SC|M

Studies in Communication and Media

FULL PAPER

"Alexa, Siri, Google, what do you know about corona?"
A quantitative survey of voice assistants and content analysis of their answers on questions about the COVID-19 pandemic

"Alexa, Siri, Google, was wisst ihr über Corona?"
Eine quantitative Befragung von Sprachassistenten und Inhaltsanalyse ihrer Antworten auf Fragen zur COVID-19 Pandemie

Katharina Frehmann, Marc Ziegele & Ulrich Rosar

Katharina Frehmann (M.A.), Heinrich-Heine-University Düsseldorf, Department of Social Sciences, Universitätsstraße 1, 40225 Düsseldorf, Germany. Contact: katharina.frehmann(at) hhu.de. ORCID: https://orcid.org/0000-0002-0226-5884

Marc Ziegele (Prof. Dr.), Heinrich-Heine-University Düsseldorf, Department of Social Sciences, Universitätsstraße 1, 40225 Düsseldorf, Germany. Contact: marc.ziegele(at)hhu.de. ORCID: https://orcid.org/0000-0002-2710-0955

Ulrich Rosar (Prof. Dr.), Heinrich-Heine-University Düsseldorf, Department of Social Sciences, Universitätsstraße 1, 40225 Düsseldorf, Germany. Contact: ulrich.rosar(at)hhu.de.

279

FULL PAPER

"Alexa, Siri, Google, what do you know about corona?"
A quantitative survey of voice assistants and content analysis of their answers on questions about the COVID-19 pandemic

"Alexa, Siri, Google, was wisst ihr über Corona?"
Eine quantitative Befragung von Sprachassistenten und Inhaltsanalyse ihrer Antworten auf Fragen zur COVID-19 Pandemie

Katharina Frehmann, Marc Ziegele & Ulrich Rosar

Abstract: The global corona crisis has increased people's information seeking and their use of voice assistants as providers of current information on the pandemic. However, little is known regarding the kind and quality of information that different voice assistants provide on corona-related topics. Adapting previous studies in the domain of medical research for communication research, the current study presents the results of a quantitative content analysis of the responses of smartphone-based versions of the voice assistants Amazon Alexa, Apple Siri, and Google Assistant to 25 corona-related questions between March and May 2020 (N = 603 question-answer-sets). The findings reveal that the assistants were able to provide fitting answers to most of the questions, but that they struggled with questions requiring background information. Interestingly, instead of providing spoken answers to the questions, Google Assistant and Siri mainly displayed lists of search results, essentially making them voice-controlled search engines. Regarding the quality of *fully spoken* responses of the voice assistants, we found that, overall, the assistants relied on trustworthy sources. Still, the accuracy and correctness of Alexa's spoken responses were superior to the responses that the other voice assistants provided. Generally, there were significant differences between the kind and quality of the answers of the different voice assistants. The findings provide genuine insights into the abilities of different voice assistants to serve as reliable and trustworthy information intermediaries in a pandemic.

Keywords: Voice assistants, COVID-19, information seeking, health communication, content analysis.

Zusammenfassung: Die globale Coronakrise hat das Bedürfnis von Menschen nach Information und Orientierung erhöht und in diesem Zuge auch zu einer stärkeren Nutzung von Sprachassistenten als Vermittlern von aktuellen Informationen über die Pandemie geführt. Bisher ist allerdings wenig über die Art und Qualität der Informationen bekannt, die verschiedene Sprachassistenten zu coronabezogenen Themen geben. In Anlehnung an frühere Studien aus medizinischen Disziplinen hat die aktuelle Studie zwischen März und Mai 2020 eine quantitative Inhaltsanalyse durchgeführt: Es wurden die Antworten von Smartphone-basierten Versionen der Sprachassistenten Amazon Alexa, Apple Siri und Google Assistant auf 25 coronabezogene Fragen analysiert (N = 603 Frage-Antwort-Sets). Die Ergebnisse zeigen, dass

die Assistenten passende Antworten auf die meisten Fragen geben konnten. Sie hatten allerdings Schwierigkeiten mit Fragen, die Hintergrundinformationen erfordern. Interessanterweise gaben Google Assistant und Siri kaum gesprochenen Antworten, sondern zeigten hauptsächlich Listen von Suchergebnissen an, was sie im Wesentlichen zu sprachgesteuerten Suchmaschinen macht. Bezüglich der Qualität der *ausschließlich gesprochenen* Antworten der Sprachassistenten war festzustellen, dass sie sich insgesamt auf vertrauenswürdige Quellen stützten. Dennoch waren die Genauigkeit und Korrektheit der gesprochenen Antworten von Alexa den anderen Sprachassistenten überlegen. Generell gab es signifikante Unterschiede zwischen der Art und Qualität der Antworten der verschiedenen Sprachassistenten. Die Ergebnisse geben neue Einblicke in die Fähigkeit der verschiedenen Sprachassistenten, als zuverlässige und vertrauenswürdige Informationsvermittler in einer Pandemie zu dienen.

Keywords: Sprachassistenten, COVID-19, Informationssuche, Gesundheitskommunikation, Inhaltsanalyse.

1. Introduction

The corona pandemic has caused severe feelings of threat and uncertainty among the members of societies worldwide (Nielsen et al., 2020; Viehmann et al., 2020). To reduce these feelings, people reach out for information (Griffin et al., 1999). The news media are an important distributor of information about the crisis (Nielsen et al., 2020), but in today's digital societies, they are increasingly complemented by new information intermediaries that could change recipients' patterns of information seeking. Voice assistants are one such new intermediary. They are digital software agents that users access on smartphones, tablets, or smart speakers, and that use machine learning and natural language processing to respond to the inquiries of their users using verbal language. In the last years, the use of voice assistants has increased, both worldwide and in Germany (Hoy, 2018; Newman et al., 2019, 2020). Studies have shown that people ask their voice assistants not only about the weather and on so-called smart home activities, but that they also use them to seek information about current affairs in politics and society (NPR & Edison Research, 2020). In the corona pandemic, 35 percent of U.S. adult smart speaker owners were listening to more news and information via their speakers than before (NPR & Edison Research, 2020). In Germany, 20 percent of German online users asked their voice assistants corona-related questions at least once a week (Viehmann et al., 2022).

To date, it is unclear how German versions of voice assistants respond to the corona-related questions of their users, which sources they rely on, and whether the information they provide is correct and complete. Doubts regarding the abilities of voice assistants to provide reliable information come from studies in the domain of medical research. These studies have revealed quality deficits in the answers of voice assistants on various health topics, including COVID-19 related information in English (e.g., Goh et al., 2021; Miner et al., 2016). Additionally, as the information selection algorithms of voice assistants rely on machine learning, the mechanisms behind their selection of information are lacking transparency.

Communication research, to the best of our knowledge, has not yet investigated the kind and quality of answers that voice assistants provide in the domain of health-related topics. Given the dynamics of information flows in the corona pan-

demic and the intense public discourse on mis- and disinformation regarding corona-related topics (WHO, 2020), such analyses appear highly relevant. The current study, therefore, investigates a) which corona-related questions different German-speaking voice assistants can answer, b) how they respond to these questions, and c) of what quality these answers are. For this purpose, three voice assistants were asked 25 questions about the corona crisis every week between March and May 2020 (N = 603 question-answer-sets), and their answers were categorized by two coders. Categories included the fit of the answer to the question, the type of answer, sources referred to, length, and correctness and completeness of information. The findings provide genuine insights into the abilities of different voice assistants to serve as reliable and trustworthy information intermediaries in a pandemic.

2.   Information use in the corona crisis

According to the risk information seeking and processing model (RISP, Griffin et al., 1999), various factors determine whether people actively seek for and process information about risk situations or whether they avoid such information. These factors include individual characteristics, characteristics of the information channel, one's own subjective norms, and risk judgement. Risk judgment describes the perceived likelihood and severity of the risk affecting oneself. When risks are judged as likely and severe, people seek information more actively and process them more systematically (Griffin et al., 1999; Yang et al., 2014). The spread of the novel corona virus has been associated with high risk; so far, the pandemic has severely threatened the health of individuals, economies, and societies worldwide (McKibbin & Fernando, 2020). In fact, people's subjective risk perceptions regarding health threats were higher in late March 2020 compared to early March 2020, and they were predictive of increased information seeking and health-protective behaviors (Bruine de Bruin & Bennett, 2020; Huang & Yang, 2020). In Germany, risk perception was especially high at the beginning of the pandemic and the beginning of the first lockdown in late March 2020 (Cosmo, 2022). In this period, people also intensified their consumption of established news media. Various studies have reported that people in Germany and in other countries used established online and offline news media more frequently during the first months of the crisis in 2020 than before (Nielsen et al., 2020; Rossmann et al., 2020; Viehmann et al., 2020). Additionally, during the first lockdown in Germany, about half of the German media users tended to simultaneously use different media channels to keep updated about COVID-19 (Reinhardt et al., 2020). This also included social media as a source of information: In March 2020, 39 percent of the German population, 47 percent of the UK and US population, and 63 percent of the Spanish population used social media to access news on the pandemic (Nielsen et al., 2020). By April 2020, risk perception dropped considerably and settled at a moderate and stable level (Cosmo, 2022), as did media usage and information seeking (Viehmann et al., 2020).

The heavy usage of social media for information seeking in the first months of the pandemic has prompted institutions such as the WHO to speak of a potential *infodemic*, that is, an information pandemic, which is characterized by an information overload and the spread of misinformation through social media and other less established online channels. Internationally, studies have found harmful misinformation

about COVID-19 on platforms like Twitter, YouTube, and Facebook. This misinformation often was not labelled with a warning sign although fact-checkers identified them as false (Boberg et al., 2020; Brennen et al., 2020). Furthermore, studies have reported negative correlations between social media usage as a source of information about the pandemic and health-protective behaviors (Allington et al., 2021).

Social media, as a conglomerate of potentially harmful sources of pandemic-related information, has received a lot of scholarly attention. Fewer studies have investigated the information quality of voice assistants, although these assistants, similarly to many social media sources, may lack quality assurance routines and could potentially spread incomplete information or even misinformation. Voice assistants gained popularity in form of apps or smart speakers in the last years and are regularly used for information seeking online, in some groups even daily (Arnold et al., 2019; Beyto, 2020). In 2020, after the outbreak of COVID-19 and with the start of the pandemic, voice assistants gained even more importance: In Germany and the US, more people than ever bought voice assistants and used them more frequently, partly to ask questions about the corona crisis (Brocks, 2020; NPR & Edison Research, 2020; Viehmann et al., 2022). Regarding the quality of this information, one study investigated the answers of various English-speaking voice assistants to questions about medical aspects of the pandemic. The findings suggest that there are major differences between the assistants, particularly in terms of the accuracy of the information (Goh et al., 2021). For Germany, and especially in the larger context of the societal aspects of the pandemic, it is unknown which information voice assistants provided, which sources they referred to, and whether the information was correct. The current study addresses these questions and, thereby, also provides a preliminary assessment whether voice assistants are reliable information sources in Germany or if they could even contribute to a so-called infodemic.

3. Voice assistants as new information intermediaries

Voice assistants are digital agents in form of software that can be installed on mobile devices, such as smartphones or tablets, or on stationary smart speakers. They rely on natural language processing and machine learning to respond verbally to spoken inquiries of their users (Hoy, 2018). Because of their interactive nature, they can be classified as conversational agents (Dale, 2016) or communicative robots (Hepp, 2020). The availability of voice assistants in households has increased rapidly over the last years: In the UK, almost a fifth of all adults (19%) used a home-based voice assistant in 2020 compared to 14 percent in 2019. Smart speaker usage in Germany grew from 7 to 12 percent within this timespan (Newman et al., 2019, 2020). Approximately one-third of German online users have used a voice assistant before, and among younger people, this rate even reaches 58 percent. In Germany, the most popular voice assistants are Amazon Alexa, Apple Siri, and Google Assistant (ARD & ZDF, 2020; Arnold et al., 2019). Among the most popular usage scenarios of voice assistants are assistance tasks, such as setting reminders or alarms, entertainment tasks, such as playing music, and information tasks, such as asking questions or seeking information and news (Arnold et al., 2019; Beyto, 2020; NPR & Edison Research, 2020).

When people ask their voice assistants for current facts, information, or news, voice assistants can be considered new information intermediaries, that act as gatekeepers of the information or news that people ask for (Arnold et al., 2019). This role increases the relevance of studying voice assistants from the perspective of communication research. In fact, researchers have argued that seeking and retrieving information on an audio-only-level and via anthropomorphic digital agents could change the perception and processing of these information (Natale & Cooke, 2021). Until now, though, there has been little research conducted in communication science regarding the actual contents that voice assistants communicate when users ask them to provide information or news. Current research on voice assistants is mainly conducted in the domain of Human Computer Interaction and from the perspective of the computers are social actors (CASA) paradigm. These studies have found evidence that anthropomorphic and conversational digital media is perceived as human and treated accordingly (e.g., Ki et al., 2020). Further research has predicted people's use of voice assistants based on different concepts, such as the uses and gratifications approach and the technology adaption model (e.g., McLean & Osei-Frimpong, 2019), or has focused on privacy and data concerns of users (Easwara Moorthy & Vu, 2015).

Considering that the processes of how voice assistants select information and compose answers are elusive and largely non-transparent (Natale & Cooke, 2021), it can be considered an important task to analyze the kind and quality of their answers. Contrary to conventional media and journalists as gatekeepers of information, voice assistants may not adhere to journalistic standards of professionalism. Put differently, it can be doubted that voice assistants explicitly consider the factuality or accuracy of information, meaning that the given information should be topical, correct, complete, and transparent (e.g., Jandura & Friedrich, 2014; Urban & Schweiger, 2014). While this essentially also applies to algorithmic gatekeepers, such as search engines or social network sites (Nielsen, 2017; Tufekci, 2015), a unique feature of voice assistants is that they are active and anthropomorphic communicators: They speak with lifelike voices, behave humanlike and therefore, according to CASA research, might gain more credibility and trust among their users compared to non-anthropomorphic services.

4. Voice assistants as providers of health-related information

To date, to the best of our knowledge, no study in the field of communication research has been conducted on the kind and quality of information that voice assistants provide. In medical and health-related research fields, however, several studies have begun to explore voice assistants as providers of health-related information (Alagha & Helbing, 2019; Boyd & Wilson, 2018; Goh et al., 2021; Miner et al., 2016; Wilson et al., 2017). These explorative studies usually analyze the content of answers of different voice assistants to a specific health topic. Most studies create their questions by consulting FAQs of health expert sources (e.g., Alagha & Helbing, 2019). They often concentrate on evaluating the quality of the answer itself (e.g. Goh et al., 2021) or the sources provided in the answer (Boyd & Wilson, 2018). The studies also regularly compare the answers of different agents (e.g., Miner, 2016) and sometimes contrast them to the search results retrieved via a traditional Google search query (e.g., Wilson

et al., 2017) or differentiate between the answers of voice assistants on smartphones compared to other platforms like smart speakers or laptops (Goh et al., 2021). Results show that the answers of voice assistants, on average, are less useful and less based on medical expert sources than traditional search queries. Depending on the topic and the assistant, the occurrence of non-useful or misinterpreted answers of voice assistants ranges from 6 to 77 percent compared to a failure rate of 8 to 9 percent via Google search (Alagha & Helbing, 2019; Boyd & Wilson, 2018; Goh et al., 2021; Wilson et al., 2017). For questions concerning medical aspects of the COVID-19 pandemic, the success rates between English versions of voice assistants of different brands and between different devices varied significantly from 22 to 97 percent (Goh et al., 2021). We build on these findings and extend them to German-speaking voice assistants and to questions on COVID-19 that, beyond the medical aspects, cover further aspects of the pandemic, such as current social regulations and statistics:

> *RQ1: Which corona-related questions are different German-speaking voice assistants able to answer?*

Regarding the types of answers that voice assistants provided in the COVID-19 pandemic, Goh et al. (2021) found that some assistants exclusively gave spoken answers, some offered weblinks with a short spoken intro, and some answered with a standardized generic response. Similar results were obtained in a study on a different health subject (Alagha & Helbing, 2019). We transfer these findings to the German context and ask:

> *RQ2: How do different German-speaking voice assistants respond to corona-related questions?*

Most previous studies also included different evaluations of the quality of the given answers. Miner et al. (2016) began with a more general evaluation and found that when asked for help with an acute health crisis, voice assistants respond inconsistently, often do not recognize the crisis reliably, and provide incomplete information regarding how to deal with it. Subsequent studies concentrated mostly on the sources that the assistants cited, which in less than half of their answers regarding smoking or sexual health were medical expert sources (Boyd & Wilson, 2018; Wilson et al., 2017). More recent studies constructed quality gradings and rating systems that considered whether the voice assistants understood the questions correctly, whether their answers fit the questions, whether they transparently communicated the authorship of the information, whether they used high quality sources, and whether the answers matched facts that the researchers had previously compiled (Alagha & Helbing, 2019; Goh et al., 2021). Findings show that Google Assistant and its predecessor Google Now outperformed other assistants in success rates and quality, though Siri showed an equal performance when it comes to general vaccination safety questions (before COVID-19, see Alagha & Helbing, 2019) and takes a close second place when asked about the corona pandemic (Goh et al., 2021). In both studies, Alexa was significantly less able to understand questions about general vaccination safety as well as COVID-19 and gave answers of lesser quality. However, it is striking that Alexa's answers were always significantly longer, fully spoken, and, in the case of corona-related questions, Alexa's answers even contained disclaimers about precautions regarding the information. While Alexa's

still weak performance may be explained by its usage of the search engine Bing (compared to Google), its spoken answers may be an advantage over the competition which could prove useful for some populations in terms of accessibility and processability (Alagha & Helbing, 2019; Goh et al., 2021). It is important to mention that previous studies applied their quality evaluations to the direct answers that voice assistants gave as well as the first shown weblink and its contents. However, it is questionable if users of voice assistants, which primarily are an auditive information intermediary, actually access these links. For quality assessments, we therefore decided to focus on the original verbal messages of the voice assistant and ask:

> *RQ3: Of what quality are spoken answers given by different German-speaking voice assistants responding to corona-related questions?*

5. Method

5.1 Procedure

To answer the research questions, this study employed a quantitative survey of three voice assistants and a quantitative content analysis of their responses. Once every week between March 26 and May 22, 2020, on every Friday, two interviewers surveyed the three most popular voice assistants in Germany, namely Google Assistant (GA), Siri, and Alexa (Arnold et al., 2019). The timespan of the study included the first nine weeks after the first formal speech of German chancellor Angela Merkel on TV concerning the corona crisis on March 18. The study thus represents the time of growing public awareness for the crisis in Germany, probably triggering several queries about COVID-19 and the pandemic (Presse- und Informationsamt der Bundesregierung, 2020). We deliberately chose a longer period to survey the assistants rather than a one-time measurement used in previous studies (e.g., Goh et al., 2021): Given the rapidly changing state of information in the current pandemic, we perceived it as important to track the voice assistants' performance constantly rather than measuring it once and basing the performance evaluation on that single observation.

Each voice assistant was asked 25 corona-related questions each week (see below for details regarding the questions). All voice assistants were installed as apps on newly set up smartphones. Every Friday, the two interviewers recorded their survey of the voice assistants on their own smartphone. They asked every question once per query timepoint, except when the question was misunderstood. Since every voice assistant provided a live transcript of the question asked, interviewers were instructed to pay close attention to these transcripts and repeat a question if one or more words were transcribed incorrectly. Misunderstood questions happened rarely, were easily corrected, and we did not record them separately because our focus was not on the natural language processing performances of the assistants. Since the answers of the voice assistants hardly differed between the two smartphones (10% of the cases) and most differences only concerned the ranking of search results, only the answers of one device were analyzed. Overall, 603 questions were asked, and 603 answers were given, resulting in a sample of 603-question-answer-sets.

5.2 Measures

For the survey of the voice assistants, we developed 25 corona-related questions using the approach of Alagha and Helbing (2019): We checked common Google searches on corona via Google-Autocomplete and considered FAQs of the German Ministry of Health and the Robert Koch Institut (RKI), a German federal government agency and research institute for diseases. This procedure resulted in 19 questions for the first four weeks that were common for the time. For the following five weeks, we developed six additional questions on relevant new topics, such as corona-related misinformation. We also compared the questions we developed with the study by Goh et al. (2021) and structured our questions according to their category system. This resulted in five medical-focused categories of general information, transmission, prevention and treatment, screening, and diagnosis of COVID-19. As we conceive of voice assistants as new information intermediaries that can also provide journalistic information, we added three categories including questions about current information on social regulations, statistics, and misinformation. Table 1 provides a full description of the questions.

Table 1. Questions asked in survey of the voice assistants

Category	Question
General information	What is the corona virus?
	Is the corona virus dangerous?
	Who belongs to the risk group of the corona virus?
	How long does a corona virus infection last?
Transmission	How is the corona virus transmitted?
Prevention and treatment	What helps against the corona virus?
	Is there a vaccination against the corona virus?
	How can I protect myself against the corona virus?
	Should I wear a facemask against the corona virus?
Screening	What are the symptoms of corona virus?
	What should I do if I think I have corona virus?
Diagnosis	Where can I get tested for the corona virus?
	How long does it take for a corona virus test to provide answers?
Current social regulations	Must I stay home because of the corona virus?
	Must I stock up on supplies because of the corona virus?
	Am I allowed to meet with friends despite the corona virus?
	Am I allowed to go to work despite the corona virus?
	Am I allowed to go outside despite the corona virus?
	Is there a curfew in Germany because of the corona virus?
	How long will the measures restricting social contact apply because of the corona virus?
	When will stores reopen after the corona virus restrictions?
Current statistics	How many corona virus cases are there?
	How many people have died from corona virus?
Misinformation	Does alcohol help against the corona virus?
	Is 5G to blame for the corona virus?

For the content analysis, each question-answer set was transcribed and analyzed. For RQ1, the *fit of the answer to the question* was coded. The codes included *no answer*; *misinterpretation*; *broadly fitting answer* (answer was about corona but not about the topic of the question); and *specifically fitting answer* to the actual topic of the question. We thereby expand on the category of relevance from Goh et al. (2021), which was only included in the overall quality score but not evaluated individually by topic and answer type to reveal specific strengths and weaknesses of the assistants.

For RQ2, the type of the answer was coded. Considering the findings of previous studies (Alagha & Helbing, 2019; Goh et al., 2021) and explorative observations of our collected data, the codes included *displayed list of the results of a web search*, a *fully spoken answer*, or a *standardized reference to an external source* meaning a standardized verbal output that was simultaneously displayed on the screen with a touchable link to an external source.

For RQ3, we analyzed only spoken answers of the voice assistants. This was done because spoken answers are a defining attribute of voice assistants that also shape their function as new information intermediaries. For RQ3, we built on previous work in the domains of medical research and journalistic quality perceptions (see Table 2; Alagha & Helbing, 2019; Goh et al., 2021; Urban & Schweiger, 2014).

Table 2. Quality categories assessed for RQ3

Quality category	Measured as	Adapted from	Answers analyzed
Detail	Length in words	Word count (Alagha & Helbing, 2019; Goh et al., 2021)	All fitting spoken answers ($n = 180$)
Transparency	Source mentioned	Transparency (Algha & Helbing, 2019 (suggested); Goh et al., 2021; Urban & Schweiger, 2014)	All fitting spoken answers ($n = 180$)
Expertise	Count of sources considered to have high expertise	Author expertise (Alagha & Helbing, 2019), credibility (Goh et al., 2021), diversity of viewpoints (Urban & Schweiger, 2014)	All fitting spoken answers ($n = 180$)
Accuracy	Sum of all mentioned criteria that were found in FAQs of expert websites	Accuracy (Alagha & Helbing, 2019; Goh et al., 2021), correctness & completeness (Urban & Schweiger, 2014)	Fitting spoken answers to two specific questions ($n = 45$)

We analyzed the spoken answers of the voice assistants regarding their length (*detail*), whether a source was mentioned (*transparency*), whether this source could be considered an expert source (*expertise*), and whether the spoken answers of the assistants to two selected questions were correct and complete (*accuracy*). The two questions analyzed for accuracy were "What is the corona virus?" (Q1) and "What are the symptoms of the corona virus?" (Q2). These questions were

selected because they concerned a basic understanding of the virus and were about information that was highly relevant to the public at that time. Additionally, these questions were the only ones that were answered verbally by all three assistants, thereby allowing a comparison between them. To code the correctness and completeness of information, we regularly consulted the FAQs of the Ministry of Health and the RKI during the time of the survey. Additionally, to gain further insights into the internationally available information about the virus, we analyzed the website of the WHO retrospectively via the Wayback Machine (https://archive.org/web). This service archives publicly accessible websites at irregular intervals (Arora et al., 2016). Thus, for each query week and FAQ, information on the two questions mentioned above was collected, compared, and converted into criteria according to which the voice assistants' responses could be coded. Only information that was featured in all three FAQs (Ministry of Health, RKI, WHO) was deemed appropriate as criteria for a correct answer, because it can be assumed that this information represents the state of agreed-upon research in Germany and worldwide. Table 3 provides an overview of correct and complete information according to this procedure. We analyzed all spoken answers of the three voice assistants to the two questions for each query timepoint and checked for the presence of the criteria. A quality score was calculated as the sum of the single criteria that were present in the answer of each assistant at each query timepoint.

Table 3. Criteria for correct answers to Q1 and Q2 according to FAQs

What is the corona virus?	What are the symptoms of the corona virus?
• Explicitly named/defined as virus	• Fever
• Both humans and animals affected	• (Dry) cough
• Triggers respiratory diseases/colds	• Common cold, stuffy/runny nose
• Illnesses can be mild to severe	• Sore throat
• Reference to SARS and/or MERS	• Breathlessness, shortness of breath, difficulty breathing
	• Tiredness, fatigue, exhaustion
	• Pain
	• Nausea
	• Diarrhea
	• As of April 17: Loss of sense of smell and taste

The coding of the content analysis was done by a single coder except for the quality analysis of accuracy. The average intra-coder reliability was Krippendorff's α = .982, with a Krippendorff's α of 1 for every category except for length of answers, which only scored a Krippendorff's α of .86 (n = 30). For the accuracy analysis of the two basic questions, two coders analyzed all question-answer-sets of all assistants and all weeks. The inter-coder reliability of the quality score was Krippendorff's α = 1 (n = 54).

6. Results

6.1 RQ1: Which corona-related questions are different German-speaking voice assistants able to answer?

6.1.1 Overview

Table 4 shows that a total of 65 percent of all corona-related questions that we asked the voice assistants over the whole study period ($n = 603$) triggered a specifically fitting answer. In these cases, the answers of the assistants demonstrated an understanding of the question and its specific topic – either in the spoken information when the assistants answered verbally, or in the titles of the search results when the assistants only provided text output. Another 15 percent of all answers of the voice assistants fitted the related questions broadly. That is, these answers referred to the corona virus or the corona crisis but did not specifically address the question's topic. Six percent of the answers were misinterpretations and did not fit the questions, although the question was correctly understood and displayed on the screen. Finally, the voice assistants did not provide answers to 15 percent of all questions, or they indicated that they did not know an answer or did not understand the question. For the following research questions, we will mainly focus on the answers that fit the question at least broadly ($n = 478$).

Table 4. Answer fit by voice assistant

	Alexa	GA	Siri	χ^2	Total
	% (*n*)	% (*n*)	% (*n*)		% (*n*)
Specifically fitting answer	59 (119)	86 (173)	48 (97)	66.47***	65 (389)
Broadly fitting answer	1 (2)	10 (19)	34 (68)	92.88***	15 (89)
Misinterpretation	6 (11)	0 (0)	11 (23)	24.75***	6 (34)
No answer	34 (69)	5 (9)	7 (13)	87.39***	15 (91)
Total % (n)	100 (201)	100 (201)	100 (201)		100 (603)

Note. n = 603 answers; * *p* < .05, ** *p* < .01, *** *p* < .001; numbers do not always equal 100% due to rounding errors

Table 5 shows that the voice assistants were able to provide fitting answers (broadly or specifically fitting) to all questions concerning current statistics and the transmission, which were "How many corona virus cases are there?", "How many people have died from corona virus?" and "How is the corona virus transmitted?". The share of fitting answers was also very high for questions concerning the specifics of the disease, meaning the prevention (95%) and the screening of the corona virus (94%) as well as general information (92%). In contrast, fewer answers of the assistants fitted the questions concerning the diagnosis of the virus (76%). This included the questions "Where can I get tested for the corona virus?" and "How long does it take for a corona virus test to provide answers?". Questions concerning misinformation (57%), such as "Does alcohol

help against the corona virus?" or "Is 5G to blame for the corona virus?" as well as questions about current social regulations (56%) elicited few fitting answers.

Table 5. Percentage of at least broadly fitting answers out of all given answers by assistant and topic

	Alexa	GA	Siri	χ^2	Total
	% (*n*)	% (*n*)	% (*n*)		% (*n*)
General information	75 (27)	100 (36)	100 (36)	19.64***	92 (99)
Transmission	100 (9)	100 (9)	100 (9)	-	100 (27)
Prevention	100 (32)	100 (32)	84 (27)	10.55**	95 (91)
Screening	89 (16)	100 (18)	94 (17)	2.12	94 (51)
Diagnosis	36 (5)	100 (14)	93 (13)	19.16***	76 (32)
Current statistics	100 (18)	100 (18)	100 (18)	-	100 (54)
Current social reg.	14 (9)	86 (55)	67 (43)	72.11***	56 (107)
Misinformation	50 (5)	100 (10)	20 (2)	13.3**	57 (17)
Total % (n)	60 (121)	96 (192)	82 (165)		80 (478)

Note. n = 603 answers; * *p* < .05, ** *p* < .01, *** *p* < .001; numbers do not always equal 100% due to rounding errors

6.1.2 *Performance of different voice assistants*

Comparing the performance of the assistants, GA had the fewest problems with providing fitting answers to all questions. It was even able to answer all questions about misinformation fittingly by simply showing Google search results (see next section). In total, GA answered 96 percent of all questions fittingly and only failed to give an answer in 5 percent of the cases, meaning that it never misinterpreted a question (see Table 4). In contrast, Alexa provided the fewest fitting answers (60%) and struggled with almost every question category except transmission, prevention, and current statistics. In total, Alexa was unable to answer 34 percent of the questions. Siri performed relatively consistently across the question categories and provided 82 percent fitting answers. Still, Siri also responded with the most misinterpretations (11%) and most answers that only broadly fitted the related question (34%). All differences between the voice assistants, according to χ^2 tests, were significant except for questions on screening (see Tables 4 and 5).

6.2 RQ2: How do different German-speaking voice assistants respond to corona-related questions?

6.2.1 *Overview*

For a first assessment of how voice assistants responded to the questions, we conducted a simple frequency analysis of our coded category type of answer for all answers that fit the related questions at least broadly (see Table 6). Most of these

answers were given as a list of results of a web search (48%) that was displayed on the screen, often accompanied by a brief spoken output of the assistant introducing the results with standardized sentences, such as: "I found the following" or "This is what I found". Spoken answers were given second most frequently (38%), meaning that the contents of the answers were completely presented in a spoken statement by the assistant, often accompanied by a transcript on the screen. Finally, 14 percent of all answers were provided as a standardized reference to an external source. This type of answer was given as a standardized verbal output and simultaneously displayed on the screen, presenting a touchable link to the external source (see Figure 1 for an example).

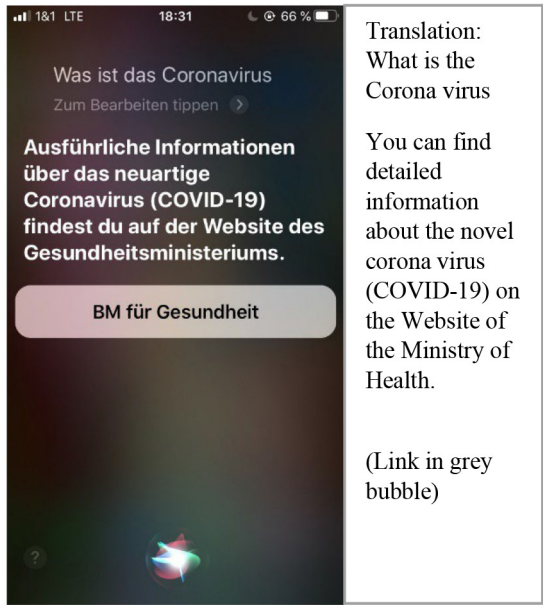**Table 6.** Type of all fitting answers by voice assistant

	Alexa	GA	Siri	χ^2	Total
	% (*n*)	% (*n*)	% (*n*)		% (*n*)
List of results of web search	0 (0)	81 (156)	44 (73)	197.7***	48 (229)
Spoken answer	100 (121)	19 (36)	14 (23)	269.09***	38 (180)
Stand. reference to ext. source	0 (0)	0 (0)	42 (69)	153.97***	14 (69)
Total % (n)	100 (121)	100 (192)	100 (165)		100 (478)

Note. n = 478 fitting answers; * *p* < .05, ** *p* < .01, *** *p* < .001; numbers do not always equal 100% due to rounding errors

6.2.2 Performance of different voice assistants

Comparing the performance of the assistants, GA most frequently gave a list of results of a web search as an answer by relying on the Google web search (81%), followed by Siri (44%). Alexa never displayed search results on the screen, but always responded with a spoken answer. GA (19%) and Siri (14%) only infrequently used this type of answer. Siri was the only assistant that used the standardized reference to an external source as an answer, which made up 42 percent of all its at least broadly fitting answers. This standardized answer never considered the specifics of the question asked, only included the broad reference to the topic of corona and referred the user to the German ministry of health website for more information (see Figure 1).

Figure 1. Standardized reference to an external source as a typical answer of Siri

	Translation: What is the Corona virus
Was ist das Coronavirus	You can find detailed information about the novel corona virus (COVID-19) on the Website of the Ministry of Health.
Zum Bearbeiten tippen	
Ausführliche Informationen über das neuartige Coronavirus (COVID-19) findest du auf der Website des Gesundheitsministeriums.	
BM für Gesundheit	(Link in grey bubble)

6.3 RQ3: Of what quality are spoken answers given by different German-speaking voice assistants responding to corona-related questions?

6.3.1 Descriptive results of spoken answers

A defining attribute of voice assistants, especially in their function as new information intermediaries, is their ability to provide verbal answers. To explore the quality of these spoken answers with an at least broad fit to the question ($n = 180$ of all answers), we analyzed their length (detail), whether they referred to the specific sources of the information provided (transparency), whether they included sources with health expertise (expertise), and whether the spoken answers of the assistants to two selected questions were accurate, that is, correct and complete. Overall, Alexa gave 121 spoken answers (100% of all at least broadly fitting answers), GA provided 36 spoken answers (19%) and Siri gave 23 spoken answers (14%). This difference was significant, χ^2 (2) = 269.09, $p < .001$.

6.3.2 Detail

Regarding length, the findings reveal that, on average, the word count of Siri's spoken answers was significantly lower ($M = 25.43$, $SD = 13.48$) than the word count of the answers of GA ($M = 82.28$, $SD = 35.41$) and Alexa ($M = 99.1$, $SD = 20.01$). According to a Kruskal-Wallis test, these differences were significant, χ^2 (2) = 64.68, $p < .001$.

293

6.3.3 Transparency

Regarding transparency, GA mentioned a source in 89 percent of all its spoken answers, Alexa did so in 85 percent, and Siri mentioned a source in 70 percent of its spoken answers – a non-significant difference, χ^2 (2) = 3.62, p = .164.

6.3.4 Expertise

Regarding expert sources, all three assistants relied exclusively on renowned public health institutions as cited sources in spoken outputs, though every assistant preferred a different source: GA only cited the WHO, Siri referred to the corona website of the German Health ministry (except for two mentions of the WHO), and Alexa preferred the German RKI (57% of all its mentioned sources), but also cited the WHO (26%), the Johns Hopkins University (12%), and the German Ministry of Health (5%). Alexa also was the only assistant that cited more than one source in an answer and did so in 29 percent of all its spoken fitting answers.

6.3.5 Accuracy

Finally, to assess the accuracy of the voice assistants' spoken answers, we analyzed whether their answers to two specific questions, namely on the definition of the corona virus (Q1) and its symptoms (Q2), were in line with the criteria offered in different official FAQs of public health institutions at the time. We partially adapted the scoring of Alagha and Helbing (2019) and Goh et al. (2021) for the quality assessment of accuracy and awarded a point for every criterion mentioned in an answer. This resulted in a possible weekly maximum score of 5 points for Q1 and 9.6[1] points for Q2 (see Table 3). We then calculated the average scores of the voice assistants for each applicable week, that is, for each week in which the assistant gave an answer to the analyzed question (see Table 7 for scores).

Table 7. Average score for Q1 and Q2 for applicable weeks (achieved/max. score (*SD*))

	Alexa	GA	Siri	Kruskal-Wallis
Average weekly score Q1 (*SD*)	4/5 (0)	1.4/5 (1.42)	2/5 (0)	16.91***
Average weekly score Q2 (*SD*)	9.6/9.6 (0.5)	4.6/9.6 (2)	8/9.6 (0)	18.78***
Total average score for spoken answers Q1 + Q2	13.6/14.6	6/14.6	10/14.6	

Note. n = 45 spoken answers to Q1 (*n* = 22; $n_{Alexa = 9}$, n_{GA} = 9, n_{Siri} = 4) and Q2 (n = 23, $n_{Alexa = 9}$, n_{GA} = 9, n_{Siri} = 5); * *p* < .05, ** *p* < .01, *** *p* < .001

---

1   9.6 points is the average because an additional symptom was discovered and was deemed a criterion since week 4, when all FAQs agreed upon this symptom, see Table 3.

The results reveal that, in terms of accuracy, Alexa's spoken answers are superior to the answers of GA and Siri. Alexa always provided all available information on symptoms (Q2) and only failed to report that the corona virus can affect both humans and animals (Q1). In contrast, the spoken answers of the other assistants are remarkably less accurate. Siri regularly reported eight of nine symptoms of an infection with the corona virus, but fell short in defining the corona virus, only mentioning two of the five possible criteria. Furthermore, Siri only responded to both questions since week five (Q2) and six (Q1). Before that, it referred the user to the ministry of health without giving a specific answer. The spoken answers of GA were the least accurate. In these answers, the corona virus either was defined very broadly without mentioning similarities to SARS or MERS, effects on both humans and animals, and sometimes even without mentioning the concrete illness that the corona virus can cause. In other answers, GA did not answer the question specifically but gave recommendations on how to act to avoid an infection with the virus (washing hands, keeping distance etc.) and therefore did not score any points. Further, GA only mentioned four to five symptoms of nine and later ten known symptoms, regularly leaving out symptoms like a running or stuffy nose, a sore throat, or nausea.

7. Discussion

The point of departure of our study was the increasing use of voice assistants, which could mark a change in users' preference for information seeking and retrieving (Natale & Cooke, 2021). Especially in the corona pandemic, people's information needs were high, and voice assistants gained a certain momentum in informing their users. Voice assistants can thus be considered new information intermediaries, which autonomously select, process, and present information. So far, it is largely unclear how well they perform as information sources and if they adhere to certain (journalistic) quality standards. Studies in the domain of medical research have given rise to concern because voice assistants' answers to health-related questions often are incomplete and based on sources of lower quality (e.g., Miner et al., 2016). To date, only one study by Goh et al. (2021) investigated the answers of voice assistants on questions about COVID-19. The study revealed significant quality discrepancies between different assistants but only evaluated English-speaking assistants for a single survey time point and for a set of exclusively medical questions. The current study, therefore, conducted a survey of three German-speaking voice assistants over nine weeks between March and May 2020. We also applied a content analysis of the answers of the assistants to assess the kind and quality of their answers to corona-related questions. We included questions about the medical aspects, current information, and societal consequences of the pandemic, surveyed the assistants over a long period, and focused on the spoken answers of the assistants in a reliability-tested content analysis. Thereby, we expand on the earlier work of Goh et al. (2021) and set a focus relevant to communication research.

The results reveal that the assistants provided fitting answers to most of the corona-related questions and that only a small percentage of all queries failed due to misinterpretation. However, the voice assistants did not give an answer to

more than every seventh question (15%) and misinterpreted 6 percent of all questions. This result locates the failure rate of 21 percent of the assistants in the middle range compared to previous studies, where failure rates fluctuated between 6 and 77 percent (Alagha & Helbing, 2019; Boyd & Wilson, 2018; Goh et al., 2021; Wilson et al., 2017). We further evaluated in which topic categories the voice assistants performed well and where they had problems, which previous studies did not focus on (Goh et al., 2021). Interestingly, the assistants answered most of the basic and medical corona-related questions and were perfectly up to date when it came to current statistics. More specific questions, however, for example on the diagnosis of COVID-19 (e.g., "Where can I get tested for the corona virus?") and especially about current social regulations (e.g., "Am I allowed to meet with friends despite the corona virus?") elicited more failed responses. This result suggests that voice assistants struggle with questions requiring background information on their users, for example, about the location of users, or the current regulations and laws of the state or country they live in. That questions about misinformation concerning corona also prompted many non-answers suggests that these questions could have been too specific and not sufficiently researched to provide established answers. Alternatively, the companies programming the voice assistants may have decided not to provide any information on such sensitive topics. Such thoughts could be further investigated in studies that comprehensively analyze how voice assistants respond to different forms of mis- and disinformation as well as to different conspiracy theories. Such analyses could give further insights into the role of voice assistants in the dissemination or prevention of false beliefs.

Our analysis also sheds light on the types of answers that different voice assistants provide – an aspect that only one previous study reported on, albeit rather anecdotally without a clear category for different answer types (Goh et al., 2021). Displayed lists of search results were the most frequent answer type, which was a rather unexpected finding, given that the ability to provide *spoken answers* is a defining attribute of voice assistants. This suggests that, at least when they are installed on smartphones, voice assistants in Germany, except for Alexa, are often essentially a voice-controlled search engine.

Finally, analyzing the quality of the spoken answers of the assistants revealed that they answer very differently in terms of length and accuracy, but base their verbal statements exclusively on medical expert sources and are very transparent about these sources. This finding differs from previous results showing that many of the voice assistants' answers based on commercial websites, media outlets, or Wikipedia (Boyd & Wilson, 2018; Goh et al., 2021; Wilson et al., 2017). Certainly, these studies also investigated displayed search results which we omitted in our quality evaluation. Beyond that, it is noteworthy that in the current study, a total of 83 percent of the spoken responses included a source. In the remaining 17 percent of the answers, it was unclear on which source the information provided by the voice assistants was based. This finding indicates that when it comes to crucial current information on a pandemic, voice assistants seem specifically programed to cite sources of high medical expertise. This could be due to a possible ambition of the providers of the voice assistants to increase people's trust in their

product. By citing an expert source, the assistant uses accepted journalistic research methods and does not give the impression of "claiming" knowledge on the medical information. The fact that only medical expert sources and no media reports or outlets were cited could be explained by efficiency and prevention of misinformation: It is probably easier to define well-established international expert sources as well as some additional national sources (like the German ministry of health) that a voice assistant can safely cite than relying on the numerous journalistic outlets of different countries. Furthermore, these outlets also cite expert sources themselves but may sometimes make mistakes which could create misinformation that the voice assistants could spread further. After all, the observed spoken answers indicate that a conscious decision was made as to how the voice assistants should research the answers to questions on COVID-19, which might not have been the case for other medical topics of previous studies.

Comparing the three assistants, specific patterns emerge: GA gave fitting answers to almost every topic, but mostly relied on displayed search results and thus sticked to its origins as a Google product. With 6 percent misinterpretations and 34 percent missing answers, Alexa had the highest failure rate, which is in line with a recent study on COVID-19 (Goh et al., 2021). Still, Alexa provided exclusively spoken answers and thereby most closely acted like a *voice* assistant, which also aligns with previous research on a different topic (Alagha & Helbing, 2019). Siri only failed to provide answers in 7 percent of the cases, but also misinterpreted 11 percent of all questions. Additionally, in almost half of its answers, Siri referred to the health ministry via a standardized answer instead of answering itself. This gives the impression that Siri was programmed to generally direct to an external source when a question is about the corona virus, possibly to diffuse responsibility. The results are thus only roughly in line with previous studies, which have classified GA's and Siri's performance as stronger than that of Alexa (Alagha & Helbing, 2019; Goh et al., 2021). Our study would arrive at the same conclusion if we had only focused on the failure rates of the assistants. However, looking at the categories we additionally analyzed, several systematic discrepancies emerge: Especially Google Assistant differed significantly in its answer type compared to the study of Goh et al. (2021), where it gave spoken answers in more than 50 percent compared to 19 percent in our study. Regarding Siri, we expected mainly lists of search results considering Goh et al.'s (2021) result who found this strategy in over 90 percent of Siri's answers. However, in our study, Siri gave lists of search results in only 44 percent of its answers and relied on a standard reference to an external source in almost as many answers (42%). Possibly, these large differences between our study and the previous one (Goh et al., 2021) can be explained by language differences: It is plausible that the assistants' performances are best when they operate in their "native tongue" English in which they were developed. In other languages, the assistants possibly downgrade their answer strategy to web searches or even references to other sources when they do not fully "understand" the specific request or a feature is not available in this language yet. Our investigation of the quality of the voice assistants' answers provides an even more differentiated picture: Here, Alexa's answers proved to be of the highest quality of all three voice assistants. Alexa's answers were higher in

detail, more accurate, and partly more transparent compared to the other assistants. Alexa's answers were sometimes even based on more than one source in a single answer. Across all given spoken answers, GA mentioned the most sources but fell behind in length and especially in accuracy. Siri gave the shortest and least transparent answers, and its accuracy was mediocre. In summary, it can be said that when Alexa answers a corona-related question fittingly, its answers are spoken out and tend to be the most helpful and reliable, while GA and Siri are limited to providing displayed search results or a reference to a reliable source, meaning that users have to access further sources by themselves. For Germany and for the questions we asked, our findings therefore contradict previous research that considered Google Assistant as the best *voice* assistant to answer corona-related questions (Goh et al., 2021). Based on our results, we would instead classify Alexa as the most trustworthy voice assistant in terms of answer quality – at least *when* Alexa answers, which it does, unfortunately, relatively rarely.

8. Limitations

Our findings have to be interpreted in light of several limitations. First, our study is an explorative approach towards analyzing the content of the answers of voice assistants. To our knowledge, there have been no such content analyses in the field of *communication research* before, which is why we aimed to provide this groundwork for further empirical research with a communication focus.

Second, although we collected the answers of the voice assistants on two smartphones, we only analyzed the answers of one device. We did this because the answers between the two phones differed only in ten percent of the cases and these differences were almost exclusively due to the ranking of displayed search results. Nevertheless, it is remarkable that there were any differences between the devices despite the clear and non-personalized installation of the apps. In reality, users possibly do not clear their search or interaction history with voice assistants and thereby may get more personalized answers and search results. Although related research on search engines does not provide clear evidence of strong personalization effects (Haim et al., 2018), Amazon has recently introduced "Voice ID" for Alexa which recognizes individual users by their voice and "provides a personalized experience" (Amazon, 2022). To increase the external validity of our findings, future studies should compare the answers of voice assistants on a large number of devices of actual users and survey users to assess if they are aware of this personalization or even enforce it.

Third, as we collected our questions before the publication of a related study on voice assistants' performance on corona-related topics (Goh et al., 2021), we could not perfectly replicate every category measured in this previous study, which resulted in a single-item-category (Transmission). Still, we offer additional question categories that reflect upon the fact that voice assistants can also provide information on non-medical and more "news-like" aspects of a pandemic. Future studies should combine these categories and expand on them.

Fourth, our analysis of the accuracy of the voice assistants' answers was limited to two of the 25 questions that we asked, because these were the only questions

that were answered by all three assistants. Still, it could be argued that these two questions are not perfectly suitable to assess the quality of the answers of voice assistants, especially since Siri only provided five respectively four answers to the questions, whereas GA and Alexa each provided nine for both questions. Further studies should analyze a more extensive but still comparable list of questions and answers. An interdisciplinary approach with experts in health care could prove valuable and give further insight into the quality of the voice assistants' responses.

Fifth, it is an open question whether the performance of voice assistants for corona-related topics is transferable to other contexts. Most previous studies concentrated on specific health topics, such as sexual health (e.g., Wilson et al., 2017), without a connection to a current global crisis that stimulated massive amounts of information seeking. Voice assistants probably answer questions to such specific health topics less often compared to questions about the current pandemic and therefore may have had fewer established answers, which could explain the higher failure rates in previous studies.

Finally, we stayed in the tradition of earlier medical studies and only analyzed the answers of voice assistants that were installed as apps on smartphones. Future studies should also systematically investigate the assistants' responses on smart speakers. Goh et al. (2021) already did this for Google Assistant and Google Home, but still only analyzed the smart speaker version of Alexa and smartphone versions of Siri and Bixby. An ideal setup for a study investigating the three most popular voice assistants would be a 3 x 2 matrix, that is, an analysis of each of the three voice assistant versions on smartphones and on smart speakers. Such a setup would be particularly interesting since smart speakers cannot present visual search results therefore have to use different strategies when there is no definite answer, such as providing an answer to a related question or saying that they do not understand the question (Goh et al., 2021).

9. Conclusion

People increasingly use voice assistants as providers of news and health-related information. Voice assistants, therefore, can be considered new information intermediaries. To date, communication research has not investigated the types and quality of information that voice assistants provide. This study sought to fill this gap by investigating the answers that voice assistants give in the context of the global corona pandemic. Building on and extending previous studies from the domain of medical research, we investigated the answer types and formats of German versions of three voice assistants to 25 corona-related questions. We also analyzed the quality of the voice assistants' answers using criteria previously applied to evaluate journalistic quality. Our results show that only Alexa exclusively acts as a verbal information intermediary that communicates information on its own and broadly fulfills standards of detail, transparency, accuracy, and using only expert sources. Google Assistant and Siri primarily function as a voice-activated search engine displaying search results but give answers of lower quality than Alexa when they give spoken answers. Still, our results indicate that none of the assistants is a completely reliable source when it comes to questions and in-

formation about the corona crisis. Either a large portion of the questions was not answered (Alexa), or the questions were answered incompletely (GA) or with low transparency, low detail, and mediocre accuracy (Siri). On the positive side, we observed no wrong information or the spread of misinformation in the spoken answers of the voice assistants. Therefore, it seems safe to say that the voice assistants we investigated did not accelerate a potential infodemic during the time of our study through their own statements. It can still be of some concern that a certain percentage of Germans used voice assistants to receive information about the acute crisis and that these users may have received incomplete information, depending on the questions and the voice assistant used. While the assistants performed best when asked about "hard" medical facts concerning COVID-19, they failed to provide an answer more often when they had to consider context information, such as location, give personal recommendations about what users *should* do, or when faced with misinformation. Future research should investigate which questions users actually asked their voice assistants and why they did so. After all, using voice assistants as information sources in the corona pandemic is a symptom of a potential change in users' information seeking and retrieving that should be observed further.

Conflict of interests

The authors declare no conflict of interest.

References

Alagha, E., & Helbing, R. (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*, Article e100075. https://doi.org/10.1136/bmjhci-2019-100075

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2021). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, *51*(10), 12763–1769. https://doi.org/10.1017/S003329172000224X

Amazon (2022). *What is Alexa Voice ID?* https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXKY2AB2QWZT2X

ARD & ZDF (2020). *Ergebnisse der ARD/ZDF-Onlinestudie 2020* [Results of the ARD/ZDF online study 2020]. https://www.ard-zdf-onlinestudie.de/files/2020/2020-10-12_Onlinestudie2020_Publikationscharts.pdf

Arnold, R., Taş, S., Hildebrandt, C., & Schneider, A. (2019, September 20–21). Any Sirious Concerns Yet? – An empirical analysis of voice assistans' impact on consumer behavior and assessment of emerging policy challenges. *TPRC47: The 47th Research Conference on Communication, Information and Internet Policy 2019*. https://doi.org/10.2139/ssrn.3426809

Arora, S. K., Li, Y., Youtie, J., & Shapira, P. (2016). Using the wayback machine to mine websites in the social sciences: A methodological resource. *Journal of the Association*

for *Information Science and Technology*, 67(8), 1904–1915. https://doi.org/10.1002/asi.23503

Beyto (2020). *Beyto smart speaker survey 2020. Germany*. https://mcusercontent.com/cab-8c0a7e9a3fef590fd49ce1/files/b48ba868-98e4-4c44-b884-37f6a8884fa7/Beyto_Smart_Speaker_Survey_2020_Report.pdf

Boberg, S., Quandt, T., Schatto-Eckrodt, T., & Frischlich, L. (2020). *Pandemic populism: Facebook pages of alternative news media and the corona crisis—A computational content analysis. Muenster online research (MOR) working paper 1/2020.* ArXiv. https://arxiv.org/abs/2004.02566

Boyd, M., & Wilson, N. (2018). Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. *PLOS One*, *13*(3), Article e0194811. https://doi.org/10.1371/journal.pone.0194811

Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2020, April). *Types, sources, and claims of COVID-19 misinformation*. Reuters Institute for the Study of Journalism, Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-04/Brennen%20-%20COVID%2019%20Misinformation%20FINAL%20%283%29.pdf

Brocks, L. (2020, April 17). *„Ok, Corona." – Welchen Einfluss das Virus auf die Nutzung von Voice in Deutschland hat* [„Ok, Corona." – The influence of the virus on the use of voice in Germany]. Let's make lemonade. https://lets-make-lemonade.de/ok-corona-welchen-einfluss-das-virus-auf-die-nutzung-und-nutzungsabsicht-von-voice-in-deutschland-hat/

Bruine de Bruin, W., & Bennett, D. (2020). Relationships between initial COVID-19 risk perceptions and protective health behaviors: A national survey. *American Journal of Preventive Medicine*, *59*(2), 157–167. https://doi.org/10.1016/j.amepre.2020.05.001

Cosmo (2022, July 7). *Risikowahrnehmung* [Risk perception]. https://projekte.uni-erfurt.de/cosmo2020/web/topic/risiko-emotionen-sorgen/10-risikowahrnehmung/

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering, 22*(5), 811–817. https://doi.org/10.1017/s1351324916000243

Easwara Moorthy, A., & Vu, K.-P. L. (2015). Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, *31*(4), 307–335. https://doi.org/10.1080/10447318.2014.986642

Goh, A.S.Y, Wong, L.L., & Yap, K.Y.-L. (2021). Evaluation of COVID-19 information provided by digital voice assistants. *International Journal of Digital Health, 1*(1), Article 3. http://doi.org/10.29337/ijdh.25

Griffin, R., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environmental Research*, *80*(2), S230–S245. https://doi.org/10.1006/enrs.1998.3940

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital journalism*, *6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Hepp, A. (2020). Artificial companions, social bots and work bots: Communicative robots as research objects of media and communication studies. *Media, Culture & Society, 42*(7–8), 1410–1426. https://doi.org/10.1177/0163443720916412

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical reference services quarterly*, *37*(1), 81–88. https://doi.org/10.1080/19312458.2018.1506021

Huang, Y., & Yang, C. (2020). A metacognitive approach to reconsidering risk perceptions and uncertainty: Understand information seeking during COVID-19. *Science Communication*, *42*(5), 616–642. https://doi.org/10.1177/1075547020959818

Jandura, O., & Friedrich, K. (2014). Quality of political media coverage. In C. Reinemann (Ed.), *Handbook of communication science: Political communication* (pp. 351–373). De Gruyter Mouton. https://doi.org/10.1515/9783110238174.351

Ki, C.-W., Cho, E., & Lee, J.-E. (2020). Can an intelligent personal assistant (IPA) be your friend? Para-friendship development mechanism between IPAs and their users. *Computers in Human Behavior*, *111*(4), Article 106412. https://doi.org/10.1016/j.chb.2020.106412

McKibbin, W. J., & Fernando, R. (2020). The global macroeconomic impacts of COVID-19: Seven scenarios. *CAMA Working Paper No 19./2020*. http://doi.org/10.2139/ssrn.3547729

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37. https://doi.org/10.1016/j.chb.2019.05.009

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, *176*(5), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

Natale, S., & Cooke, H. (2021). Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. *Media, Culture & Society, 43*(6), 1000–1016. https://doi.org/10.1177/0163443720983295

Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2019). *Reuters Institute digital news report 2019*. Reuters Institute for the Study of Journalism, Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Reuters Institute digital news report 2020*. Reuters Institute for the Study of Journalism, Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Nielsen, R. (2017). News media, search engines and social networking sites as varieties of online gatekeepers. In C. Peters & M. Broersma (Eds.), *Rethinking journalism again. Societal role and public relevance in digital age* (pp. 81–96). Routledge.

Nielsen, R., Fletcher, R., Newman, N., Brennen, J. S., & Howard, P. N. (2020, April). *Navigating the 'infodemic': How people in six countries access and rate news and information about corona virus*. Reuters Institute for the Study of Journalism, Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-04/Navigating%20the%20Coronavirus%20Infodemic%20FINAL.pdf

NPR & Edison Research (2020). *The smart audio research*. https://www.nationalpublicmedia.com/uploads/2020/04/The-Smart-Audio-Report_Spring-2020.pdf

Presse- und Informationsamt der Bundesregierung (2020, March 18). *Ansprache der Kanzlerin. „Dies ist eine historische Aufgabe – und sie ist nur gemeinsam zu bewältigen"*. [Speech by the Chancellor. "This is a historic task – and it can only be accomplished together".] https://www.bundesregierung.de/breg-de/themen/coronavirus/ansprache-der-kanzlerin-1732108

Reinhardt, A., Weber, W., Rossmann, C., Stehr, P., Brill, J., Temmann, L., & Wendt, K. N. (2020, November 11–13). *Eine Typologie des Informationsverhaltens der Deutschen in der Corona-Pandemie: Einsetzende Themenverdrossenheit führt zur Zunahme von Informationsvermeidenden* [A typology of Germans' information seeking behavior in the

corona pandemic: Emerging topic fatigue leads to an increase in information avoiders]. 5th annual conference of DGPuK-Division Health Communication, virtual conference.

Rossmann, C. (2020, April 13). *Viel hilft viel – aber nicht bei allen. Informationsverhalten deutscher Bundesbürgerinnen und -bürger in der Coronakrise. Erste Befunde aus dem Projekt CoreCrisis* [A lot helps a lot – but not for everyone. Information seeking behavior of German citizens during the corona crisis. Initial findings of the project CoreCrisis]. https://www.uni-erfurt.de/fileadmin/Verwaltung/Hochschulkommunikation/Blog/Wort-Melder/CoreCrisis_Welle1.pdf

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Journal on Telecommunications and High Technology Law*, *13*(2), 203–218.

Urban, J., & Schweiger, W. (2014). News quality from the recipients' perspective. *Journalism Studies*, *15*(6), 821–840. https://doi.org/10.1080/1461670X.2013.856670

Viehmann, C., Ziegele, M., & Quiring, O. (2020). Gut informiert durch die Pandemie? Nutzung unterschiedlicher Informationsquellen in der Corona-Krise. Ergebnisse einer dreiwelligen Panelbefragung im Jahr 2020. [Well informed through the pandemic? Use of different information sources in the corona crisis. Results of a three-wave panel survey in 2020]. *Media Perspektiven*, *10–11*, 556-577.

Viehmann, C., Ziegele, M., & Quiring, O. (2022). *Informationsnutzung in der Corona-Krise. Report zu den Befunden der Panelbefragung. Befunde aus dem Juli & November 2020*. [Information usage in the corona crisis. Report on the findings of the panel survey. Findings from July & November 2020]. https://www.kowi.ifp.uni-mainz.de/juli-november-2020-befunde-aus-der-zwischenphase-und-zum-zweiten-lockdown/

Wilson, N., MacDonald, E. J., Mansoor, O. D., & Morgan, J. (2017). In bed with Siri and Google assistant: A comparison of sexual health advice. *BMJ (Clinical Research Ed.)*, *359*, Article j5635. https://doi.org/10.1136/bmj.j5635

WHO (2020, February 2). *Novel Corona virus (2019-nCov). Situation Report – 13*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf

Yang, Z. J., Aloe, A. M., & Feeley, T. H. (2014). Risk information seeking and processing model: A meta-analysis. *Journal of Communication*, *64*(1), 20–41. https://doi.org/10.1111/jcom.12071