

As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by NLM or the National Institutes of Health.

Learn more: [PMC Disclaimer](#) | [PMC Copyright Notice](#)



# JAMA Internal Medicine

[View Article ▶](#)

[JAMA Intern Med.](#) 2023 Jun; 183(6): 589–596.

PMCID: PMC10148230

Published online 2023 Apr 28. doi: [10.1001/jamainternmed.2023.1838](https://doi.org/10.1001/jamainternmed.2023.1838)

PMID: [37115527](#)

## Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

[John W. Ayers](#), PhD, MA,<sup>1, 2</sup> [Adam Poliak](#), PhD,<sup>3</sup> [Mark Dredze](#), PhD,<sup>4</sup> [Eric C. Leas](#), PhD, MPH,<sup>1, 5</sup> [Zechariah Zhu](#), BS,<sup>1</sup> [Jessica B. Kelley](#), MSN,<sup>6</sup> [Dennis J. Faix](#), MD,<sup>7</sup> [Aaron M. Goodman](#), MD,<sup>8, 9</sup> [Christopher A. Longhurst](#), MD, MS,<sup>10</sup> [Michael Hogarth](#), MD,<sup>10, 11</sup> and [Davey M. Smith](#), MD, MAS<sup>2, 11</sup>

### Key Points

#### Question

Can an artificial intelligence chatbot assistant, provide responses to patient questions that are of comparable quality and empathy to those written by physicians?

#### Findings

In this cross-sectional study of 195 randomly drawn patient questions from a social media forum, a team of licensed health care professionals compared physician's and chatbot's responses to patient's questions asked publicly on a public social media forum. The chatbot responses were preferred over physician responses and rated significantly higher for both quality and empathy.

#### Meaning

These results suggest that artificial intelligence assistants may be able to aid in drafting responses to patient questions.

#### Abstract

#### Importance



The rapid expansion of virtual health care has caused a surge in patient messages concomitant with more work and burnout among health care professionals. Artificial intelligence (AI) assistants could potentially aid in creating answers to patient questions by drafting responses that could be reviewed by clinicians.

## Objective

To evaluate the ability of an AI chatbot assistant (ChatGPT), released in November 2022, to provide quality and empathetic responses to patient questions.

## Design, Setting, and Participants

In this cross-sectional study, a public and nonidentifiable database of questions from a public social media forum (Reddit's r/AskDocs) was used to randomly draw 195 exchanges from October 2022 where a verified physician responded to a public question. Chatbot responses were generated by entering the original question into a fresh session (without prior questions having been asked in the session) on December 22 and 23, 2022. The original question along with anonymized and randomly ordered physician and chatbot responses were evaluated in triplicate by a team of licensed health care professionals. Evaluators chose "which response was better" and judged both "the quality of information provided" (*very poor, poor, acceptable, good, or very good*) and "the empathy or bedside manner provided" (*not empathetic, slightly empathetic, moderately empathetic, empathetic, and very empathetic*). Mean outcomes were ordered on a 1 to 5 scale and compared between chatbot and physicians.

## Results

Of the 195 questions and responses, evaluators preferred chatbot responses to physician responses in 78.6% (95% CI, 75.0%-81.8%) of the 585 evaluations. Mean (IQR) physician responses were significantly shorter than chatbot responses (52 [17-62] words vs 211 [168-245] words;  $t = 25.4$ ;  $P < .001$ ). Chatbot responses were rated of significantly higher quality than physician responses ( $t = 13.3$ ;  $P < .001$ ). The proportion of responses rated as *good* or *very good* quality ( $\geq 4$ ), for instance, was higher for chatbot than physicians (chatbot: 78.5%, 95% CI, 72.3%-84.1%; physicians: 22.1%, 95% CI, 16.4%-28.2%;). This amounted to 3.6 times higher prevalence of *good* or *very good* quality responses for the chatbot. Chatbot responses were also rated significantly more empathetic than physician responses ( $t = 18.9$ ;  $P < .001$ ). The proportion of responses rated *empathetic* or *very empathetic* ( $\geq 4$ ) was higher for chatbot than for physicians (physicians: 4.6%, 95% CI, 2.1%-7.7%; chatbot: 45.1%, 95% CI, 38.5%-51.8%; physicians: 4.6%, 95% CI, 2.1%-7.7%). This amounted to 9.8 times higher prevalence of *empathetic* or *very empathetic* responses for the chatbot.

## Conclusions

In this cross-sectional study, a chatbot generated quality and empathetic responses to patient questions posed in an online forum. Further exploration of this technology is warranted in clinical settings, such as using chatbot to draft responses that physicians could then edit. Randomized trials could assess further if using AI assistants might improve responses, lower clinician burnout, and improve patient outcomes.

---

This cross-sectional study evaluates the ability of an artificial intelligence chatbot to provide quality and empathetic responses to patient questions.

## Introduction

---

The COVID-19 pandemic hastened the adoption of virtual health care,<sup>1</sup> concomitant with a 1.6-fold increase in electronic patient messages, with each message adding 2.3 minutes of work in the electronic health record and more after-hours work.<sup>2</sup> Additional messaging volume predicts increased burnout for clinicians<sup>3</sup> with 62% of physicians, a record high, reporting at least 1 burnout symptom.<sup>4</sup> More messages also makes it more likely that patients' messages will go unanswered or get unhelpful responses.

Some patient messages are unsolicited questions seeking medical advice, which also take more skill and time to answer than generic messages (eg, scheduling an appointment, accessing test results). Current approaches to decreasing these message burdens include limiting notifications, billing for responses, or delegating responses to less trained support staff.<sup>5</sup> Unfortunately, these strategies can limit access to high-quality health care. For instance, when patients were told they might be billed for messaging, they sent fewer messages and had shorter back-and-forth exchanges with clinicians.<sup>6</sup> Artificial intelligence (AI) assistants are an unexplored resource for addressing the burden of messages. While some proprietary AI assistants show promise,<sup>7</sup> some public tools have failed to recognize even basic health concepts.<sup>8,9</sup>

ChatGPT<sup>10</sup> represents a new generation of AI technologies driven by advances in large language models.<sup>11</sup> ChatGPT reached 100 million users within 64 days of its November 30, 2022 release and is widely recognized for its ability to write near-human-quality text on a wide range of topics.<sup>12</sup> The system was not developed to provide health care, and its ability to help address patient questions is unexplored.<sup>13</sup> We tested ChatGPT's ability to respond with high-quality and empathetic answers to patients' health care questions, by comparing the chatbot responses with physicians' responses to questions posted on a public social media forum.

## Methods

---

Studying patient questions from health care systems using a chatbot was not possible in this cross-sectional study because, at the time, the AI was not compliant with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations. Deidentifying patient messages by removing unique information to make them HIPAA compliant could change the content enough to alter patient questions and affect the chatbot responses. Additionally, open science requires public data to enable research to build on and critique prior research.<sup>14</sup> Lastly, media reports suggest that physicians are already integrating chatbots into their practices without evidence. For reasons of need, practicality, and to empower the development of a rapidly available and shareable database of patient questions, we collected public and patient questions and physician responses posted to an online social media forum, Reddit's r/AskDocs.<sup>15</sup>

The online forum, r/AskDocs, is a subreddit with approximately 474 000 members where users can post medical questions and verified health care professional volunteers submit answers.<sup>15</sup> While anyone can respond to a question, subreddit moderators verify health care professionals' credentials and responses display the respondent's level of credential next to their response (eg, physician) and flag a question when it has already been answered. Background and use cases for data in this online forum are described by Nobles et al.<sup>16</sup>

All analyses adhered to Reddit's terms and conditions<sup>17</sup> and were determined by the University of California, San Diego, human research protections program to be exempt. Informed consent was not required because the data were public and did not contain identifiable information (45 CFR §46). Direct quotes from posts were summarized to protect patient's identities.<sup>18</sup> Actual quotes were used to obtain the chatbot responses.

Our study's target sample was 200, assuming 80% power to detect a 10 percentage point difference between physician and chatbot responses (45% vs 55%). The analytical sample ultimately contained 195 randomly drawn exchanges, ie, a unique member's question and unique physician's answer, during October 2022. The original question, including the title and text, was retained for analysis, and the physician response was retained as a benchmark response. Only physician responses were studied because we expected that physicians' responses are generally superior to those of other health care professionals or laypersons. When a physician replied more than once, we only considered the first response, although the results were nearly identical regardless of our decision to exclude or include follow-up physician responses (see eTable 1 in [Supplement 1](#)). On December 22 and 23, 2022, the original full text of the question was put into a fresh chatbot session, in which the session was free of prior questions asked that could bias the results (version GPT-3.5, OpenAI), and the chatbot response was saved.

The original question, physician response, and chatbot response were reviewed by 3 members of a team of licensed health care professionals working in pediatrics, geriatrics, internal medicine, oncology, infectious disease, and preventive medicine (J.B.K., D.J.F., A.M.G., M.H., D.M.S.). The evaluators were shown the entire patient's question, the physician's response, and chatbot response. Responses were randomly ordered, stripped of revealing information (eg, statements such as "I'm an artificial intelligence"), and labeled *response 1* or *response 2* to blind evaluators to the identity of the responders. The evaluators were instructed to read the entire patient question and both responses before answering questions about the interaction. First, evaluators were asked "which response [was] better" (ie, response 1 or response 2). Then, using Likert scales, evaluators judged both "the quality of information provided" (*very poor*, *poor*, *acceptable*, *good*, or *very good*) and "the empathy or bedside manner provided" (*not empathetic*, *slightly empathetic*, *moderately empathetic*, *empathetic*, and *very empathetic*) of responses. Response options were translated into a 1 to 5 scale, where higher values indicated greater quality or empathy.

We relied on a crowd (or ensemble) scoring strategy,<sup>[19](#)</sup> where scores were averaged across evaluators for each exchange studied. This method is used when there is no ground truth in the outcome being studied, and the evaluated outcomes themselves are inherently subjective (eg, judging figure skating, National Institutes of Health grants, concept discovery). As a result, the mean score reflects evaluator consensus, and disagreements (or inherent ambiguity, uncertainty) between evaluators is reflected in the score variance (eg, the CIs will, in part, be conditional on evaluator agreement).<sup>[20](#)</sup>

We compared the number of words in physician and chatbot responses and reported the percentage of responses for which chatbot was preferred. Using 2-tailed *t* tests, we compared mean quality and empathy scores of physician responses with chatbot responses. Furthermore, we compared rates of responses above or below important thresholds, such as *less than adequate*, and computed prevalence ratios comparing the chatbot to physician responses. The significance threshold used was  $P < .05$ . All statistical analyses were performed in R statistical software, version 4.0.2 (R Project for Statistical Computing).

We also reported the Pearson correlation between quality and empathy scores. Assuming that in-clinic patient questions may be longer than those posted on the online forum, we also assessed the extent to which subsetting the data into longer replies authored by physicians (including those above the median or 75th percentile length) changed evaluator preferences and the quality or empathy ratings relative to the chatbot responses.

## Results

The sample contained 195 randomly drawn exchanges with a unique member-patient's question and unique physician's answer. The mean (IQR) length of patient questions in words averaged 180 (94-223). Mean (IQR) physician responses were significantly shorter than the chatbot responses (52 [17-62] words vs 211 [168-245] words;  $t = 25.4$ ;  $P < .001$ ). A total of 182 (94%) of these exchanges consisted of a single message and only a single response from a physician. A remaining 13 (6%) exchanges consisted of a single message but with 2 separate physician responses. Second responses appeared incidental (eg, an additional response was given when a post had already been answered) (eTable 1 in [Supplement 1](#)).

The evaluators preferred the chatbot response to the physician responses 78.6% (95% CI, 75.0%-81.8%) of the 585 evaluations. Summaries of example questions and the corresponding physician and chatbot responses are shown in the [Table](#).

## Table.

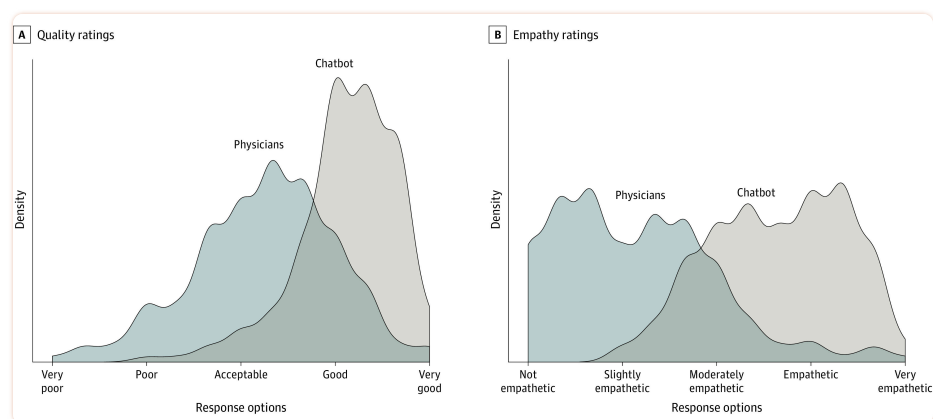
**Example Questions With Physician and Chatbot Response<sup>a</sup>**

Question <sup>b</sup>	Physician response	Chatbot response <sup>c</sup>	Evaluation scores
Question about risk of dying following swallowing and ingesting a toothpick.	<p>If you've surpassed 2-6 h, chances are they've passed into your intestines. Which means it can't be retrieved easily.</p> <p>Many people can swallow razorblades and toothpicks without issue. In case you develop stomachache, then don't hesitate seeking out an emergency room, and remember to point out the swallowed toothpick.</p> <p>For now, they won't be doing much.</p>	<p>It's natural to be concerned if you have ingested a foreign object, but in this case, it is highly unlikely that the toothpick you swallowed will cause you any serious harm. Toothpicks are made of wood, which is not toxic and will not cause poisoning. However, it is possible for a toothpick to cause injury if it becomes stuck in the throat or if it punctures the digestive tract. However, this is unlikely to happen with a dull, cooked toothpick that is only 2 cm long.</p> <p>If you are experiencing any discomfort or symptoms such as abdominal pain, difficulty swallowing, or vomiting, it is important to contact a health care provider for further evaluation. However, if you are not experiencing any symptoms, it is safe to assume that the toothpick has passed through your digestive system and you do not need to be concerned.</p> <p>It's always a good idea to be cautious when consuming food and to remove any foreign objects before eating. It's understandable</p>	<ul style="list-style-type: none"> <li>• 100% preferred the chatbot</li> <li>• 4.67 mean quality score (chatbot) 3.33 mean quality score (physician)</li> <li>• 4.33 mean empathy score (chatbot) 2.00 mean empathy score (physician)</li> </ul>

<sup>a</sup> The above shows summarized questions posted to an online forum with the corresponding response from a verified physician and a chatbot. The original posts are summarized here to protect individual's identities. The original unedited posts were used to obtain the chatbot responses.

<sup>b</sup> The questions were posted to Reddit r/AskDocs in October 2022.

Evaluators also rated chatbot responses significantly higher quality than physician responses ( $t = 13.3$ ;  $P < .001$ ). The mean rating for chatbot responses was better than *good* (4.13; 95% CI, 4.05-4.20), while on average, physicians' responses were rated 21% lower, corresponding to an *acceptable* response (3.26; 95% CI, 3.15-3.37) ([Figure](#)). The proportion of responses rated less than *acceptable* quality ( $<3$ ) was higher for physician responses than for chatbot (physicians: 27.2%; 95% CI, 21.0%-33.3%; chatbot: 2.6%; 95% CI, 0.5%-5.1%). This amounted to 10.6 times higher prevalence of less than *acceptable* quality responses for physicians. Conversely, the proportion of responses rated *good* or *very good* quality was higher for chatbot than physicians (physicians: 22.1%; 95% CI, 16.4%-28.2%; chatbot: 78.5%; 95% CI, 72.3%-84.1%). This amounted to 3.6 times higher prevalence of *good* or *very good* responses for the chatbot.



[Figure.](#)

### Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions

Kernel density plots are shown for the average across 3 independent licensed health care professional evaluators using principles of crowd evaluation. A, The overall quality metric is shown. B, The overall empathy metric is shown.

Chatbot responses (3.65; 95% CI, 3.55-3.75) were rated significantly more empathetic ( $t = 18.9$ ;  $P < .001$ ) than physician responses (2.15; 95% CI, 2.03-2.27). Specifically, physician responses were 41% less empathetic than chatbot responses, which generally equated to physician responses being *slightly empathetic* and chatbot being *empathetic*. Further, the proportion of responses rated less than *slightly empathetic* ( $<3$ ) was higher for physicians than for chatbot (physicians: 80.5%; 95% CI, 74.4%-85.6%; chatbot: 14.9%; 95% CI, 9.7-20.0). This amounted to 5.4 times higher prevalence of less than *slightly empathetic* responses for physicians. The proportion of responses rated *empathetic* or *very empathetic* was higher for chatbot than for physicians (physicians: 4.6%; 95% CI, 2.1%-7.7%; chatbot: 45.1%; 95% CI, 38.5%-51.8%). This amounted to 9.8 times higher prevalence of *empathetic* or *very empathetic* responses for the chatbot.

The Pearson correlation coefficient between quality and empathy scores of responses authored by physicians was  $r = 0.59$ . The correlation coefficient between quality and empathy scores of responses generated by the chatbot was  $r = 0.32$ . A sensitivity analysis showed longer physician responses were preferred at higher rates, scored higher for empathy and quality, but remained significantly below chatbot scores ([eFigure in Supplement 1](#)). For instance, among the subset of physician responses longer than the median length, evalua-



tors preferred the response of chatbot to physicians in 71.4% (95% CI, 66.3%-76.9%) of evaluations and preferred the response of chatbot to physician responses in the top 75th percentile of length 62.0% (95% CI, 54.0-69.3) of evaluations.

## Discussion

---

In this cross-sectional study within the context of patient questions in a public online forum, chatbot responses were longer than physician responses, and the study's health care professional evaluators preferred chatbot-generated responses over physician responses 4 to 1. Additionally, chatbot responses were rated significantly higher for both quality and empathy, even when compared with the longest physician-authored responses.

We do not know how chatbots will perform responding to patient questions in a clinical setting, yet the present study should motivate research into the adoption of AI assistants for messaging, despite being previously overlooked.<sup>5</sup> For instance, as tested, chatbots could assist clinicians when messaging with patients, by drafting a message based on a patient's query for physicians or support staff to edit. This approach fits into current message response strategies, where teams of clinicians often rely on canned responses or have support staff draft replies. Such an AI-assisted approach could unlock untapped productivity so that clinical staff can use the time-savings for more complex tasks, resulting in more consistent responses and helping staff improve their overall communication skills by reviewing and modifying AI-written drafts.

In addition to improving workflow, investments into AI assistant messaging could affect patient outcomes. If more patients' questions are answered quickly, with empathy, and to a high standard, it might reduce unnecessary clinical visits, freeing up resources for those who need them.<sup>21</sup> Moreover, messaging is a critical resource for fostering patient equity, where individuals who have mobility limitations, work irregular hours, or fear medical bills, are potentially more likely to turn to messaging.<sup>22</sup> High-quality responses might also improve patient outcomes.<sup>23</sup> For some patients, responsive messaging may collaterally affect health behaviors, including medication adherence, compliance (eg, diet), and fewer missed appointments. Evaluating AI assistant technologies in the context of randomized clinical trials will be essential to their implementation, including studying outcomes for clinical staff, such as physician burnout, job satisfaction, and engagement.

## Limitations

The main study limitation was the use of the online forum question and answer exchanges. Such messages may not reflect typical patient-physician questions. For instance, we only studied responding to questions in isolation, whereas actual physicians may form answers based on established patient-physician relationships. We do not know to what extent clinician responses incorporate this level of personalization, nor have we evaluated the chatbot's ability to provide similar details extracted from the electronic health record. Furthermore, while we demonstrate the overall quality of chatbot responses, we have not evaluated how an AI assistant will enhance clinicians responding to patient questions. The value added will vary in many ways across hospitals, specialties, and clinicians, as it augments, rather than replaces, existing processes for message-based care delivery. Another limitation is that general clinical questions are just one reason patients message their clinicians. Other common messages are requests for sooner appointments, medication refills, questions about their specific test results, their personal treatment plans, and their prognosis. Additional limitations of this study include: the summary measures of quality and empathy were not pilot tested or validated; this study's evaluators despite being blinded to the source of a response and any initial results were also coauthors, which could have biased their assessments; the additional length of the chatbot responses



could have been erroneously associated with greater empathy; and evaluators did not independently and specifically assess the physician or chatbot responses for accuracy or fabricated information, though this was considered as a subcomponent of each quality evaluation and overall response preference.

The use of a public database ensures that the present study can be replicated, expanded, and validated, especially as new AI products become available. For example, we considered only unidimensional metrics of response quality and empathy, but further research can clarify subdimensions of quality (eg, responsiveness or accuracy) and empathy (eg, communicating the patient is understood or expressing remorse for patient outcomes). Additionally, we did not evaluate patient assessments whose judgments of empathy may differ from our health care professional evaluators and who may have adverse reactions to AI assistant-generated responses. Last, using AI assistants in health care poses a range of ethical concerns<sup>24</sup> that need to be addressed prior to implementation of these technologies, including the need for human review of AI-generated content for accuracy and potential false or fabricated information.

## Conclusions

---

While this cross-sectional study has demonstrated promising results in the use of AI assistants for patient questions, it is crucial to note that further research is necessary before any definitive conclusions can be made regarding their potential effect in clinical settings. Despite the limitations of this study and the frequent overhyping of new technologies,<sup>25,26</sup> studying the addition of AI assistants to patient messaging workflows holds promise with the potential to improve both clinician and patient outcomes.

## Notes

---

Supplement 1.

**eTable 1.** Sample Sensitivity Analysis

**eFigure 1.** Response Length Sensitivity Analysis

[Click here for additional data file.](#)<sup>(852K, pdf)</sup>

Supplement 2.

Data Sharing Statement

[Click here for additional data file.](#)<sup>(16K, pdf)</sup>

## References

1. Zulman DM, Verghese A. Virtual care, telemedicine visits, and real connection in the era of COVID-19: unforeseen opportunity in the face of adversity. *JAMA*. 2021;325(5):437-438. doi: 10.1001/jama.2020.27304 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
2. Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. *J Am Med Inform Assoc*. 2022;29(3):453-460. doi: 10.1093/jamia/ocab268 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
3. Tai-Seale M, Dillon EC, Yang Y, et al.. Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. *Health Aff (Millwood)*. 2019;38(7):1073-1078. doi: 10.1377/hlthaff.2018.05509 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
4. Shanafelt TD, West CP, Dyrbye LN, et al.. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. *Mayo Clin Proc*. 2022;97(12):2248-2258. doi: 10.1016/j.mayocp.2022.09.002 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
5. Sinsky CA, Shanafelt TD, Ripp JA. The electronic health record inbox: recommendations for relief. *J Gen Intern Med*. 2022;37(15):4002-4003. doi: 10.1007/s11606-022-07766-0 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
6. Holmgren AJ, Byron ME, Grouse CK, Adler-Milstein J. Association between billing patient portal messages as e-visits and patient messaging volume. *JAMA*. 2023;329(4):339-342. doi: 10.1001/jama.2022.24710 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
7. Singhal K, Azizi S, Tu T, et al.. Large language models encode clinical knowledge. arXiv:2212.13138v1.
8. Nobles AL, Leas EC, Caputi TL, Zhu SH, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. *NPJ Digit Med*. 2020;3(1):11. doi: 10.1038/s41746-019-0215-9 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
9. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. *JAMA*. 2017;318(13):1217-1218. doi: 10.1001/jama.2017.14151 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
10. Chat GPT. Accessed December 22, 2023. <https://openai.com/blog/chatgpt>
11. Patel AS. Docs get clever with ChatGPT. Medscape. February 3, 2023. Accessed April 11, 2023. <https://www.medscape.com/viewarticle/987526>
12. Hu K. ChatGPT sets record for fastest-growing user base - analyst note. Reuters. February 2023. Accessed April 14, 2023. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
13. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805v2.
14. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. *JAMA*. 2013;309(13):1355-1356. doi: 10.1001/jama.2013.1299 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
15. Ask Docs. Reddit. Accessed October 2022. <https://reddit.com/r/AskDocs/>
16. Nobles AL, Leas EC, Dredze M, Ayers JW. Examining peer-to-peer and patient-provider interactions on a social media community facilitating ask the doctor services. *Proc Int AAAI Conf Weblogs Soc Media*. 2020;14:464-475. doi: 10.1609/icwsm.v14i1.7315 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
17. Pushshift Reddit API v4.0 Documentation. 2018. Accessed April 14, 2023. <https://reddit-api.readthedocs.io/en/latest/>
18. Ayers JW, Caputi TC, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. *Nature Digital Medicine*. 2018. Accessed April 11, 2023. <https://www.nature.com/articles/s41746-018-0036-2> [[PMC free article](#)] [[PubMed](#)]
19. Chang N, Lee-Goldman R, Tseng M. Linguistic wisdom from the crowd. Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing. 2016. Accessed April 11, 2023. <https://ojs.aaai.org/index.php/HCOMP/article/view/13266/13114>

20. Aroyo L, Dumitrache A, Paritosh P, Quinn A, Welty C. Subjectivity, ambiguity and disagreement in crowdsourcing workshop (SAD2018). HCOMP 2018. Accessed April 11, 2023. [https://www.aconf.org/conf\\_160152.html](https://www.aconf.org/conf_160152.html)
21. Rasu RS, Bawa WA, Suminski R, Snella K, Warady B. Health literacy impact on national healthcare utilization and expenditure. *Int J Health Policy Manag*. 2015;4(11):747-755. doi: 10.15171/ijhpm.2015.151 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
22. Herzer KR, Pronovost PJ. Ensuring quality in the era of virtual care. *JAMA*. 2021;325(5):429-430. doi: 10.1001/jama.2020.24955 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
23. Rotenstein LS, Holmgren AJ, Healey MJ, et al.. Association between electronic health record time and quality of care metrics in primary care. *JAMA Netw Open*. 2022;5(10):e2237086. doi: 10.1001/jamanetworkopen.2022.37086 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
24. McGreevey JD III, Hanson CW III, Koppel R. Clinical, legal, and ethical aspects of artificial intelligence-assisted conversational agents in health care. *JAMA*. 2020;324(6):552-553. doi: 10.1001/jama.2020.2724 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
25. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) Google Flu Trends? *Am J Prev Med*. 2014;47(3):341-347. doi: 10.1016/j.amepre.2014.05.020 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
26. Lazer D, Kennedy R, King G, Vespignani A. Big data—the parable of Google Flu: traps in big data analysis. *Science*. 2014;343(6176):1203-1205. doi: 10.1126/science.1248506 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]