://doi.org/10.1177/2374373521998847

Sharkiya SH (2023) Quality communication can improve patient-centred health outcomes among older patients: a rapid review. BMC Health Serv Res 23:886. https://doi.org/10.1186/s12913-023-09869-8

The New York Times (2025). What is hMPV? Experts explain the virus surging in China. https://www.nytimes.com/2025/01/07/health/hmpv-virus-china.html (Accessed: 21.01.2025)

The Washington Post (2025) China's respiratory outbreak draws attention to human metapneumovirus. https://www.washingtonpost.com/world/2025/01/12/china-hmpv-outbreak-health/347cb00c-d0d0-11ef-9835-51843d9371d6_story.html (Accessed: 21.01.2025)

van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA, Osterhaus AD (2001) A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat Med 7(6):719–24. https://doi.org/10.1038/89098

Vermeir P, Vandijck D, Degroote S, Peleman R, Verhaeghe R, Mortier E, Hallaert G, Van Daele S, Buylaert W, Vogelaers D (2015) Communication in healthcare: a narrative review of the literature and practical recommendations. Int J Clin Pract 69(11):1257–67. https://doi.org/10.1111/ijcp.12686

Warren CJ, Payne NG, Edmonds VS, Voleti SS, Choudry MM, Punjani N, Abdul-Muhsin HM, Humphreys MR (2025) Quality of chatbot information related to benign prostatic hyperplasia. Prostate 85(2):175–180. https://doi.org/10.1002/pros.24814

WebFX (2025) Readability Test Tool. https://www.webfx.com/tools/read-able/ (Accessed: 21.01.2025)

World Health Organization (2025) Human metapneumovirus (hMPV) infection – Questions and Answers. https://www.who.int/news-room/questions-and-answers/item/human-metapneumovirus-(hmpv)-infection (Accessed: 18.01.2025)

Yau JY, Saadat S, Hsu E, Murphy LS, Roh JS, Suchard J, Tapia A, Wiechmann W, Langdorf MI (2024) Accuracy of prospective assessments of 4 large language model chatbot responses to patient questions about emergency care: experimental comparative study. J Med Internet Res 26:e60291. https://doi.org/10.2196/60291

Zarour M, Alenezi M, Ansari MTJ, Pandey AK, Ahmad M, Agrawal A, Kumar R, Khan RA (2021) Ensuring data integrity of healthcare information in the era of digital health. Healthc Technol Lett 8(3):66–77. https://doi.org/10.1049/htl2.12008