Replies to Queries in Gynecologic Oncology by Bard, Bing and the Google Assistant

Edward J. Pavlik *, Dharani Ramaiah , Taylor Rives , Allison Swiecki-Sikora , Jamie Land

Posted Date: 8 March 2024

doi: 10.20944/preprints202403.0522.v1

Keywords: accuracy; virtual assistants; Google; Google Bard/Gemini; Microsoft Bing-Copilot; gynecologic; oncology

 doi:10.20944/preprints202403.0522.v1

Article

Replies to Queries in Gynecologic Oncology by Bard, Bing and the Google Assistant

Edward J. Pavlik *, Dharani D. Ramaiah, Taylor A. Rives, Allison L Swiecki-Sikora and Jamie M. Land

Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Chandler Medical, Center-Markey Cancer Center, University of Kentucky College of Medicine, Lexington, KY 40536-0293, USA; edward.pavlik@uky.edu (EJP)

* Correspondence: edward.pavlik@uky.edu; Tel.: +1-(859)-323-3830

Abstract: 1) Background: When women receive a diagnosis of a gynecologic malignancy, they can have questions about their diagnosis or treatment that can result in voice queries to virtual assistants. Recent advancement in artificial intelligence (AI) has transformed the landscape of medical information accessibility. The Google virtual assistant (VA) outperformed Siri, Alexa, and Cortana in voice queries presented prior to the explosive implementation of AI in early 2023. The efforts presented here focus on determining if advances in AI in the last 12 months improved the accuracy of Google VA responses related to gynecologic oncology. (2) Methods: Previous questions were utilized to form a common basis for queries prior to 2023 and responses in 2024. Correct answers were obtained from the *UpToDate* medical resource. Responses related to gynecologic oncology were obtained using Google VA, as well as the generative AI chatbots Google Bard/Gemini and Microsoft Bing-Copilot. (3) Results: The AI narrative responses varied in length and positioning of answers within the response. Google Bard/Gemini achieved an 87.5% accuracy rate, while Microsoft Bing-Copilot reached 83.3%. In contrast, the Google VA's accuracy in audible responses improved from 18% prior to 2023 to 63% in 2024. (4) Conclusion: While the accuracy of the Google VA has improved in the last year, it underperformed Google Bard/Gemini and Microsoft Bing-Copilot so that there is considerable room for further improved accuracy.

Keywords: accuracy; virtual assistants; Google; Google Bard/Gemini; Microsoft Bing-Copilot; gynecologic; oncology

1. Introduction

The Internet is a well-known source of health information [1] which can be accessed by voice technology that searches the Internet, and addressed with a voice reply by a virtual assistant (VA). We recently reported on the performance of the Google VA, Siri (Apple), Cortana (Microsoft), and Alexa (Amazon) to general information queries and those specifically related to gynecologic oncology [2]. The most correct audible replies (83.3% correct) were generated by the Google VA for general queries unrelated to gynecologic oncology, as well as for those related to gynecologic oncology (18.1% correct). An explosive introduction of artificial intelligence into search engines occurred in the year that followed our report, with the release of ChatGPT-3.5 (November 30, 2022), ChatGPT-4 (March 14, 2023) and ChatGPT-4 Turbo (November 2023) developed by Open AI [3]. It is well-recognized that the ChatGPT family has been developed to sound coherent and not necessarily to be factually accurate [3]. For example, in the clinical setting Chat GPT was reported to have an accuracy of 60.3% in forming an accurate initial differential diagnosis [4]. This performance is probably linked to the degree to which information is updated to be current. Note that the most recent update of the ChatGPT knowledgebase was in April 2023 [5]. Google Bard is a generative AI chatbot introduced in March of 2023 that utilizes its own large language AI model called Gemini. [6] Google Bard/Gemini pulls current information from the Internet and is available in over 40 languages. Microsoft has discontinued Cortana and launched Bing Chat known as Copilot within its Edge browser in February

 doi:10.20944/preprints202403.0522.v1

2023. It uses Prometheus, which is its own large language model that was built on the OpenAI GPT-4 foundation [7]. Present day searches on the Google Chrome browser look distinct from Google Bard/Gemini searches; however, Chrome searches can display a button to generate an "AI-powered overview" or can return generative AI results that are visually distinct from those returned by Google Bard/Gemini . The focus of this paper was to determine the extent to which utilization of AI in the last year has improved the performance of the Google VA with regard to answers for previously examined questions specific to gynecologic oncology, and to make comparisons to the AI driving Google's Bard/Gemini and Microsoft's Bing-Copilot. These efforts are important because inaccuracies in healthcare information can lead to misinterpretations by patients that can cause them to reject or withdraw from therapies that might have proven effective. In the present paper, we determined if the accuracy of responses by the Google VA to questions related to gynecologic oncology have improved since AI implementation.

2. Materials and Methods

Google VA (version 15.4.35.29.arm64) was accessed on Smartphones. Google's Bard/Gemini and Microsoft's Bing-Copilot were accessed on personal computers running Windows 10 Enterprise 64 bit (version 10.0.19045 build 19045.3930 with experience pack 1000.19053.1000.0) & Windows 11 Enterprise 64 bit (version 10.0.22631 build 22631 with experience pack 1000.22684.1000.0). Bard was accessed running the Gemini family of models, including Pro 1.0 with the "Bard" service rebranded as "Gemini." Microsoft Edge (version 121.0.2277.110) was used with Bing-Copilot being continuously updated without specifying a version. Questions specific to gynecologic oncology were exactly as previously used and reported [2]. Each question was queried five times. Evidence-based answers to queries were obtained from *UpToDate*, a subscription-based online service used as a point-of-care decision-support resource by clinical caregivers [8]. Queried responses were evaluated as percent of responses that were correct and in terms of location in the narrative response. Application of the correct answer from *UpToDate* is shown in Figure 1 for a Google Bard/Gemini AI narrative query response with the correct *UpToDate* information in green font. The word counts were obtained by copy and paste into Microsoft Word for the entire Google Bard/Gemini AI narrative query response (463 words, Figure 1) and for the number of words that preceded the text containing the correct information obtained from *UpToDate* (180 words, top box, Figure 1). The percent of narrative defined the location of the correct information in the Google Bard/Gemini AI narrative query response and was calculated as 180 / 463 = 38.9% (Figure 1). These evaluations were applied to query responses obtained from both Google Bard/Gemini and Microsoft Bing-Copilot. The results collected here were summarized with descriptive statistics using Winstat (version 2012.1)

 doi:10.20944/preprints202403.0522.v1

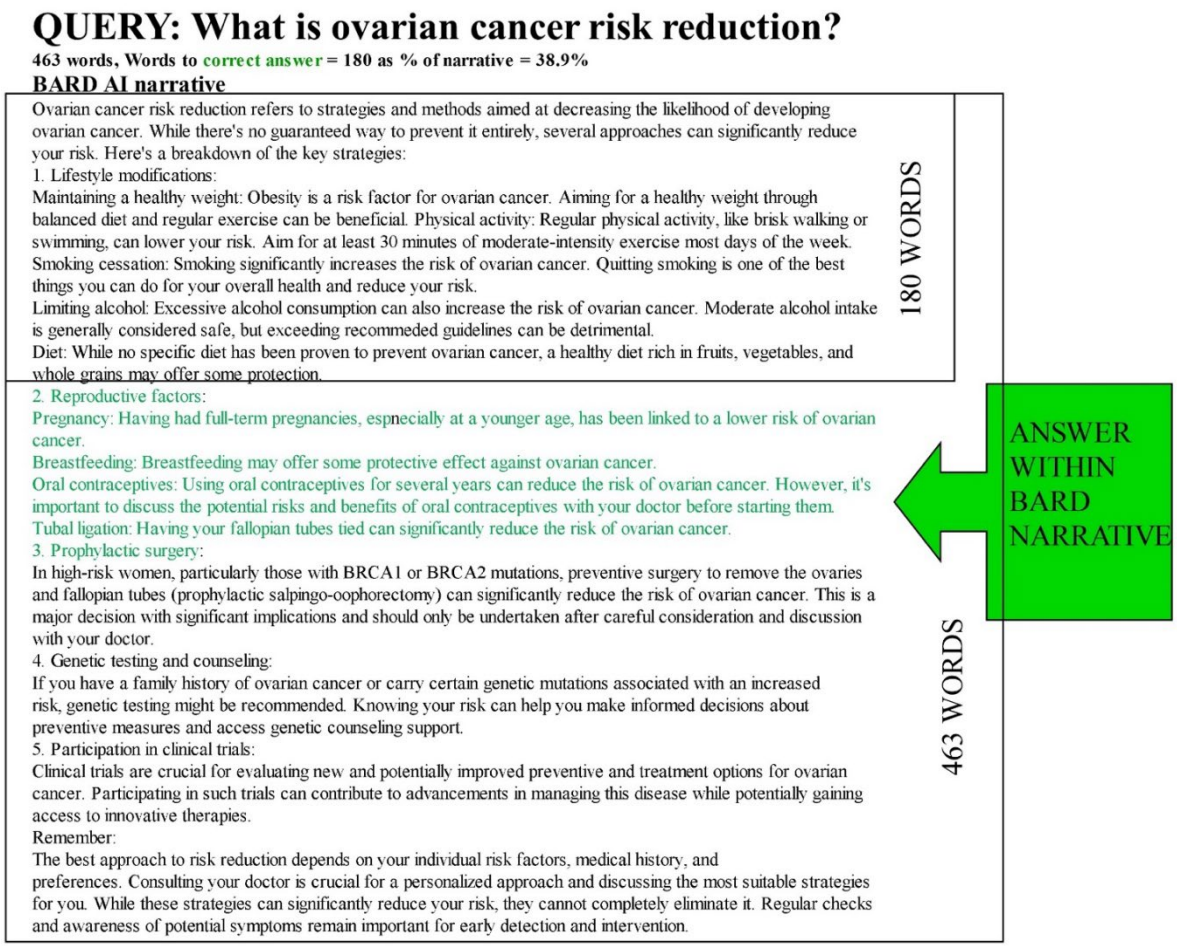

Figure 1. Determination of the position of a correct answer within the AI-generated narrative. The narrative was examined for the answer provided by UpToDate and marked with green font. The total narrative word count was determined (463 words = A) and word count preceding the correct answer was determined (180 words = B). The % of narrative is the position of the correct answer in the AI narrative and was determined as [(B/A)x 100].

3. Results

The questions posed as gynecologic oncology-related queries are listed in Table 1, as well as links to correct answers in *UpToDate*. In re-evaluating Google VA efforts, we first evaluated Google Bard/Gemini in order determine the degree to which the accuracy of Bard's AI narrative response was mirrored in the Google VA. Narrative length responses from Bard/Gemini changed in each repeat of a query, ranging from 32-39% of the mean narrative response length for each question, Table 1. Next, similar evaluations were made using Microsoft Bing-Copilot on the Edge browser in order to determine how similar narrative responses originating from a different large language model would be to Bard/Gemini's. The narrative response length from Microsoft Bing-Copilot also varied for each repeat of a query similarly to those from Google's Bard/Gemini. Microsoft Bing-Copilot allowed voice queries without voice recognition enabling the query, so that a mouse click on a microphone icon was needed to initiate voice recognition and supplied voice replies along with the text narrative. On Google's Bard/Gemini, a voice query could also be submitted by clicking on a microphone icon; however, the text narrative from Google's Bard/Gemini was not accompanied with a voice reply. Thus, in terms of operation Microsoft Bing-Copilot was very similar to Google VA except for voice recognition alone being able to enable the query using the Google VA.

 doi:10.20944/preprints202403.0522.v1

Table 1. Gynecologic Oncology-related Queries to Google Bard/Gemini. Narrative length returned to five repeat queries is shown as mean ± SEM with the range within parentheses. Sources of correct answers from *UpToDate* are hyperlinked for each question. Links were accessed on 30 January 2024.

Query #	Question	Correct Answer Link
1	What is stage I ovarian cancer? [308.6±15.6 (247,330)]	Answer
2	What is stage II ovarian cancer? [349.6±15.5 (310,394)]	Answer
3	What is stage III ovarian cancer? [344.6±20 (307,421)]	Answer
4	What is stage IV ovarian cancer? [338.6±27.7 (252,425)]	Answer
5	What is stage IC1 ovarian cancer? [309.6±22.1 (223,347)]	Answer
6	What is stage IIIA1 ovarian cancer? [355±16.2 (325,414)]	Answer
7	What is stage IVB ovarian cancer? [371.8±25.1 (303,436)]	Answer
8	What are the subtypes of epithelial ovarian cancer? [296.8±15.7 (365,350)]	Answer
9	What is screening for ovarian cancer? [324±23.2 (264,384)]	Answer
10	What are the screening recommendations for ovarian cancer? [318.4±15.3 (284,359)]	Answer
11	What are ways to prevent ovarian cancer? [387.8±23 (318,439)]	Answer
12	What are the symptoms of ovarian cancer? [313.8±22.9(237,365)]	Answer
13	What is hereditary ovarian cancer? [320±24.9 (256,386)]	Answer
14	What is ovarian cancer risk reduction? [372.2±43.1 (252,518)]	Answer
15	What is screening for cervical cancer? [335.4±38.2 (236,423)]	Answer
16	What are the screening recommendations for cervical cancer? [260.8±15.2 (219,350)]	Answer
17	What are the options for a 20-year-old woman requesting a Pap smear? [383±23.8 (334,473)]	Answer
18	What is the HPV vaccine? [355.8±19.2 (301,412)]	Answer
19	What are the ages for HPV vaccination? [229.6±21.4 (186,304)]	Answer
20	What are the three dose HPV vaccine recommendations? [258.2±24.8 (197,344)]	Answer
21	What are borderline epithelial tumors of the ovary? [349.6±21.9 (294,413)]	Answer
22	What is carcinosarcoma of the ovary? [322.6±7.4 (303,344)]	Answer
23	What are high-grade serous tumors of the ovary? [311.4±15.5 (277,368)]	Answer
24	What is stage IB endometrial cancer? [338.4±13.4 (295,366)]	Answer

Next, Google Bard/Gemini and Microsoft Bing-Copilot were compared directly. Performance was similar, although Google's Bard/Gemini had a higher percentage of correct responses (87.5% Bard/Gemini, 83.3% Microsoft Bing-Copilot) (Table 2). Out of five repeated queries for each of the 24 questions, there were a small number of instances where a correct response disappeared from the narrative after being present in the preceding narratives for that query. Accordingly, if all five queries failed to generate a narrative with a correct response, then the response was assigned as "Correct = NO" for that question.

Using the approach presented in Figure 1, the position of a correct answer within each AI-generated narrative was determined and summarized as lying in the 1st decile, 1st quartile or 2nd quartile of the total word count for the narrative response of either Google Bard/Gemini or Microsoft Bing-Copilot (Table 2). Responses that occurred early (1st decile), later (1st quartile) or even later (2nd quartile) in the word count of narrative responses were not significantly different (P = 0.496) between Google Bard/Gemini (66.7%, 23.8%, 9.5%) and Microsoft Bing-Copilot (50%, 30%, 20%). Thus, Google

 doi:10.20944/preprints202403.0522.v1

Bard/Gemini and Microsoft Bing-Copilot performed similarly in accuracy (% correct responses) and positioning the correct information within their respective AI-generated narratives.

Table 2. Responses to Gynecologic Oncology-related Questions by Google Bard/Gemini via the Chrome browser and Microsoft Bing-Copilot via the Edge browser.

	Question:	**Google-Bard/Gemini Correct (YES/NO)**	**MS-Bing-CoPilot Correct (YES/NO)**
1.	What is stage I ovarian cancer?	YES (3.3%, 1st decile)	YES (5.3%, 1st decile)
2.	What is stage II ovarian cancer?	YES (1.4%, 1st decile)	YES (0.4%, 1st decile)
3.	What is stage III ovarian cancer?	YES (8.7%, 1st decile)	YES (14.5%, 1st quartile)
4.	What is stage IV ovarian cancer?	YES (4.2%, 1st decile)	YES (4.2%, 1st decile)
5.	What is stage IC1 ovarian cancer?	YES (28%, 2nd quartile)	YES (25.6%, 2nd quartile)
6.	What is stage IIIA1 ovarian cancer?	YES (9.1%, 1st decile)	YES (19%, 1st quartile)
7.	What is stage IVB ovarian cancer?	YES (7.1%, 1st decile)	YES (11.6%, 1st quartile)
8.	What are the subtypes of epithelial ovarian cancer?	YES (8.4%, 1st decile)	YES (27.6%, 2nd quartile)
9.	What is screening for ovarian cancer?	YES (9.9%, 1st decile)	YES (16%, 1st decile)
10.	What are the screening recommendations for ovarian cancer?	YES (0.3%, 1st decile)	YES (0.5%, 1st decile)
11.	What are ways to prevent ovarian cancer?	YES (32%, 2nd quartile)	YES (24.4%, 1st quartile)
12.	What are the symptoms of ovarian cancer?	YES (12.8%, 1st quartile)	YES (25.9%, 2nd quartile)
13.	What is hereditary ovarian cancer?	YES (4.6%, 1st decile)	YES (8.2%, 1st decile)
14.	What is ovarian cancer risk reduction?	YES (22%, 1st quartile)	YES (28.3%, 2nd quartile)
15.	What is screening for cervical cancer?	YES (18.2%, 1st quartile)	YES (13.7%, 1st decile)
16.	What are the screening recommendations for cervical cancer?	NO	NO
17.	What are the options for a 20-year-old woman requesting a Pap smear?	YES (9.7%, 1st decile)	YES (10.2%, 1st decile)
18.	What is the HPV vaccine?	YES (0.3%, 1st decile)	YES (0.5%, 1st decile)
19.	What are the ages for HPV vaccination?	NO	NO
20.	What are the three dose HPV vaccine recommendations?	NO	NO
21.	What are borderline epithelial tumors of the ovary?	YES (3.9%, 1st decile)	YES (17.9%, 1st quartile)
22.	What is carcinosarcoma of the ovary?	YES (13.5%, 1st decile)	NO
23.	What are high-grade serous tumors of the ovary?	YES (17.3%, 1st quartile)	YES (10.9%, 1st quartile)
24.	What is stage IB endometrial cancer?	YES (13.7%, 1st quartile)	YES (0.5%, 1st decile)
Total number of correct responses		21	20
Percent correct responses		87.5%	83.3%%

Finally, when Google VA was re-queried in 2024 with the same questions used before 2023, where an accuracy rate of 18% correct auditory responses was found [**Error! Bookmark not defined.**], performance increased to 63% (Figure 2). While the accuracy of the Google VA has improved, it

 doi:10.20944/preprints202403.0522.v1

remains less than the accuracy of audible replies to queries in gynecologic oncology returned by Google Bard/Gemini and Microsoft Bing-Copilot.

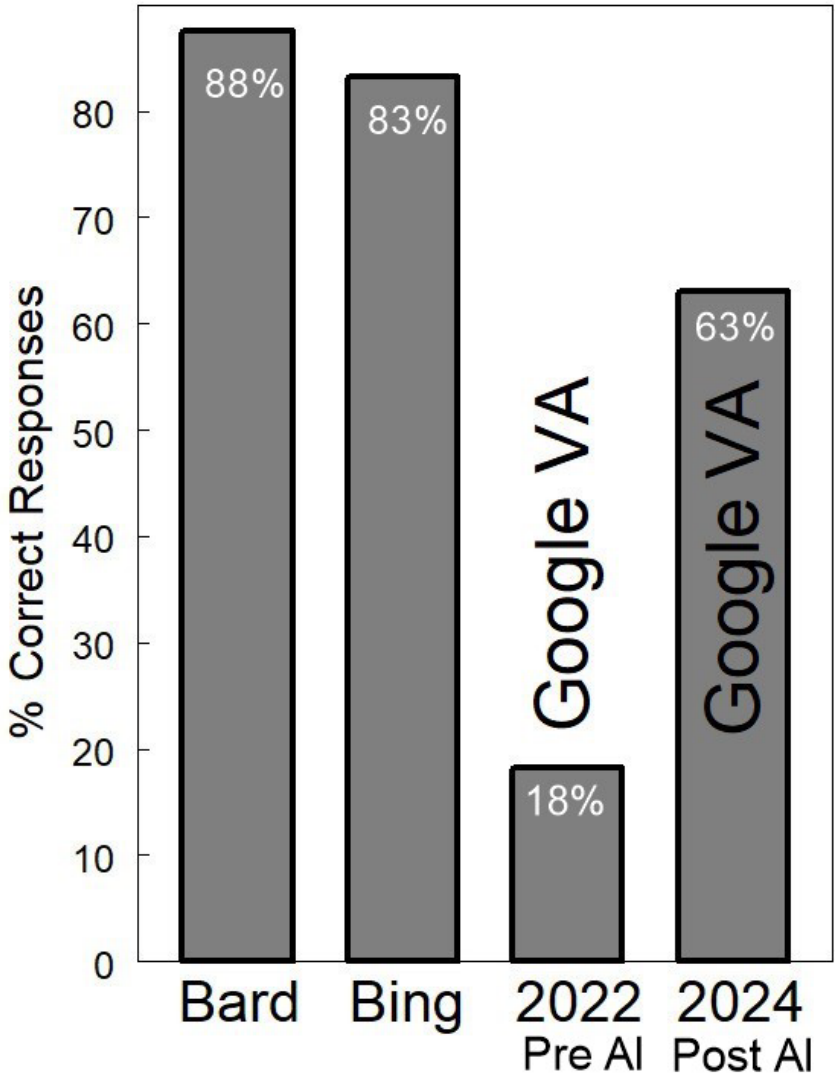

Figure 2. Comparative Performance of Google Bard/Gemini, Microsoft Bing-Copilot, and Google virtual assistant (VA) in 2022 and 2024.

4. Discussion

In summary, replies by the Google VA to gynecologic oncology-related queries have improved considerably in the last year, but are not as accurate as narrative replies from Google Bard/Gemini or Microsoft Bing-Copilot. In the year following the announcements on narrative AI heralding their various advantages and improvements, reports have also been published questioning chatbot accuracy. ChatGPT-3.5 $_{\text{turbo-0301}}$ performed poorly in providing accurate cancer treatment recommendations, and generated outputs that were not concordant with NCCN recommendations a third of the time [9]. In answering "Questions to Ask About Your Cancer," recommended by the American Cancer Society, both ChatGPT-3.5 and Bing answered correctly less than 80% of the time [10]. ChatGPT-4 was found to be capable of correctly diagnosing only 57% of complex clinical cases [11]. However, on a board-style neurology examination ChatGPT4 was able to answer 85% correct on over 1900 questions in behavioral, perceptive, and psychological–related areas, using confident language for both correct as well as incorrect answers [12]. Furthermore, when asked to provide the energy content of 222 food items, both ChatGPT-3.5 and ChatGPT-4 provided accurate answers less than half of the time [13]. From experiences related to the research presented in this paper, several

 doi:10.20944/preprints202403.0522.v1

characteristics of the narrative responses should be noted. For the answers returned in narratives ending with statements of "Additional Resources," "Resources", "For more information", "Sources", "Here are some helpful resources for additional information", "Here are some resources you might find helpful", and "Resources for further information", these resources most often originated from .org and .gov sources in Google Bard/Gemini, while it seemed that in Microsoft Copilot origination in .com sources was more common. No narrative included a reference to *UpToDate*, suggesting that sources of information behind pay walls are excluded from information gathering by Google Bard/Gemini and Microsoft Bing-Copilot. The degree to which this exclusion applies to other narrative chat AI is an open question. From a clinical standpoint because *UpToDate* has long-standing universal acceptability as a point-of-care decision support resource that is very frequently updated, it is difficult to expect acceptability of narratives that exclude *UpToDate* from their narrative responses to queries, no matter how confident, well-presented and high quality the narrative text is. It should be pointed out that Google Bard/Gemini does qualify their narrative results: "This is for informational purposes only. This information does not constitute medical advice or diagnosis." However, the accuracy of the information is not perfect. Microsoft Bing-Copilot's narrative replies: "It is important to consult with healthcare professionals for accurate information." This statement leaves unanswered as to whether healthcare professionals should look elsewhere than Microsoft Bing-Copilot for their information. The strengths of the present paper include: 1) use of the same 24 question set that was previously evaluated in order to make precise comparisons to current performance by the Google VA, Google Bard/Gemini and Microsoft Bing Copilot, 2) utilization of a continuously updated point-of-care decision-support resource compiled from more than 440 journals by clinicians as the source of correct answers [8], 3) evaluations of the Google VA in comparison to Google Bard/Gemini and Microsoft Bing Copilot, 4) multiple examinations of the length of each response to a query, 5) identification of how deeply imbedded the correct response was within a narrative response, 6) use of speech-to-text conversion to find and validate correctness of answers from the Google VA. The weakness of this work involves: 1) inability to determine how future-stable responses by the Google VA, Google Bard/Gemini and Microsoft Bing Copilot are for either accuracy or inaccuracy, 2) exclusion of efforts to broadly assess the findings here to parallel queries in clinical areas other than gynecologic oncology, 3) exclusion of efforts to determine if and when voice activation will be active for Google Bard/Gemini and Microsoft Bing Copilot so that further improved accuracy will become available. The voice-activated VA will be most valuable when users cannot manually enter information, and this will be very helpful to the elderly, to physically-impaired individuals, and to those that are bed-ridden. In situations where a resource like *UpToDate* cannot be utilized to evaluate the accuracy of AI responses, a degree of risk in the trustworthiness of responses to queries must be anticipated. For example, in queries about drug cross reactions or about poisons, the consequences of inaccuracies can be deadly. There are multiple ways that can provide users with greater confidence about the accuracy of AI search results. The use of a "custom attribution engine" as announced by Adobe would allow users to verify AI findings through source citation [14]. This type of approach should allow users to interpret narrative results in terms of source information and determine if there is any distortion in the AI narrative. Importantly, for narratives that contain clinical information, it is important to gauge the results against information resources that are frequently updated through efforts involving a large number of medical experts, like *UpToDate* or the National Comprehensive Cancer Network (nccn.org).

Taken together, there is ample evidence that echoes our findings that there is room for improved accuracy in the Google VA, Google Bard/Gemini and Microsoft Bing Copilot. Inaccuracy can pose dangers in medical settings as it can be stated in a manner in which the VA or the AI narrative is very confident and convincing. While a recent editorial acknowledges that for use in publication, large language models present concerns regarding authorship, originality, factual inaccuracies, and "hallucinations"or confabulations, the stated key to their acceptability is that authors take complete responsibility for the content and properly acknowledge the use of LLMs" [15]. In short, the accuracy of clinically-related publication remains the responsibility of the authors. As reported recently, an airline was unsuccessful in establishing that it was not responsible on its own corporate site for

 doi:10.20944/preprints202403.0522.v1

information provided by a chatbot [16]. Thus, the precedent has been set for legal responsibility and liability related to chatbot utilization.

5. Conclusions

This paper underscores the need for improvements in the accuracy of information related to gynecologic oncology supplied by the Google VA, Google Bard/Gemini and Microsoft Bing-Copilot as considered here. Although it has been reported that Google's Bard/Gemini has been found to provide erroneous information related to discoveries made by the James Webb Space telescope [17],[18], while Microsoft's AI-powered Bing-Copilot has also susceptible to providing false information [19],[20], the work that is presented here updates the state of accuracy and reliability to early 2024.

We conclude that audible replies by the Google VA to gynecologic oncology-related voice queries still have appreciable room for improving accuracy. Overall, we advise that patients exercise caution in the use of VAs that provide information in gynecologic oncology.

Author Contributions: Conceptualization, E.J.P. and J.M.L.; methodology, J.M.L.; software, E.J.P. and J.M.L.; validation, E.J.P.; formal analysis, E.J.P.; investigation, J.M.L., D.D.R., T.A.R., A.L.S-S., K.Q., ; resources, E.J.P.; data curation, E.J.P.; writing—original draft preparation, E.J.P..; writing—review and editing, J.M.L., D.D.R., T.A.R., A.L.S-S., K.Q., and E.J.P.; visualization, J.M.L.; supervision, E.J.P.; project administration, E.J.P.; funding acquisition, E.J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not have external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data will be made available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fox, S.; Duggan, M. Health Online 2013. Pew Research Center 2013. Available online: https://www.pewresearch.org/internet/ 2013/01/15/health-online-2013/ (accessed on 13 February 2024).
2. Land, J.M.; Pavlik, E.J.; Ueland, E.; Ueland, S.; Per, N.; Quick, K.; Gorski, J.W.; Riggs, M.J.; Hutchcraft, M.L.; Llanora, J.D.; et al. Evaluation of Replies to Voice Queries in Gynecologic Oncology by Virtual Assistants Siri, Alexa, Google, and Cortana. BioMedInformatics 2023, 3, 553-562. https://doi.org/10.3390/biomedinformatics3030038
3. Brandl, R.; Ellis, C. Tooltester. ChatGPT Statistics 2024 - All the latest statistics about OpenAI's chatbot. Available online: https://www.tooltester.com/en/blog/chatgpt-statistics/
4. Rao, A.; Pang, M.; Kim, J.; Kaminen, M.; Lie, W. Prasad, AK.; Landman, A.; Dreyer, K.; Succi, MD.
5. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study
6. J Med Internet Res 2023;25:e48659
7. doi: 10.2196/48659
   Kan, M. ChatGPT's Knowledge Base Finally Extends Beyond 2021. PC 11/06/2023 https://www.pcmag.com/news/chatgpts-knowledge-base-finally-extends-beyond-2021 (accessed on 13 February 2024)
8. Ortiz, S. ZDNET What is Google Bard? Here's everything you need to know. Feb. 7, 2024. https://www.zdnet.com/article/what-is-google-bard-heres-everything-you-need-to-know/ (accessed on 13 February 2024)
9. Microsoft Copilot. Wikipedia. https://en.wikipedia.org/wiki/Microsoft_Copilot (accessed on 13 February 2024)
10. UpToDate. Wikipedia. https://en.wikipedia.org/wiki/UpToDate (accessed on 13 February 2024)
11. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, Bitterman DS. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. JAMA Oncol. 2023 Oct 1;9(10):1459-1462. doi: 10.1001/jamaoncol.2023.2954
12. Janopaul-Naylor, J.R.; Koo, A.; Qian, D.C.; McCall, N.S.; Liu, Y.; Patel, S.A. Physician Assessment of ChatGPT and Bing Answers to American Cancer Society's Questions to Ask About Your Cancer. Am J Clin Oncol. 2024 Jan 1;47(1):17-21. doi: 10.1097/COC.0000000000001050

 doi:10.20944/preprints202403.0522.v1

13. Shea, Y.F.; Lee, C.M.Y.; Ip, W.C.T.; Luk, D.W.A.; Wong, S.S.W. Use of GPT-4 to Analyze Medical Records of Patients With Extensive Investigations and Delayed Diagnosis. JAMA Netw Open. 2023 Aug 1;6(8):e2325000. doi: 10.1001/jamanetworkopen.2023.25000
14. Schubert, M.C.; Wick, W.; Venkataramani, V. Performance of Large Language Models on a Neurology Board-Style Examination. JAMA Netw Open. 2023 Dec 1;6(12):e2346721. doi: 10.1001/jamanetworkopen.2023.46721
15. Hoang, Y.N.; Chen, Y.L.; Ho, D.K.N.; Chiu, W.C;. Cheah, K.J.; Mayasari, N.R.; Chang, J.S. Consistency and Accuracy of Artificial Intelligence for Providing Nutritional Information. JAMA Netw Open. 2023 Dec 1;6(12):e2350367. doi: 10.1001/jamanetworkopen.2023.50367
16. Irwin, K. Adobe's AI Assistant Can Summarize PDFs, PowerPoints for You. PC Magazine 02 February 20 2024 https://www.pcmag.com/news/adobes-ai-assistant-can-summarize-pdfs-powerpoints-for-you?utm_source=email&utm_campaign=whatsnewnow&zdee=gAAAAABlzjXaj__BrV5eo-f2qF7sDrWWKEtkmH2G19SKAH7kdCyO2QZWMpWIrsyK4J9MUmTylA0dKjizA-4gORPeq8yq9vplTV5ka-cM6LR2CdQFveA56U4%3D (accessed on 21 February 2024)
17. Koller, D.; Beam, A.; Manrai, A.; Ashley, E.; Liu, X.; Gichoya, J.; Holmes, C.; Zou, J.; Dagan, N.; Wong,T.Y.; Blumenthal, D.; Kohane, I. Why We Support and Encourage the Use of Large Language Models in NEJM AI Submissions. NEJM AI 2023;1(1) DOI: 10.1056/AIe2300128
18. Cecco, L. Air Canada ordered to pay customer who was misled by airline's chatbot. The Guardian 21 February 2024 https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit (accessed on 21 February 2024)
19. Miao, H. Alphabet Stock Drops 8% After Google Rollout of AI Search Features. The Wall Street Journal. Available online: https://www.wsj.com/livecoverage/stock-market-news-today-02-08-2023/card/alphabet-stock-drops-after-google-parent- introduces-ai-search-features-wgCJG3IDoSbfL3SgyrNI (accessed on 13 February 2024).
20. Mihalcik, C. Google ChatGPT Rival Bard Flubs Fact About NASA's Webb Space Telescope. Available online: https://www.cnet.com/science/space/googles-chatgpt-rival-bard-called-out-for-nasa-webb-space-telescope-error/ (accessed on 13 February 2024)
21. Hao, K. What Is ChatGPT? What to Know About the AI Chatbot That Will Power Microsoft Bing. The Wall Street Journal. 16 May 2023. Available online: https://www.wsj.com/articles/chatgpt-ai-chatbot-app-explained-11675865177?st=q4wbp2 (accessed on 13 February 2024)
22. ercfh1zo3&reflink=share_mobilewebshare (accessed on 13 February 2024)
23. Quach, K. Microsoft's AI Bing Also Factually Wrong, Fabricated Text During Launch Demo. The Register. Available online: https://www.theregister.com/2023/02/14/microsoft_ai_bing_error/ (accessed on 13 February 2023).