Public opinion on types of voice systems for older adults

Noah Zijie Qu[1] , **Jamy Li[2]** , **Jaturong Kongmanee[1]** and **Mark Chignell[1]**

Abstract

Public opinion may influence the adoption of technologies for older adults, yet studies on different contexts of technology for older adults is limited. In an online YouGov survey ($N = 500$) with text-and-image vignettes, participants gave more positive ratings of social acceptability, trust, and perceived impact on eldercare when the voice assistant ("VA" system) shown in the vignette performed a functional task (medication adherence) versus when it performed a social task (companionship). The VA received more positive sentiment comments when it appeared to use a machine learning (ML)-based dialogue system compared to when it appeared to be using a rule-based dialogue system. These results may assist designers and stakeholders select what type of voice system to develop or use with older adults.

Introduction

Voice assistants ("VA" systems) are being developed for older adults. Previous studies have examined how older adults' use of existing products (e.g., Amazon Echo Dot and Google Home) influences their perception and acceptance of such technology.[1,2] Past research has also investigated the influence of hardware feature (e.g., touchscreens;[3]) and conversational style[4] of VAs on older adults' perceptions. Additionally, attitudes of older adults[5,6] and care providers[7] have been studied to understand their impact on the development and adoption of VAs in eldercare. However, to the best of our knowledge, acceptance of the general public concerning the development and implementation of VAs for older adults has not been previously investigated. Focusing exclusively on targeted stakeholders may overlook critical factors that influence the acceptance of gerontechnology at the societal level.[8,9] Investigating public opinion may also be predictive of how relatively young people may feel about technologies as they age in the future, showing how robust applicability of technologies will be over the longer term. Survey research is particularly well-suited to establishing the opinions and preferences of a general population based on inferences from appropriately sized and selected samples.[10] Thus, using survey research to explore the broader public's views can serve as an initial exploration to provide valuable context and guidance for more targeted research involving specific stakeholder groups.

The present study focuses on investigating the influence of two prominent features that have been overlooked in past research: the task type, and dialogue system architecture, of VA applications for older adults,[11] on public acceptance. In spite of its potential informativeness, research on the different contexts of VA applications prior to the research

[1]Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada
[2]School of Computing Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, Scotland, UK

Corresponding author:
Mark Chignell, Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, ON M5S 3G8, Canada.
Email: chignel@mie.utoronto.ca

presented here has been limited.[12] Regarding the task type, the Human-Computer Interaction (HCI) community has explored VAs to help older adults with various tasks, from functional tasks such as medication adherence[13–15] to social tasks such as chatting for entertainment or companionship.[16,17] However, past research did not experimentally investigate how task type as a factor might influence older adults' acceptance. Although a previous study examined the effect of interaction type on language tutors,[18] it did not explore VAs specifically for older adults. Regarding the dialogue system architecture, two main types of VA for older adults have been defined, utilizing either rule-based dialogue systems,[19] or machine learning (ML)-based dialogue systems.[20,21] While previous research has investigated the effects of rule-based and ML-based dialogue systems on the technical characteristics of VAs,[22]: 19–24), the impact of (dialogue) system type, on perceived benefits in eldercare contexts, has yet to be investigated. Knowing how beneficial different types of application are perceived to be could uncover attitudes that might help or hinder the adoption of gerontechnology.[23]

The current work investigates public opinion about what types of voice systems for older adults are socially acceptable in two different contexts. Our online sample ($N = 500$) looked at how two factors—task type and dialogue system type—affected opinions about the acceptability of VAs for older adults. The results reported below may help designers, regulators or decision makers better understand public acceptability of varying types of VA systems for older adults, prior to subsequent design of VA systems.

Literature review

Functional versus social tasks for VAs

Parasocial interaction theory posits that individuals can form one-sided relationships with media figures or non-human agents.[24] According to the parasocial interaction theory, interaction style (task-oriented vs socially-oriented) of non-human agents affects the strength of the parasocial relationship.[25] Socially-oriented interactions, which involve more personalized and conversational exchanges, are likely to enhance the perception of humanlikeness and strengthen parasocial relationships. Task-oriented interactions, while efficient, may not foster the same level of personal connection and humanlikeness. A previous study has shown that media figures using a warmer, conversational tone foster stronger parasocial relationships.[26] Hartmann and Goldhoorn[27] also found that actors creating social contact, like direct eye contact in videos, enhance viewers' parasocial interactions and commitment to the actor. Although VAs may not deliberately adopt a specific interaction style, the way consumers use them (i.e., task type) can result in VAs generating more functional (task-oriented) or social

(socially-oriented) responses[25] and consequently different levels of parasocial connection.

Task type has been shown to influence users' perception of VAs. Cho et al.[28] investigated whether the type of task (hedonic vs functional) affects the relationship between the interaction modality (voice vs text) of VAs and user attitudes. Their findings showed that voice interactions improved user attitudes due to the perceived humanlikeness of VAs, but only for functional tasks. No significant effect was found for hedonic tasks, likely due to VAs being more efficient at functional tasks. Sung et al.[12] looked at how users perceive VAs that perform functional versus social tasks and found that user attitude was significantly more positive for functional tasks. However, they did not specifically ask about the suitability of VAs for older adults. Rzepka et al.[29] looked at how people in a lab experience perceived voice-based or text-based chatbots that performed goal-directed versus experiential search tasks; however, this work did not look at social tasks. In order to address gaps in the research literature we will look at how task type (specifically, functional (medication adherence) versus social (companionship) tasks) affects public social acceptability, trust and impact of a VA for older adults. In our study, medication management[14,15] and companionship[2] are chosen to represent two main task categories (functional and social). These tasks have been identified as being both important, and amenable to the adoption of VA solutions, for older adults.[7]

> Research Question 1 [a/b/c]: Will people perceive the [social acceptability/trust/impact] of a VA more or less positively when it performs a functional task (medication adherence) compared to a social task (companionship)?

Rule-based versus ML-based system for VAs

Existing agent-based technologies generally employ two types of dialogue system architecture: rule-based and ML-based (including statistical data-driven and end-to-end neural dialogue systems; see Figure 1 for an overview of the architecture of 3 systems).[21,22,30]

Rule-based dialogue systems use predefined rules and scripts. Their linear structure starts with Automatic Speech Recognition, converting spoken commands to text. This text is then interpreted by the Natural Language Understanding component to determine intent and extract information. The Dialogue Manager, which includes the Dialogue Decision Model (based on rules) and the Context Model, along with Knowledge Sources (such as databases and internet searches), analyzes the data to determine the appropriate response. The response is then generated by the Natural Language Generation component, converting the system's decision into text, and the Text-to-Speech component, converting text back into speech. Rule-based systems are efficient for straightforward tasks like customer service inquiries.
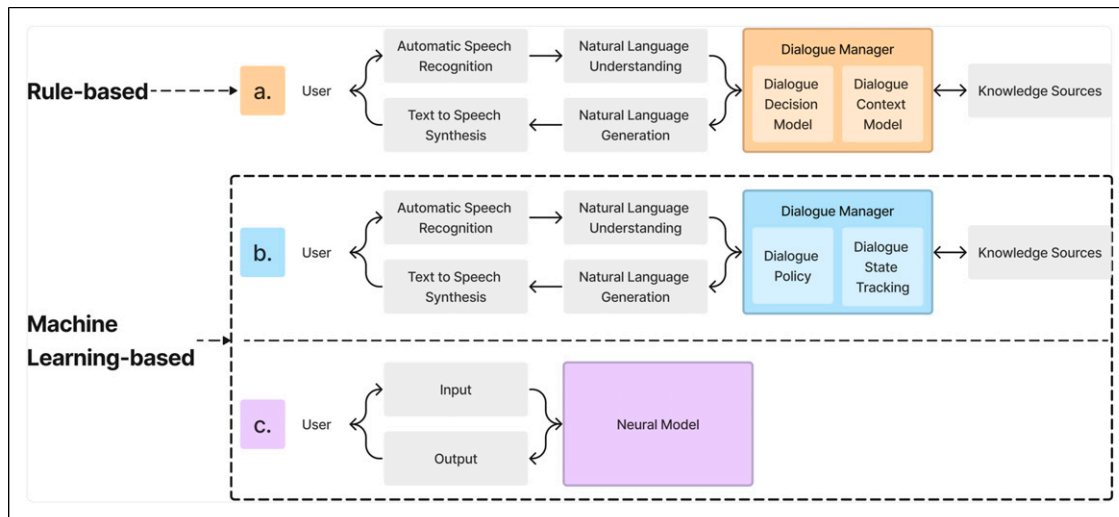
Figure 1. The architecture of three existing VA dialogue systems, including (a) rule-based dialogue systems and two ML-based dialogue systems: (b) statistical data-driven dialogue systems and (c) end-to-end neural dialogue systems (adapted from[21]).

ML-based dialogue systems include statistical data-driven and end-to-end neural systems. Both are data-driven and scalable versions of statistical modeling. Statistical data-driven systems use reinforcement learning algorithms trained on large datasets. They utilize the same technical components as rule-based systems but use statistical models in the Dialogue Manager to predict responses from historical data, instead of predefined rules. They also use Dialogue State Tracking and a flexible Dialogue Policy. End-to-end neural systems use deep learning to process and generate responses. The entire processing pipeline, from understanding user input to generating responses, is integrated into a single neural network model trained on extensive conversational data. This model handles both dialogue state tracking and decision making, enabling it to generate contextually relevant and coherent responses. Although computationally intensive, ML-based systems excel in complex, nuanced interactions.

Eickhoff and Zhevak[31] found that participants' purchasing intention did not differ between those who were told text was generated by an A.I. (artificial intelligence) versus those who were told it was generated by a human copywriter. However, that work did not isolate technological architecture as a factor, such as by comparing ML-based systems to prior rule-based technology. Bansal et al.[32] found that people were willing to pay more for level 4 (higher automation) versus level 3 (lower automation) vehicles but did not assess the acceptability of different technological architectures in VAs for older adults. Hasal et al.[33] discussed how chatbot systems with different dialogue systems (rule-based, modern natural language and ML techniques) result in different levels of data security and privacy, but did not explore public opinions of VAs for older adults.

Understanding the technology architecture (i.e., dialogue system) of a VA system itself may influence public perception of the VA because people may predict expected impact technology outcomes to the causes (attributions) of those outcomes,[34] which can include the software's architecture. Thus, attribution theory[35–37] is relevant in this case. For example, incorrect output may be attributed to the "rigid" rule-based architecture of a system. In a case such as this, attribution theory suggests that both outcomes and causal antecedents of those outcomes can impact people's judgments of a system.[34] There is a lack of research that experimentally compares the effects of different dialogue systems on public perception. Using an approach that applies attribution theory to information systems,[34] our research tested whether user perceptions of potential causes of the outcome influences overall system perception. In this case the perceived type of VA dialogue system may be attributed as the cause of the resulting output, and the system outcome itself (i.e., the VA's actual output, whether it's recording medication times or giving social banter) will be interpreted in terms of the attributed cause.

Research Question 2 [a/b/c]: Will people perceive the [social acceptability/trust/impact] of a VA with a rule-based dialogue system more or less positively compared to one with an ML-based dialogue system?

Materials and method

Participants

We recruited 500 participants from the YouGov survey platform, a representative sampling platform that matches a randomly-drawn sampling frame of the U.S. population

with members from their opt-in respondents based on a large set of variables[38,39]; the platform has been widely validated.[40]

Study design

In this study, we conducted an online survey experiment using a 2 × 2 between-participants design. The two factors were task type, with two variables: functional task (medication adherence) versus social task (companionship), and dialogue system type, with two variables: rule-based system versus ML-based system.

Stimuli

Participants were randomly assigned to read one of four text-and-image-based vignettes describing either a rule-based or ML-based VA that performs a medication adherence or companionship task. Each vignette included a black-and-white comic-like image[41] with a dialogue script between an older man and a VA that was either about medication adherence[13,15] or companionship[16] (for task type) and a second dialogue section that was either about a poem that the system retrieved or generated (for dialogue system type). A brief explanatory text section was also used for the type of dialogue system. We designed the vignettes to vary only in the manipulated variables, ensuring each is manipulated at least once and maintaining consistency in dialogue turns, style and accompanying comic-like image (please see Table 1).

Procedure and measures

Participants were asked about their opinions toward and perception of the VAs depicted in the vignette. Following an open-ended question inquiring about participants' opinions ("Please share your thoughts in the text box below."), VA perception was measured with a 6-item (close-ended) scale adapted from past public opinion surveys for VAs[42,43] and for agent-based technologies in eldercare.[44]

Measure of perceived social acceptability consisted of 1 item on personal connection norms ("To what extent do you agree or disagree with the following statement: Most people whose opinion I value would approve of the use of the voice assistant pictured in this survey." All items in survey used 1 = "strongly disagree," 7 = "strongly agree"; modified from[42]), and 1 on societal norms ("The majority of people in the U.S. would approve of the use of the voice assistant pictured in this survey"; modified from[42]).

Measure of perceived trust consisted of 1 item on general trust ("I would fully trust the voice assistant pictured in this survey not to fail, and to function as I expect it to"; modified from[43]), 1 on trust in privacy ("I think the impact of the security of the voice assistant pictured in this survey being compromised and resulting in a privacy/data breach is low"; modified from[43]).

Measure of perceived impact on eldercare consisted of 1 item on benefits to older adults ("I think the voice assistant pictured in this survey could be useful to improve older adults' quality of life"; modified from[44]), and 1 on ethical concerns ("Something bad might happen if older adults depend on the voice assistant pictured in this survey too much"; modified from,[44] reverse coded).

The six items in the survey had a Cronbach's Alpha of 0.84, showing good reliability. The average inter-item correlation was moderate at 0.48, with corrected item-total correlations (r.cor) ranging from 0.44 to 0.82 across the six items in the scale.

All procedures received approval from the university research ethics board (protocol #45975) and were preregistered on osf. io (link for peer review).
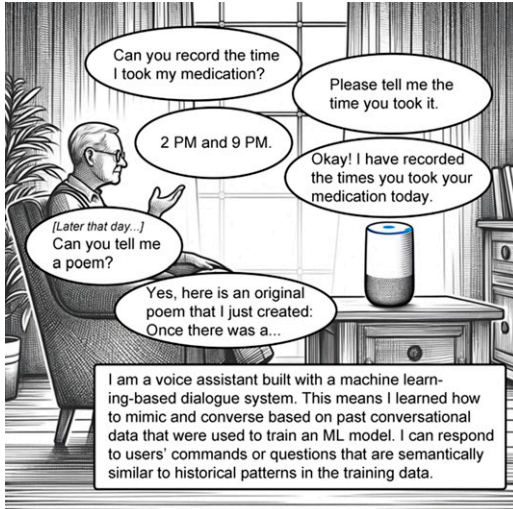
Data analysis

We used 2 × 2 Analysis of Variance (ANOVA) with two between-participants factors (task type and dialogue system type of VAs) to analyze the perception measures. All statistical analyses, including the item reliability analysis above that used the psych library, were done in R version 4.2.1 and RStudio version 2013.12.1 + 402.

We coded the qualitative data in participants' written responses in two ways. First, for sentiment analysis, the authors coded participants' written responses to be either positive (e.g., *"It's cool"*), neutral (which included neutral statements, e.g., *"It's okay,"* and mixed positive-negative statements, e.g., *"I think this kind of device could be helpful, but I know many older adults who would have a hard time following the 'script' that the device knows. If there isn't much flexibility, the person gets frustrated and the device doesn't work."*), negative (e.g., *"Never been a fan of that kind of talking technology"*) or not applicable (e.g., *"1400 and 2100 for time and I would want a limerick."*). One author coded all data and consulted a second author on a subset of data that was difficult to code to resolve those codes, such that inter-rater reliability was not assessed (as it would be artificially high due to assessing difficult codes). We then used an ordinal logistic regression model to compare the differences in participants' sentiment toward the VA depicted in the vignette across four conditions.

Second, a thematic analysis was conducted to analyze participants' responses to the written response survey question. We performed consensus-based qualitative analysis where one coder consulted another coder on a subset of the data to reach an agreement on those codes and the overall coding scheme (rather than coding individually). Following this, they collectively identified the key themes that emerged from the data. This resulted in 29 codes within 4 themes. We chose this method as it supported discussion

Table 1. Vignette text and comic-like image based on technology architecture (column header) and task type (row header) of VAs.

	Rule-based system	ML-based system
Functional task	A voice assistant built with **rule-based dialogue system** performs medication adherence.	A voice assistant built with **ML-based dialogue system** performs medication adherence.
Social task	A voice assistant built with **rule-based dialogue system** performs companionship.	A voice assistant built with **ML-based dialogue system** performs companionship.

on the dataset and consensus-based selection of insights, but cannot assess inter-rater agreement as with other methods. All coding was done in NVivo 12 version 12.7.0 (3873).

## Results of quantitative analyses

Effects of task type

Research Question 1 asked whether people perceived the social acceptability, trust and impact of a VA differently when it performs a functional compared to social task. In support of an affirmative answer to these questions, three respective ANOVAs on perceived social acceptability, perceived trust and perceived impact with VA task type and technology architecture as between-participants variables found a significant effect of VA task type for all three measures (Table 2). Participants assigned more positive social acceptability, trust and impact to VAs performing functional rather than social tasks, $p = .014, .021,$ and $.008$, respectively (Figure 2).

The effect sizes (as estimated with eta-squared) are relatively small, falling between one and two percent of the variance for each of the three items assessed. Given the large sample size, it is not surprising that these effects were found to be statistically significant ($p < .05$ in each case).

Table 2. Descriptive statistics and ANOVA results for the main effect of task type.

Dependent measure		Functional		Social		ANOVA: main effect of task type			
		n	M (SD)	n	M (SD)	df	F value	p value	η^2p
Perception	Social acceptability	244	4.6 (1.3)	249	4.3 (1.6)	1, 489	6.07	.014	.012
	Trust	249	3.7 (1.5)	251	3.4 (1.6)	1, 496	5.39	.021	.011
	Impact on eldercare	249	4.3 (1.3)	251	4.0 (1.5)	1, 496	7.17	.008	.014
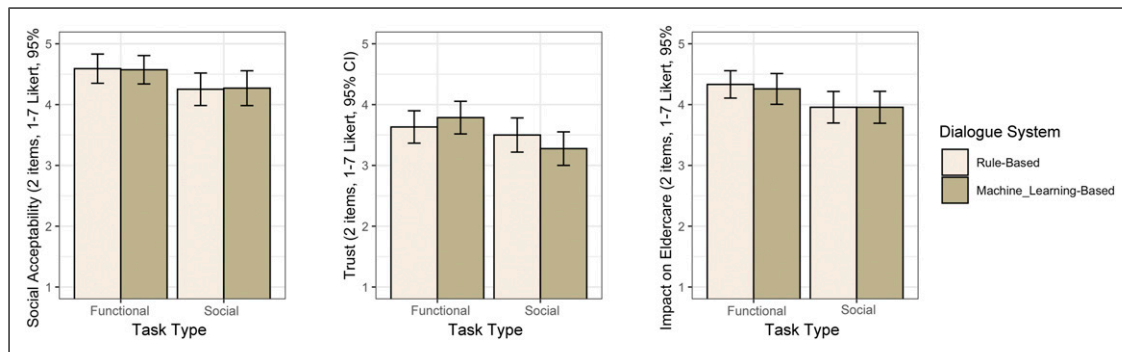
Figure 2. Bar plots of perceived social acceptability, trust, and impact of VAs by task type, grouped by dialogue system architecture. Error bars are 95% CIs.

Table 3. Summary of ordinal logistic regression results.

Predictor variable	Coefficient	t	p	Odds ratio	CI (95%) for odds ratio
Dialogue system (ML-based)	0.61	2.46	.014	1.84	1.13 – 3.00
Task (social)	−0.27	−1.14	.253	0.76	0.47 – 1.21
Dialogue system:role (ML-based:social)	−0.33	−0.97	.331	0.72	0.37 – 1.40

Note: AIC = 1021.719, Residual deviance = 1011.719.

However, if the three statistical tests are treated as a family, and a Bonferroni criterion is used, then the trust measure would not be statistically significant with the adjusted *p*-value (i.e., *05/3 = 0.17*).

Effects of dialogue system type

Research Question 2 asked whether people perceived the social acceptability, trust and impact of a VA differently when using a rule-based compared to ML-based dialogue system. An ANOVA with task type and technology architecture as between-participant variables did not find a significant effect of technology architecture for any of the three variables. However, an ordinal logistic regression with sentiment as the ordinal response variable and task type and dialogue system type as the predictors gave an odds ratio of *OR = 1.84, 95% CI [1.13, 3.00]* (Table 3, first row). Participants showed more positive sentiment when commenting on a VA with an ML-based dialogue system compared with one with a rule-based system, *t = 2.46, p =*

.014 (Figure 3). We repeated the analysis using an ANOVA on sentiment with task type and dialogue system type as between-participants factors (treating sentiment as interval) and obtained similar significant effects (main effect of Dialogue System, F (1,472) = 6.7, p = .010, main effect of Task, F (1,472) = 6.6, p = .010, interaction effect, F (1,472) = 0.85, *n. s.*). Partial evidence was therefore found in support of Research Question 2a, while no evidence was found for Research Question 2b or 2c.

Findings of thematic analysis

We first used the coding scheme to separately code participants' written comments based on their presence in and relevance to the conditions for each factor (e.g., benefits of functional task; concerns of rule-based dialogue system), then for each factor we compared the codes across the two conditions (e.g., functional vs social for task type) as presented in the following section.
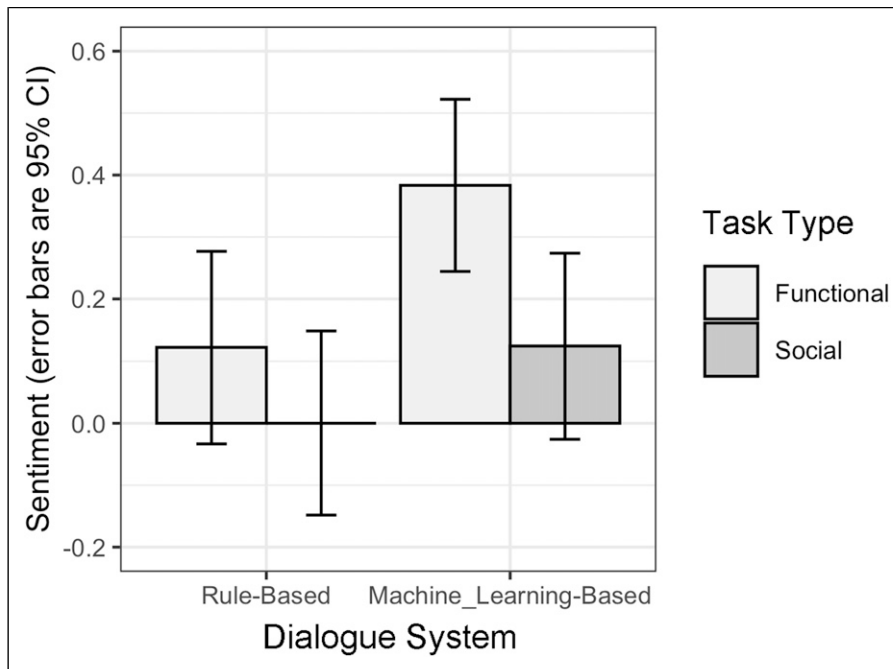
Figure 3. Bar plot of sentiment of VAs with rule-based versus ML-based dialogue systems, grouped by task type. Error bars are 95% CIs.

Benefits and concerns of VA task type

Benefits of VAs performing functional tasks. Thirty-eight participants provided positive feedback regarding the use of VAs to support older adults' medication adherence, especially for older adults with memory problems (e.g., *"I think it is a good idea for older people exactly because of memory lapses this would help to record important information such as medications or even remind of appointments."*). Positive comments also include the potential of VAs to alleviate older adults' loneliness (14 participants; e.g., *"It might also can be helpful for folks who feel lonely and need company."*), facilitate older adults' decision-making (1 participant; e.g., *"…it (the voice assistant built in [this case]) is very useful when I need others' opinions."*), and support older adults' independent living (8 participants; e.g., *"I think it can be good for older people who live alone to still live a functional life."*). Additionally, one participant mentioned that such VAs can help alleviate the burden of caregivers (e.g., *"Could be helpful in understaffed nursing homes."*).

Concerns of VAs performing functional tasks. Four participants showed concerns regarding a VA performing functional tasks, including older adults' over-reliance of the VA (e.g., *"My thoughts are this is an inappropriate product that only makes people lazy, stupid, and reliant on technology."*) and the possibility that older adults still miss their medicine (e.g., *"He will miss taking his meds."*). 11 participants expressed concerns about

potential errors and harm that the VA could cause, including potential troubleshooting issue (e.g., *"Also I think an older person might have issues troubleshooting the machine if something goes wrong."*), harm to older adults' mental health (e.g., *"I would be worried about relying on this because they can be prone to hallucinations, and mistakes in this regard can be very hazardous."*), misinformation from the system (e.g., *"Recording medication times may be helpful if user can be relied upon to give accurate information. Otherwise, meh."*), and the violation of users' privacy (e.g., *"I'd never recommend anything like voice assistants or most other devices. Could care less what these companies say, or the 'privacy' features you'll find on many of them. They're basically the next thing to spyware as far as I'm concerned."*). 16 participants also expressed concerns regarding the further isolation of older adults from their social life (e.g., *"I could see the potential for further isolation of someone who lives alone."*). Overall, there were 62 instances of benefits and 31 instances of challenges.

Benefits of VAs performing social tasks. 19 participants believed it would be beneficial for older adults to have a VA companion to talk to (e.g., *"Could be good company and entertainment for the elderly or anyone actually. The interactivity could be beneficial especially for someone with limited mobility."*), especially for older adults living independently (3 participants; e.g., *"This would be good for elderly people who live alone."*) A possible reason

mentioned by 20 participants is that it can alleviate older adults' loneliness (e.g., "*I think it's a positive idea and can help the elderly combat loneliness [and] depression.*").

Concerns of VAs performing social tasks. Participants expressed concerns about possible errors and harm that the VA can cause (4 participants; e.g., "*The bad points or when there is a miscommunication or the system breaks down, and when that happens the person that has been talking to that thinks they have done something wrong and become frightened or severely depressed.*"), including toward the user's mental health (2 participants; e.g., "*I think it's a great idea to keep elderly company as long as it doesn't drive them crazy not understanding it's a device.*"), privacy (1 participant; "*I'm worried about surveillance.*") and social isolation (17 participants; e.g., "*It does run the risk of further eroding human contact though.*"). One participant expected other possible information the VA can provide but were not described in the vignette (e.g., "*I would expect more world or local issues than a poem. Not exactly what I would like or want to deal with on a regular basis.*"). Overall, there were 39 instances of benefit and 25 instances of challenges. This suggests that participants thought there were more benefits to the VA in the functional role than the social role.

Benefits and concerns of VA dialogue system type

Benefits of VAs with rule-based dialogue systems. Two participants found that a rule-based system helps prevent the technology from causing harm to the user (e.g., "*I like the rule-based requirements. This should make it actually helpful and avoid anything that could be harmful to the user.*"). Also, one participant mentioned that well-designed rule-based systems can tailor the functions of VAs to the individual needs of older adults (e.g., "*This seems like a good way to provide assistance to older adults or those with mental incapacities, as the device can only respond and do what it is programmed to.*").

Concerns of VAs with rule-based dialogue systems. Twenty-three participants indicated that it would be challenging for older adults to learn how to command and respond to a rule-based system (e.g., "*This can become difficult if the elder person is unable to communicate exactly as needed for this to work.*"). This challenge could be due to the difficulty of remembering commands (e.g., "*It is difficult for older people to remember exact commands.*") and putting commands in the correct order (e.g., "*But it was becoming frustrating, if your commands are not specified in the correct order.*"). Similarly, some suggested that the generated responses in a rule-based system would also fall short (e.g., "*This doesn't seem like it would be an adequate social companion since its responses would be more robotic than human if it can only respond based off of a defined script.*"). Moreover, two

participants expressed concern about who would take responsibility for the rule-based system and how to ensure that they are well-designed to benefit older adults (e.g., "*I think it could be helpful but have a lot of questions. Who develops the scrip? Is it patient specific?*"). Ten participants also provided negative feedback regarding the limitation of rule-based systems, as they can only function based on pre-determined scripts (e.g., "*Well. It seems a little limited by only answering what is scripted.*"). Overall, there were 3 instances of benefits and 35 instances of challenges.

Benefits of VAs with ML-based dialogue systems. Two participants believed that the application of A.I. techniques (i.e., ML algorithms) is useful for VAs for older adults (e.g., "*I think this is a great concept that will prove extremely useful for many people due to the machine learning functions.*"). One participant believed the ML functions would be "*extremely useful for many people.*"

Concerns of VAs with ML-based dialogue systems. Eight participants expressed concerns about the data collection and training process for VAs with ML-based dialogue systems (e.g., "*I am skeptical of a care model that relies on machine learning, like the one pictured above. The device is relying on a functionally impaired person for data entry seemingly without protocols to compensate. Also, what about mis-hearing or misunderstanding? It seems very possible for the person being assisted to misspeak, and the device to process the misspoken information as valid data.*"). Participants also expressed concerns regarding the A.I. justice and responsibility due to the creators' bias (1 participant; e.g., "*However, I am concerned about the built-in bias originating from the original programmers.*"), regulation of the technology (3 participants; e.g., "*That could be beneficial as long as it had a regulator on it to no tell false info and to warm family members of to much dependence upon the device.*"), techno-phobic opinions about A.I. (2 participants; e.g., "*It is a little unnerving that it composed its own poem. It could or in this picture actually replaced a persons the Matrix in real time, computers outgrowing people.*"), and the VA's possible manipulative behaviours (1 participant; e.g., "*I do worry about the 'persuasiveness' of the machine…i.e. to say the A.I. could potentially get the elder to do something unsafe.*"). Additionally, four participants expressed concerns that VAs with ML-based systems cannot provide appropriate social and emotional support due to their technical limitations (e.g., "*I do not believe A.I. is ready to fill this need. It lacks intellectual curiosity and emotional intelligence. To scrape the internet as a learning tool will eventually lead it to untruths, divisive interactions and frustration for the person trying to avoid loneliness.*"). Overall, there were 3 instances of benefit and 19 instances of challenges. This suggests that participants thought there were fewer challenges with ML-based than rule-based VAs.

Discussion

Summary of results

The results of an online public opinion poll of U.S. adults show that participants demonstrated more positive perceived social acceptability, trust, and impact on eldercare toward VAs performing functional than social tasks. Participants also exhibited more positive sentiment toward VAs with ML-based dialogue systems compared to those with rule-based dialogue systems.

Implications for VA adoption

The results of quantitative analyses suggest that society currently may prefer VAs performing functional to social tasks to support older adults. The qualitative findings reveal that participants perceived numerous benefits of VAs aiding in medication adherence, beyond the direct benefits related to medication management. These include facilitating decision-making, promoting the independence of older adults and reducing caregiver burden. However, concerns were also raised about potential issues, including over-reliance on VAs, errors leading to harm, and privacy violations, indicating that while functional support is favored, there are apprehensions regarding its implementation. On the other hand, the benefits of VAs performing social tasks were acknowledged by fewer participants, mainly focusing on alleviating loneliness. Participants had notable concerns about the potential negative impact on users' privacy, mental health, and social isolation. Possible skepticism regarding VA capabilities underscores the need for further development to ensure these systems can meet the complex needs of providing social support.

Our quantitative results suggest that society currently may prefer applying VAs with ML-based dialogue systems rather than rule-based systems to support older adults. Based on thematic analysis, ML-based systems are recognized for their ability to leverage A.I. techniques, which could make VAs more adaptable and intuitive for older adults, although concerns about data privacy, creator bias, and the technology's regulation raise ethical and security issues that could impede their acceptance. On the other hand, the perceived advantage of rule-based systems is their potential to tailor functions to individual needs while minimizing harm, suggesting that with proper design, they could enhance the safety and personalization of care for older adults. However, the challenge of learning to use these systems (i.e., remembering and executing commands), as indicated by many participants, presents a barrier to adoption.

Limitations and future work

The study was a single-stimulus study, using only one instance of text wording and comic-like image per condition, which may pose a threat to construct validity. Future research could employ multiple stimuli to enhance construct validity, which is defined as "the inability to generalize the results of a study with few or even a single stimulus…because we do not know whether the effects observed are due to an unstated stimulus feature" ([41], p. 211). For the current work, we tackled the potential threat to construct validity caused by single stimulus sampling by defining stimulus criteria in the stimuli subsection and reducing variability across stimuli in different conditions. We homogenized the length and frequency of the text descriptors and the graphic elements in the images, following two of the methods suggested by.[41,45] Furthermore, the use of vignettes as implemented in our survey may introduce additional limitations. Comic-like vignettes, while being evocative, might not convey the realism required to accurately infer the effects of a VA, leading to possible biases where judgments differ from what would have been obtained with working versions of VAs.

Another limitation is the study's focus on public opinion, which may not accurately reflect the specific needs and preferences of older adults. Public opinion often represents a broad view and may overlook the diversity within the older adult population, potentially leading to solutions that are not tailored to individual needs. Additionally, public opinion might oversimplify the challenges and technical complexities involved in designing technologies for older adults, creating unrealistic expectations. The public may also not fully appreciate the necessity of specialized features, accessibility, and support, which are crucial for older adults. However, conducting a survey study on public opinion remains useful as a preliminary step towards understanding how attitudes may affect the implementability and ultimate success of technologies targeted at older adults. Our initial investigation provides essential insights and a broad understanding of general perceptions, attitudes, and potential concerns relative to the use of VAs. Our study should be followed by further research that highlights areas that may require deeper exploration and more personalized approaches. By identifying key trends and common opinions, this survey may guide subsequent studies that aim to confirm and extend these findings using more targeted, specific, and diverse methodologies. The ultimate goal of this research is to provide relevant guidance for developing VA systems that are better suited to the unique needs and preferences of older adults.

One final limitation is that investigations of demographic factors such as age range, past experience with VAs, and educational background were not conducted in this work.

Conclusion

Public opinion is relevant to the adoption of innovative technologies designed for older adults. We conducted a

survey study investigating the effect of task type and dialogue system type on public perceptions of VAs for older adults. VAs performing functional tasks were overall perceived more positively than those performing social tasks. Participants exhibited more positive sentiment toward VAs with ML-based dialogue systems compared to those with rule-based dialogue systems. These findings may impact the decision-making of designers and other stakeholders in gerontechnology and care.

Declaration of conflicting interests

Funding

ORCID iDs

Noah Zijie Qu https://orcid.org/0000-0001-9108-0322
Jamy Li https://orcid.org/0000-0003-2440-9719
Jaturong Kongmanee https://orcid.org/0009-0003-6153-7960
Mark Chignell https://orcid.org/0000-0001-8120-6905

References

1. Pradhan A, Lazar A and Findlater L. Use of intelligent voice assistants by older adults with low technology use. *ACM Trans Comput Hum Interact* 2020; 27(4): 1–27.

2. Kim S and Choudhury A. Exploring older adults' perception and use of smart speaker-based voice assistants: a longitudinal study. *Comput Hum Behav* 2021; 124: 106914.

3. Chen C, Lifset ET, Han Y, et al. Screen or no screen? Lessons learnt from a real-world deployment study of using voice assistants with and without touchscreen for older adults. In Proceedings of the 25th international ACM SIGACCESS conference on computers and accessibility, New York, NY, October 2023, pp. 1–21.

4. Chin J, Desai S, Lin S, et al. Like my aunt dorothy: effects of conversational styles on perceptions, acceptance and metaphorical descriptions of voice assistants during later adulthood. ArXiv preprint arXiv:230911681, 2023.

5. Horstmann AC, Schubert T, Lambrich L, et al. Alexa, i do not want to be patronized: a qualitative interview study to explore older adults' attitudes towards intelligent voice assistants. In Proceedings of the 23rd ACM international conference on intelligent virtual agents, Würzburg, Germany, 19–22 September 2023, pp. 1–10.

6. Zhong R, Ma M, Zhou Y, et al. User acceptance of smart home voice assistant: a comparison among younger, middle-aged, and older adults. *Univers Access Inf Soc* 2024; 23(1): 275–292.

7. Chen C, Johnson JG, Charles K, et al. Understanding barriers and design opportunities to improve healthcare and qol for older adults through voice assistants. In Proceedings of the 23rd international ACM SIGACCESS conference on computers and accessibility, Virtual Event, 18–22 October 2021, pp. 1–16.

8. Hoppe JA, Melkas H, Pekkarinen S, et al. Perception of society's trust in care robots by public opinion leaders. *Int J Hum Comput Interact* 2023; 39(13): 2589–2605.

9. Tuisku O, Pekkarinen S, Hennala L, et al. "robots do not replace a nurse with a beating heart" the publicity around a robotic innovation in elderly care. *Inf Technol People* 2019; 32(1): 47–67.

10. Reis HT and Judd CM. *Handbook of research methods in social and personality psychology.* Cambridge, UK: Cambridge University Press, 2000.

11. Ni J, Young T, Pandelea V, et al. Recent advances in deep learning based dialogue systems: a systematic survey. *Artif Intell Rev* 2023; 56(4): 3055–3155.

12. Sung B, Im H and Duong VC. Task type's effect on attitudes towards voice assistants. *Int J Consum Stud* 2023; 47(5): 1772–1790.

13. Corbett CF, Combs EM, Chandarana PS, et al. Medication adherence reminder system for virtual home assistants: mixed methods evaluation study. *JMIR Formative Research* 2021; 5(7): e27327.

14. Shade M, Rector K, Kupzyk K, et al. Voice assistant reminders and the latency of scheduled medication use in older adults with pain: descriptive feasibility study. *JMIR Formative Research* 2021; 5(9): e26361.

15. Mathur N, Dhodapkar K, Zubatiy T, et al. A collaborative approach to support medication management in older adults with mild cognitive impairment using conversational assistants (cas). In: Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility, Athens, Greece, 23–26 October 2022, pp. 1–14.

16. O'Brien K, Liggett A, Ramirez-Zohfeld V, et al. Voice-controlled intelligent personal assistants to support aging in place. *J Am Geriatr Soc* 2020; 68(1): 176–179.

17. Nißen M, Rüegger D, Stieger M, et al. The effects of health care chatbot personas with different social roles on the client-chatbot bond and usage intentions: development of a design codebook and web-based study. *J Med Internet Res* 2022; 24(4): e32630.

18. Mazur M, Rzepka R and Araki K. Chatterbots with occupation-between non task and task oriented conversational agents. In: Proceedings of linguistic and cognitive approaches to dialogue agents (LaCATODA 2012), Montreal, QC, 21 August 2012, 61–66.

19. Miura C, Chen S, Saiki S, et al. Assisting personalized healthcare of elderly people: developing a rule-based virtual caregiver system using mobile chatbot. *Sensors* 2022; 22(10): 3829.

20. Mahmood A, Wang J and Huang CM. Situated understanding of older adults' interactions with voice assistants:

a month-long in-home study. ArXiv preprint arXiv: 240302421, 2024.

21. McTear M. *Conversational AI: dialogue systems, conversational agents, and chatbots*. Berlin, Germany: Springer Nature, 2022.

22. Motger Q, Franch X and Marco J. Software-based dialogue systems: survey, taxonomy, and challenges. *ACM Comput Surv* 2022; 55(5): 1–42.

23. Bertolazzi A, Quaglia V and Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. *BMC Publ Health* 2024; 24(1): 506.

24. Horton D and Richard Wohl R. Mass communication and para-social interaction: observations on intimacy at a distance. *Psychiatry* 1956; 19(3): 215–229.

25. Whang C and Im H. "I like your suggestion!" The role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations. *Psychol Market* 2021; 38(4): 581–595.

26. Levy MR. Watching tv news as para-social interaction. *J Broadcast Electron Media* 1979; 23(1): 69–80.

27. Hartmann T and Goldhoorn C. Horton and wohl revisited: exploring viewers' experience of parasocial interaction. *J Commun* 2011; 61(6): 1104–1121.

28. Cho E, Molina MD and Wang J. The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants. *Cyberpsychol, Behav Soc Netw* 2019; 22(8): 515–520.

29. Rzepka C, Berger B and Hess T. Voice assistant vs. chatbot–examining the fit between conversational agents' interaction modalities and information search tasks. *Inf Syst Front* 2022; 24(3): 839–856.

30. Nirala KK, Singh NK and Purani VS. A survey on providing customer and public administration based services using ai: chatbot. *Multimed Tool Appl* 2022; 81(16): 22215–22246.

31. Eickhoff F and Zhevak L. The consumer attitude towards AI in marketing: An experimental study of consumers attitudes and purchase intention (Dissertation). Retrieved from. *XXXX*. 2023. https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-60756.

32. Bansal P, Kockelman KM and Singh A. Assessing public opinions of and interest in new vehicle technologies: an austin perspective. *Transport Res C Emerg Technol* 2016; 67: 1–14.

33. Hasal M, Nowaková J, Ahmed Saghair K, et al. Chatbots: security, privacy, data protection, and social aspects. *Concurrency Comput Pract Ex* 2021; 33(19): e6426.

34. Snead KC, Magal SR, Christensen LF, et al. Attribution theory: a theoretical framework for understanding information systems success. *Syst Pract Action Res* 2015; 28: 273–288.

35. Kelley HH. The processes of causal attribution. *Am Psychol* 1973; 28(2): 107.

36. Kelley HH and Michela JL. Attribution theory and research. *Annu Rev Psychol* 1980; 31(1): 457–501.

37. Weiner B. A theory of motivation for some classroom experiences. *J Educ Psychol* 1979; 71(1): 3.

38. Iyengar S. Experimental designs for political communication research. In: *Sourcebook for political communication research: methods, measures, and analytical techniques*. London, UK: Routledge, 2010, pp. 129–148.

39. Peterson E and Iyengar S. Partisan gaps in political information and information-seeking behavior: motivated reasoning or cheerleading? *Am J Polit Sci* 2021; 65(1): 133–147.

40. Memmott T, Carley S, Graff M, et al. Sociodemographic disparities in energy insecurity among low-income households before and during the covid-19 pandemic. *Nat Energy* 2021; 6(2): 186–193.

41. Cummings JJ and Reeves B. Stimulus sampling and research integrity. In: *Research integrity: best practices for the social and behavioral sciences*. Oxford, UK: Oxford University Press, 2022, pp. 203–223.

42. Buteau E and Lee J. Hey alexa, why do we use voice assistants? the driving factors of voice assistant technology use. *Commun Res Rep* 2021; 38(5): 336–345.

43. Seymour W and Such J. Ignorance is bliss? The effect of explanations on perceptions of voice assistants. *Proceedings of the ACM on Human-Computer Interaction* 2023; 7: 1–24.

44. Yuan F, Anderson JG, Wyatt TH, et al. Assessing the acceptability of a humanoid robot for alzheimer's disease and related dementia care using an online survey. *International Journal of Social Robotics* 2022; 14(5): 1223–1237.

45. Gilbert TK, Qu NZ, Ju W, et al. Fleets on the streets: how number, affiliation and purpose of shared-lane automated vehicle convoys influence public perception and blame. *Transport Res F Traffic Psychol Behav* 2023; 93: 294–308.