Users' perceptions of voice assistants (VAs): The interplay of trust, intelligence, information accuracy, and usefulness

Dania Bilal. *University of Tennessee, dania@utk.edu*
Li-Min Huang, *National Taiwan University, cassandralmh@ntu.edu.tw*

Abstract

Voice Assistants (VAs) are AI-powered programs that utilize machine learning, voice recognition, and natural language processing to answer user questions and perform various tasks. Despite the numerous studies of VAs, scant research has examined the factors users employ when evaluating the effectiveness of their VAs. This study examined college students' VAs' usage, perceptions of trust, intelligence, usefulness, and information accuracy. It investigated the interplay between and among the participants' perceptions of these factors in their VAs. This study's findings revealed that while most participants highly trusted their VAs' information accuracy, their conceptions of how they functioned were lacking. One-third of the participants' VAs failed to provide accurate information, resulting in frustration and confusion. Although we found an interplay between and among users' perceived VAs' information accuracy, trust, intelligence, and usefulness, a significant correlation existed between users' perceived trust and the information accuracy it provided. The VAs' accents influenced the participants' trust in the accuracy of the information. This study's findings have implications for designing user-centered VAs, improving users' interactions, and empowering users with AI literacy skills.

Keywords: voice assistants (VAs), user perceptions, college students.

Introduction

Voice Assistants (VAs) have become ubiquitous and part of everyday life. They are AI-enabled and perform various tasks using natural language processing, speech recognition, and Large Language Models (LLMs). Apple's Siri, Google's Assistant, and Amazon's Alexa are the most popular in the United States. There are 4.2 billion VAs in use as of 2024; this number is expected to reach 8.4 billion by the end of 2025 (https://www.demandsage.com/voice-search-statistics). Users interact with VAs using text or speech inputs (voice) to shop online, find information, and perform routine tasks, such as obtaining traffic updates, news, and weather, as well as operating and controlling smart home devices.

Currently, 41% of adults in the United States and 58% globally have used voice search daily. (https://www.demandsage.com/voice-search-statistics). Users may converse with their VAs using text or speech input (voice). VAs retrieve answers to users' questions from web search engines, publicly available sources such as Wikipedia, and encyclopaedias, Questions and Answers (Q&A) platforms such as Quora reference service, and their providers' knowledge bases, such as data repositories. (Wheatley & Hervieux, 2024). VAs are available on computers, mobile devices, and cars and are compatible with many smart home devices. (Terzopoulos & Satratzemi, 2020). Despite the numerous studies on VAs, scarce research has

examined users' perceived effectiveness of their VAs, including information accuracy, trust, intelligence, and usefulness, as well as what they value most. We also lack knowledge of the interplay between and among these factors from the users' perspectives. This study aims to fill this gap. Its findings have implications for designing user-centered VAs, improving users' interactions, and empowering users with AI literacy skills.

Background and Literature Review

Studies revealed that users' perceptions of trust in their VAs were influenced by various factors, including but not limited to failures, lack of feedback, usefulness, content credibility, VAs' expertise, interaction quality, and social interaction and presence (Lee et al., 2021). Research that examined users' perceived intelligence of their VAs revealed that understanding speech input, feedback output, and acceptance of blame when virtual agents committed errors symbolized intelligence (Baughan et al., 2023). Other studies have shown that the VAs' comprehension of voice commands, learning, and human-like abilities impact users' judgment of intelligence. (Poushneh, 2021). The voice accent of VAs was also a factor in judging the reliability and accuracy of the information; mainly, users perceived the British accent to provide more accurate information than the American English accent (Pycha & Zellou, 2024).

Trust in VA
Users' negative experiences, such as failures in VAs, impact their trust and willingness to use them. Baughan et al. (2023) elicited data from Amazon Mechanical Turk (http://mturk.com) workers and created a dataset of 199 failures experienced by 107 participants. They also interviewed 12 users to collect their stories of failures in VAs and surveyed 268 participants to evaluate the impact of failures on their trust in using VAs. The authors found that certain failures, such as a lack of response, misunderstanding speech, misunderstanding the context or the request, truncating a question asked, and providing an incomplete answer, were among those that caused frustration and influenced their willingness to use their VAs in the future.

Studies have explored users' trust in VAs, but few have focused on healthcare AI-enabled VAs. In a survey of 300 users, Zhan et al. (2024) identified factors influencing their trust in VAs. Their findings revealed that the users' perceived usefulness, content credibility, relative service quality, privacy risks, and safety significantly influenced their trust and willingness to use their VAs in the future.

Wienrich et al. (2021) focused on users' perception of VA's expertise and its impact on trust. The authors surveyed 40 university students to examine their perceptions of the Amazon Echo Dot's expertise and trustworthiness. They found that the perceived expertise of the VA and the VA's provider influenced users' trust. Other studies examined users' trust in VAs, including Nasirian et al. (2017) and Lee et al. (2021), whose study findings showed that users' trust was more significantly affected by the quality of their interaction with VAs than information quality and system quality and influenced their intention to use their VAs in the future.

Pitardi and Marriott's (2021) survey of 466 users' perceptions of VAs' trust revealed that human-like qualities, including warmth, competence, and social interaction, were crucial for developing trust in their VAs. Other studies considered the impact of gender and voice pitch on developing trust. For example, Tolmeijer et al. (2021) conducted an online experiment with 234 participants, exploring their perceptions of gender stereotyping on two tasks by analyzing the influence of the VA's pitch (low versus high), gender (women and men), and the interaction between these factors. They also incorporated a gender-ambiguous voice and compared their trust formation. The authors did not find significant differences in trust formed

towards a gender-ambiguous voice compared to gendered voices. Trust also varies by the intellectual maturity of the user group, such as children versus adults. For example, Girouard-Hallam and Danovitch's (2022) study findings showed that children's trust in VAs varied by age and task type. Children trusted VAs to answer factual, simple questions, but resorted to humans for personal questions.

VA's Intelligence

From the human-computer interaction perspective, a VA's intelligence is measured by its ability to understand a user's speech command in natural language and provide relevant answers to questions. (Moussawi & Koufaris, 2019). User perceptions of VAs' intelligence may vary based on the types of questions or tasks, whether they are simple or routine compared to complex, the quality of the answers provided, and the acknowledgment of errors committed. Mahmood et al. (2022) surveyed 37 VA users and elicited data regarding their experiences with AI agents and the errors committed in responding to their questions. Their study's findings showed users judged the AI agents who accepted the blame for their errors as more intelligent, effective, and preferred than those who ignored the blame.

Bawack (2021) surveyed 278 users to measure the perceived intelligence of their VAs. The author found that learning and understanding voice commands significantly influenced user intention to adopt VAs. Poushneh (2021) recruited 275 users and randomly assigned them to the three voice assistants: Microsoft's Cortana, Google's Assistant, and Amazon's Alexa. The author found that the VA's functional intelligence (e.g., effectiveness, efficiency, reliability, and usefulness of information) significantly influenced their interactions with the VAs. In addition, the users' perceived emotional intelligence of the VA, including its ability to exhibit human-like attributes, humor, and modesty) significantly enhanced their interaction experience.

VA's Information Quality

From the information science perspective, assessing the quality of digital information includes credibility, authority, accuracy, usefulness, timeliness or currency, and objectivity (Stvilia et al., 2009). These assessment criteria may not all apply to information quality in VAs, primarily because users interact with their VAs through speech or voice commands. The lack of visual cues in VAs' responses to questions makes it challenging to process information and assess its quality (Brewer, 2023). Specifically, VAs should understand users' speech, questions, and context to provide accurate information. Brewer (2023) involved 30 participants to assess how they evaluated the quality of information on health topics that their VAs retrieved. Participants interacted with five scenarios they played on their devices, and ambiguous pre-recorded VA responses were provided for each scenario. The participants were asked to assess the perceived information quality and the criteria they used to assess it. The study findings revealed that the participants used limited criteria or cues to assess the information quality and were unsure of how to address unhelpful VA responses. The author concluded that VA responses to questions differ from information retrieved by search engines in response to users; queries are due to a lack of visual cues.

Studies revealed that users' perceptions of trust in their VAs were impacted by various factors, including but not limited to failures, lack of feedback, usefulness, content credibility, VAs' expertise, interaction quality, and social interaction and presence. Research that examined users' perceived intelligence of their VAs found that understanding speech input, feedback output, and acceptance of blame when virtual agents committed errors symbolized intelligence. VA's voice accent was a factor in judging the reliability and accuracy of the information. Research revealed that users perceived certain English language accents as influencing their judgment of information quality and accuracy. For example, Pycha and Zellou (2024) recruited 139 native American English speakers. They listened to statements in either an American English or British English accent and rated each statement's accuracy based on the voice accent and level of accuracy (high vs. low). The authors found that participants perceived the information provided by the

British accent voice as more credible than that delivered by the American English accent voice, suggesting that the VA voice is more influential than the content or accuracy of the statement.

Dambanemuya and Diakopoulos (2021) evaluated the information quality of news questions in VAs. Using Amazon's Alexa, the authors compiled a dataset of questions (queries) by topic and another by phrases they extracted from Google's top 20 US trending topics and daily search trends. They found that most of the responses Alexa understood were relevant, predominantly accurate, and timely. However, 60.4% of the responses lacked information sources. The responses varied by query phrasings (formulation) and query categories (e.g., news, sports). Well-formulated queries yielded more accurate information than poorly formulated ones, making it challenging for users to assess the quality of the information. Evaluating news information quality should consider users' accents, pronunciation accuracy, and tone. However, the authors did not suggest exposing users to AI literacy training to help them phrase/formulate good queries.

The reviewed literature revealed that various factors influence users' perceived effectiveness of their VAs, including understanding speech inputs or voice commands and context of the questions, accuracy and information quality of the responses, learning, ability to take blame for errors, possessing human-like attributes, responding in a timely manner, accents, intelligence, usefulness, and reliability of information. However, little research has examined the interplay of key factors in assessing VAs' effectiveness from the users' perspectives. This study aimed to address this gap by posing the following research questions.

RQ1: How do college students' perceived trust, information accuracy, usefulness, and intelligence influence the evaluation of their VAs?

RQ2: What is the interplay between and among college students' perceived trust, information accuracy, usefulness, and intelligence of their VAs?

Method

We employed a structured survey instrument comprising both closed-ended and open-ended questions to collect both quantitative and qualitative data. The survey instrument was adapted and expanded from Bilal and Barfield's (2021a, 2021b) previous work. The findings reported in this paper are based on a subset of that survey and focus on the following aspects: (1) demographic information such as age, gender, ethnicity and VA usage (e.g., frequent use, type of VAs, tasks performed), (2) students' perceived trust in the accuracy of information VAs provided and perceived intelligence, and (3) students' perceived usefulness and unusefulness of VAs based on their satisfactory and/or unsatisfactory experiences. The original survey had been pilot-tested in earlier studies (Bilal & Barfield, 2021a, 2021b) and revised; it was further refined for this study. Survey items related to trust, information accuracy, and intelligence were adapted from Bilal and Barfield's (2021a, 2021b), while the items on usefulness and unusefulness of VAs were developed specifically for this study and were not part of the original instrument.

The participants were undergraduate students—mainly freshmen and sophomores—attending a research-intensive university in the United States. These students were enrolled in two mandatory courses, including research participation as a credit-earning requirement. To facilitate participant recruitment, we utilized the SONA system, a widely recognized, subscription-based platform used by universities nationwide for managing research involvement (https://www.sona-systems.com). Through SONA, students can select from among the posted research studies to participate in, earning course credits in return. Participants are not directly invited or sampled; they voluntarily choose from a list of available research studies or projects. A faculty member oversees the SONA platform at the university where the participants were recruited and

where the study took place. The faculty member approves the studies, ensuring that researchers do not have direct access to the participants or system.

After obtaining approval from the university's Institutional Review Board (IRB), we contacted the faculty member managing the SONA platform. We provided the faculty member access to the survey's URL for review. The faculty member approved and, upon our request, posted the survey in SONA from November 2021 to April 2022. Students enrolled in two communication studies public speaking courses with a research participation assignment offered for credit could then take the survey posted on the SONA platform. Those students could not be invited or sampled to participate in this study. There were 1,700 students enrolled in both courses and they all had access to the survey in SONA. We received (n=632) responses, but not all were valid. After examining the responses, 150 responses were valid and used for analysis.

The survey instrument employed a responsive design layout, allowing students to view identical layouts on both desktops and mobile devices and select the layouts of interest to them for completion. It was adapted and expanded from Bilal and Barfield's (2021a, 2021b) study, which was created using Qualtrics software, available to faculty, staff, and students through the designated university's Office of Information Technology. The original survey consisted of 57 questions, collecting demographic data, the type of VAs owned, experience with using VAs, perceptions of embodied voice interfaces (EVIs), preferences for gendered voice interfaces, perceived age of the voice, information accuracy, trust in VAs, and their perceived intelligence.

The 2021 studies recruited workers (participants) from Amazon Mechanical Turk (http://mturk.com). In the present study, we focused on college students and elicited their perceptions of trust in information accuracy, intelligence, usefulness, and unusefulness of their VAs. The findings in the next section are based on these responses.

Findings

Demographics Information
The participants ranged from 18 to 34 years, with an average age of 19.45 years. Among the 150 participants, 96 (64%) identified as female, 53 (35.33%) as male, and one individual (0.67%) did not disclose their gender. The majority (n = 126, 84%) self-identified as White, while 9 (6%) identified as Black/African American, 9 (6%) as Asian, and 6 (4%) as part of other racial or ethnic groups (e.g., Middle Eastern, Multiracial). All participants were native English speakers who had completed high school in the United States.

VA Usage
Regarding smartphone ownership, 147 (98%) of the participants reported using an Apple iPhone, with the remaining 2% using either Google Pixel or Android devices. Regarding the VAs, 146 (97.33%) had Apple's Siri installed, and 7 (4.67%) had Google Assistant. Of the most preferred VA, 9 participants (60.67%) mentioned Apple's Siri, 48 (32%) favored Amazon's Alexa, and 11 (7.33%) indicated Google Assistant. On average, participants reported using VAs for 4.34 years and interacting with them nearly 15 times weekly.

RQ1a. Perceived Trust in VAs' Information Accuracy
Participants were asked to rate their trust in the accuracy of the information provided by their VAs using a single survey item on a five-point Likert scale, ranging from one (not at all trusted) to five (highly trusted).

More than two-thirds of the participants expressed high trust in their VAs' information accuracy, while one-third rated their trust as average, somewhat untrusted, or not at all trusted. This finding is consistent with that in Bilal and Barfield's (2021b).

RQ1b. Perceived VAs' Intelligence

Participants were asked to rate their perceived intelligence of the VAs on a five-point Likert scale, with 1 indicating "far below average" and 5 indicating "far above average." More than half of the participants rated their VAs' intelligence as "above average" or "far above average," and around one-third provided an average rating. Fewer than 10% of the participants rated their VAs' intelligence as "below average" or "far below average." Compared with Bilal and Barfield's (2021b) study findings, college students in this study tended to perceive their VAs as more intelligent.

RQ1c. Perceived VA Usefulness and Unusefulness

In another single survey item, we elicited the participants' perceptions of the usefulness of their VAs by asking about their most satisfactory experiences using an open-ended question. Nearly half of the participants attributed the perceived usefulness of their VAs to the ability to process their requests accurately. For example, participant 33 (P33) noted, "*When I use Siri to set a reminder, it is always done on the correct dates and times very accurately*." Approximately one-third of the participants highlighted VAs' capability of providing accurate and relevant information as a key factor contributing to the perception of usefulness, as P28 mentioned, "*It answers my questions when asking about sports questions or actors when watching tv shows*". Other participants indicated that the convenience and enjoyment their VAs brought into their daily lives were important factors in shaping the perceived usefulness of their VAs, as P22 shared, "*I told my Google Home it was my birthday and it sang happy birthday to me. It was cute!*"

The participants' perceptions of unusefulness were gathered through another open-ended survey question, where participants reflected on their most dissatisfactory experiences with VAs. The participants felt their VAs were not useful when they: 1) processed their requests inaccurately (reported by 36%), For example, P30 said, "*I wanted to call my brother Josh and it called my friend John.*"; 2) misunderstood requests or queries (noted by 25%), as P77 noted, "*sometimes when I ask for a specific song, Alexa either misunderstands or doesn't understand at all*"; 3) provided irrelevant information (indicated by 20%), as P148 shared, "*When I asked for the weather in Knoxville when I was home in Cleveland, and it gave information on not Knoxville.*"; and 4) for other factors, such as misunderstanding speech or accents, issues with VAs' settings, or going off on thier own, as P74 mentioned, "*It was around 3 AM and my Google was in the other room from me and nobody else was talking. And all the sudden it said recipes for cookies here you go. It was very creepy*".

### RQ2. The Interplay Between and Among Users' Perceptions

To examine the interplay between participants' perceptions of VAs' usefulness/unusefulness, perceived trust in VAs' information accuracy, and perceived VAs' intelligence, we performed a one-way Multivariate Analysis of Variance (MANOVA). Before conducting the test, we used G*Power (version 3.1.9.7) to calculate the required sample size before conducting the MANOVA. Based on a medium effect size ($f^2 = 0.15$), a power level of 0.80, two groups (perceived usefulness and perceived unusefulness), and two response variables (perceived trust and perceived intelligence), the required sample size was 68. Our actual sample consisted of 150 participants, which exceeds the minimum requirement.

The results revealed no statistically significant effect of perceived usefulness/unusefulness on trust in information accuracy or perceived intelligence, $F_{(16, 264)} = 0.869$, $p = 0.605$. This insignificant effect is further supported by Wilk's Lambda test ($\Lambda = 0.902$), suggesting that the explanatory variables only accounted for a small proportion of the variance, with a partial (η^2 of 0.05). These results indicate that

participants' perceptions of VAs' usefulness or unusefulness would not impact their trust in information accuracy or their assessment of VAs' intelligence (see Figures 1 & 2).
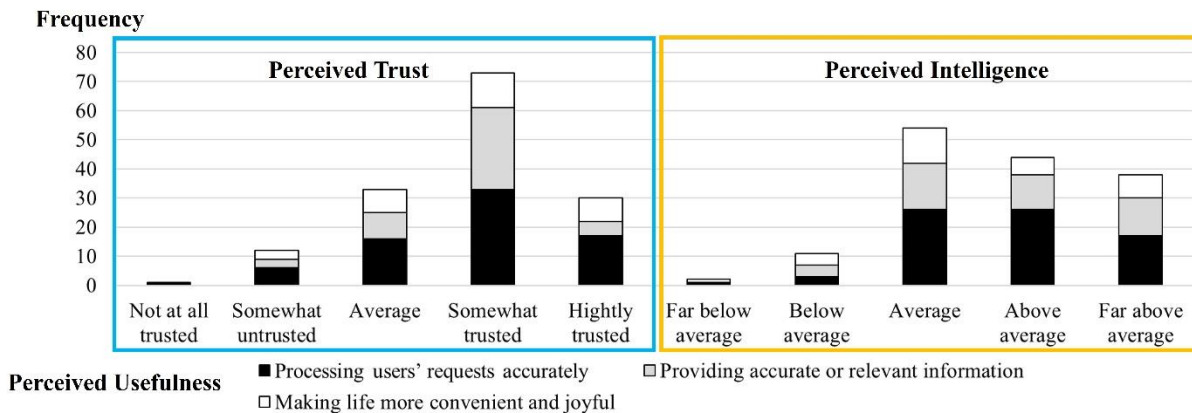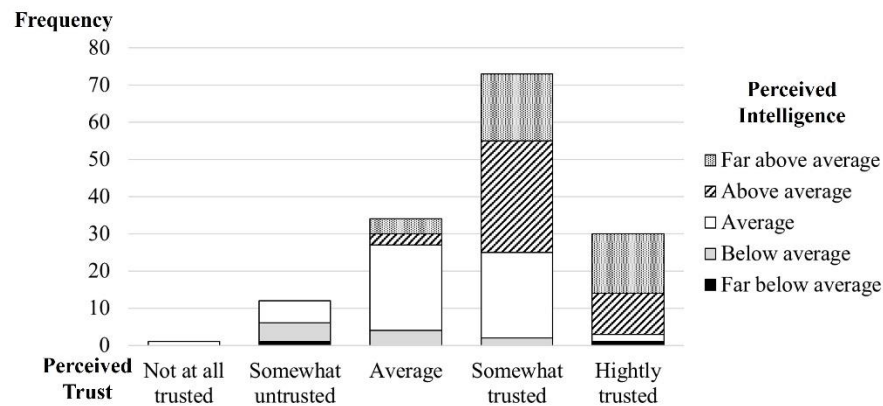

Figure 1. The Interplay of Perceived Usefulness, Perceived Trust in VA's Information Accuracy, and Perceived VA's Intelligence
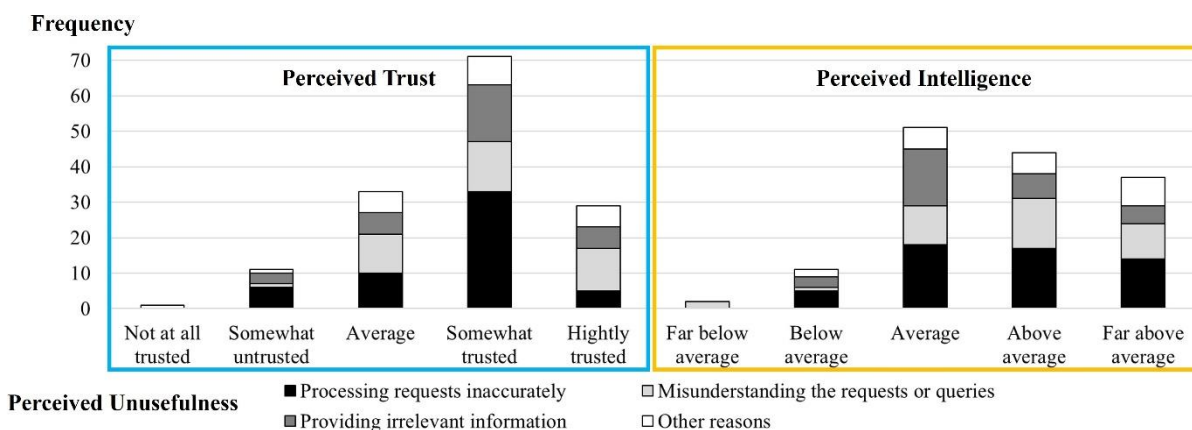
Figure 2. The Interplay of Perceived Unusefulness, Perceived Trust in VA's Information Accuracy, and Perceived VA's Intelligence

In addition, we performed a Spearman's rho correlation test to identify a relationship between the participants' perceived VAs' intelligence and their trust in the accuracy of the information it provided. We used G*Power to calculate the required sample size. Based on a two-tailed model, a medium effect size ($\rho = 0.3$) and a power level of 0.80, the required sample size was 84, which is smaller than our actual sample. We found a significant positive correlation between the two variables ($r\_s = .561$, $p < .001$), indicating that those who perceived their VAs as more intelligent were also more likely to trust the accuracy of the information they provided (see Figure 3).

Figure 3. The Interplay of Perceived VA's Intelligence and Perceived Trust in VA's Information Accuracy

Discussion

Users' trust in the accuracy of information from VAs varies by the type of tasks they perform, such as simple versus complex. The finding showed that some participants praised their VAs for providing accurate information on simple tasks (e.g., weather, traffic) but used other sources (e.g., Wikipedia) to find information on complex ones. This finding has been confirmed in previous studies, including those that involved children. For example, Girouard-Hallam and Danovitch's (2022) found that children used their VAs to ask simple questions but resorted to family members for other questions. Despite using AI to empower VAs, there remains a need for training VAs on datasets that include complex queries. In this study, we did not observe the participants' voice search behaviors and strategies in VAs (using phrases, the length of phrases, or requests) to evaluate their interaction behaviors, successes, or failures using evidence-based data.

Judging VA intelligence varies across published research. However, from the human-computer interaction viewpoint, a VA's perceived intelligence is based on its ability to understand the user's voice commands or speech in natural language and provide useful information that meets their needs (Moussawi & Koufaris, 2019). In this study, we found that some participants perceived their VAs as intelligent because they successfully retrieved the requested information and provided accurate responses, which aligns with this definition of VAs' intelligence, particularly concerning the provision of "useful information" and meeting users' needs. However, the fact that most participants highly trusted their VAs' information accuracy because they used the internet or AI and ML to find information, as P3, for example, indicated, "*I expect my Digital Assistant to provide me with accurate information pretty high because it is connected to the internet which has most all information,*" and P19, who noted, "*due to the fact that Siri is AI, Siri is very intelligent. She cannot relay incorrect information to me due to the fact that she has access to the internet,*" are problematic. These views were shared by most participants and are based on misconceptions and a limited understanding of how VAs work and the type of content available on the internet (as accurate), suggesting a need for formal AI literacy training.

While it is true that VAs utilize AI and ML to process and retrieve relevant information, their ability to comprehend the participants' natural language speech failed in some instances, causing frustration and confusion, which led to the perception that VAs are unintelligent. This finding confirms the interplay between the participants' perceived rust in their VAs and their judgment of their intelligence. It also concurs with previous studies, which showed that failures and errors impacted users' perceptions of trust in their

VAs (Baughan et al., 2023; Mahmood et al., 2022), usefulness, content credibility (e.g., Zhan et al., 2024), and interaction quality (e.g., Lee et al., 2021; Nasirian et al., 2017).

On On the other hand, the participants who highly trusted their VAs valued the accurate information they consistently provided to simple questions, as P55 mentioned: "*I ask simple things like what is the weather and it gets that data from the weather channel nearby so it should be very accurate,*" thus, associating trust with information accuracy. The participants whose trust in their VAs was moderate or minimal attributed the issues to their VA's ineffectiveness, including a lack of understanding of voice commands or questions, retrieving incorrect responses, failure to process requests, not responding to questions, pausing or having "lull moments," and "going off" on their own. For example, P67 noted, "*They can only answer questions if asked in the most perfect format that fits their coding, which rarely happens. I usually can never successfully get the results I am looking for from an AI.*" Despite advances in AI and related *domains*, these issues are still persistent, confirming the findings from previous studies, particularly regarding VAs' misunderstanding of voice inputs or commands, including accents (Baughan et al., 2023; Lee et al., 2021; Nasirian et al., 2017).

The novel finding revealed by this study was a significant correlation between the participants' perceived intelligence of their VAs and their trust in the accuracy of the information. This suggests that these two variables are more important to the participants than the usefulness of their VAs. Since we employed a survey to elicit participants' perceptions of various factors impacting the perceived effectiveness of their VAs, future studies should further explore these factors through usability studies, incorporating observations of interactions and interviews to develop deeper insights into user behaviors, successes, and experiences with VAs.

Limitations

This study collected data from the participants (i.e., college students) using an online survey developed using the Qualtrics software. Surveys have limitations since they rely on self-reported data, which could be prone to inaccuracy. We invited participants to volunteer for individual interviews to triangulate the data. However, we were unable to conduct the interviews due to the participants' time constraints and scheduling issues. In future research, triangulating research methods — such as surveys, observations, focus groups, and individual interviews — would help overcome the limitations of using a single approach and provide more reliable data.

The students participated in this study to earn research credit and fulfill their course requirements. Since we did not have access to the students enrolled in the two required courses for research credit, we were unable to select a random sample. Thus, the findings may not be generalizable to the population of college students. The lack of diversity in the participant pool is another limitation of this study. limitation of the study.

Conclusions

Most participants had misconceptions regarding their VAs, believing that since they are connected to the internet and use online information and AI to find answers to their questions, they provide accurate information. It appears that the participants relied on their experiential knowledge, including trial and error, when using their VAs and lacked formal AI literacy training. Educators and information professionals should lead the way in developing and offering AI literacy programs, including but not limited to how VAs

work, how to evaluate responses to queries, and how to phrase effective questions or requests in natural language. The failures participants experienced require system designers' intervention to provide design improvements that support users' interactions and enhance their cognitive and emotional experiences. As Generative AI (GenAI) is becoming ubiquitous and continues to evolve, leveraging the capabilities of this technology, particularly Large Language Models (LLMs), natural language processing, and speech recognition, it is expected that the design and functionalities of VAs will improve to better support user interactions and experience (UI&UX).

This study advances research in VAs, particularly concerning users' conceptions of how VAs function and their *blind trust* in internet information and AI-enabled VAs as sources of accurate information. It also enhances our understanding of the factors that users consider most significant when interacting with VAs, including trust, information accuracy, and intelligence. The ability of VAs to process not only simple requests but also complex ones and understand speech input is crucial for enhancing the user's experience.. The findings from this study bring attention to the role of formal AI literacy and training and the design of user-centered VAs.

Acknowledgements

References

Bawack, R. (2021). How perceived intelligence affects consumer adoption of AI-based voice assistants: An affordance perspective. In *PACIS 2021 Proceedings*, Article 178. AIS Electronic Library. https://aisel.aisnet.org/pacis2021/178/

Baughan, A., Wang, X., Liu, A., Mercurio, A., Chen, J., & Ma, X. (2023). A mixed-methods approach to understanding user trust after voice assistant failures. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). https://doi.org/10.1145/3544548.3581152

Bilal, D., & Barfield, J. K. (2021a). Voice Switching in Voice-Enabled Digital Assistants (VDAs). In *Human-Computer Interaction. Theory, Methods and Tools: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, Proceedings, Part I 23* (pp. 507-520). Springer International Publishing. https://doi.org/10.1007/978-3-030-78462-1_39

Bilal, D., & Barfield, J. K. (2021b). Hey there! What do you look like? User voice switching and interface mirroring in voice-enabled digital assistants (VDAs). *Proceedings of the Association for Information Science and Technology,* 58(1), pp. 1–12. https://doi.org/10.1002/pra2.431

Brewer, R. (2023). Understanding voice-based information uncertainty: A case study of health information seeking with voice assistants. *Journal of the Association for Information Science and Technology*, pp. 1–17. https://doi.org/10.1002/asi.24854

Dambanemuya, H. K., & Diakopoulos, N. (2021). Auditing the information quality of news-related queries on the Alexa voice assistants. In J. Nichols (Ed.), *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 83. ACM. https://doi.org/10.1145/344915

Girouard-Hallam, L. N., & Danovitch, J. H. (2022). Children's trust in and learning from voice assistants. *Developmental Psychology, 58*(4), pp. 646–661. https://doi.org/10.1037/dev0001318

Lee, O., Ayyagari, R., Nasirian, F., & Ahmadian, M. (2021). Role of interaction quality and trust in the use of AI-based voice-assistant systems. *Journal of Systems and Information Technology, 23*(2), pp. 154–170. https://doi.org/10.1108/JSIT-07-2020-0132

Mahmood, A., Fung, J. W., Won, I., & Huang, C. (2022). Owning mistakes sincerely: Strategies for mitigating AI errors. In *CHI' 22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, Article 578, pp. 1–11. ACM. https://doi.org/10.1145/3491102.351756

Moussawi, S., & Koufaris, M. (2019). Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation. In T. Bui (Ed.), *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 115–124). AIS Electronic Library. http://hdl.handle.net/10125/59452

Nasirian, F., Ahmadian, M., & Lee, O. (2017). AI-based voice assistant systems: Evaluating from the interaction and trust perspectives. In *Proceedings of the 23rd Americas Conference on Information Systems* (Vol. 1, pp. 46–55). AIS Electronic Library.

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but…Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology and Marketing, 38*(4), pp. 626–642. https://doi.org/10.1002/mar.21457

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services, 58*, 102283. https://doi.org/10.1016/j.jretconser.2020.102283

Pycha, A., & Zellou, G. (2024). The influence of accent and device usage on perceived credibility during interactions with voice-AI assistants. *Frontiers in Computer Science, Human-Media Interaction, 6*, pp. 1–9. https://doi.org/10.3389/fcomp.2024.1411414

Stvilia, B., Mon, L., & Yi, Y. J. (2009). A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology, 60*(9), pp. 1781–1791. https://doi.org/10.1002/asi.21115

Terzopoulos, G., & Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. *Informatics in Education, 19*(3), pp. 473–490. https://doi.org/10.15388/infedu.2020.21

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Marco, J., Leimeister, M., & Bernstein, A. (2021). Female by default? Exploring the effect of voice assistant gender and pitch on trait and trust attribution. In Y. Kitamura, A. Quigley, K. Isbister, & T. Igarashi (Eds.), *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 455, pp. 1–7). ACM. https://doi.org/10.1145/3411763.3451623

Wienrich, C., Reitelbach, C., & Carolus, A. (2021). The trustworthiness of voice assistants in the healthcare context: Investigating the effect of perceived expertise on the trustworthiness of voice assistants, providers, data receivers, and automatic speech recognition. *Frontiers in Computer Science, 3*, 685250, pp. 1–12. https://doi.org/10.3389/fcomp.2021.685250

Wheatley, A., & Herviewx, S. (2024). Comparing generative artificial intelligence tools to voice assistants using reference interactions. *Journal of Academic Librarianship, 50*(102942). https://doi.org/10.1016/j.acalib.2024.102942

Zhan, X., Abdi, N., Seymour, W., & Such, J. (2024). Healthcare voice AI assistants: Factors influencing trust and intention to use. In J. Nichols (Ed.), *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), Article 62, pp. 1–37. https://doi.org/10.1145/363733

<div align="center">

Appendix. Survey Instrument

</div>

The original and comprehensive version of this survey was created by Dania Bilal and Jessica Barfield (2021a-b). Additional questions were added to the survey by Dania Bilal and Li-Min Huang for this study. The survey was created in Qualtrics software. The layout of this survey in Qualtrics differs from the layout shown here.

Part 1: Demographic questions

1. What is your age in years?   _____

2. What gender do you identify with?
- ☐ *Male*
- ☐ *Female*
- ☐ *Non-binary / third gender*
- ☐ *Prefer not to say*

3. What race or ethnicity do you most identify with?
- ☐ *White*
- ☐ *Black or African American*
- ☐ *American Indian or Alaska Native*
- ☐ *Asian*
- ☐ *Native Hawaiian or Pacific Islander*
- ☐ *Other*
- ☐ *Prefer Not To Say*

4. In which country did you attend high school?

Part 2: Screening questions

1. Which of the following items do you personally own? Select all that apply.
- ☐ Bluetooth Speaker
- ☐ Digital Assistant (e.g., Google Home, Amazon Alexa)
- ☐ Sound Bar
- ☐ Smart TV
- ☐ Smartphone
- ☐ High Speed Internet
- ☐ Laptop Computer
- ☐ Desktop Computer

2. Which Digital Assistant(s) do you own? Select all that apply.
- ☐ Google Home
- ☐ Apple Siri/ Apple Homepod
- ☐ Amazon Alexa
- ☐ Microsoft Cortana
- ☐ Other. Please specify the Digital Assistant that you own. _____

Part 3: Background information

1. What type of mobile phone do you own?
- ☐ *Apple iPhone*
- ☐ *Google Pixel*
- ☐ *Windows*
- ☐ Other Android

2. Which Digital Assistant do you use the most?
- ☐ Apple Siri
- ☐ Amazon Alexa
- ☐ Google Assistant
- ☐ Microsoft Cortana

3. Which Digital Assistant do you have on your mobile phone? Select all that apply.
- ☐ Google Assistant
- ☐ Apple Siri
- ☐ Microsoft Cortana None/ Other

4. What is your English language proficiency level?
- ☐ Beginner (Basic)
- ☐ Intermediate (More than Basic)
- ☐ Advanced (Proficient)
- ☐ Native (Fluent)

Part 4: Experiences and perceptions of digital assistants

1. For the remainder of the survey, we will be asking you about your most frequently used Digital Assistant. Please type in the model/name/brand of the Digital Assistant you will be referring to for the rest of this survey.

_____

*2. Describe an experience you had with your Digital Assistant when you were **very satisfied** with the answer or outcome.* _____

*3. Describe an experience you had with your Digital Assistant when you were **very dissatisfied** with the answer or outcome.* _____

4. How long have you been using your Digital Assistant? _____Years _____Months

5. How often do you use your Digital Assistant? (Number of times Per Week)_____

6. To what extent do you trust your Digital Assistant to provide you with accurate information?
- ☐ *1: Not at all*
- ☐ *2*
- ☐ *3*
- ☐ *4*
- ☐ *5: Very high*

7. Please explain your answer to the previous question.

_____

8. How intelligent is your Digital Assistant?
- ☐ Far above average
- ☐ Somewhat above average
- ☐ Average
- ☐ Somewhat below average
- ☐ Far below average

9. What makes your Digital Assistant intelligent/unintelligent? Please be specific.

_____

Part 5: Follow-up interview recruitment

For this project, we are selecting a number of participants for a 1-on-1 observation of using their digital assistant and an interview.

The interview will be conducted with the researchers and will likely be held via Zoom to gather additional insights about your experience with Digital Assistants.

The observation and interview will last approximately 30 minutes and will be scheduled at your convenience. The observation and interview will be recorded and used for this study only. You may be eligible for an additional research credit by participating in an interview and observation.

Are you interested in participating in the observation and interview?
- ☐ *Yes*
- ☐ *No*