






Preliminary exploration of ChatGPT-4 shows the potential of generative artificial intelligence for culturally tailored, multilingual antimicrobial resistance awareness messaging

Oluwaseun Akinyede, DVM, MPH^{1*} ; Valerii Yustyniuk, DVM, PhD¹ ; Sylvester Ochwo, DVM, PhD¹ ; Mabel Aworh, DVM, MPH, PhD² ; Melinda Wilkins, DVM, MPH, PhD, DACVPM¹ 

¹Center for Animal Health and Food Safety, Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN

²Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC

*Corresponding author: Dr. Akinyede (oakinyed@jhsph.edu)

Objective

Antimicrobial resistance (AMR), a global threat driven by factors such as improper antimicrobial use in humans and animals, is projected to cause 10 million annual deaths by 2050. For behavior change, public health messages must be tailored for diverse audiences. Generative AI may have the potential to create culturally and linguistically suited AMR awareness messages. This study assesses ChatGPT-4's capability for crafting such content.

Methods

4 veterinary public health professionals from diverse linguistic and cultural backgrounds identified top AMR contributors and audiences in their countries. A fifth person developed and refined ChatGPT-4 prompts to create AMR awareness content in US English, Ukrainian, Luganda, Ugandan English, Yoruba, and Nigerian Pidgin, using behavior change models. The content was rated for accuracy, applicability, language, cultural fit, originality, clarity, persuasiveness, and overall quality.

Results

ChatGPT-4 created 2 content types (long and short) per language for social media, television ads, and WhatsApp. Quality ranged from poor to excellent. Shorter content outperformed longer ones. Performance varied across languages, with abysmal results for Yoruba and excellent for Pidgin. Problematic issues like simplistic language and inappropriate terminology were identified.

Conclusions

ChatGPT-4 has the potential to generate content and training aids. However, the varied quality requires professional verification. Future research should optimize prompts and incorporate expert and audience insights for better results. These preliminary findings should be interpreted cautiously due to the small sample size and subjectivity.

Clinical Relevance

ChatGPT-4 can quickly create tailored content for global AMR awareness. More research is needed to explore generative AI for One Health messaging.

Keywords: antimicrobial resistance (AMR), generative AI, ChatGPT and GPT-4, behavior change intervention, AMR awareness messaging

Antimicrobial-resistant infections are a growing public health threat estimated to cause 10 million human deaths annually by 2050.¹ Antimicrobial-resistance (AMR) occurs naturally but is accelerated by a combination of several human-related factors including improper antimicrobial prescription

practices for humans and animals, use of antimicrobials to prevent disease and promote growth in farmed animals, incomplete use of prescribed antimicrobials, inadequate infrastructure, and poor regulatory frameworks.²

Proper education and persuasion of different groups can lead to behavior change, which is one of many strategies necessary to slow the rise of AMR. However, there is no one-size-fits-all approach for public health interventions and communications addressing AMR awareness. Different countries and

Received September 29, 2024

Accepted January 7, 2025

Published online January 29, 2025

doi.org/10.2460/ajvr.24.09.0283

© 2025 THE AUTHORS. Published by the American Veterinary Medical Association as an Open Access article under Creative Commons CCBY-NC license.

cultures around the world have varying contexts, and effective communication for behavior change should account for differences in context, languages, and cultural sensitivities.

For effective public health messaging, behavior-change models such as the Health Belief Model, should be incorporated. These models, developed by social psychologists, are theories that provide a

Table 1—Rubric for rating AI-generated antimicrobial resistance (AMR) awareness content.

Criteria	Poor (1)	Fair (2)	Average (3)	Good (4)	Excellent (5)
Content accuracy	Contains significant factual errors or misinformation.	Includes some inaccuracies but also some correct information.	Mostly accurate but may contain minor errors or omissions	Demonstrates a high level of accuracy, with few or no errors.	Meticulously accurate and could rival that produced by an expert in the field.
Applicability/context	Largely irrelevant or fails to address the context outlined in the prompt.	Partially addresses the context but lacks relevance or applicability.	Adequately addresses the context but may lack depth or specificity.	Effectively addresses the context and provides relevant information.	Highly applicable and tailored precisely to the context outlined in the prompt.
Language use	Contains numerous grammatical errors, unclear language, and inappropriate use of terminology.	Generally understandable but may lack clarity or consistency.	Language is clear and mostly grammatically correct but could be improved for precision.	Language is polished, coherent, and effectively communicates ideas.	Language is flawless, exhibiting mastery of grammar, vocabulary, and style.
Cultural relevance	Lacks cultural sensitivity and fails to incorporate relevant cultural nuances.	Some attempt is made to address cultural elements, but they are not fully integrated or may be inaccurate.	Cultural elements are addressed to some extent, but improvements could enhance cultural relevance.	Demonstrates cultural awareness and effectively incorporates relevant cultural nuances.	Cultural elements are seamlessly integrated, demonstrating deep understanding and sensitivity.
Creativity/originality	Lacks creativity and originality, presenting conventional or clichéd ideas.	Some attempt at creativity is evident, but ideas are largely derivative or unremarkable.	Shows moderate creativity, offering some original perspectives or approaches.	Demonstrates creativity and innovation, presenting fresh insights or novel solutions.	Exceptionally creative and original, offering unique perspectives or groundbreaking ideas.
Clarity/coherence	Disjointed and difficult to follow, lacking clear organization or logical flow.	Some coherence is present, but overall structure and organization are weak.	Generally clear and coherent but may contain occasional lapses in structure.	Content is well-organized and logically structured, facilitating easy comprehension.	Impeccably structured, with seamless transitions and a clear, logical progression of ideas.
Engagement/persuasiveness	Fails to engage the reader and lacks persuasive elements or compelling arguments.	Some attempt at engagement is made, but the output lacks persuasiveness or fails to hold the reader's interest.	Moderately engaging and presents persuasive arguments with some effectiveness.	Effectively captivates the reader's attention and persuasively presents compelling arguments.	Highly engaging and persuasive, holding the reader's interest throughout and convincingly conveying key points.
Relevance to prompt	Deviates significantly from the prompt and fails to address the specified topic or objectives.	Some relevance to the prompt is evident, but output may contain tangential or unrelated content.	Generally aligns with the prompt and addresses the specified topic or objectives adequately.	Closely adheres to the prompt, demonstrating clear relevance and alignment with the specified topic or objectives.	Precisely addresses the prompt, showing deep understanding and thorough exploration of the specified topic or objectives.
Overall quality	Fundamentally flawed and fails to meet basic standards of quality in multiple respects.	Output exhibits several shortcomings but still possesses some redeeming qualities or potential for improvement.	Meets basic standards of quality and demonstrates competence in addressing the task or topic.	Output is of high quality, showcasing excellence in most aspects and effectively fulfilling the task or topic requirements.	Output is outstanding in every aspect, surpassing expectations and demonstrating exceptional quality and proficiency.

framework for understanding and influencing health-related behaviors.³

Cultural nuances and behavior change models form the foundation of successful public health interventions because they can determine the effectiveness of health communication strategies. By tailoring messaging to resonate with diverse groups, interventions can lead longer lasting behavior transformations, which in turn can contribute to improved health outcomes.

Culturally tailored public health campaigns can improve engagement by addressing specific beliefs and practices in communities.³ Segmenting public health interventional messages based on cultural characteristics, enhances their impact and relevance.^{3,4}

The recent explosive growth in the field of AI offers vast potential for AI to simplify and quicken processes, make knowledge more accessible, and improve our understanding of the world. In veterinary medicine, AI has been used for various purposes⁵ including diagnostic imaging^{6,7} and clinical documentation.⁸

Generative AI can create content targeted at different literacy levels, languages, and geographic locations.⁹ As of the time of submission of this article, there are currently no published studies on the use of generative AI for creating tailored AMR awareness content. This brief informal study aimed to explore the potential of generative AI for developing AMR messaging communications in different languages, geared at different audiences.

Methods

This exploratory study involved 4 veterinary public health professionals selected through convenience sampling based on their diverse cultural and linguistic backgrounds, geographical representation, and lived experiences. Three participants were affiliated with the same institution, and those who migrated from countries outside the US currently reside in the US. Each participant had veterinary public health experience in their home country. The participants included 1 professional from Nigeria, fluent in Pidgin English and Yoruba; 1 from Ukraine, fluent in Ukrainian; 1 from Uganda, fluent in Luganda and English; and 1 from the US, fluent in English. All participants had extensive experience in their home countries and a deep understanding of the cultural nuances within their regions. Each held veterinary

degrees, along with 1 or 2 advanced degrees, such as an MPH or PhD.

Each participant identified some key drivers of AMR based on prior experiences and understanding of the context in their home countries and selected a target audience within their country for whom they wanted ChatGPT-4-generated awareness content. A fifth individual, also a veterinary public health professional, developed and iteratively refined ChatGPT-4 prompts based on the input provided by the participants. The prompts were designed to generate AMR awareness messages in US English, Ukrainian, Luganda, Ugandan English, Yoruba, and Nigerian Pidgin English. The content was tailored to suit the specific audiences identified by the participants and generated for different media platforms (social media, WhatsApp, and TV ads). ChatGPT-4 was instructed to incorporate cultural nuances and apply appropriate behavior change models for each language and country.

The final prompts were run in ChatGPT-4 to produce both short-form and long-form content, which was subsequently evaluated by the participants. Two types of content were created for each language: short-form and long-form. Short-form content was generated for Facebook and Instagram videos across all languages. For long-form content, LinkedIn posts were generated for US English and Ukrainian; WhatsApp posts for Ugandan English, Luganda, and Nigerian Pidgin English; and television ads were generated for Yoruba.

A quality rating rubric was developed by the fifth person, focusing on criteria such as content accuracy, applicability and context, language use, cultural relevance, creativity, clarity and coherence, persuasiveness, relevance to the prompt, and overall quality. Participants independently rated the outputs as Poor, Fair, Average, Good, or Excellent for each criterion and provided general feedback. The rubric and definitions for each rating level are presented (**Table 1**).

The participants were also given the corresponding ChatGPT-4 prompts and were asked to suggest modifications if necessary. The fifth person analyzed the ratings by calculating averages for each criterion.

The ChatGPT-4 prompts and their outputs are presented (**Supplementary Material S1**).

Table 2—Selected AMR drivers and target audience.

Country	Chosen target audience	Key driver of AMR
US	General public	Use of antibiotics in animal feed as growth promoters and for disease prevention in intensive farming.
Ukraine	Clinicians (both physicians and veterinarians)	Overprescribing by physicians. Over-the-counter selling of antibiotics. Violation of antibiotic administration protocols: overuse of antibiotics (administration of antibiotics in cases where it is not required: viral infection, aseptic surgeries, etc), inappropriate use “watch and “reserve” antibiotics.
Uganda	Livestock farmers	Limited surveillance and monitoring. Limited access to veterinary services, leading to inappropriate antibiotic use in livestock.
Nigeria	Livestock farmers	Lack of awareness about the dangers of misuse of antibiotics. Over-the-counter selling of antibiotics for humans and animals. Use of antibiotics in feed.

Results

Participants identified some key drivers of AMR and target audiences in their respective countries to inform the development of customized ChatGPT-4 content. The selected target audiences and drivers for each participant are provided (Table 2).

ChatGPT-4 employed various behavior change models depending on the country including the Social Cognitive Theory, Health Belief Model, and Theory of Planned Behavior. The specific models chosen for each language are detailed (Table 3).

There was a noticeable variation in quality ratings across languages. The Yoruba content demonstrated the weakest performance. Although it contained correct Yoruba words, both the long-form and short-form outputs lacked coherence making the sentences difficult to comprehend. As a result, the output was rated poorly due to its lack of clarity and meaning.

In contrast, Nigerian Pidgin English content received the highest ratings, excelling across various criteria. ChatGPT-4 showed a remarkable understanding of cultural nuances and effectively incorporated slang leading to highly relevant and engaging content.

Table 3—Behavior change models selected by ChatGPT-4.

Country	Behavior change model
US English	Social Cognitive Theory
Ukrainian	Theory of Planned Behavior
Ugandan English	Health Belief Model, the Social Cognitive Theory, and the Theory of Planned Behavior
Luganda	Health Belief Model, the Social Cognitive Theory, and the Theory of Planned Behavior
Nigerian Pidgin English	Health Belief Model and the Social Cognitive Theory
Yoruba	Social Cognitive Theory

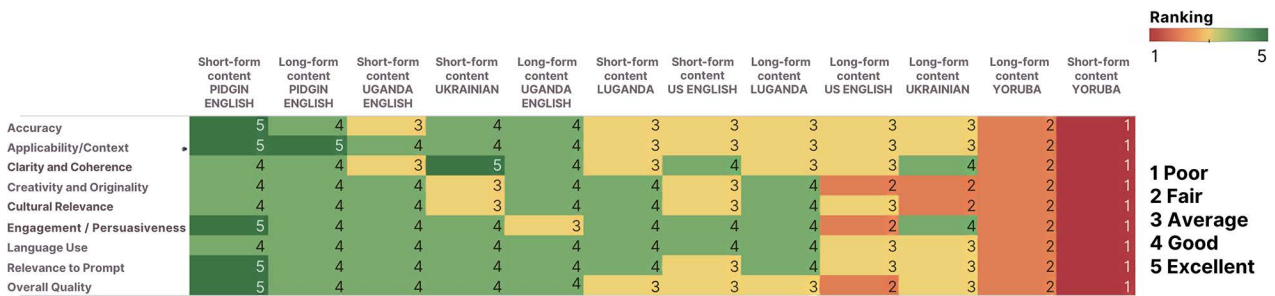


Figure 1—Individual ratings of AI-generated antimicrobial resistance (AMR) awareness content. Long-form and short-form content crafted by ChatGPT-4 for AMR messaging were rated across different criteria as Poor, Fair, Average, Good, or Excellent.

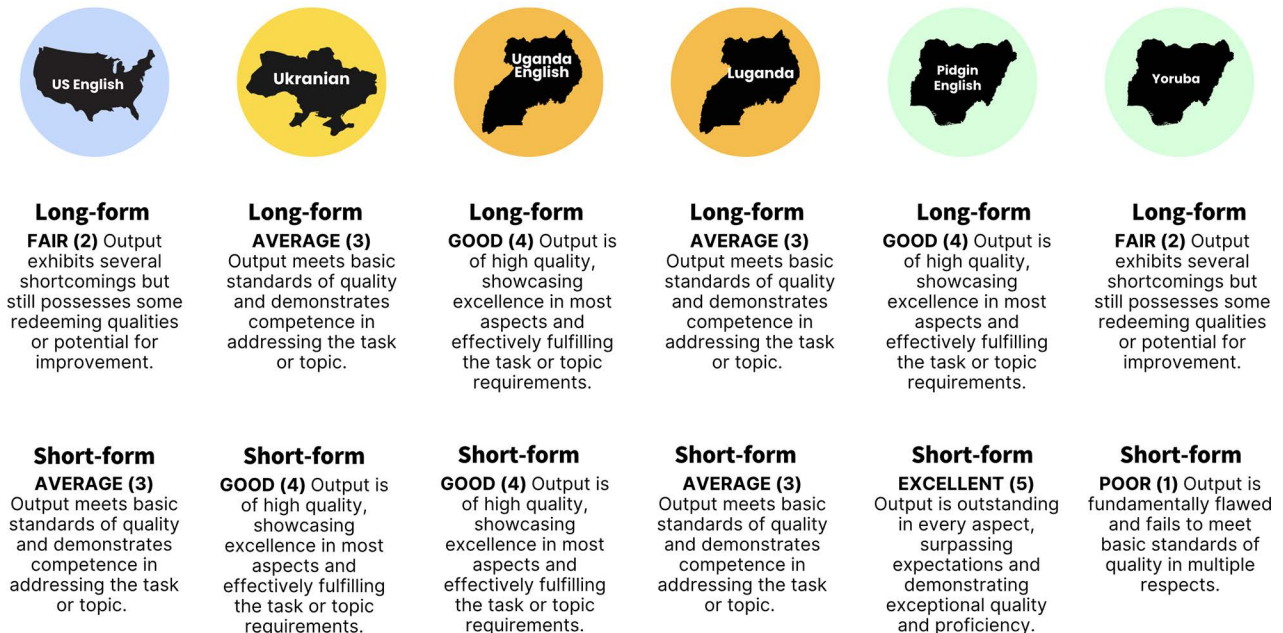


Figure 2—Overall quality of AI-generated AMR awareness content. Short-form content was generated by ChatGPT-4 for Facebook and Instagram videos across all languages. For long-form content, LinkedIn posts were created for US English and Ukrainian; WhatsApp posts for Ugandan English, Luganda, and Nigerian Pidgin English; and TV ads for Yoruba. The image shows the overall rating as well as the definition of that rating from the rubric. Detailed information on the definitions for each criterion is provided (Table 1).

The Ukrainian content received mixed ratings. The long-form LinkedIn content was rated Average, as it provided valuable information but was too simplistic and lacked the complexity and terminology expected by clinicians. In addition, some words were inaccurately translated, and the content fell short in cultural sensitivity. The short-form Ukrainian content was rated Good with helpful scenarios that could have been more tailored to the audience.

The US English content received a Fair rating for the long-form content and an Average rating for the short-form content. The long-form content used language that seemed oversimplified and included inappropriate patriotic and militaristic terms (like “fight” and “battle,” along with grammatically awkward phrases (such as “every pill not taken unnecessarily”).

The Ugandan English content was rated Good for both formats. The long-form content was accurate but too lengthy for the target audience, making it difficult to maintain focus. The short-form message was well-constructed but lacked practical applicability with respect to generating video clips that effectively demonstrate bacterial resistance development.

The Luganda content was rated Average for both formats. Although the long-form message was accurate; it was deemed excessively long for the target audience.

Overall, short-form content generally performed better than long-form content across most languages, with Nigerian Pidgin English emerging as the most successful in terms of quality.

The individual quality ratings (**Figure 1**) and the overall quality ratings (**Figure 2**) of the ChatGPT-4-generated content across all languages are presented.

Discussion

This study demonstrated that ChatGPT has the potential to bridge language and cultural barriers, by quickly generating AMR awareness and behavior change communication tailored to diverse audiences. However, the results were inconsistent across languages, suggesting that the AI model could have undergone more comprehensive training for some languages compared to others. While we did not evaluate the training data directly, it is well established that a key limitation of generative AI is that the quality of its output often depends on the data used for training.⁹ As a result, it may perform better in languages and contexts where it has received more in-depth training.

The effectiveness of AI-generated content is also influenced by the quality of prompts, which depends on the knowledge and expertise of the individual crafting them. Prompt engineering is a skill that improves with experience.¹⁰ Outputs can vary significantly based on the skill level of the individual creating the prompt. Involving a behavioral scientist or public health communication expert to select an appropriate model to be used in the prompt rather than relying on ChatGPT-4's model selection may also yield different results.

Although there was no published study on generative AI for tailored AMR messaging at the time

of article submission, some research has explored its application in other public health contexts. For instance, a recent study¹¹ demonstrated ChatGPT's capacity to adapt health messages for low-literacy audiences, showcasing its potential to create audience-specific content that retains key messages while simplifying complex language. Similarly, another study¹² found that ChatGPT and 2 other large language models could generate medical literature for urology patients.

Both studies demonstrated strengths and flaws highlighting the importance of professional oversight in ensuring the accuracy of AI-generated content. These existing findings support the potential utility of generative AI in public health.

Key strengths of our study on AMR messaging include the application of a standardized rubric for evaluating content across multiple languages, which enhanced the consistency and reliability of the assessment process. By evaluating ChatGPT-4's capability to generate appropriate content across several languages, rather than English alone, the present study provides insight into the model's versatility in producing linguistically diverse and accurate content. In addition, assessing ChatGPT-4's ability to create content across different formats and media platforms enhances our understanding of its capacity and effectiveness in delivering diverse types of communication. Furthermore, identifying its language-specific limitations enhances our understanding of the shortcomings of generative AI, which can inform future improvements in AI applications for AMR and One Health research. This study contributes to the emerging and relatively unexplored field of AI in One Health messaging and may inspire future research at the intersection of human, animal, and environmental health.

Limitations of this study include a small sample size, which resulted from the use of convenience sampling. Another limitation was the evaluation by veterinary public health professionals with varied levels of experience, which may have influenced their perceptions of content quality. In addition, the selection of key AMR drivers and target audiences was subjective and not grounded in a comprehensive literature review, potentially affecting the relevance of the outputs. Moreover, the behavior change models were chosen by ChatGPT-4 without direct input from a behavioral scientist, which could have impacted the effectiveness of the content. Involving experts in behavioral science might yield more refined and impactful messaging. Furthermore, the AI-generated messages were not directly evaluated by the target audiences, which could have provided additional insights into overall effectiveness.

While generative AI holds promise for streamlining AMR and One Health messaging, several ethical concerns must be considered, such as the potential for spreading misinformation, copyright infringement, use of unreliable data sources, and potential legal implications. These risks highlight the need for ongoing oversight and continuous review

by professionals to ensure that the AI-generated content is accurate, culturally appropriate, and ethically sound.

Future studies should expand on this preliminary exploration by including multiple evaluators per language to minimize subjectivity. Additional segmentation based on demographic factors like gender and age can further improve the specificity and relevance of the outputs. Testing scenarios where ChatGPT-4 selects the behavior change model versus scenarios where a behavioral scientist provides a predefined model could offer insights into how much human expertise impacts the effectiveness of AI-generated content. Increasing the number of participants, using and testing systematic prompt variations, and conducting statistical comparisons on these prompt variations can provide more in-depth insights. In addition, future research should involve the target audience to assess content understanding and effectiveness.

This study highlights the potential of generative AI in crafting AMR messaging but emphasizes the need for rigorous oversight, strategic prompt design, and collaboration with behavioral and communication experts to maximize impact and ensure responsible use. To maintain ethical and responsible practices, all AI-generated outputs should be thoroughly reviewed and validated by qualified professionals.

Acknowledgments

None reported.

Disclosures

The authors have nothing to disclose. No AI-assisted technologies were used in the generation of this manuscript.

Funding

The authors have nothing to disclose.

ORCID

O. Akinyede  <https://orcid.org/0000-0003-3439-4178>
V. Yustyniuk  <https://orcid.org/0000-0003-1387-1756>
S. Ochwo  <https://orcid.org/0000-0001-5704-2082>
M. Aworh  <https://orcid.org/0000-0003-4213-6929>
M. Wilkins  <https://orcid.org/0000-0002-4701-734X>

References

1. O'Neill J. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. London: review on antimicrobial resistance. Wellcome Trust and UK Government. 2014. Accessed September 1, 2024. https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf
2. Salam MA, Al-Amin MY, Salam MT, et al. Antimicrobial resistance: a growing serious threat for global public health. *Healthcare (Basel)*. 2023;11(13):1946. doi:10.3390/healthcare11131946
3. Odongo M. Health communication campaigns and their impact on public health behaviors. *J Commun*. 2024;5(2):55–69. doi:10.47941/jcomm.1980
4. Kaur VP. Social and behavior change communication. *Int J Adv Nurs Manag*. 2022;10(1):53–56. doi:10.52711/2454-2652.2022.00014
5. Chu CP. ChatGPT in veterinary medicine: a practical guidance of generative artificial intelligence in clinics, education, and research. *Front Vet Sci*. 2024;11:1395934. doi:10.3389/fvets.2024.1395934
6. Banzato T, Wodzinski M, Burti S, Vettore E, Muller H, Zotti A. An AI-based algorithm for the automatic evaluation of image quality in canine thoracic radiographs. *Sci Rep*. 2023;13(1):17024. doi:10.1038/s41598-023-44089-4
7. Li S, Wang Z, Visser LC, Wisner ER, Cheng H. Pilot study: application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. *Vet Radiol Ultrasound*. 2020;61(6):611–618. doi:10.1111/vru.12901
8. AI for busy veterinarians. ScribbleVet. 2023. Accessed September 27, 2024. <https://www.scribblevet.com/>
9. Bharel M, Auerbach J, Nguyen V, DeSalvo KB. Transforming public health practice with generative artificial intelligence. *Health Aff (Millwood)*. 2024;43(6):776–782. doi:10.1377/hlthaff.2024.00050
10. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. *J Med Internet Res*. 2023;25:e50638. doi:10.2196/50638
11. Ayre J, Mac O, McCaffery K, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. *J Gen Intern Med*. 2024;39(4):573–577. doi:10.1007/s11606-023-08469-w
12. Pompili D, Richa Y, Collins P, Richards H, Hennessey DB. Using artificial intelligence to generate medical literature for urology patients: a comparison of three different large language models. *World J Urol*. 2024;42(1):455. doi:10.1007/s00345-024-05146-3

Supplementary Materials

Supplementary materials are posted online at the journal website: avmajournals.avma.org.