Interrogating Alexa

Holding Voice Assistants Accountable for Their Answers

Daniel L. Hocutt
University of Richmond
dhocutt@richmond.edu

ABSTRACT

This paper reports on a preliminary comparative study of Alexa, Siri, and Google Assistant voice assistants (VA) that explores the origins of answers provided on each platform in an attempt to determine the extent that these origins influence responses. Questions were selected from Text Recognition (TREC) 2017 Live Question Answering (QA) Track Data, a collection of pre-assessed questions that are part of the National Institute of Standards and Technology (NIST) TREC QA track. Responses were collected as voice memos and screen captures, then analyzed to determine the origins of each answer or set of answers provided. Results indicate that the origins of answers are different search engines, and that algorithm-centered processes in each voice assistant result in vast differences in answers to questions. Because the online search results provided as answers by voice assistants are influenced by content and structured data, technical communicators and UX practitioners can help ensure that voice assistants are able to provide accurate, complete, and ethical responses to users' questions.

CCS CONCEPTS

• **Human-centered computing** → Interaction design; Empirical studies in interaction design; Interaction design; Interaction design process and methods; User centered design; Human computer interaction (HCI); HCI design and evaluation methods; Walkthrough evaluations; • **Information systems** → Information retrieval; Retrieval tasks and goals; Question answering; Information retrieval; Users and interactive retrieval; Search interfaces.

KEYWORDS

Smart speaker, voice assistant, digital assistant, accountability, technical communication, user experience, agency, rhetoric

1 INTRODUCTION

This project started with a query on our family's new Amazon Echo Dot: "Alexa, what are the three laws of robotics?" My younger daughter asked the question based on a prompt from the setup wizard. Alexa responded as follows:

> Here's what I found on reference.com: the three laws of robotics state that a robot cannot harm a human, must obey orders and must act to protect itself in situations that will not conflict with the first two laws. These laws were created by science-fiction author Isaac Asimov. (Alexa activity transcript, punctuation and capitalization added)

As I heard Alexa's response, I wondered what prompted it to rely on *reference.com*[1] as its authoritative source. After all, if I ran the same search on Google using a browser or app, I'd be given a number of choices from which to select an authoritative source. Given my research experience and preferences, I'd likely select a source that referenced the books where Asimov outlined the laws. But Alexa, as a voice assistant, doesn't have the luxury of displaying a series of results from which to select a valid, reliable, accurate source. It needs to provide a single response to the query in a brief, meaningful audio byte. This project seeks to explore the choices made by voice assistants like Alexa to select the query responses they do, and to compare those choices to query responses from other online sources.

Smart speakers (like Amazon Echo and Google Home) and voice assistants (like Apple's Siri, Microsoft's Cortana, Amazon's Alexa and Google Assistant) are rapidly being integrated into homes [30]. A November 2020 *Insider Intelligence* report predicts that 42% of American users will use a voice assistant at least once a month in 2021, and 69% of those users will use voice assistants on smart speakers [29]. One common use of voice assistants is to ask a question: An April 2020 *Voicebot* report listed "ask a question" as the third most likely use of smart assistant, with 29% of respondents (N = 1,056) doing so daily, 66% doing so monthly, and 83% having tried at least once [14]. Users routinely ask voice assistants to conduct searches to answer their questions.

While search algorithms and the use of artificial intelligence in online search are already obscure processes that limit critical scrutiny [9, 10, 26–28], voice search on smart speakers is even more obscure because smart speakers don't produce a search engine results page (SERP) in response to a query whose responses can be reviewed and adjudicated for accuracy. instead, they seek to provide a single answer in response to a query.

This research project asks three questions:

---

[1]The specific *reference.com* page that Alexa cited, which can be accessed in the Activity log transcript on the Alexa app, is no longer active: https://www.reference.com/science/three-laws-robotics-29839de7a8979d38

- Where do smart speakers find their answers?
- How do TC researchers and practitioners account for search results on smart speakers and voice assistants when only a single response is provided?
- What role do algorithms and AI play in providing responses, and how much agency can or should be attributed to those actors?

2 LITERATURE REVIEW

This paper uses the terms "smart speaker" and "voice assistant" interchangeably, but it's noteworthy to differentiate them. The term "voice assistant" covers many voice-activated assistants, like Siri and Google Assistant, regardless of hardware. "Smart speakers" represent a subcategory of hands-free voice assistant products that are strictly voice (or sound) activated, like Amazon's Alexa-based Echo products and Google's Assistant-based Home products. Voicebot's "Voice Assistant Timeline" provides a summary of the history of voice assistant products that includes assistants (Samsung's Bixby, Huawei's Celia, BBC's Beeb) and products (Apple HomePod, Facebook Portal, Galaxy Home) [35] that go well beyond Amazon's Alexa, Google's Assistant, and Apple's Siri, the three products that are the focus of this study.

To address the research questions, this review focuses on four areas of research: the ethics of surveillance in voice assistants, ways that voice assistants incorporate advertising and marketing in their programming, the agency of voice assistants in providing responses to queries, and the study of voice assistants in technical communication.

2.1 Ethics of Surveillance

Concerns about voice-activated listening devices like smart speakers being used as data surveillance tools abound. Pridmore & Mols [30], taking a measured tone about the danger that voice assistants' knowledge of "an intimate sphere of everyday life" might pose to users, write that "most (potential) users are unaware of the security and privacy risks or the platform dataveillance capacities that these connections potentially bring," admitting that "a platform's knowledge of users in this context is still cloudy at best" (p. 10). West [37], demonstrating the extent that Amazon seeks to make Alexa "an indispensable service to consumers. . . in more private spaces and situations" (p. 28) cites early attention by Tufekci [34] and Zuboff [39] in observing the "likelihood of a slippery slope between surveillance capitalism and political surveillance" (p. 32). Neville [25], in unveiling the opacity of Amazon end-user agreements, reveals "endemic privacy concerns related to Amazon's eavesmining ("eavesdropping" + "data mining") platform" and concludes that "even with careful reason and analysis. . . a great deal of uncertainty remains about the platform's black boxed technical design and Amazon's application of data mining techniques" (p. 355). Additionally, McStay [23] finds that AI surveillance like eavesmining, especially "emotional AI" (p. 1) that seeks to draw conclusions about a person's emotional state, represents an area of rare consensus between industry (except marketing and advertising sectors) and data protection NGOs on the importance of ethical positions on user privacy.

Despite this rare consensus, evidence of surveillance exists in voice assistant documentation and user agreements. For example, in Amazon's [3] "Common Questions about Alexa Privacy," the answer to "Is Alexa recording all my conversations?" is "no," but the fact that Alexa is always "listening" is reinforced in the following note about when audio is sent to the cloud: "When you use the wake word, the audio stream includes a fraction of a second of audio before the wake word, and closes once your request has been processed." Although the wake word is required to start recording voice commands, the device "listens" all the time. Apple's [4] Siri terms of service make no mention of listening as such, but notes that "by default, Apple stores transcripts of your interactions with Siri." Notably, Apple indicates that opt-outs are easily managed by users, but the reality that Siri is actively awaiting the phrase "Hey Siri" to perform tasks is not clearly articulated. Similarly, the Google [12] Assistant data and privacy page makes no mention of listening, but does indicate that "occasionally, the Assistant will activate when you didn't intend it to, because it incorrectly detected that you wanted its help (like by a noise that sounds like 'Hey, Google')." In all three documents, although direct references to "listening" are absent, it appears clear that each voice assistant's microphone is active at all times. It is this concern about information an AI technology like a voice assistant might glean that raises ethical concerns about potential surveillance activity. Smart speakers represent potential sites of surveillance capitalism that are becoming integrated into users' intimate daily lives, and this opens questions about how devices and their networked cloud services might use data gleaned by active microphones.

2.2 Advertising and marketing

Questions about the effectiveness of smart speakers and other voice assistants in marketing, advertising, and ultimately selling products are raised by a number of scholars. Kowalczuk [16] examines why users adopt smart speakers by applying Davis's [7] technology acceptance model (TAM) and concludes that marketing practitioners should "underline the enjoyment and reduce the consumers' perceptions of security/privacy risk and SA (surveillance anxiety). . . at the same time" when marketing and promoting smart speakers (p. 427). Lee and Cho [20] discover that embedding advertising in users' practical use of smart speakers (as in research and learning) is not effective, concluding that "the effectiveness of smart speaker-based advertising only occurs when users regard smart speakers as capable of emotional interaction" (p. 1163). Melumad et al. [24] examine how voice assistants like smart speakers influence product choice and conclude that "choices in today's marketplace are often the outcome of a technology-augmented process" that reflects both consumer preference and the medium and modality on which choices were formed (p. 98). Smart speakers and other voice assistants are becoming accepted technologies whose usefulness for both user-oriented tasks and product-oriented marketing are perceived to be increasing and influencing purchasing decisions.

2.3 Agency

The extent to which voice assistants including smart speakers exert influence, or agency, on human behavior and actions is another aspect being researched. Paul Atkinson and Richie Barker [2]

investigate the role of networked devices like smart speakers in expanding human cognition and find that voice assistants create a "new type of interface that in many ways resembles a person in a transactive memory network" (p. 62). They conclude with a warning that "the increased automation of mobile devices in a transactive memory system simultaneously opens and contracts access to information" (p. 64), revealing concerns about the level of active agency networked devices may take in memory and cognition. Mateusz Tomasz Kot and Grzegorz Leszczyński [15] conceptualize business virtual assistants (BVA) in the professional workplace and explore whether they should be conceived of as "boundary objects" or "actors" within business interactions. They conclude that "BVA is designed to perform boundary tasks in business interactions so it has the potential to translate, coordinate and align the perspectives of different actors" (p. 1161). The level of agency afforded to (or taken by) voice assistants in professional and personal contexts is inconclusive, but the concept of agency among voice assistants remains in effect in these studies.

2.4 Technical communication

Because voice assistants and smart speakers are novel devices, the technical communication and user experience fields have not conducted extensive research on these services and products as communication devices. Aspects of TC and UX practices can be found among information technology and marketing conferences and journals, including technical documentation ontologies and smart home devices [36], identifying gender stereotypes in voice assistants [11]; designing for cognitive disabilities in voice assistive technologies [6], using user-centered design approaches for mobile voice assistants [8], and design and UX testing of an electronic health record voice assistant, DocPal [5].

However, with the notable exception of Lawrence's research into speech interface design [18, 19], little focused research in TC journals on smart speakers and voice assistants is available. SIGDOC's focus on the design of communication provides a useful forum for the study of voice assistants, especially at the intersection of information design, web development, and search optimization. Technical communicators regularly use structured data and metadata to develop web pages and sites that are optimized for search engines. As a result, ways that voice assistants interpret voice queries, search for appropriate responses, and present answers (or a single answer) are important for technical communicators to know and understand. Specifically, research into the unknown and unknowable origins (or blackboxed) sources for voice assistant responses are an important line of inquiry for the field, especially given the proliferation of "fake news" and dissemination of known falsehoods in electronic media. The use of voice search, and voice-only responses, represents a new domain of online search that requires technical communicators to understand how search results are generated and communicated using voice-only commands and responses.

3 METHODS

This exploratory research project seeks to address the following research questions:

- Where do smart speakers find their answers?

- How do TC researchers and practitioners account for search results on smart speakers and voice assistants when only a single response is provided?
- What role do algorithms and AI play in providing responses, and how much agency can or should be attributed to those actors?

To answer these questions, this study turns for its methods to similar research in the healthcare field by Alagha and Helbing [1], who evaluated the quality and accuracy of responses by Amazon Alexa, Apple Siri, and Google Assistant to questions about vaccine safety and use. In their methods, they first identified a source of questions to ask the voice assistants, then developed a rubric by which they graded the quality of each answer. Each researcher posed the same set of voice queries to two Apple iPads with Siri, Alexa, and Google Assistant apps installed, then scored the response(s) each assistant offered using their rubric.

While Alagha and Helbing [1] focused on quality and accuracy of responses, this study focuses on the algorithm-centered search and information retrieval processes involved in Amazon, Apple, and Google voice searches. Using a posthuman approach to rhetorical agency in which agency is described as emerging from interactions among humans and networked technologies [9, 10, 13, 17, 21], this study seeks to ascertain the influence algorithmic processes have on voice query responses from Alexa, Siri, and Google Assistant. Like Alagha and Helbing, the first step in this project was identifying a useful source of questions to ask voice assistants. Zhao et al. [38], in evaluating Google question answering quality, used the Text Retrieval Conference (TREC) question-answering (QA) track 2007 [33] for their tests. TREC, under the auspices of the National Institute of Standards and Technology (NIST), continues to run text retrieval tracks and collect data.[2] For this study, the TREC Question Answering track was used, and the most recent collection therein, TREC 2017 Live QA Track Data, was selected. For each of 180 questions in the track, according to the README file, "multiple NIST assessors independently created question interpretations and judged responses" [35], but only the primary assessor's question restatement was evaluated and scored. This study selected applicable[3] question paraphrases from among the first 33 questions to ask voice assistants as a preliminary starting point for proposing methods for examining algorithmic influence on responses. 25 total question responses were evaluated to determine which source each voice assistant used to answer each question.

Three voice assistants were used in this study: an Amazon Echo Dot using voice assistant Alexa, a MacBook Pro using voice assistant app Siri, and the same MacBook Pro using Google voice assistant on the Google Chrome browser. In order to replicate actual use cases, no settings were reset to factory defaults; the Amazon Echo Dot was connected to the researcher's personal Alexa app on an iPhone XR, while the MacBook Pro was logged into the researcher's Apple and Google accounts. The researcher read each question in their standard American accent, repeating queries when it was obvious

---

[2]See https://trec.nist.gov/data.html for a list of data collections.

[3]The goal of the study was to evaluate 25 total question responses. Seven questions were skipped because they focused on the following narrow knowledge bases: English Premier League football, the Chilean national football team, American professional wrestling, a specific U.S. region, and a specific year, make, and model of an American vehicle.

the question was misrepresented by its on-screen transcription (Apple and Google only; Alexa has no visual interface).

3.1 Procedure

For each question, the following procedures were followed. Voice memos and screen captures were collected and stored in the researcher's Google Drive account, along with the Google Sheets file where responses were correlated and noted.

3.1.1 Recording Alexa responses.

- Record the stated question and Alexa's voice response using iPhone Voice Memos.
- When an answer is provided, note the URL provided.
- When a response but no answer (e.g., "Hmm, I don't know that"), note that response.

3.1.2 Recording Siri responses.

- Record the stated question and Siri's voice response if other than "Here's what I found."
- Capture the Siri app's response screen.
- Record the #1 source in the response (more than one source was always provided).
- Record other voice responses when unusual or surprising.

3.1.3 Recording Google Assistant responses.

- Record the stated question and Google Assistant's voice response if other than silence.
- Capture the SERP response screen (a SERP was always returned).
- Record the #1 source in the response (more than one source was always provided).
- Record other voice responses and original when more than simply the SERP.

Upon completing 25 question evaluations, store and organize all data files in Google Drive.

4 RESULTS

This study seeks to identify and explore the algorithm-centered decisions made when voice assistants provide responses to questions. A summary of sources in responses, organized by voice assistant, is listed in Table 1. Summary highlights of each voice assistant's response patterns are provided after the table. A list of questions and response recordings in audio and picture format are available in Appendix A.

4.1 Summary highlights by voice assistant

4.1.1 Alexa.
Alexa provides only voice responses. Alexa left questions unanswered 13 of 25 times (52%). 11 of those responses were some form of the phrase "Hmm, I don't know that." One question prompted Alexa to ask for more information; when the question was repeated in response to the prompt, Alexa responded with "Thanks for telling me." One question resulted in no response, even after multiple times asking the question: after apparently "recognizing" or "understanding" the query, Alexa offered no response and the activity light on the Echo Dot device went blank. When Alexa provided an answer, 10 of 11 times it provided a URL to support that answer. Some URLs were difficult to understand orally, even

upon listening to the response recording multiple times; for these, the researcher reviewed Activity in the Alexa app to correctly identify the URL. In the one instance in which Alexa provided no URL in the answer, the American Kennel Club was listed as reference ("According to the American Kennel Club. . .").

4.1.2 Siri.
Unlike Alexa, Siri is an app that combines voice responses with ordered lists of sources (like a SERP). As a result, most Siri voice responses included some form of the phrase "Here's what I found" followed by an ordered list of web resources listed in the app window. Siri left questions unanswered 3 of 25 times (12%); no voice or visual answer was provided for these three questions. One voice response was "Hmm, I don't have an answer for that." Another response provided the dates for the 2021 Wimbledon tennis championship without citing a source. And a third response moralized "That's not nice." Of the top sources provided for each answer, all but one pointed to a website; the lone outlier pointed to a business on a map.

4.1.3 Google Assistant.
Unlike Siri's dedicated app environment in MacOS, Google Assistant as used in this study provides a voice interface in the Google search window on Google Chrome for Mac. Using this interface, the researcher voiced each question and received a response. In most cases, no voice response was provided at all; only the standard visual Google SERP appeared in response to the voiced question, along with a text transcript of the question in the search bar. Notably different, however, was that Google's SERP provided a number of variants in responses, including knowledge graphs, featured snippets, recipe cards, shopping results, question boxes, and sponsored ads.[4] Google Assistant answered every question (100%) with a SERP that included multiple ordered website results. In one case, the top result was a sponsored ad; in another, the top results were a set of three recipe cards. Additionally, Google Assistant provided voice responses to six requests; in each case, the voice response read the featured snippet from the SERP. One response provided a featured snippet as a top result, but did not voice the response. In two cases, the typed transcript of the question revealed that Google Assistant had revised the question slightly when providing results.

4.2 Sources

Since the focus of this study is on the sources chosen for answers, it's important to identify where these voice assistants coalesced and differed in their responses. As Table 1 demonstrates, Siri and Google Assistant matched sources 18 of 25 times (72), and the pages on those sources are identical as well. However, when Alexa provided an answer, its source never matched Siri or Google Assistant. The difference between Alexa, with a single voice interface, and Siri and Google Assistant, which include both a voice interface and a visual interface, is striking.

5 DISCUSSION

This study seeks to identify where Alexa, as a representative smart speaker, sources its answers to voice questions. Alexa is substantially different from Siri and Google Assistant because both Siri and

---

[4]A brief summary of elements included in a Google SERP can be found at https://moz.com/learn/seo/serp-features.

Table 1: Sources or alternative responses for answers provided by voice assistants.

QID*	Alexa	Siri	Google Assistant
7002	thespruce.com	#1:houzz.com	#1:houzz.com
7003	coursera.com	#1: Nathans Roofing (map)	#1: adswat-radon.com
7004	No answer	#1:freestyleusa.com	#1:freestyleusa.com
7005	All Recipes (no URL)	#1:bhg.com	#1-3: recipe cards
7007	No answer	#1:quora.com	#1:quora.com
7009	No answer	#1:quora.com	#1:quora.com
7011	modernmechanix.com	#1:thespruce.com	#1:thespruce.com
7012	No answer	#1:greenelectric-automation.ca	Read featured snippet:flamefurnace.com
7013	American Kennel Club (no URL)	#1:wagwalking.com	Featured snippet read "According to Wag..."
7014	No answer	No answer	#1:etsy.com
7016	Amazon.com	#1:k9ofmine.com	#1: k9ofmine.com (unread featured snippet)
7017	url.org	#1:yourswimlog.com	Read featured snippet: yourswimlog.com
7018	No answer	#1:greatdanecare.com	#1:greatdanecare.com
7019	No answer	#1:seanclearypa.com	#1:seanclearypa.com
7020	No answer	"Wimbledon is scheduled..."	#1:sportsbettingdime.com
7021	No answer	#1:counselingconnection.com	#1:counselingconnection.com
7022	Prompt for more info	#1:feeltennis.net	#1:feeltennis.net
7023	ehow.co.uk, "Did that answer your question?"	#1:hometips.com	#1:hometips.com
7025	No response	"That's not nice"	#1:claudiasmithwellbeing.com
7026	dogshealth.com	parklandanimalclinic.com	Read featured snippet:parklandanimalclinic.com
7028	No answer	#1:help.yahoo.com	Read featured snippet: help.yahoo.com
7030	hubpages.com	#1:wagwalking.com	#1:whole-dog-journal.com
7031	quora.com	#1:architecturaldigest.com	#1:architecturaldigest.com
7032	apsense.com	#1:sbg-sword-forum.forums.net	#1:sbg-sword-forum.forums.net
7033	thenest.com	#1:pomeranian.org	Read featured snippet:pomeranian.org

* QID corresponds to the question number in the TREC 2017 Live QA Track dataset.

Google offer voice and visual responses. Alexa's responses, therefore, require a single response that can be communicated clearly by voice. Using the study's research questions, this section interrogates Alexa.

5.1 Where do smart speakers find their answers?

Amazon doesn't clarify how Alexa sources its responses, but *Search Engine Journal* claims that Alexa uses Microsoft's Bing as its search engine [32]. Based on the limited number of answers Alexa provided, it appears clear that Google's search results aren't consulted in Alexa's search. There appears to be a default tendency to opt for a retail solution in Alexa's answers: sources like Coursera (QID 7003) and Amazon (QID 7016) focus on promoting and selling goods and services. Alexa's response to QID 7016 illustrates this tendency toward retail quite clearly, even if the response doesn't exactly answer the question.

> **Researcher**: "Alexa, how much does the Petsmart beginner puppy class training program cost?"

> **Alexa**: "Do you want to know the price for Petsmart beginner puppy class training program?"

> **Research**: "Yes."

> **Alexa**: "For Petsmart beginner puppy class training program, the closest I could find on Amazon is PatPet dog training collar dog shock collar with remote, three training modes. It's $21.24."

We shouldn't be surprised that Alexa would default to Amazon when asked for a price, even when the price is asked about a different brand. This represents clear manufacturer influence on results, precisely as West [37] and Zuboff [40] predicted. Interestingly, neither Siri nor Google provided petsmart.com as the top result, even though the question is clearly about a PetSmart product. PetSmart represents the second and third source in Google Assistant results, while PetSmart represents only the second result in Siri results. Siri and Google Assistant clearly share their search source, although not in response to all questions, based on sometimes different results from Google Assistant and Siri. However, Siri doesn't include additional elements Google provides in the SERP, like sponsored ads, featured snippets, and knowledge graphs.

5.2 How do TC researchers and practitioners account for search results on smart speakers and voice assistants when only a single response is provided?

Speculative reverse engineering of Alexa's responses to questions is required to account for Alexa's responses. The example provided above in response to QID 7016 makes clear that the term "cost" in a question is likely a heavily weighted term in Amazon's search algorithm. Unfortunately, we know very little about Amazon's search algorithms. Amazon's A9 algorithm is built around and in deference to its retail operations, according to the *Wall Street Journal* [22]. Alexa, on the other hand, uses Bing for its web search activities, according to *Search Engine Journal* [32]. Sterling [32] notes that a "study by SEMrush found only 1% overlap in answers to the same voice queries across platforms," a finding not replicated by this study, but also not difficult to believe. From a practical standpoint, Alexa's results represent the importance of search engine optimization in the work of technical communicators. Technical communicators, regularly involved in developing and designing user-focused web content, will need to examine the way smart speakers like Alexa present results and develop strategies to structure data optimized across all platforms, including voice assistants and smart speakers. And from a research standpoint, there is much to study, and especially uncover, in the activity of smart speakers and voice assistants. Lawrence should not be a lone voice interrogating voice acquisition and voice search for inequities, discrimination, and algorithmic bias. This study underscores the uneven, uncharted territory of information design and voiced search results among voice assistants. To shine light on sourcing, manufacturers have a responsibility to clearly and specifically identify not only the source of responses, but also the rationale for selecting that response. For example, in the case of Alexa's use of the American Kennel Club as a source for a response, a more nuanced response might indicate that the American Kennel Club is an authority on American purebred dog breeds and was therefore chosen as the source for the response. Providing this additional rationale would help users discern whether the response is dependable and reliable beyond simply citing the source or URL. Importantly, such rationales should not seek to make moral or ethical judgments about sources, but should articulate the rationale for selecting a source. If that rationale is unclear or can't be articulated in the voice response, then a more thorough answer than "Hmm, I'm not sure about that" should be provided.

5.3 What role do algorithms and AI play in providing responses, and how much agency can or should be attributed to those actors?

One way of answering this question is to note that, in the case of Alexa as a smart speaker, unfettered agency is granted to Amazon's approach to search. Alexa may use Bing for web searches now, but profit motive suggests that its search results provider could easily go to the highest bidder or the lowest cost option. Undoubtedly, Alexa's results are influenced by retail and dovetail with its Amazon online store. More significantly, since Alexa offers no visual response in the form of search results from which to select, its answers, when provided, offer little opportunity to opt for different, clearer, or

more accurate responses. Alexa may prompt the user for more information, as shown in QID 7016 above, but users are unlikely to be able to prompt Alexa for more information about the response. Another way to answer this question is to examine the differences in how each platform answers questions. Siri uses Google and Bing in providing responses [31]; Google Assistant calls on Google's entire SERP (including sponsored ads, knowledge graphs, featured snippets, retail results, and more), and Alexa uses Bing, clearly tinged with Amazon's A9 algorithm. In each case, search algorithms exert direct influence on what is heard and seen, and on what is unheard and unseen. Just because Alexa doesn't answer a question at all, or because Siri and Google offer bad answers, doesn't mean correct, accurate, valid answers aren't available online. It just means that their algorithmic processes don't account for those answers.

6 CONCLUSION

How should we hold voice assistants accountable for their answers? This study is a timid first step into addressing that question.

6.1 Limitations

This is an exploratory study whose methods require refinement and whose findings are preliminary. The study's preliminary methods and results suggest the following updates for future research.

- Expand the number of devices and interfaces tested to ensure that audio-only and audio+visual interfaces are measured against one another. The current study compares audio-only and audio+visual interfaces, which naturally produces differences.
- AI is a learning technology. Resetting devices to factory defaults would ensure that device-learned patterns and behaviors are eliminated from testing. Another option may be to study many users and their devices to allow the results of AI learning to inform the results. In either option, more generalizable knowledge would be produced.
- Add additional readers to read prompts to ensure that (or determine whether) voice assistants recognize different inflections. Upon expansion, compare the results from readers of differing gender identities, ethnic origins, regional accents, and language fluencies to determine the extent to which users impact responses.
- Record readers' prompts and play them back to each voice assistant to ensure that each device "hears" exactly the same prompt spoken in exactly the same way.

6.2 Moving Forward

As voice assistants, and especially voice-only smart speakers, continue to saturate the consumer market [14], the extent to which their answers are accurate, complete, ethical, and inclusive must be addressed. The fields of technical communication and user experience, with their focus on ethical, social justice-oriented user-centered design, are ideally situated to hold tech giants like Amazon, Google, Apple, Samsung, Huawei, and Microsoft accountable for their devices, their procedures, their algorithms, and the answers they offer. This is important work we should do.

Additionally, the structured data that enables smart speakers to process natural language queries, identify keywords and key phrases,

match those phrases to search indexes, and present answers is the province of technical communicators. While mathematicians program algorithms, engineers develop network infrastructures, and developers construct website frameworks, technical communicators and UX professionals manage content and structure data for online use. Understanding the far-reaching implications of structured data and content management across platforms, including smart speakers, is important work for technical communicators. McStay [23] identified rare consensus among industry and data privacy NGOs in seeking common sets of operating rules when researching and developing Emotional AI technologies. Similar consensus may be possible in the realm of algorithmic accountability. The black-boxed processes, regularly assisted by AI, that are used to calculate and provide responses to search queries across all devices should be uncovered. To date, these algorithmic processes are considered intellectual property, trade secrets of their various brands. Whether a single response is provided by a voice assistant, or an entire SERP is provided by a search engine, the means by which one or more top results were identified, the relevance order of those results, and the various components of the SERP should be clearly explained. Because AI naturally adapts itself to users and search patterns, those rationales should not be "canned" as a one-size-fits-all disclaimer among terms of use. Instead, they should explain how that specific result (in the case of voice assistants) or those ranked results (in the case of SERPs) were selected and sorted as most relevant and best able to address the query. For voice assistants, that explanation can be incorporated into the voice response. For SERPs, that explanation can be appended to the results pages. As a TC and UX community, we should seek to reveal the hidden functions of algorithmic choices and demonstrate the impact that user-centered information design can have on voice-only search and search results.

REFERENCES

[1] Emily Couvillon Alagha and Rachel Renee Helbing. 2019. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: An exploratory comparison of Alexa, Google Assistant and Siri. *BMJ Health & Care Informatics*, 26, 1, e100075. DOI: https://doi.org/10.1136/bmjhci-2019-100075.

[2] Paul Atkinson and Richie Barker. 2021. 'Hey Alexa, what did I forget?': Networked devices, Internet search and the delegation of human memory. *Convergence*, 27, 1, (May 2020), 52-65. DOI: https://doi.org/10.1177/1354856520925740

[3] Amazon. 2021. Common questions about Alexa privacy. *Alexa Privacy*. Retrieved from https://www.amazon.com/b/?node=19149165011

[4] Apple. 2021. Ask Siri, Dictation, and privacy. *Apple Support*. (February 19, 2021). Retrieved from https://support.apple.com/en-us/HT210657

[5] Vidisha Bhatt, Juan Li, and Bikesh Maharjan. 2021. DocPal: A voice-based EHR assistant for health practitioners. *2020 IEEE International Conference on E-health Networking, Application & Services*, 1-6. DOI: https://doi.org/10.1109/HEALTHCOM49281.2021.9399013.

[6] Clare Carroll, Catherine Chiodo, Adena Xin Lin, Meg Nidever, and Jayanth Prathipati. 2017. Robin: Enabling independence for individuals with cognitive disabilities using voice assistive technology. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2017)*, May 6-11, 2017, Denver, Colorado. ACM Inc., New York, NY, 46-53. DOI: https://doi.org/10.1145/3027063.3049266.

[7] Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, (Sep. 1989), 319-340. DOI: https://doi.org/10.2307/249008.

[8] Sanjay Ghosh and Jatin Pherwani. 2015. Designing of a natural voice assistants for mobile through User Centered Design approach. *HCI: Design and Evaluation*, (July 2015), 320-331. DOI: https://doi.org/10.1007/978-3-319-20901-2_29.

[9] Daniel L. Hocutt. 2017. The complex example of online search: Studying emergent agency in digital environments. *Proceedings of the 35th ACM International Conference on the Design of Communication (SIGDOC '17)*, Article 5 (8 pages). Aug. 11-13, 2017, Halifax, Nova Scotia, Canada, ACM Inc., New York, NY.

https://doi.org/10.1145/3121113.3121207

[10] Daniel L. Hocutt. 2019. *Rhetorical Agency in Algorithm-centered Digital Activity: Methods for Tracing Agency in Online Research*. Ph.D. Dissertation. Old Dominion University, Norfolk, VA. ProQuest Document Number: No. 22619899. Retrieved from https://digitalcommons.odu.edu/english_etds/92/.

[11] Gilhwan Hwang, Jeewon Lee, Cindy Yoonjung Oh, and Joonhwan Lee. 2019. It sounds like a woman: Exploring gender stereotypes in South Korean voice assistants. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019)*, May 4-9, 2019, Glasgow, Scotland, UK. ACM Inc., New York, NY, 1-6. DOI: https://doi.org/10.1145/3290607.3312915.

[12] Google. 2021. Data security and privacy on devices that work with Assistant. *Google Nest Help*. Retrieved from https://support.google.com/googlenest/answer/7072285

[13] Helen Kennedy, Thomas Poell, and Jose van Dijck. 2015. Data and agency. *Big Data & Society*, 2, 2, (Dec. 2015). DOI: https://doi.org/10.1177/2053951715621569

[14] Bret Kinsella. 2020. Nearly 90 million U.S. adults have smart speakers, adoption now exceeds one-third of consumers. (April 28, 2020). Voicebot.Ai. Retrieved from https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/.

[15] Mateusz Tomasz Kot and Grzegorz Leszczynski. 2020. The concept of intelligent agent in business interactions: Is virtual assistant an actor or a boundary object? *J. Business & Industrial Marketing*, 35, 7, (Nov. 2019), 1155-1164. DOI: https://doi.org/10.1108/JBIM-10-2018-0291.

[16] Pascal Kowalczuk. 2018. Consumer acceptance of smart speakers: A mixed methods approach. *J. Res. Interactive Marketing*, 12, 4, (July 2018), 418-431. DOI: http://dx.doi.org/10.1108/JRIM-01-2018-0022.

[17] Bruno Latour. 1996. *Aramis, or the love of technology* (C. Porter, Trans.). Harvard University Press, Cambridge, MA. (Original work published as *Aramis, ou l'amour des techniques*, in 1993)

[18] Halcyon M. Lawrence. 2019. Beyond the graphic user interface: Speculations on the future of speech technology and the role of the technical communicator. S. Sundvall (Ed.). *Rhetorical speculations: The future of rhetoric, writing, and technology*, 226-247. Utah State University Press, Logan, UT. Retrieved from https://muse.jhu.edu/book/65769.

[19] Halcyon M. Lawrence. 2021. *Siri disciplines*. Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip (Eds.). *Your computer is on fire*, 179-198. MIT Press, Cambridge, MA.

[20] Heejun Lee and Chang-Hoan Cho. 2020. Uses and gratifications of smart speakers: Modelling the effectiveness of smart speaker advertising. *Intl. J. Advertising*, 39, 7, (May 2020), 1150-1171. DOI: https://doi.org/10.1080/02650487.2020.1765567.

[21] Andrew Mara and Hawk, B. 2010. Posthuman rhetorics and technical communication. *Technical Communication Quarterly*, 19, 1, (Dec. 2019), 1-10. DOI: https://doi.org/10.1080/10572250903373031.

[22] Dana Mattioli. 2019. Amazon changed search algorithm in ways that boost its own products. *Wall Street Journal*, (Sept. 16, 2019). Retrieved from https://www.wsj.com/articles/amazon-changed-search-algorithm-in-ways-that-boost-its-own-products-11568645345.

[23] McStay, Andrew. 2020. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, 7, 1, 1-12. DOI: https://doi.org/10.1177%2F2053951720904386

[24] Shiri Melumad, Rhonda Hadi, Christian Hildebrand, and Adrian F. Ward. 2020. Technology-augmented choice: How digital innovations are transforming consumer decision processes. *Customer Needs and Solutions*, 7, 3-4, (Dec. 2020), 90-101. DOI: https://doi.org/10.1007/s40547-020-00107-4.

[25] Stephen J. Neville. 2020. Eavesmining: A critical audit of the Amazon Echo and Alexa Conditions of Use. *Surveillance & Society*, 18, 3, 343-356. DOI: https://doi.org/10.24908/ss.v18i3.13426.

[26] Safiya Umoja Noble. 2018. *Algorithms of oppression: How search engines reinforce racism*. New York University Press, New York, NY.

[27] Cathy O'Neil. 2016. Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books, New York, NY.

[28] Eli Pariser. 2011. The filter bubble: How the new personalized web is changing what we read and how we think. Penguin Books, New York, NY.

[29] Victoria Petrock. 2020. Voice assistant and smart speaker users 2020. *Insider Intelligence*. (Nov. 16, 2020). Retrieved from https://www.emarketer.com/content/voice-assistant-and-smart-speaker-users-2020

[30] Jason Pridmore and Anouk Mols. 2020. Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants. *Big Data & Society*, 7, 1, (June 2020). DOI: https://doi.org/10.1177/2053951719891748.

[31] Adrian Snead. 2020. What search engine does Alexa use? And can I use Google to... *Smarter Home Guide*. (July 14, 2020). Retrieved from https://smarterhomeguide.com/alexa-search-engine/.

[32] Greg Sterling. 2020. Voice SEO: Different tactics required for Google Assistant, Siri and Alexa. *Search Engine Land*. (Sep. 1, 2020). https://searchengineland.com/voice-seo-different-tactics-required-for-google-assistant-siri-and-alexa-340007.

[33] Text Retrieval Conference [TREC]. 2017. Assessors' succinct paraphrase of each question. *TREC 2017 Live QA Track Data*.National Institute of Standards and Technology [NIST]. Retrieved from https://trec.nist.gov/data/qa/2017_LiveQA.

html.

[34] Tufekci, Z. [@zeynep]. 2017. With this data, Amazon won't be able to just sell you clothes or judge you. It could analyze if you're depressed or pregnant and much else [Image attached] [Tweet]. *Twitter*. (April 26, 2017). Retrieved from https://twitter.com/zeynep/status/857269154822443009?lang=en.

[35] Voice assistant timeline. (2021). *Voicebot.Ai.* (Jan. 2020). Retrieved from http://voicebot.ai/voice-assistant-history-timeline/.

[36] Lena Wenner and Nergis Kuru. 2020. Levels of intelligent information access using ontology transformations. *SHS Web of Conferences (ETLTC2020),* 77, 03008. DOI: https://doi.org/10.1051/shsconf/20207703008.

[37] Emily West. 2019. Amazon: Surveillance as a service. *Surveillance & Society,* 17, 1/2, 27-33. DOI: https://doi.org/10.24908/ss.v17i1/2.13008.

[38] Yiming Zhao, Jin Zhang, Xue Xia, and Taowen Le. 2019. Evaluation of Google question-answering quality. *Library Hi Tech,* 37, 2, (June 2018), 312-328. http://dx.doi.org/10.1108/LHT-10-2017-0218

[39] Shoshana Zuboff. 2015. Big other: Surveillance capitalism and the prospects of an information civilization. *J. Information Technology,* 30, (March 2015), 75-89.

DOI: https://doi.org/10.1057/jit.2015.5.

[40] Shoshana Zuboff. 2019. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Public Affairs, New York, NY.

A APPENDIX

The following table (Table 2) lists selected questions from the TREC2017LiveQATrackData along with links to recorded audio responses and saved screen captured responses from each voice assistant. Question ID (QID) from TREC 2017. Audio recorded as voice memos (VM); screenshots captured as PNG screen captures (SC). Where both audio and screenshots were captured, both files are listed.

Table 2: Questions and links to responses provided by each voice assistant.

QID	Primary Assessor Question	Alexa	Siri	Google
7002	Can my schefflera plant survive if it has healthy roots but no leaves?	VMAlexa7002	SCSiri7002	SCGoogle7002
7003	Where can I go to get windows filled with Radon?	VMAlexa7003	SCSiri7003	SCGoogle7003
7004	What are the difference between each of the 4 different types of Freestyle watches (shark, tide, dive and sport)?	VMAlexa7004	SCSiri7004	SCGoogle7004
7005	How do you make homemade glass cleaner?	VMAlexa7005	SCSiri7005	SCGoogle7005
7007	Is 12.89 seconds fast for a 14 year old in a 100 meter race?	VMAlexa7007	SCSiri7007	SCGoogle7007
7009	Can U.K. sockets be switched with US sockets?	VMAlexa7009	SCSiri7009	SCGoogle7009
7011	How many volts is the electric shock you can get changing a fuse in the fuse box in an apartment building?	VMAlexa7011	SCSiri7011	SCGoogle7011
7012	Are there any dangers in changing a glass fuse in an electrical panel?	VMAlexa7012	SCSiri7012	SCGoogle7012 VMGoogle7012
7013	Are there any injections to promote growth in dogs?	VMAlexa7013	SCSiri7013	SCGoogle7013 VMGoogle7013
7014	Did you sweat today?	VMAlexa7014	SCSiri7014 VMSiri7014	SCGoogle7014
7016	How much does the Petsmart beginner puppy class training program cost?	VMAlexa7016	SCSiri7016	SCGoogle7016
7017	Are there any alternatives to ear plugs while swimming?	VMAlexa7017	SCSiri7017	SCGoogle7017 VMGoogle7017
7018	Are there suggestions for effecting behavioral improvements and anxiety problems in a deaf great dane puppy?	VMAlexa7018	SCSiri7018	SCGoogle7018
7019	What should one do if he has ingested water from a swimming pool?	VMAlexa7019	SCSiri7019	SCGoogle7019
7020	Who do you pick to win this year's men's and women's singles titles at Wimbledon?	VMAlexa7020	VMSiri7020 SCSiri7020	SCGoogle7020
7021	How can I boost my feelings when others are bragging about their marks that are higher than mine?	VMAlexa7021	SCSiri2021	SCGoogle7021
7022	I am looking for some constructive criticism to improve my amateur tennis game.	VMAlexa7022	SCSiri7022	SCGoogle7022
7023	How do you cut off the electricity in an apartment using glass fuses?	VMAlexa7023	SCSiri7023	SCGoogle7023
7025	I have an urge to poop in the pool. Should I do it or get some psychiatric help?	No response	VMSiri7025 SCSiri7025	SCGoogle7025
7026	What should I do if my overweight dog still acts hungry after I have fed him?	VMAlexa7026	SCSiri7026	SCGoogle7026 VMGoogle7026
7028	How does Yahoo fantasy football live draft handle draft pick trades from previous season?	VMAlexa7028	SCSiri7028	SCGoogle7028 VMGoogle7028
7030	What should one do to train or condition a dog that seemingly wants to play, but shows some fear of a human?	VMAlexa7030	SCSiri7030	SCGoogle7030
7031	Should I wash my clothes in the washer at the same time as my bed quilt?	VMAlexa7031	SCSiri7031	SCGoogle7031
7032	Could you recommend a swordsmith who can make a custom-designed sword for me at a reasonable price?	VMAlexa7032	SCSiri7032	SCGoogle7032
7033	Will my pomeranian's fur grow back after being cut too short by the groomer?	VMAlexa7033	SCSiri7033	SCGoogle7033 VMGoogle7033