

# Comparative Performance of ChatGPT and Bard in a Text-Based Radiology Knowledge Assessment

Canadian Association of  
Radiologists' Journal  
2023, Vol. 0(0) 1–7  
© The Author(s) 2023



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/08465371231193716  
[journals.sagepub.com/home/caj](https://journals.sagepub.com/home/caj)



Nikhil S. Patil<sup>1</sup> , Ryan S. Huang<sup>2</sup> , Christian B. van der Pol<sup>3</sup> , and  
Natasha Larocque<sup>4</sup> 

## Abstract

**Purpose:** Bard by Google, a direct competitor to ChatGPT, was recently released. Understanding the relative performance of these different chatbots can provide important insight into their strengths and weaknesses as well as which roles they are most suited to fill. In this project, we aimed to compare the most recent version of ChatGPT, ChatGPT-4, and Bard by Google, in their ability to accurately respond to radiology board examination practice questions. **Methods:** Text-based questions were collected from the 2017-2021 American College of Radiology's Diagnostic Radiology In-Training (DXIT) examinations. ChatGPT-4 and Bard were queried, and their comparative accuracies, response lengths, and response times were documented. Subspecialty-specific performance was analyzed as well. **Results:** 318 questions were included in our analysis. ChatGPT answered significantly more accurately than Bard (87.11% vs 70.44%,  $P < .0001$ ). ChatGPT's response length was significantly shorter than Bard's (935.28 ± 440.88 characters vs 1437.52 ± 415.91 characters,  $P < .0001$ ). ChatGPT's response time was significantly longer than Bard's (26.79 ± 3.27 seconds vs 7.55 ± 1.88 seconds,  $P < .0001$ ). ChatGPT performed superiorly to Bard in neuroradiology, (100.00% vs 86.21%,  $P = .03$ ), general & physics (85.39% vs 68.54%,  $P < .001$ ), nuclear medicine (80.00% vs 56.67%,  $P < .01$ ), pediatric radiology (93.75% vs 68.75%,  $P = .03$ ), and ultrasound (100.00% vs 63.64%,  $P < .001$ ). In the remaining subspecialties, there were no significant differences between ChatGPT and Bard's performance. **Conclusion:** ChatGPT displayed superior radiology knowledge compared to Bard. While both chatbots display reasonable radiology knowledge, they should be used with conscious knowledge of their limitations and fallibility. Both chatbots provided incorrect or illogical answer explanations and did not always address the educational content of the question.

## Résumé

**Objectif :** Le robot conversationnel Bard de Google, un concurrent direct de ChatGPT, a récemment été lancé. La compréhension des performances relatives de ces deux robots conversationnels peut mettre en lumière leurs forces et leurs faiblesses ainsi que les fonctions pour lesquelles ils sont les plus adaptés. Ce projet avait pour objectif de comparer les réponses de Bard et de la version la plus récente de ChatGPT, à savoir ChatGPT-4, à une série de questions pratiques tirées d'un examen en radiologie. **Méthodes :** Les questions écrites utilisées provenaient des annales 2017-2021 de l'examen Diagnostic Radiology In-Training (DXIT) de l'American College of Radiology. Ces questions ont été soumises à ChatGPT-4 et Bard, et leurs temps de réponse ainsi que la longueur et l'exactitude de leurs réponses respectives ont été analysés. Les performances relatives aux sous-spécialités ont également été examinées. **Résultats :** Cette comparaison a été effectuée sur 318 questions. Les réponses de ChatGPT étaient plus exactes que celles de Bard (87,11 % contre 70,44 %,  $P < ,0001$ ). Les réponses de ChatGPT étaient considérablement plus courtes que celles de Bard (935,28 ± 440,88 caractères contre 1 437,52 ± 415,91 caractères,  $P < ,0001$ ). Le temps de réponse de ChatGPT était nettement plus long que celui de Bard (26,79 ± 3,27 secondes contre 7,55 ± 1,88 seconde,  $P < ,0001$ ). Les performances de ChatGPT étaient supérieures à celles de Bard en neuroradiologie (100,00 %

<sup>1</sup> Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada

<sup>2</sup> Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

<sup>3</sup> Department of Diagnostic Imaging, Hamilton Health Sciences, Juravinski Hospital and Cancer Centre, Hamilton, ON, Canada

<sup>4</sup> Department of Radiology, McMaster University, Hamilton, ON, Canada

## Corresponding Author:

Natasha Larocque, Department of Radiology, Hamilton General Hospital, 237 Barton St. East Hamilton, ON L8L 2X2, Canada.

Email: [natasha.larocque@medportal.ca](mailto:natasha.larocque@medportal.ca)

contre 86,21 %,  $P = ,03$ ), en radiologie générale et en physique (85,39 % contre 68,54 %,  $P < ,001$ ), en médecine nucléaire (80,00 % contre 56,67 %,  $P < ,01$ ), en radiologie pédiatrique (93,75 % contre 68,75 %,  $P = ,03$ ), et en échographie (100,00 % contre 63,64 %,  $P < ,001$ ). Aucune différence notable entre les performances de ChatGPT et celles de Bard n'a été notée en ce qui concerne les autres sous-spécialités. **Conclusion** : ChatGPT a fait preuve de connaissances en radiologie supérieures à celles de Bard. Bien que les deux robots conversationnels semblent démontrer des connaissances correctes en radiologie, il est nécessaire de les utiliser en ayant conscience de leurs limites et de leurs défauts. Les robots ont également tous deux justifié leurs réponses avec des explications incorrectes ou illogiques, et n'ont pas toujours pris en compte le côté éducatif de la question.

## Keywords

Google, Bard, radiology, artificial intelligence, chatbot, ChatGPT

## Introduction

Artificial intelligence (AI) chatbots have utility in a variety of contexts, and their applications within medicine are a particularly exciting current area of investigation. ChatGPT (Chat Generative Pre-trained Transformer) by OpenAI is currently the most popular chatbot and was released in a preliminary form last year, in November 2022.<sup>1</sup> A direct competitor was recently released by Google, named Bard.<sup>2</sup> Similar to ChatGPT, Bard is an interactive AI chatbot that engages in conversation in response to human input. However, unlike ChatGPT, Bard also has real-time access to the internet and a unique communication model named the Language Model for Dialogue Applications (LaMDA) instead of ChatGPT's Generative Pre-training Transformer.<sup>3</sup>

In recent months, there have been a number of studies published regarding the potential role of ChatGPT in radiology. Potential uses include assistance with report generation, as a study tool for trainees, improving differential diagnoses, data analysis, determining suitable imaging options, and as a source of information for patients.<sup>4-6</sup> Studies earlier this year found that ChatGPT-4 performed better than ChatGPT-3.5 in answering a sample of 150 questions designed to be similar to the Canadian Royal College and American Board of Radiology examinations.<sup>7,8</sup> Given the novelty of Bard by Google, to the best of our knowledge, there are only 2 published articles comparing ChatGPT to Bard, both of which found ChatGPT to be superior.<sup>9,10</sup> In this project, we aimed to compare the most recent version of ChatGPT, ChatGPT-4, and Bard in their ability to accurately respond to radiology board examination questions. Understanding the relative performance of these different chatbots can provide important insight into the strengths and weaknesses of each of these chatbots and which roles they are most suited to fill.

## Methods

Questions were collected from the American College of Radiology's Diagnostic Radiology In-Training (DXIT)

examination.<sup>11</sup> The DXIT examination is an annual exam prepared by the ACR used to simulate the American Board of Radiology (ABR) Core exam. The exam is administered in many radiology residency programs internationally and serves as a formative assessment tool for residents in training. The literature suggests that it can serve as a predictor for residents' future performance on the Canadian Royal College examination or the ABR Core examination.<sup>12,13</sup> DXIT exam questions were collected through a publicly available, open-source flashcard software.<sup>14</sup> Consent was obtained from the ACR to conduct our analysis using these publicly available questions. A 5-year sample of DXIT examination questions, between 2017 and 2021, was used in this study. These questions were categorized by year and by exam subsections including neuroradiology, mammography, general & physics, nuclear medicine, pediatric radiology, interventional radiology, gastrointestinal radiology, genitourinary radiology, cardiac radiology, chest radiology, musculoskeletal radiology, and ultrasound. Questions were entered into each chatbot using the same methodology as described by Gilson et al. (2023).<sup>15</sup> Only text-based questions were included in our analysis.

Data was collected using ChatGPT-4 and Bard on June 11, 2023. Each AI chatbot was reset before entering another question. The primary outcome of this study was the comparative performance of ChatGPT and Bard in answering DXIT questions. The secondary outcomes of this study included the chatbots' performance in each subspecialty, the length of responses, the response time, and the proportion of questions answered with an explanation. Additionally, 10% of our questions and their respective responses by ChatGPT and Bard were selected for more detailed analysis by 2 radiologists at a tertiary care center (N.L., C.V.P.). Specifically, each response by each chatbot was read in detail and the proportions of responses which contained hallucinations/fabrications, illogical or incorrect answer rationales, or do not accurately address the educational content of the questions, were recorded. Data was collected independently, and conflicts were resolved through discussion. Paired t-tests were used to compare means and chi-squared tests were used to compare proportions. A  $P$ -value threshold of .05 was used to determine

**Table 1.** ChatGPT-4 and Bard Response Characteristics and Accuracy in Answering American College of Radiology DXIT Exam Questions.

Category	Questions	Question Length (Characters)	ChatGPT-4 Accuracy (%)		Bard Accuracy (%)	Accuracy P-Value (Chi-Squared Test)		ChatGPT-4 Response Length (Characters)	Bard Response Length (Characters)	Response Length P-Value (t Test)	ChatGPT-4 Response Time (Seconds)	Bard Response Time (Seconds)	Response Length P-Value (t Test)
			ChatGPT-4 Accuracy (%)	Bard Accuracy (%)		ChatGPT-4 Response Length (Characters)	Bard Response Length (Characters)						
Neuroradiology	29	207.28 (72.68)	100.00	86.21	.03*	840.90 (426.35)	1443.52 (415.88)	<.0001*	26.26 (2.69)	7.20 (1.81)	<.0001*		
Mammography	19	234.26 (58.05)	84.21	68.42	.14	787.63 (447.38)	1454.95 (442.34)	<.001*	28.06 (4.05)	7.65 (1.78)	<.0001*		
General & physics	89	305.22 (145.80)	85.39	68.54	<.001*	1022.38 (453.50)	1490.69 (406.58)	<.0001*	26.89 (3.42)	7.75 (1.89)	<.0001*		
Nuclear medicine	30	274.67 (103.73)	80.00	56.67	<.01*	947.30 (486.57)	1321.57 (374.86)	<.001*	26.35 (3.14)	8.33 (2.10)	<.0001*		
Pediatric radiology	16	195.81 (48.96)	93.75	68.75	.03*	764.63 (330.04)	1368.88 (547.91)	<.001*	26.37 (3.53)	7.51 (2.07)	<.0001*		
Interventional radiology	26	217.08 (90.04)	88.46	80.77	.32	952.31 (510.00)	1538.31 (446.90)	<.0001*	26.06 (3.22)	7.33 (1.78)	<.0001*		
Gastrointestinal radiology	29	190.00 (61.58)	89.66	79.31	.17	901.93 (423.01)	1427.66 (322.21)	<.0001*	25.95 (3.01)	7.20 (1.59)	<.0001*		
Genitourinary radiology	11	225.09 (72.15)	72.73	63.64	.53	1048.82 (338.28)	1373.09 (342.2)	.04*	29.74 (3.06)	7.66 (3.02)	<.0001*		
Cardiac radiology	16	188.69 (59.27)	75.00	68.75	.59	915.50 (3.48)	1537.94 (692.44)	<.001*	27.71 (3.48)	7.07 (1.43)	<.0001*		
Chest radiology	6	230.67 (97.04)	100.00	83.33	.27	816.33 (303.77)	1492.33 (295.28)	<.01*	26.24 (1.17)	6.93 (1.45)	<.0001*		
Musculoskeletal radiology	25	184.52 (71.06)	80.00	64.00	.10	945.48 (394.93)	1326.40 (316.33)	<.001*	27.56 (3.12)	7.78 (1.94)	<.0001*		
Ultrasound	22	237.23 (66.20)	100.00	63.64	<.001*	944.91 (518.11)	1371.95 (352.66)	<.01*	25.99 (2.63)	7.07 (1.56)	<.0001*		
Total	318	242.01 (109.40)	87.11	70.44	<.0001*	935.28 (440.88)	1437.52 (415.91)	<.0001*	26.79 (3.27)	7.55 (1.88)	<.0001*		

\* signifies a statistically significant difference

**Question:** Which of the following is a complication of umbilical venous lines?

A. Renal vein thrombosis  
B. Portal vein thrombosis  
C. Pulmonary embolism  
D. Iliac vein aneurysms

↓

**ChatGPT Response:** The correct answer is:

B. Portal vein thrombosis

Umbilical venous lines, also known as umbilical catheters, are commonly used in neonatal intensive care units for various medical purposes. However, one of the potential complications associated with umbilical venous lines is the development of portal vein thrombosis. This occurs when a blood clot forms in the portal vein, which is the main vessel that carries blood from the gastrointestinal tract and spleen to the liver. Portal vein thrombosis can lead to serious complications such as liver damage and impaired blood flow to the liver.

↓

**Inaccurately Addressing the Question's Educational Purpose:**

Despite providing the correct answer, the explanation provided by ChatGPT does not touch on the educational purpose of the question which is to test the veins that connect to the umbilical vein. Umbilical venous catheters pass through the umbilical vein, left portal vein, ductus venosus, into the inferior vena cava. The remainder of the answer choices are incorrect as they have no anatomical connection to the umbilical vein.

**Figure 1.** Sample ChatGPT response demonstrating a correct answer with an answer rationale that does not accurately address the educational content of the question.

statistical significance in this study. Microsoft Excel was used for all analyses.

## Results

There was a total of 318 text-based DXIT exam questions between 2017 and 2021 that were included in our analysis. The average question length was  $242.01 \pm 109.40$  characters. There were 29 (9.12%) neuroradiology questions, 18 (5.66%) mammography questions, 89 (27.99%) general & physics questions, 30 (9.43%) nuclear medicine questions, 16 (5.03%) pediatric radiology questions, 26 (8.18%) interventional radiology questions, 29 (9.12%) gastrointestinal radiology questions, 11 (3.46%) genitourinary radiology questions, 16 (5.03%) cardiac radiology questions, 6 (1.89%) chest radiology questions, 25 (7.86%) musculoskeletal radiology questions, and 22 (6.92%) ultrasound questions. [Table 1](#) outlines question characteristics, response characteristics, and the accuracy of both chatbots separated by subspecialty.

[Supplementary Table 1](#) shows sample responses by ChatGPT and Bard for correct and incorrect responses.

ChatGPT answered more accurately than Bard (87.11% vs 70.44%,  $P < .0001$ ). ChatGPT's response length was shorter than Bard's ( $935.28 \pm 440.88$  characters vs  $1437.52 \pm 415.91$  characters,  $P < .0001$ ). ChatGPT's response time was longer than Bard's ( $26.79 \pm 3.27$  seconds vs  $7.55 \pm 1.88$  seconds,  $P < .0001$ ).

ChatGPT performed superiorly to Bard in neuroradiology, (100.00% vs 86.21%,  $P = .03$ ), general & physics (85.39% vs 68.54%,  $P < .001$ ), nuclear medicine (80.00% vs 56.67%,  $P < .01$ ), pediatric radiology (93.75% vs 68.75%,  $P = .03$ ), and ultrasound (100.00% vs 63.64%,  $P < .001$ ). In the remaining subspecialties, there were no differences between the chatbots. In each subspecialty-specific analysis, ChatGPT's response length was shorter, and its response time was longer ([Table 1](#)).

32 questions (10%) were randomly selected for closer analysis by 2 staff radiologists. This analysis revealed that both chatbot responses for these questions did not contain

hallucinations or fabrications. 19% of ChatGPT's responses contained illogical and/or incorrect answer rationales. 6% of ChatGPT's responses did not accurately address the educational content of the question. An example is provided in [Figure 1](#). 38% of Bard's responses contained illogical and/or incorrect answer rationales. 9% of Bard's responses did not accurately address the educational content of the question.

## Discussion

Overall, ChatGPT performed superiorly to Bard in responding to DXIT exam questions. There were several differences between these chatbots in their accuracy and response characteristics. Both chatbots consistently provided explanations for their responses, although this was the case regardless of whether the answer was correct or incorrect. Bard generally provided lengthier responses, but answered quicker, possibly due to lower rates of usage currently. While performance varied for both chatbots across subspecialties, a similar trend to the primary analysis was established, with ChatGPT performing superiorly. Possibly, ChatGPT-4 performed superiorly to Bard as there is variation in their respective training data as well as the strength of their models in answering medical questions. Recently published studies suggest that ChatGPT outperforms Bard in answering lung cancer questions and neurosurgery oral boards preparation questions, which is consistent with our analysis.<sup>9,10</sup>

ChatGPT-4 and Bard both display reasonable radiology knowledge. The role of AI in radiology spans many domains including image interpretation, tumour staging, report generation, workflow improvements, medical writing, and research.<sup>4,16-18</sup> However, at this stage, human input remains vital in ensuring quality and optimal patient outcomes.<sup>19</sup> Although AI chatbots continue to be updated, with multimodal input anticipated to arrive later this year, currently, their role is more limited. Specifically, AI chatbots' primary role currently is likely as a study tool for trainees or as an information resource for patients, and even then, their role may be limited. As illustrated in [Figure 1](#), a chatbot may provide the correct answer, but the explanation may not appropriately address the educational content of the question. Similarly, our analysis of the quality of answers suggested that both chatbots also occasionally provided illogical or incorrect answer rationales. This can be particularly deceptive for trainees or patients who may be fooled by the confidence with which AI chatbots communicate. As a result, the way in which trainees are educated and assessed will also need to adapt.<sup>20,21</sup> Particularly, an increasingly large focus should be placed on higher-level thinking questions as outlined in Bloom's taxonomy.<sup>22</sup> It is also important to educate radiologists and radiology trainees about the strengths and weaknesses of AI in radiology, a topic that they are currently underexposed to.<sup>23,24</sup>

There are drawbacks to AI chatbots and their potential role in radiology. As mentioned previously, the rationale chatbots provide for their answers can serve as a useful tool for trainees

and patients looking to gain knowledge, but a significant downside arises when erroneous responses are justified with confidence. Particularly, for a patient or trainee who does not have the prerequisite knowledge or experience, it would be easy to be led astray by the responses from these chatbots. Particularly in a field such as radiology, and more broadly medicine, the dangers of inaccuracy can be devastating. Furthermore, ChatGPT and Bard are designed to provide unique responses, even to the same prompts, suggesting that there may be times when questions may be answered correctly or incorrectly depending on the content of previous conversations with the chatbot. AI chatbots also have a tendency to make assumptions when insufficient information is provided which can have significant repercussions in the context of medical care.<sup>25</sup> Hallucinations, omissions, and errors have all been documented with the use of ChatGPT.<sup>26,27</sup> Our analysis of a subset of chatbot responses demonstrated that both chatbots contain errors in their answer explanations and do not always address the intended educational content of the question, with Bard performing worse than ChatGPT. Without a comprehensive understanding of the nuances and subtlety in radiology, AI chatbots should be used with caution. Lastly, thorough investigation into the safety and efficacy of AI technology both prior to adoption and after adoption is vital. A study by van Leeuwen et al. (2021) showed that of 100 AI products, only 36 had peer-reviewed evidence regarding efficacy.<sup>28</sup>

Limitations of our study include the following. Only text-based questions were included in our analysis which is not reflective of radiology as a field that is inherently image based. Further, this led to disproportionate subspecialty representation. For example, there was an increased representation of questions from the general & physics subsection and decreased representation from the chest radiology subsection. That being said, by focusing on text-based questions, our study provides specific insights into AI chatbots' radiology knowledge, which is separate from interpretation skills. In the future, multimodal input is anticipated for both these chatbots, and a repeat assessment and that point will provide important insights. Another limitation of our study is that given the sample size, there may be selection bias in the questions included in this study. A lack of long-term evaluation of these chatbots precludes an understanding of the reliability, consistency, and potential improvements of these chatbots. As the precise data used as training input for ChatGPT and Bard is unknown, it is possible that the publicly available dataset of DXIT questions we used may have been included in the chatbot training process. In a separate project, we found that in some cases, these chatbots cited the specific resource the question was based on (ACR appropriateness criteria), but nevertheless, still provided an incorrect answer.<sup>29</sup> On the other hand, further improvements in chatbot performance may be possible if more radiology specific content is provided as part of the training process in future updates to these models.

In conclusion, our study highlights some key differences between ChatGPT and Bard in radiology knowledge. Chatbot assumptions, confidence in response explanations, and are important factors to consider when choosing the optimal chatbot. Furthermore, future investigations into direct comparisons between AI chatbot performance and radiologist or radiology trainees' performance would be useful. As AI chatbots, and more broadly AI models, continue to develop and learn, continual reassessment of their strengths, weakness, and role in radiology is warranted.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Nikhil S. Patil  <https://orcid.org/0000-0003-3929-0482>  
 Ryan S. Huang  <https://orcid.org/0000-0002-3404-5376>  
 Christian B. van der Pol  <https://orcid.org/0000-0002-1718-2619>  
 Natasha Larocque  <https://orcid.org/0000-0003-0449-9438>

### Supplemental Material

Supplemental material for this article is available online.

### References

1. ChatGPT. OpenAI; 2023. Accessed June 14, 2023. <https://chat.openai.com/>
2. MeetGoogle Bard. 2023. Accessed June 14, 2023. <https://bard.google.com/>
3. LaMDA. Our breakthrough conversation technology. *Meta*; 2023. Accessed June 14, 2023. <https://blog.google/technology/ai/lamda/>
4. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. 2023;104(6):269-274. doi:10.1016/J.DIII.2023.02.003
5. Elkassem AA, Smith AD. Potential use cases for ChatGPT in radiology reporting. *Am J Roentgenol*. 2023. Published online April 5, 2023. doi:10.2214/AJR.23.29198
6. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an adjunct for radiologic decision-making. *medRxiv*. 2023;221:12023-4. Published online February 7. doi:10.1101/2023.02.02.23285399
7. Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning. *Radiology*. 2023;307(5):e230987. Published online May 16, 2023. doi:10.1148/RADIOL.230987
8. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology Board-style examination: insights into current strengths and limitations. *Radiology*. 2023;307(5):e230582. Published online May 16, 2023. doi:10.1148/RADIOL.230582
9. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology*. 2023;307(5):e230922. doi:10.1148/RADIOL.230922
10. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral Boards preparation question Bank. *Neurosurgery*. 2023. Published online June 12, 2023. doi:10.1101/2023.04.06.23288265
11. Diagnostic radiology in-training (DXIT) exam. American College of radiology; 2023. Accessed June 10, 2023. <https://www.acr.org/Lifelong-Learning-and-CME/Learning-Activities/In-Training-Exams/Diagnostic-Radiology-In-Training-Exam>
12. Patel MD, Hunt KN, Benefield T, et al. The relationship Between ACR diagnostic radiology in-training examination scores and ABR Core examination outcome and performance: a multi-institutional study. *J Am Coll Radiol*. 2020;17(12):1663-1669. doi:10.1016/J.JACR.2020.04.032
13. Orton TH, McInnes M. Can American College of Radiology in-training examination scores be used to predict Canadian radiology licensing examination results? A respective study. *BMC Med Educ*. 2013;13(1):1-7. doi:10.1186/1472-6920-13-17/FIGURES/5
14. ACR radiology DXIT in-training in-service questions - Anki-Web. Anki; 2023. Accessed June 10, 2023. <https://ankiweb.net/shared/info/720532200>
15. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. doi:10.2196/45312
16. Jin D, Harrison AP, Zhang L, et al. Artificial intelligence in radiology. *Artif Intell Med*. 2021;265-289. Published online January 1, 2021. doi:10.1016/B978-0-12-821259-2.00014-4
17. Mohan C. Artificial intelligence in radiology – are we treating the image or the patient? *Indian J Radiol Imag*. 2018;28(2):137. doi:10.4103/IJRI.IJRI\_256\_18
18. Zand A, Sharma A, Stokes Z, et al. An exploration into the use of a chatbot for patients with inflammatory bowel diseases: retrospective cohort study. *J Med Internet Res*. 2020;22(5):e15589. doi:10.2196/15589
19. Kitamura FC. ChatGPT is shaping the future of medical writing But still requires human judgment. *Radiology*. 2023;307:e230171. Published online February 2, 2023. doi:10.1148/RADIOL.230171
20. Mese I. The impact of artificial intelligence on radiology education in the wake of coronavirus disease 2019. *Korean J Radiol*. 2023;24(5):478. doi:10.3348/KJR.2023.0278
21. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer*. 2021;7(4):e27850. doi:10.2196/27850
22. Lourenco AP, Slanetz PJ, Baird GL. Rise of ChatGPT: it may Be time to reassess how we teach and test radiology residents.

- Radiology*. 2023;307:e231053. Published online May 16, 2023. doi:[10.1148/RADIOL.231053](https://doi.org/10.1148/RADIOL.231053)
23. Khafaji MA, Safhi MA, Albadawi RH, Al-Amoudi SO, Shehata SS, Toonsi F. Artificial intelligence in radiology: are Saudi residents ready, prepared, and knowledgeable? *Saudi Med J*. 2022;43(1):53. doi:[10.15537/SMJ.2022.43.1.20210337](https://doi.org/10.15537/SMJ.2022.43.1.20210337)
  24. Barreiro-Ares A, Morales-Santiago A, Sendra-Portero F, Souto-Bayarri M. Impact of the rise of artificial intelligence in radiology: what do students think? *Int J Environ Res Publ Health*. 2023;20(2):1589. doi:[10.3390/IJERPH20021589/S1](https://doi.org/10.3390/IJERPH20021589/S1)
  25. Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords. *Radiology*. 2023; 307(2):2023. doi:[10.1148/RADIOL.230163](https://doi.org/10.1148/RADIOL.230163)
  26. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13): 1233-1239. doi:[10.1056/NEJMSR2214184](https://doi.org/10.1056/NEJMSR2214184).
  27. Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. *Can Assoc Radiol J*. 2023;8465371231171125. Published online April 20, 2023. doi:[10.1177/08465371231171125](https://doi.org/10.1177/08465371231171125)
  28. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol*. 2021;31(6):3797-3804. doi:[10.1007/S00330-021-07892-Z](https://doi.org/10.1007/S00330-021-07892-Z).
  29. Patil NS, Huang R, van der Pol C, Larocque N. Using AI chatbots as a radiologic decision-making tool for liver imaging: do ChatGPT and Bard communicate information consistent with the American College of Radiology Appropriateness Criteria? *J Am Coll Radiol*. 2023;31(6):3797-3804. Published online 2023. <https://doi.org/10.1016/j.jacr.2023.07.010>