Healthcare Voice Assistants: Factors Influencing Trust and Intention to Use

XIAO ZHAN, King's College London, United Kingdom
NOURA ABDI, Liverpool Hope University, United Kingdom
WILLIAM SEYMOUR, King's College London, United Kingdom
JOSE SUCH, King's College London, United Kingdom & Universitat Politecnica de Valencia, Spain

Voice assistants such as Alexa, Google Assistant, and Siri, are making their way into the healthcare sector, offering a convenient way for users to access different healthcare services. Trust is a vital factor in the uptake of healthcare services, but the factors affecting trust in voice assistants used for healthcare are under-explored and this specialist domain introduces additional requirements. This study explores the effects of different functional, personal, and risk factors on trust in and adoption of healthcare voice assistants (HVAs), generating a partial least squares structural model from a survey of 300 voice assistant users. Our results indicate that trust in HVAs can be significantly explained by functional factors (usefulness, content credibility, quality of service relative to a healthcare professional), together with security, and privacy risks and personal stance in technology. We also discuss differences in terms of trust between HVAs and general-purpose voice assistants as well as implications that are unique to HVAs.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; • **Security and privacy** → **Trust frameworks**; • **Applied computing** → *Health care information systems.*

Additional Key Words and Phrases: healthcare voice assistants, trust in voice assistants, amazon alexa, google assistant, apple siri, healthcare technologies

ACM Reference Format:
Xiao Zhan, Noura Abdi, William Seymour, and Jose Such. 2024. Healthcare Voice Assistants: Factors Influencing Trust and Intention to Use. 1, 1 (January 2024), 35 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Voice assistants such as Alexa, Google Assistant and Siri are now found in millions of smart speakers, phones, and other devices. Popular because of the hands-free, convenient experience they offer, voice assistants are mainly used to play music, search for information, and control smart IoT devices in the home [11, 47, 156]. However, the capabilities of voice assistants are continuously expanding to cover other uses and enahnce user experience. Capabilities added by third parties (often called 'skills' or 'actions') continue to grow, with the number of Alexa skills surpassing 100,000 [172]. Beyond the main uses mentioned above, there are many other popular skill categories including ride hailing (e.g. Uber), checking and sending emails (e.g. Myemail), and online banking.

One particular domain that is gaining attention is the application of voice assistants in healthcare, which we refer to as Healthcare Voice Assistants (HVAs) [151, 154]. In particular, HVAs offer more and more healthcare services such as

Authors' addresses: Xiao Zhan, King's College London, United Kingdom, xiao.zhan@kcl.ac.uk; Noura Abdi, Liverpool Hope University, United Kingdom, abdin@hope.ac.uk; William Seymour, King's College London, United Kingdom, william.1.seymour@kcl.ac.uk; Jose Such, King's College London, United Kingdom and & Universitat Politecnica de Valencia, Spain, .

Manuscript submitted to ACM

virtual assessments for patients, diagnostic suggestions, and health and lifestyle tips [63]. One of the key advantages of HVAs is that they are a cost-effective alternative to in-person or telephone consultations, especially at a time when the number of healthcare professionals is not enough to meet the needs of patients [119]. They also offer convenience and efficiency when patients need to travel for long distances to see a doctor [119], or when situations like the Covid-19 pandemic make it difficult for patients to be attended to physically [23]. In fact, many healthcare providers around the world are already using HVAs in some way, such as the UK's National Health Service which offers information via Alexa.[1] The number of US citizens that used voice assistants for healthcare related purposes has nearly tripled since 2019.[2] Beyond what is currently available through contemporary voice assistants, researchers have been exploring ways to deliver a wide range of healthcare services remotely through HVAs [101, 188]. However, HVAs also pose challenges. Privacy and data security are critical due to the sensitivity of health information involved [152] and the absence of robust AI healthcare regulations potentially magnifies these issues, inviting misuse or data misinterpretation [141]. There is also a risk of inaccuracies in HVA diagnoses and advice, potentially leading to poor health outcomes [1]. It is therefore crucial to address these issues as the field progresses.

As HVAs continue to develop and improve, a question arises as to whether people will trust and adopt them. The multifaceted nature of trust in both technology and general-purpose voice assistants has been extensively studied (see Section 2 for a comprehensive review). This body of work has substantiated the correlation between trust and user acceptance [7, 9, 108, 110, 115, 135, 150, 163], and includes comprehensive studies across diverse domains on the effect of functional [32, 33, 41, 122, 122, 148, 187, 190], personal [9, 37, 37, 122], and risk [53, 71, 121, 146] factors in trust, thereby reinforcing the paramount importance of trust within the technological sphere. However, there is a need to study trust specifically in the context of HVAs because the healthcare domain is known to bring a range of additional challenges [14, 24]. For instance, despite the shared influence of certain factors, such as the *reliability of services and safety*, it is evident that human trust in healthcare and trust in other technologies exhibit notable disparities in their respective levels of confidence [128]. As described above, healthcare data is also more sensitive than other types of data, as shown by previous academic works [3, 127], and is often subject to additional protection under privacy regulations such as the GDPR in Europe[3] and HIPPA in the United States.[4] Furthermore, it has been shown that trust in traditional healthcare providers such as physicians [67, 78, 186] cannot be directly applied to services delivered by HVAs [185]. This is because interacting with a voice assistant is fundamentally distinct from interacting with a human, even if voice assistants exhibit several human-like characteristics (like voice interaction) [4]. In addition, AI algorithms embedded in HVAs lead to decision-making processes that may not be apparent to the users or that are not adequately explained compared to healthcare professionals [66, 134, 183]. Finally, issues related to aspects that affect trust may be more pre-eminent. For instance, incomplete or incorrect health-related information, such as first-aid instructions or medication recommendations may cause patients harm [159].

As a field, CSCW has shown a keen interest in digital healthcare in particular studying its socio-technical dimensions. This encompasses the collaborative efforts inherent in healthcare AI and the considerations steered by policy perspectives [59, 145]. The study of patient trust in Healthcare AI in particular aligns with these core themes. Besides, Healthcare AI is a multidisciplinary field, from medical professionals, AI developers, data scientists and other key

[1] https://www.gov.uk/government/news/nhs-health-information-available-through-amazon-s-alexa

[2] See the report https://voicebot.ai/2022/01/07/voice-assistant-use-in-healthcare-nearly-tripled-over-two-years-and-demand-still-outstrips-supply-new-report/ for details.

[3] The General Data Protection Regulation (GDPR) in the EU designates healthcare data as 'special category', requiring additional protections (Article 9 GDPR) https://gdpr.eu/tag/gdpr/

[4] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) sets the standard for sensitive patient data protection https://www.hhs.gov/hipaa/index.html.

stakeholders [99, 145]. Their collective efforts are needed for the development, implementation, and refinement of healthcare solutions, in particular those focusing on the role of trust. There are also CSCW studies examining communication dynamics between patients and providers [144]. These studies emphasize the importance of integrating diverse stakeholder perspectives to foster a holistic understanding of the relationship between patients and providers. Equally crucial is the unique situations and expectations of specific stakeholder groups, like the shared experiences and challenges faced by patients, forming a cohesive community actively engaging with AI technologies.

Given that the healthcare sector serves as a vital domain for human-AI collaboration [99] and considering trust as the cornerstone fostering this collaboration [13, 50, 82, 85, 184], it is vital to extensively research and investigate in order to further enhance and supplement this specific research field. The socio-technical implications of HVAs, which bridge the gap between technology and social healthcare practices [59], further emphasize the importance of understanding trust dynamics in this context. Finally, the two central tenets delineated in our paper, 'trust' and 'adoption', inherently embody a collaborative essence [100, 131]. This dual-faceted approach not only seeks to enhance user sentiment and satisfaction but also addresses the developmental requirements discernible to HVA developers. To the best of our knowledge, this work is the first to examine factors that affect trust and intention to use in HVAs. We study the effect of different functional (Anthropomorphism, Effort Expectancy, Perceived Usefulness, Perceived Content Credibility, and Perceived Relative Service Quality), personal (Stance in Technology, Familiarity, Technology Attachment, and Social Influence), and risk (Security Risk, Privacy Risk, and Perceived Substitution Risk) factors on trust in HVAs and, in turn, the effect of this trust on peoples' intention to use HVAs. The selection of our factors is based on those with a confirmed substantial impact on trust in broader contexts such as general voice assistants, technology at large, and the healthcare industry, in order to measure their effect in the context of HVAs (as we detail in Section 3). More specifically, this study aims to answer the following research questions:

RQ1 How do functional, personal and risk factors influence users' trust in healthcare voice assistants?
RQ2 How does trust in healthcare voice assistants influence users' intention to use them?

To answer these questions, we developed a model of factors affecting trust in HVA and intention to use. In order to test the model we administered a questionnaire to 300 participants which included questions measuring these latent functional, personal, and risk variables alongside trust and intention to use. The results were analysed using partial least squares structural modelling (PLS-SEM) and show how functional factors (anthropomorphism, perceived usefulness, effort expectancy, content credibility, and relative service quality), risk factors (security risk and privacy risk), and personal factors (stance in technology) substantially explain trust in HVAs. They also suggest that trust in HVAs explains intention to use HVAs. Based on these results we derive some practical insights and recommendations to help design and develop the next-generation of HVAs, promoting user trust and encouraging adoption.

2 RELATED WORK

2.1 Trust in Technology

2.1.1 Trust Scales and Measurement. The topic of trust has been extensively investigated, but there is not yet a completely agreed definition of what it is and what it means. Trust has often been described as "a willingness to ascribe good intentions to and have confidence in the words and actions of other people" in social science [62], while trust in technology usually refers to "beliefs that a specific technology has the attributes necessary to perform as expected in a given situation in which negative consequences are possible." [121]. Trust is recognized as a central focus for the CSCW discipline over the past several decades [100]. Trust is recognized as a pivotal factor influencing human-AI collaboration [82, 85], and in particular, previous work published at CSCW [85] has underscored the positive effects of research on trust in fostering human-AI collaboration. In order to study the different ways in which other factors can affect trust, a critical step is to be able to measure trust in a particular technology, in our case, HVA. Previous literature proposed different metrics and scales to measure trust in technology [22, 93, 121, 130]. More recently, and in the context of general-purpose voice assistants such as Amazon Alexa, Cho et al. [39] used a scale to measure trust, adapting it from [93], where they consider users' perceptions about the assistant's competence, integrity, and benevolence, which were suggested by previous studies to be key components of trust [22, 93, 121, 130]. As we explain later on, we adapt the scale proposed in [39] to measure trust in our study.

2.1.2 Adoption. As studied in early CSCW literature [131], the dynamics of collaborative work are often embedded in the processes of technology adoption and adaptation. This suggests that even if the focus seems individualistic, the broader collaborative context is often at play. In terms of the models to study adoption, one of the most well-known and used models is the Technology Acceptance Model (TAM). The model was first proposed by Davis in 1989, and it suggests that an individual's acceptance of technology is determined by two key factors: perceived usefulness and perceived ease of use. Subsequent models built on TAM (e.g. TAM2 [179], TAM3 [178], and UTAUT [180]) have extended the original TAM by incorporating additional variables such as social influence [178–180], system characteristics [178], and facilitating conditions [178, 180] to explain the acceptance and use of technology. However, recent works [114, 123] have criticised their effectiveness as being highly context-dependent, and that they are too simplified for emerging technologies. It is therefore now common to combine them with other theories and attributes that influence users' usage behaviour, as we elaborate in the next section, to compensate for their limitations in focusing more on some of the functional factors.

2.1.3 Trust and Adoption. Investigating users' trust in new technologies and products is of great significance, because user acceptance and adoption are substantially and significantly influenced by trust [7, 9, 108, 110, 115, 135, 150, 163]. In particular, with the recent development of Generative AI like ChatGPT, it has also contributed to the significance of trust when adopting new technologies [6, 42]. In all of these studies, trust has shown a positive effect on users' behavioral intentions to use a technology in domains such as e-commerce [163], mobile payments [115], IoT devices [9], and ChatGPT [162]. This has also been shown recently for general-purpose voice assistants [7, 108, 110, 135, 150]. For instance, Liao et al. [110] found that users' attitudes towards adoption were rooted in whether they trusted voice assistants, and Pitardi et al. [150] showed that trust is one of the key contributors to foster adoption of Google Assistant. More recently, the importance of trust for adoption of voice assistants was also shown when assistants are used for specific tasks, such as in an education setting [7].

2.2 Factors Affecting Trust

HCI studies examining how different factors affect users' trust in technology have identified three main types of factors that affect trust, including functional, personal and risk factors. Functional factors refer to the characteristics of the technologies themselves [32, 33, 41, 122, 122, 148, 187, 190], personal factors refer to the individual user, their social circle and their experiences [9, 37, 37, 122], and risk factors refer to the harms that could be derived from the technology [53, 71, 121, 146]. In a related vein, Knowles et al. [90] emphasized trust as crucial for collaboration and introduced design patterns to enhance trust. While not empirically validated like the aforementioned models, these patterns highlight the importance of factors such as usability (functional factor) and socialization (personal factor).

Regarding the functional factors that affect trust in technology, previous literature has uncovered factors such as the usefulness of the technology and how easy the technology is to use [31], as well as its reliability [109, 117] and the interfaces designed for interaction [143]. These all manifest slightly differently in voice assistants, particularly interface design and reliability. With regards to interface design, Anthropomorphism [35, 158, 176] has been shown to influence trust, with studies suggesting that voice alone facilitates human connection with the assistant [79, 138]. Regarding reliability, and partly because of how voice assistants operate, the broader concept of content credibility [113], which includes reliability as well as other aspects such as accuracy, completeness and authenticity of the information, is a critical factor to foster trust in voice assistants, that is, the credibility of what the voice assistant utters back to the user seems to affect how much the user trusts the assistant.

In addition to functional elements, whether users consider technology to be trustworthy depends on personal factors. This includes individual factors and social factors. Individual factors, such as user predisposition to try new technologies [5], user familiarity with an understanding of a technology [37, 95], and even the users' emotional attachment [197] and psychological connection to a technology [168]. Social factors also seem to play a role, with works like [69] showing that social influence, the degree that a customer's social group (e.g. family, friends, etc.) believes that using a technology is relevant, affects trust in that technology.

Finally, the effect of risk on trust has long been studied [121], and shown to manifest across different domains, including online transactions [146], autonomous vehicles [71], and the use of Electronic Health Care Records (EHCR) by physicians [53]. Security and privacy risks, in particular, have been shown to reduce the benefits of using general-purpose voice assistants [9, 96, 123], directly impacting the trust users' have in voice assistants, with users stopping using voice assistants all together as a result, or restricting their use of voice assistants only to functionality they perceive to be less risky [2, 19].

2.3 Healthcare Voice Assistants

2.3.1 HVA available in Commercial Voice Assistants. Commercial voice assistants nowadays incorporate healthcare capabilities as Amazon Alexa *Skills* or Google Home *Actions* [44]. In 2020, researchers grouped the spectrum of healthcare services provided by voice assistants into four levels: information-, assistance-, assessment-, and support-level services [160, 161]. Of these, information-level services (providing healthcare-related information to users) are one of the most common and basic functions and are already widely embedded in most commercially available voice assistants. There are many Alexa skills, for example, that provide healthcare information such as descriptions/or instructions on symptoms, medications, and how to use medical facilities (e.g., Dr. A.I., WebMD, and Mayo Clinic). Examples of assistance-level services include the ability for VAs to assist users in setting up reminders for medication or physical activity (i.e. tracking their symptoms) and providing appointment notifications [160, 161]. Some of these features are also available in contemporary VAs like the My Children's Enhanced Recovery After Surgery skill developed by Boston Children's Hospital, which reminds parents and caregivers about information regarding their post-op appointments. Support for functionality within the assessment and support layers ranges from scarce to non-existent. Assessment level skills, for example, would enable voice assistants to recognise changes in mood or health conditions, but commercial voice assistants are only currently able to access and view diagnostic results and records (e.g. Alexa Skill Sugarmate). Similarly, the support level includes the ability to prescribe and prioritise care/treatment but users have thus far only been able to refill their prescriptions[5].

[5]This feature was available for Alexa owners in 2019 - https://www.dailymail.co.uk/sciencetech/article-7727961/Alexa-owners-use-virtual-assistant-refill-prescriptions.html

2.3.2 Research on HVAs. Extensive research has been conducted recently on the use of voice assistants for healthcare. For example, Dojchinovski et al. [49] saw the potential for VAs to become an important part of the healthcare system for heart disease. The VAs could communicate directly with the user through their special voice interface and then, retrieve the required information from the medical cloud and transmit it back to the user. In this case, users would highly benefit from using a VA-based healthcare system to perform tasks such as checking ECG readings, scheduling appointments, recording treatments and contacting doctors. Other strategies and prototypes for coping with various healthcare issues have been discussed as well, such as supporting home care [20], and improving patients' Type2 diabetes management [29, 36]. Another stream of research on HVA has focused on evaluating the performance of voice assistants in healthcare-related tasks. Google Assistant, Amazon Alexa, Apple Siri, Microsoft Cortana and other voice commercial assistants are often tested and compared [8, 91, 137, 193]. Overall, the ability of VAs to provide health-related advice is very limited and performed unfavourably to identify the user's queries for healthcare information [8, 91, 137, 193]. Finally, there has been some recent work examining how differences in modality (i.e., voice vs. text) and device type (i.e., smartphone vs. smart home device) affect user perceptions when retrieving sensitive health information from voice assistants [38], finding that voice led to more social presence and that device did not matter. In addition, attention was paid to the behaviour and perception of older people using voice assistants for the specific purpose of seeking healthcare information [28, 199]. They found that older people's evaluation of the VA was influenced by communication styles, anthropomorphic settings, and individual difference [199], and that they appeared to use voice assistants primarily to confirm information obtained from other sources (e.g., laptop, phone) [28]. While previous work [145] emphasized the importance of addressing trust issues, especially amidst evolving healthcare workflows, we are unaware of any prior research specifically targeting user trust in HVAs and the influencing factors.

3 HYPOTHESES

Trust is a complex phenomenon that can be affected by many different factors. As discussed in Section 2.2, there are a number of personal, functional, and risk factors that have been shown to have an impact on users' trust in technology. Therefore, in this paper, we consider factors belonging to these categories that were shown by previous works to have a significant effect on trust adapted to the case of HVAs. In particular, and as we detail below, we consider: i) factors that have been shown to affect trust on general-purpose voice assistants such as Amazon Alexa and Google Assistant [9, 10, 37, 60, 150, 182] ; ii) factors that have been shown to affect trust in technology in general [33, 41, 65, 187, 190]; and iii) factors that are specific to trust in the healthcare domain [32, 148, 196]. We do not include factors that, while significant in prior works are not relevant or do not translate to the case of HVAs. For instance, *Hedonic Benefits* is one of the factors that seem to affect trust in general-purpose voice assistants [123, 150], but that only applies when the purpose of use is hedonistic [189], and it is not aligned with the relatively more serious, purposeful, and specific context of seeking health care. Also, we are interested in factors that affect trust rather than the measurement of trust as a construct [9], so we adapt already existing scales to measure trust [39] and focus on the factors that influence it. Figure 1 summarizes our hypothesized research model, the different factors that may affect trust in HVA and its relationship to intention to use HVA. We now detail the categories, factors, and our hypotheses.

3.1 Functional Factors

The extensive literature on trust in technology identified and subsequently confirmed across different domains that users' trust in a technology is influenced by whether the technology delivers on the functionality promised to complete the tasks or services they are supposed to offer [122, 192]. This normally refers to both design and performance

Table 1. Summary of factors considered in the model. (H1-5 are functional factors, H6-9 personal factors, and H10-12 risk factors.)

H#	Factors	Description	Supporting Source
H1	Anthropomorphism	Anthropomorphism in voice assistants involves human-like characteristics, creating a social connection with users. Voice-based interaction deepens this bond between users and the assistant.	[55, 79, 129, 138, 150, 171, 175]
H2	Effort Expectancy	Effort expectancy is the predicted mental and physical activity required to use a technology. It affects trust in technology, and challenges in ease of use can lead to anxiety and distrust.	[9, 33, 37, 48, 89, 97, 104, 111, 150, 155, 177, 181]
H3	Per. Usefulness	is about device efficiency in daily tasks. In healthcare, it means technology facilitating services. Usefulness affects attitudes and adoption. In HVAs, useful responses to symptoms are vital for user satisfaction.	[9, 30, 46, 48, 64, 105, 148]
H4	Per. Content Credibility	Content credibility is crucial for user trust. In healthcare, users rely on accurate information from virtual assistants. Inaccurate information can be harmful.	[12, 14, 24, 39, 41, 72, 109, 117, 122, 133, 159, 169, 190].
H5	Per. Relative Service Quality	This perception compares healthcare technology to professionals. Patients value professional care but may trust technology for specific diagnoses if it offers greater accuracy.	[18, 72, 86, 112, 153, 198]
H6	Familiarity	Familiarity represents the level of understanding of an object. When users are familiar with a technology, they are more likely to trust it compared to unfamiliar ones in various domains such as e-commerce, IoT devices, and voice assistants.	[10, 37, 65, 95, 116]
H7	Tech Attachment	This factor, aka emotional attachment, reflects the psychological connection between individuals and a technology.	[37, 149, 168, 197]
H8	Social Influence	Social influence plays a significant role in shaping users' trust and perceptions of a product, as they are influenced by the behaviors and opinions of others in their social networks.	[9, 69, 102, 170]
H9	Stance in Tech	Stance in technology reflects an individual's openness to trying new innovations, with a positive correlation found between willingness to experiment with new technologies and acceptance and trust in their services.	[5, 37, 122]
H10	Security Risk	Security risk in technology, especially in healthcare, can undermine trust due to potential consequences of flaws and incidents such as errors, malware, and hacking.	[10, 54, 72, 87, 106, 133]
H11	Privacy Risk	Privacy risk arises when users perceive a potential threat of data misuse or unauthorized access, particularly in the sensitive healthcare domain, impacting trust in voice assistants.	[3, 10, 51, 60, 70, 84, 127, 150, 174, 182]
H12	Per. Substitution Risk	Perceived substitution risk in healthcare technologies involves the concern of technology replacing traditional healthcare professionals, impacting trust in AI applications.	[56, 72, 86, 133, 136, 140]
H13	Intention to Use	Intention to use reflects the connection between trust and technology adoption, with a strong correlation between users' trust, attitudes, and actual use of technology.	[7, 9, 34, 108, 110, 115, 135, 150, 163]

issues [192]. When talking about voice assistants, functional factors take particular manifestations, and have been thoroughly examined both qualitatively and quantitatively in the previous study [40, 47, 158]. For instance, and as further detailed below, because of the very nature of their interaction modality, they offer a more anthropomorphic experience than other technologies. Another example, and in particular when talking about HVA, is that users may compare the healthcare service offered by an HVA relative to what users normally get from healthcare professionals.

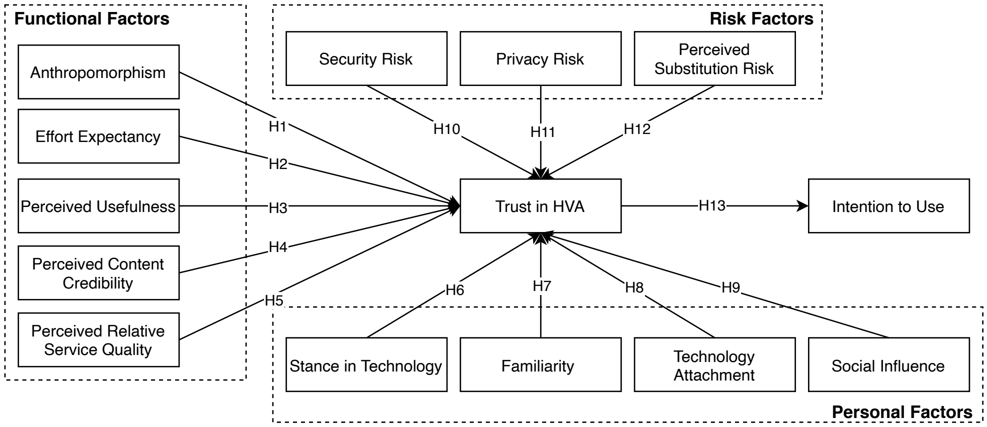

Fig. 1. Hypothesized Research Model

3.1.1 Anthropomorphism (A). This factor refers to the extent that voice assistants show human-like characteristics and are therefore perceived and treated by users in ways that may be closer to how they would interact with another human [129]. The more human-like characteristics voice assistants display the more it seems humans connect with them [55]. Starting with the modality of interaction, voice alone facilitates human connection with the assistant [79, 138], and it may make users feel they are interacting with a social entity [175]. In fact, previous research on general-purpose voice assistants showed that their anthropomorphic characteristics have a positive effect on users' trust in voice assistants [150, 175]. We hypothesize this to also be the case in HVA, particularly as healthcare is a domain where human-like characteristics in the technology used may contribute to a satisfactory experience [171].

H1 Anthropomorphism positively influences Trust in HVAs.

3.1.2 Effort Expectancy (EE). This factor is the predicted mental and physical activity and skill required to use a technology [181]. In general, it has been shown across different domains that effort expectancy, or its flip side 'ease of use', affect trust in technology [9, 33, 48, 97, 111, 150]. For instance, one case is that of older adults, who find it difficult to use digital technologies [104, 177], which has been shown to trigger anxiety and distrust [89, 104]. There is no previous work on the relationship of effort expectancy and voice assistants (not even in general-purpose ones). However, given the negative role that effort expectancy has been shown to play in similar technologies before, such as social robots for service delivery [37], we hypothesize that effort expectancy also plays a role in trust in HVA, particularly it can negatively affect it. This would also be consistent with work on mobile healthcare technologies that showed that the reverse of effort expectancy has a positive effect on intention to use [155].

H2 Effort expectancy negatively influences Trust in HVAs.

3.1.3 Perceived Usefulness (PU). In general, this factor refers to "the degree to which the device makes life more efficient and helps carry out daily tasks." [30]. In other words, and particularly considering the healthcare domain, this factor refers to whether and how technology helps users to seek and get healthcare services in an efficient and effective way. According to Davis[48], usefulness can influence attitudes towards a particular technology, and in particular user acceptance [46, 64] and adoption [9] of technologies. The contribution of usefulness to trust has been demonstrated, especially in the healthcare domain. For instance, Peng et al. [148] confirmed that the usefulness of online information and suggestions provided by physicians has a positive effect on patient-physician trust. Similarly, when using HVAs, the usefulness of responses from the HVA to user queries describing symptoms is suggested to be of great significance [105], and it might cause differences in the user's feelings and judgements. Based on all of this, we formulate the following hypothesis:

H3 Perceived Usefulness positively influences Trust in HVAs.

3.1.4 Perceived Content Credibility (PCC). This factor represents the credibility of the information provided by the technology, and it is often identified as a key factor influencing users' trust in technology in general [41, 122, 190]. The credibility of information is known to, in turn, be composed of different aspects, such as its accuracy, reliability (including completeness, consistency, and authenticity), and authoritativeness (the information was generated by experts) [12, 14, 39, 109, 117, 169]. These are all critical aspects of information in the healthcare context because users may blindly follow the instructions and advice given by HVAs, such as what medication to take and how often to take it [24]. Inaccurate information is therefore likely to result in the user being harmed [24, 72, 133, 159].

H4 Perceived Content Credibility positively influences Trust in HVAs.

3.1.5 Perceived Relative Service Quality (PRSQ). This factor refers to the quality of a healthcare service a technology offers *compared* to the same service offered by a healthcare professional. It has been used by previous research as an important indicator of the satisfaction healthcare technologies can bring compared to traditional healthcare services [18, 112, 198]. When it comes to trust, research showed that patients are reluctant to trust AI technologies because they value the quality of care provided by healthcare professionals [153]. In addition, qualitative studies on trust in AI technologies for particular types of diagnoses (like cancer) [72, 86] suggested that some patients might trust the technology if it provides more accuracy than healthcare professionals. We therefore hypothesize that PRSQ affects trust and that the higher the PRSQ (that is the higher the perception of quality of the service when compared with a healthcare professional) the higher the trust.

H5 Perceived Relative Service Quality positively influences Trust in HVAs.

3.2 Personal Factors

Users may have differing attitudes toward the same technology [107, 126, 194]. Trust is no exception, and it can be affected by individual and social factors. For instance, an individual may be more receptive to new technology than others (e.g. early adopters), and another individual may also (or instead) be influenced by the opinions of their (online and/or offline) social circle. Therefore, in addition to functional factors, we also consider personal factors.

3.2.1 Familiarity (F). This factor refers to the degree of understanding of an object [95]. Research has shown that the more familiar users are with a technology the more likely they are to trust it in comparison to other technologies they do not know about [116]. In fact, it has been shown to increase trust by reducing users' uncertainty on the technology in e-commerce [65], IoT devices [10], and voice assistants [37]. Similarly, in this study, we hypothesize that familiarity with HVA influences the trusts users have on them [37].

H6 Familiarity positively influences Trust in HVAs.

3.2.2 Technology Attachment (TA). This factor, also known as emotional attachment [197], represents the psychological connection between a person and a technology [168]. This connection is known to promote a positive view of the technology in question [149]. In particular, in general-purpose voice assistant, TA is known to significantly affect trust [37].

H7 Technology Attachment positively influences Trust in HVAs.

3.2.3 Social Influence (SI). Users are influenced by the behaviors and opinions of others in their social networks, according to the theory of social influence [102]. They tend to believe the same as the majority in those networks [69]. This is consistent with Social Identity Theory [170], which suggests that following social norms makes people feel integrated. As a result, users' perceptions, and in particular their trust towards a product, can vary [9], based on the trust their social connections have. In this study, we hypothesize that this factor reflects the propensity of users to trust HVAs, to capture the influence their social network may have on it.

H8 Social Influence (positively or negatively) influences Trust in HVAs.

3.2.4 Stance in Technology (ST). This factor is the extent to which an individual would be open to try out a new innovation or technology [5]. Many existing studies have found that people who are more willing to experiment with and try new technologies are more inclined to accept and trust the services that these technologies offer [37, 122]. It is therefore reasonable to hypothesize that such a group of individuals would likely be open towards HVAs.

H9 Stance in technology positively influences Trust in HVAs.

3.3 Risk Factors

The use of technologies usually comes with some risks [103, 118, 124, 125, 157, 194]. These risks may also affect the trust that users have in the technologies. For instance, if the perceptions of risk is high, research has shown that that can lead to distrust [9, 94]. We consider the following risk factors and the effect they may have on trust in HVA.

3.3.1 Security Risk (SR). This factor relates to how unsafe and unprotected users feel when using a technology. Previous research on other healthcare technologies showed that SR is negatively associated with trust [10]. This is, in part, because healthcare is a critical domain, where flaws in the technology can lead to serious issues [54, 72, 87, 106, 133]. In addition, the digital healthcare sector has recently reported high number of incidents due to error, malware, and hacking[6]. We therefore hypothesize that security risk can have a negative influence in trust in HVA.

H10 Security Risk negatively influences Trust in HVAs.

3.3.2 Privacy Risk (PR). When users believe there is a potential risk for their data being misused or accessed by unauthorised parties, their trust may be affected [10, 60, 150, 182]. In the healthcare domain, data can be successfully linked to individuals [84, 174] even if it has been anonymised and scrubbed of all identifiers [70]. As suggested in [51], the more sensitive the information the more the likelihood of users having concerns when interacting with their voice assistants. Healthcare data is very sensitive [3, 127], so hypothesize that PR negatively affects trust in HVA.

H11 Privacy Risk negatively influences Trust in HVAs.

3.3.3 Perceived Substitution Risk (PSR). This factor is typically considered in healthcare technologies, and it captures the risk that traditional healthcare professionals may be replaced by technology in the future [56], and that this could even lead to professional healthcare job loss [72, 133, 140]. There is also a risk that clinicians may lose skills that are increasingly carried out by technology (e.g. around diagnosis), resulting in increased errors when complex cases are escalated for human review [86]. All of this risk of substitution has been shown to affect users trust in AI technologies applied to healthcare in general [136], so we hypothesize this to be also the case in HVA.

H12 Perceived substitution risk negatively influences Trust in HVAs.

[6]See for instance the, 2022 Data Breach Investigations Report, which shows that from the cases investigated, 67% of them reported to have healthcare related issues - https://www.verizon.com/business/resources/reports/dbir/

3.4 Intention to Use

As introduced in Section 2.1.3, the relationship between trust and adaption has been extensively studied. Across technologies, there seems to be a clear connection between users trusting and adopting technology [7, 9, 108, 110, 115, 135, 150, 163]. This has also been shown for the case of general-purpose voice assistants [7, 108, 110, 135, 150]. One of the proxies to attitudes towards adoption is *intention to use* (or intention to adopt) [7, 34, 110]. In fact, previous research did show a strong connection between behavioural intentions and actual use of specific technologies [34]. In accordance to this, we use in this paper intention to use as a proxy for adoption and hypothesize that:

H13 Trust in HVAs positively influences users' intention to use them.

4 METHODOLOGY

In order to test our hypotheses, we followed a quantitative methodology, by designing and administering a questionnaire with items measuring the factors introduced before as well as trust in HVAs and users' intention to use HVAs, and analyzing the results using Partial Least Squares Structural Equation Modelling (PLS-SEM). Our study was approved by our institution's IRB.

4.1 Survey Development

We created a questionnaire that contains three parts, which can be found in Appendix A, B, and C. The first part is a description of HVAs, so participants could understand what we mean by them. The second part includes the questions regarding the model and hypothesis introduced in the previous section in the form of items representing each of the factors, which were measured using a five-point Likert scale, with answer choices ranging from 'strongly disagree' (1) to 'strongly agree' (5). To ensure reliability and validity, the items were adapted from existing scales for each of the factors considered [37, 39, 56, 68, 81, 123]. To do this, items were rephrased where appropriate to fit the study context (that is HVA). For example, one of the items for the factor *Effort Expectancy* from [37] was rephrased from "*Using AI devices takes too much of my time*" to "*Using healthcare voice assistants would take too much of my time*". Every set of items for each factor was taken from the same source, so that all items measured the same underlying construct. The third and final part of the questionnaire includes three post-survey questions about users' perceptions of future use of HVAs, organisations they trust to design HVAs, and expectations for building trusted HVAs.

To make sure that the questionnaire was effective and that participants understood the questions, we conducted two pilot studies with a total of 30 participants each. The data from the pilots was used only to improve the final survey and excluded from the final analysis. The pilots helped us refine the items, e.g. we were able to identify that some of the reversely-coded items (which we reversed for data quality purposes, as explained later) were confusing, and we adequately rephrased them.

4.2 Model Evaluation

PLS-SEM was used for the statistical analysis of the second part of the questionnaire (items). This is because PLS-SEM has been found to be appropriate when the analysis is to test a theoretical framework from a prediction perspective and the structural model is complex and includes many factors [77]. In addition, PLS-SEM does not require large samples to perform well as some other SEM approaches do [76, 77]. We used the implementation of PLS-SEM in the SmartPLS 3.3 software [43], following the recommended two-step procedure [76] of evaluating the measurement (outer) model first before estimating the structural (inner) model.

4.3 Post-Survey Questions

While statistical analysis of the model shows the factors that demonstrate statistical significance in influencing trust and intention to use, it is also important to understand participants' reasoning around the inclusion of certain specific factors and the exclusion of others. Consequently, we undertook a qualitative analysis to enhance our comprehension of the factors encompassed within the model as suggested by prior work [150, 195]. This qualitative exploration not only enabled us to delve deeper into the considered factors but also presented opportunities to identify additional factors that were not initially incorporated in our model, thus augmenting the existing body of literature in this domain.

For the third and final part of the questionnaire, responses to the open-ended questions were first converted into a text format, and then analyzed using thematic analysis following Braun and Clarke [26], involving the 6 steps of familiarization, coding, generating themes, reviewing themes, defining and naming themes and writing up [26]. Individual coding was conducted by two researchers, and codes were recognized based on the ideas and feelings expressed by participants. Researchers then reviewed the codes created and discussed the results until reaching agreement. Similarly, the researchers then developed themes individually before coming together to discuss disagreements and converge on a final set of themes.

4.4 Data Collection

The final questionnaire was implemented using Qualtrics[7] and participants were recruited using Prolific[8]. All participants were voice assistant users aged 18 or older. The beginning of the questionnaire explained the study in detail, requested consent for data collection, and gathered information about participants' prior use of their voice assistant, including for healthcare purposes.

To ensure a large enough sample, we followed two methods and retained the highest minimum number obtained. First, a minimum viable sample size was calculated using the "10-times rule" [75, 173], which suggests the minimum sample should be 10 times the maximum number of arrowheads pointing at a latent variable (a formal term employed in PLS, synonymous with the 'factors' used in our study) anywhere in the PLS path model (12 arrows points to *Trust in HVA*). This gave a minimum size of $10 * 12 = 120$ for each stage of data collection.

Second, to ensure that the sample also had sufficient statistical power to accurately detect true effects and minimize Type II errors we used the method proposed in Hair et al. [73]. This builds on Cohen's power tables for least squares regression and relies on a table listing minimum required sample sizes based on three elements (the maximum number of arrows pointing at a latent variable, the significance level used, and the minimum R^2/smallest effect size that a researcher wishes to detect in the study in the model) [45]. When the maximum number of arrows is 12, we need 181 observations to achieve a statistical power of 80% for detecting R^2 values of at least 0.1 (with a 5% probability of error). Therefore, 181 observations would be needed as a minimum for the study.

4.5 Data Quality and Reliability

To ensure reliability and quality of the data collected, it is essential to employ rigorous measures for participant selection and questionnaire design. We employed *three* widely recognized and frequently used measures in the questionnaire [83, 88, 120, 142, 147]:

[7] https://www.qualtrics.com
[8] https://prolific.co

- Recruitment of *high-reputation* participants with at least 100 submissions and an approval rate of 95% or more on the Prolific recruitment platform [147, 167].
- Use of two attention check questions to identify low-effort responses and filter those participants that failed either check from the analysis. These are essential for identifying inattentive respondents, promoting participants' concentration and involvement, and improving the reliability of the data collected [83, 120, 142].
- Applying the Simple Non-differentiation Method after reverse-coding six questionnaire items [88, 191] to the responses in order to identify *'straight lining'* by participants, excluding those who consistently selected the same response option across multiple questions (i.e., clicking in a straight line down the survey questions).

Table 2. Demographics of the survey participants.

		Participants	%			Participants	%
Gender	Male	115	38.3	Employment	Full time	115	38.3
	Female	182	60.7		Part-time	36	12
	Prefer not to say	3	1		Unemployed	17	5.7
Age	18-24	79	26.3		Other	14	4.7
	25-34	116	38.7		Retired/Homemaker	16	5.3
	35-44	61	20.3		Prefer not to say	102	34
	45-54	30	10	Student	Yes	81	27
	55-64	12	4		No	142	47.3
	65 +	3	1		Prefer not to say	77	25.7
Assistant used	Amazon Echo/Alexa	142	47.3	Healthcare use of assistant	Total	147	49
	Amazon Echo/Alexa + others	51	17		Information about illness/drug	71	23.7
	Google Home	75	25		Diagnosis based on symptoms	63	21
	Google Home + others	10	3.3		Monitor & manage chronic disease	9	3
	Microsoft Invoke/Cortana	6	2		Treatment Plan	3	1
	Apple HomePod/Siri	8	2.7		Others	1	0.3
	Other	8	2.7				

5 RESULTS

5.1 Participants

Overall, 355 participants were recruited for the study. Of these, 43 were removed from the analysis for failing one of the two attention checks, and 12 participants were eliminated after straight lining. As a result, 300 participants were retained for data analysis, which is higher than the 181 participants we had established as minimum in Section 4.4. The breakdown of demographics for the participants is reported in Table 2, and the full dataset is available online at https://osf.io/pdj3q/?view_only=a6fdef25221e4a569de661ec87571ce2 (link anonymised for review).

5.2 Model Evaluation

5.2.1 Measurement Model Analysis. We tested the internal consistency and discriminant validity of the proposed research model. The component reliability (CR) and average variance extracted (AVE) for each construct were examined to ensure that they met the threshold criteria for internal consistency.[9] The results are in Table 3, which shows that the CR values are all above 0.70 and all AVE values are above 0.50. Therefore, there are no convergent validity

[9]CR is used to gauge the consistency of survey items in assessing a certain unobservable concept. It measures how well these items work together. In contrast, AVE determines how accurately these items represent the underlying concept, or essentially, how much they 'hit the target'.

Table 3. Convergent and discriminant validity results.

	α	CR	AVE	A	SI	EE	F	IU	PCC	PRSQ	PU	PR	SR	TA	T	ST	PSR
A	0.804	0.885	0.719	0.848													
SI	0.853	0.900	0.693	0.437	0.833												
EE	0.870	0.920	0.794	-0.325	-0.340	0.891											
F	0.832	0.898	0.746	0.212	0.190	-0.084	0.864										
IU	0.776	0.869	0.689	0.572	0.502	-0.401	0.259	0.830									
PCC	0.761	0.862	0.675	0.574	0.486	-0.427	0.152	0.619	0.822								
PRSQ	0.802	0.883	0.717	0.488	0.480	-0.343	0.171	0.611	0.523	0.846							
PU	0.788	0.862	0.611	0.653	0.555	-0.549	0.193	0.685	0.690	0.597	0.781						
PR	0.907	0.941	0.842	-0.549	-0.392	0.321	-0.120	-0.596	-0.593	-0.497	-0.627	0.918					
SR	0.885	0.929	0.813	-0.553	-0.418	0.309	-0.190	-0.582	-0.587	-0.496	-0.657	0.687	0.902				
TA	0.840	0.902	0.754	0.301	0.239	-0.137	0.348	0.352	0.264	0.294	0.267	-0.271	-0.275	0.869			
T	0.836	0.891	0.672	0.682	0.546	-0.497	0.236	0.736	0.749	0.651	0.736	-0.731	-0.730	0.342	0.820		
ST	0.795	0.880	0.711	0.556	0.402	-0.337	0.222	0.573	0.554	0.486	0.636	-0.515	-0.518	0.242	0.675	0.843	
PSR	0.815	0.888	0.726	0.011	0.008	-0.047	-0.005	0.048	0.054	0.038	0.097	-0.009	-0.004	-0.081	0.073	-0.016	0.852

issues [16]. This was also confirmed with traditional Cronbach's α measure,[10] where all α values in our scale are greater than 0.7 (between 0.713 and 0.923) as reported in the same table. Furthermore, to assess discriminant validity, Fornell and Larcker [61] suggested assessing the AVE values of each construct with the inter-correlation scores within the correlations table. They recommended that the AVE score must exceed the scores of the inter-correlations. As it can be seen in Table 3, the AVE values for all of the constructs are higher than the square of their inter-construct correlations,thus confirming no discriminant validity issues [61].

In terms of construct reliability, the outer loadings analysed in our study are summarised in Table 7, and all of the constructs have individual construct reliability values that are much larger than the preferred level of 0.7, which according to Hair et al. indicates that the construct explains more than 50 per cent of the item's variance, thus providing acceptable item reliability, except ST1 (0.479), SR2 (0.256), PR1 (0.410), A3 (0.569), PCC4 (0.357), PCC5 (0.365) and PSR2 (0.612), which were therefore removed and are not included in the table.

5.2.2 Structural Model Analysis. We then examined the model's predictive capabilities and the relationships between the constructs. We first examined whether there were collinearity issues in the structural model. By calculating the variance inflation factor (VIF) values for the factors,[11] all VIF values were below the minimum recommended of 3 [92], with values from 1.389 to 2.780, thus confirming that multi-collinearity was not violated.

Table 4. PLS-SEM Analysis Results: Hypotheses testing results.

Factor Type		Hypothesis	β value	T Statistics	p value	Supported?
Functional Factor	H1	Anthropomorphism ->Trust in HVA	0.076	2.030	0.042	✓
	H2	Effort Expectancy ->Trust in HVA	-0.064	2.120	0.034	✓
	H3	Perceived Usefulness ->Trust in HVA	0.277	6.089	<0.001	✓
	H4	Perceived Content Credibility ->Trust in HVA	0.159	4.782	<0.001	✓
	H5	Perceived Relative Service Quality ->Trust in HVA	0.107	3.187	0.001	✓
Personal Factor	H6	Familiarity ->Trust in HVA	0.028	1.154	0.249	x
	H7	Technology Attachment ->Trust in HVA	0.038	1.341	0.180	x
	H8	Social Influence ->Trust in HVA	0.023	0.751	0.453	x
	H9	Stance in Technology ->Trust in HVA	0.111	3.568	<0.001	✓
Risk Factor	H10	Security Risk ->Trust in HVA	-0.138	3.149	0.002	✓
	H11	Privacy Risk ->Trust in HVA	-0.173	5.135	<0.001	✓
	H12	Perceived Substitution Risk ->Trust in HVA	0.032	1.116	0.264	x
Trust Impact	H13	Trust in HVA ->Intention to Use	0.737	24.597	<0.001	✓

[10]Cronbach's α tells us how closely related a set of questions are as a group. Higher values (typically over 0.7) suggest the questions are effectively measuring the same concept.
[11]VIF measures how much the input factors in a model influence each other. In situations where factors are closely related, indicated by a high VIF.

Next, a bootstrapping with 5000 sub-samples in SmartPLS with a “path” weight scheme was performed. The results, as shown in Table 4, illustrate support for the majority of hypotheses and indicate the importance of the relationships between the constructs. Among the thirteen hypotheses we considered, all five functional and design factors (H1 - H5), two risk factors (H10, H11) and one personal factor (H9) showed a significant influence on Trust in HVAs. In addition, H13 was proved, demonstrating that trust in HVA positively affects users’ intention to use HVAs. Each of these supported hypotheses has a T-statistic greater than 2, indicating that that the path coefficient obtained is also statistically significant [74, 77].

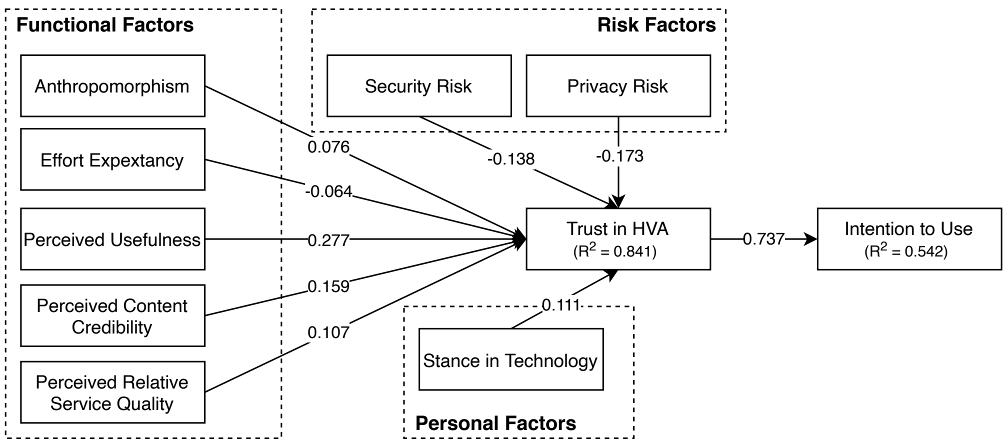

Fig. 2. Final Research Model with Path Coefficients (Only significant paths are retained.)

Figure 2 shows the final model only with significant paths. The R^2 was 0.841 and 0.542 for *Trust in HVA, and Intention to Use*, respectively, which represents the explanatory power of the construct. We can say that, the *Intention to Use* is **moderately** explained by *Trust in HVA*, which in turn is **substantially** explained by its antecedent factors[12]. Finally, the bootstrapped Standardized Root Mean Square Residual (SRMR) is the most frequently used measure to determine goodness-of-fit while using SmartPLS [57]. In this study, the SRMR value is 0.053, below the 0.08 threshold, suggesting an adequate model fit.

5.2.3 Controlling for HVA Use and Demographics. We also wanted to control for additional variables introduced by our sample of participants beyond the factors we hypothesized (as common in PLS-SEM analysis) [165]. Following guidance on the reporting of the use of control variables [15], we were particularly interested in our post-hoc analysis in the previous use of voice assistants for healthcare as well as other demographics reported by our participants (see Table 2). Although all our participants had experience using a voice assistant, not all of them were users of HVA (30.7% used HVA). This was important for properly modelling intention to use HVAs, something not possible if all participants were already using them. However, it was also important to understand whether that may have influenced trust, so we added to the model a control variable of previous HVA use on Trust in HVA. We also added control variables to capture the voice assistant they normally use (e.g. Amazon Alexa, Google Assistant, etc.) following [158], as well as general demographics like age [17] and gender [132] which may also have an influence in trust in technology in general. The resulting model (which is reported in the Appendix F) showed no significant differences in terms of the factors

[12]As suggested by Hair et al. [74, 77], R^2 values of 0.75, 0.50 and 0.25 can be considered substantial, moderate and weak.

previously analyzed; those significantly and substantially influencing trust in HVA remained so, and those that were not significant remained so as well. In terms of the control variables, only gender was statistically significant, suggesting women to be more inclined to trust in HVAs (Sample Mean = 0.107, p <0.001).

5.3 Post-survey Results

From the qualitative analysis of open-ended questions, we elicited themes that helped us better understand and supplement our structural model: 1) *Security and Privacy (associated with H10 and H11), 2) Anthropomorphism (associated with H2), 3) Other functional-related (associated with H2-5), 4) Financial Causes, 5) Discrimination Issues, 6) Develop Organisation, and 7) Substitution Risk (associated with H12).* While the model had already identified the first three themes, themes 4–6 emerged as *novel* findings from the thematic analysis. Moreover, the factor, *Perceived Substitution Risk*, which has not demonstrated statistical significance and therefore has been excluded from the model, is also not backed up by the results derived from the thematic analysis. The themes and codes can be found in Appendix D. 62 responses were excluded due to the provision of non-substantive or irrelevant responses to the posed questions, e.g., "I don't know" or "I am not sure".

5.3.1 Security and Privacy (retained). Security and privacy was a common theme mentioned by participants when asked about the use of healthcare voice assistants, they have shared several concerns including data misuse. In particular P113 (*Amazon Echo/Alexa&Google Home, none*) mentioned, "I would feel uncomfortable discussing personal issues that Amazon employees could hear". When asked about the future development of healthcare AI assistants, several mentioned improving security features; P7 (*Amazon Echo/Alexa&Google Home, Seek diagnosis results & Ask illness or drug information*) said "security improvement" and P178 (*Amazon Echo/Alexa, Ask illness or drug information*) said "A good approach regarding the security of the data collected". Although some participants are open to the idea of adopting HVAs, there are still some concerns regarding trust.

5.3.2 Anthropomorphism (retained). Participant responses helped us understand the reasons why factors, such as Anthropomorphism, were ultimately retained in the model. More specifically, 24.4% participants felt that a trustworthy HVA should firstly meet the basic ability to *"understand the user's description and have a natural conversation (P27, Google Home, Seek diagnosis results & Ask illness or drug information)*" with them, and secondly, at a higher level, *"they should be real and communicate like a human, less robot. This requires them to have more human-like features or customised voices. (P60, Amazon Echo/Alexa, Ask illness or drug information)*". In addition to this, considering that discussing a health condition is a serious and potentially emotionally draining matter, participants resonated strongly with the idea that a trustworthy HVA should be able to show compassion and empathy, with four participants mentioning that *"Although they are honest, it should have the empathy that doctors have when they say things like someone is going to die. (P16, Amazon&Microsoft Cortana, Seek diagnosis results&monitor a chronic disease; P56, Amazon Echo/Alexa, Seek diagnosis results; P103, Amazon Echo/Alexa, Seek diagnosis results & Ask illness or drug information; P223, Amazon Echo/Alexa&Apple Homepod, none)*".

5.3.3 Other functional-related (retained). Other functionality related factors were also mentioned such as the accuracy of the diagnosis provided by the HVA, P176 *(Google Home, Seek diagnosis results & Ask illness or drug information)* mentioned "Quality and accuracy of physical diagnosis, particularly in medical areas where people will feel uncomfortable seeing a doctor about". Others have mentioned the knowledge of HVAs and their ability to provide accurate responses, P252

(Google Home, none) "I would want to make sure they are consistent and knowledgeable". Usability was also been mentioned, P235 *(Amazon Echo/Alexa, Ask illness or drug information)* "easy ways to access health services".

5.3.4 Financial Causes (new). Price seems to be what will sway users to use and rely on HVAs particularly in places where healthcare is not provided as a public, free service. A participant points out that *"As for the cost, they should consider a broader group of users to make it affordable for everyone.(P242, Google Home, Seek diagnosis results & Ask illness or drug information&monitor a chronic disease)"*. For this, the majority of those impacted seems are not wealthy individuals, like one participant said: *"HVAs that give healthcare access to under-privileged populations would win more people support."* Apart from that, users are willing to trust and use HVAs as long as they are not set up to make a profit (*"An AI that is not influenced by profit motives but can supply cost information when requested (P209, Amazon Echo/Alexa, Ask illness or drug information)*".).

5.3.5 Discrimination Issues (new). Another factor to increase trust and adoption would be the **reducing discrimination** in the health care delivery process. Some users (2%) mentioned that they had received discrimination in traditional healthcare settings for different reasons and had experienced discomfort during their visits to doctors, so they wished that the HVAs could *"focus only on the facts (P235, Amazon Echo/Alexa, none)*". They were very confident in the usage of HVAs and one participant expressed that *"HVAs should help reduce the bias that many doctors have towards patients so that a real diagnosis could happen and help the patient (P292, Google Home, Seek diagnosis results & Ask illness or drug information)*".

5.3.6 Develop Organisation (new). Finally, an interesting result comes from the analysis of the **trusted organisations** that develop HVAs. In asking participants who would develop a HVA, it was concluded that the UK's NHS (i.e. the UK's National Health Service) had a much higher reputation among participants than big companies such as Google and Apple, while Amazon supporters were the least of these options (see Figure 3). Beyond the above parties, the most common type participants recommended were private tech companies, followed by hospitals and academic institutions.

5.3.7 Substitution Risk (excluded from the final model). . With regards to reasons why some factors were not retained in the model, the most common theme was about substitution risk. In particular, 22.3% participants do not feel that doctors will be replaced by HVAs in the future. On the contrary, they believe that HVAs should complement the work of doctors, *"I think it should be a tool used by doctors to help them, and not something used to replace them (P43, Google Home, none*)." The interesting thing is when HVAs work alone, especially on cases of serious illness, participants don't trust them to work well. But they trust the HVAs if their information and decisions can be monitored by a doctor (*"It is more reliable that they work alongside doctors, and decisions they made should be further validate by a real doctor (P300, Amazon Echo/Alexa, Ask illness or drug information).*").

6 DISCUSSION

We now summarize the contributions of our model and then compare it to previous research on trust in general-purpose voice assistants. After this, we discuss the main implications for HVAs and provide recommendations for developers and researchers who want to explore HVAs.

This study makes critical theoretical and practical contributions by firstly explaining user trust in HVAs through factors captured in three important aspects *(personal, functional, and risk)*. These dimensions are recognized as pivotal elements in the overarching framework of human-computer interaction, particularly in the context of fostering collaborative relationships between humans and AI entities [82, 85, 145]. Specifically, the model provided a statistically

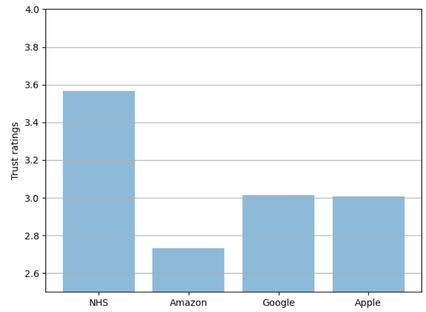

Fig. 3. Most trusted organizations to develop HVA.

based assessment of the degree to which the various factors influence trust, in addition to testing which factors increase or decrease it. According to the results, in each of the categories of personal, functional and risk factors, at least one factor was shown to have an impact on trust. This is a crucial step towards a holistic model to design trustworthy HVAs with a concrete understanding of the antecedents that foster trust. From the factors we considered, the most significant ones that affect user trust in HVAs are usefulness, information credibility, service quality, and security/privacy, exploring how they might be designed for in current and future products. We learned potential reasons why certain factors were retained in the model and were meaningful in explaining trust, while others were not (for instance because participants thought substitution was not really a risk in how they thought HVAs should work). In addition, we learned more about the users and tried to discover aspects that have not been taken into account in terms of trust (for instance who builds the HVA, if it helps mitigate discrimination and if it is affordable).

Our research, especially its results, offers profound insights that resonate with the broader society and, more specifically, the intricate healthcare sector. This field, characterized by its complexity, epitomizes the essence of multi-stakeholder collaboration, a theme central to the ethos of the CSCW community [59, 99, 144, 145]. By delving into the dynamics of trust and adoption, we provided AI product providers with a clearer roadmap for product refinement, emphasizing areas that foster enhanced collaboration [82, 85]. This collaborative lens can also help policymakers and third parties understand patient-centric concerns, notably privacy and data safety. Furthermore, when compared with general-purpose voice assistants, our findings accentuate the distinct collaborative and social complexity of the healthcare domain. In doing so, our work not only addresses societal needs but also aligns with and contributes to the core values and interests of the CSCW community, emphasizing the significance of cooperation.

6.1 Comparing Trust Between HVA and General-purpose Voice Assistants

Following the views expressed in previous literature [3, 127, 128], this study reveals the necessity of studying trust in the healthcare context, especially compared with trust in general-purpose technologies, in this case HVA in comparison with general-purpose voice assistants. To better compare the results of this study with previous work on trust in *general-purpose* voice assistants, we first summarize the outcomes of the factors being validated in literature on trust in general-purpose voice assistants in Table 5. Based on these previous works and our results, we highlight four main differences in terms of trust in HVA and trust in general-purpose voice assistants.

First, we considered factors that have not been studied in prior work on general-purpose voice assistants but that, as explained throughout this paper, have been shown to be key in the context of healthcare technologies (perceived content credibility, perceived relative service quality and perceived substitution risk). In this regard, we show that two functional factors, perceived content credibility and perceived service quality do in fact influence trust in HVAs, but that perceived substitution risk does not. According to the qualitative results that, the current use of HVAs is regarded as a complement to existing healthcare services rather than a replacement, something we discuss further in the next section on practical insights.

Table 5. Support for factors influencing trust in general-purpose voice assistants. N/C means not considered in the literature.

Category	Factor	Supported	Rejected
Functional	Anthropomorphism	[37, 60, 150]	-
	Effort Expectancy	[9, 37, 150]	-
	Perceived Usefulness	-	[150]
	Perceived Content Credibility	N/C	N/C
	Perceived Relative Service Quality	N/C	N/C
Personal	Familiarity	[10, 37]	-
	Technology Attachment	[37]	-
	Social Influence	[9, 37]	-
	Stance in Tech	[37]	-
Risk	Security Risk	[9, 10]	-
	Privacy Risk	[10, 182]	[150]
	Perceived Substitution Risk	N/C	N/C

Second, previous work on general-purpose voice assistants has not found a direct link between perceived usefulness and trust [150]. In contrast, this study shows that for HVAs, perceived usefulness is actually the factor that contributes the most towards substantially explaining trust. The reason for this may be that HVAs have a specific purpose, a hypothesis which is consistent with previous work on healthcare technology that also found perceived usefulness to influence trust [148]. On the other hand, general purpose assistants tend to have more varied purposes that are not found in the healthcare context, such as hedonic benefits and entertainment [150].

Third, among the personal factors, only Stance in Tech proved to be directly linked to trust in HVA. Even the social influence factor, where a device owner's relationship may affect user perceptions in general settings [194], seems not to hold true in the distinct environment of healthcare. This may be because what leads people to trust VAs in general, i.e. familiarity and attachment to the platform [10, 37], being recognised or influenced by others [9], are not as important in the perceptions of users as other factors, such as the usefulness and quality evidenced in the results of our model. It may also be that not all participants actually have people in their social networks that use HVAs or they may not be familiar with it. However, when controlling in the model for HVA use (recall that half of our participants used HVA) it was shown as not significant.

Finally, prior work on general-purpose voice assistants is inconclusive as to the effect of privacy risks on trust, with some studies suggesting that privacy risks influence trust [10, 182] and others suggesting that it does not [150]. We show that perceived privacy risks are very important for trust in HVAs, being the second most significant factor in the final model. One explanation for this is that healthcare data is known to be very sensitive in general [3, 127], and previous work on privacy and voice assistants in particular found that users considered the flows of this data across the voice assistant ecosystem as the least appropriate [3].

6.2 Implications Unique to Healthcare Voice Assistants

Enhancing usefulness in HVAs through participatory design. Perceived usefulness plays the most important role in the user's trust in HVAs. This suggest that engagement with patients throughout the process of HVA development is crucial, from ideation through design, implementation, and evaluation. A common and well established means of achieving highly usable solutions in healthcare and other areas of HCI is participatory design methods [80, 166], and beyond this it is important to engage with the communities that are the target of an intervention throughout the design and development process. It has been shown, for instance, that as a response to the lack of usability of some diabetes technologies for many patients, users have begun hacking their way into commercial devices to better serve their needs (e.g. do-it-yourself automated insulin delivery systems [27]). Beyond usefulness as an objective it is also important that users *perceive* HVAs to be useful, particularly as potential users may have had negative experiences with voice assistants in the past.

Balancing service quality and roles in HVAs. The quality of service relative to what a health professional would offer also contributes substantially and significantly to trust in HVA. When considering this, there are two key roles HVA could play that may make it easier or harder to offer a quality of service that is higher relative to clinical staff. The first of these roles would be HVAs as a *replacement* for in-person care, implying a challenging goal of matching the diagnostic, analytical, and social/communication skills. It has been recommended that HVAs should be more compassionate [171] and exhibit more non-verbal communication skills and emotional reactions [4]. This also requires that HVAs are able to personalise responses to different people; behaving like a human healthcare professional requires empathy when facing serious cases, as well as observing the user's personality traits in other consultation situations. This is an area where development is required. At present, research reports mixed results and a lack of testing in real-world clinical environments [66, 134, 183]. The second, and potentially easier, role for HVAs would be as a *complement* to existing healthcare services. In these scenarios, HVAs might be restricted to triage, giving limited advice on less complex conditions and handing off other cases to human practitioners. This is a much simpler task, lowering the cost of service delivery whilst being much easier to adequately perform relative to a medical professional (as an example, matching the advice given by a clinician for a common cold is much easier than for type 2 diabetes). In the qualitative results, it becomes evident that the ideal scenario for patients involves the integration of healthcare AI but supervised and regulated by physicians. This not only exemplifies the need for harmonious collaboration between doctors and cutting-edge technology for the betterment of patient care, as highlighted in [59, 99, 145], but also underscores the importance of human-AI interaction in fostering trust between patients and their physicians.

Privacy complexities in HVAs and design choices. While privacy is a key concern in general-purpose voice assistants, the introduction of healthcare data presents additional complexities. Many research prototypes and currently available HVAs are built with or on top of general-purpose commercial voice assistants [160]. The security and privacy of voice assistants in general is an active area of research—see [52] for a systematic review on the topic. How users' perceive security and privacy of these underlying platforms is very important for HVAs, as users consider healthcare-related data more sensitive and prefer to share less information with voice assistants about health [3]. As the information they share with them is often protected by additional regulation (such as HIPAA in the U.S.), the security of those HVAs is therefore highly dependent on the security of the voice assistants on top of which they are built. To accommodate this, certain platforms such as Alexa have additional requirements for health-related skills[13] — skills are the name given to the third-party functionality that can be added to Alexa to add capacities to it. Furthermore, attitudes towards health information privacy have notably shifted over time, particularly in the context of chronic illnesses such as diabetes [139]. For instance, the emergence of health-related complications within a patient's medical condition may

[13]https://developer.amazon.com/en-US/docs/alexa/custom-skills/requirements-for-hipaa-eligible-skills.html

result in a diminished emphasis on privacy. Additionally, the assumption of new responsibilities, such as the care of children afflicted with chronic illnesses, may prompt individuals to reassess their privacy considerations. Given the evolving concerns, especially in chronic illnesses like diabetes, it's imperative to address these in the design of voice assistants handling sensitive health data.

An alternative option is to create standalone HVAs (instead of on top of general-purpose voice assistants), mitigating the risks and negative perceptions related to voice assistant platforms. However, the vast amount of data and expertise required in order to create the speech recognition and NLP processing services required to operate a successful voice assistant means that, in practice, these services will likely be provided by third parties, who may even be the same organisations providing general-purpose voice assistants. While there may be differences between, e.g., creating an HVA skill to run on top of Amazon Alexa and creating a standalone HVA using Amazon Lex[14], that difference may be difficult to communicate to users.

Trust dynamics and the importance of compliance in development. An often overlooked point in previous research, is who are expected to make the HVAs? For some mainstream technology companies (Amazon, Google, and Apple), user trust in VAs when they are used for no specific purpose correlates with trust in their parent company, and Amazon has a higher level of "trust reputation" than Google [158] in the minds of users. However in terms of providing health-related services, Amazon was less trusted than Google/Apple in our results. This may be explained by the fact that platforms collaborate with different partners that provide/develop health applications, and the strength of the platform in regulating the compliance of these applications with existing privacy and security policies can also affect user trust building. Evidence suggests that breaches are common in Amazon Alexa skills for health care, with 86.36% missing disclaimers when providing medical information, and 30.23% storing users' physical or mental health data without approval [164]. Beyond big tech companies, public health organizations such as the UK's NHS are trusted more to develop HVA. This is probably because the NHS system is largely funded by the UK government and is under the jurisdiction and oversight of the Department of Health[15]. Meanwhile, given that the services provided by the NHS are free to UK citizens, this reinforces users' belief that the purpose of NHS is to help people, improve the healthcare environment and save lives, rather than earning money. This is in line with the argument made by users that they trust non-profit organisations more as for-profit companies often do not have the most stringent track record in protecting data and may sell these data without user awareness. Therefore, handing the company the keys to people's private health information raises red flags. Furthermore, each country will have its own clinical guidelines, and as we mentioned before, some countries often have additional regulations governing the use of health information (HIPAA in the U.S.). In this regard, each healthcare system would potentially suggest different organisations who would develop/manage such an assistant. This might be redundant for VAs that provide common functions (e.g. ordering a taxi, managing the To-do List), but it is necessary for HVAs. Moreover, deeper integration between large companies with established program frameworks and hospitals with specialist physicians and public credibility should be encouraged as well.

Discrimination and cost. Last but not least, other factors identified in our post-survey questionnaire should also be considered for trust in HVA in practice, such as the potential for HVA to be more or less discriminatory than health professionals and the impact this may have on users. Participants seem to think that HVA would discriminate and have less biases and preconceptions than health professionals. However, it is well-established that AI-based systems

[14] A "fully managed artificial intelligence (AI) service with advanced natural language models to design, build, test, and deploy conversational interfaces in applications" https://aws.amazon.com/lex

[15] https://www.nhs.uk/

may not necessarily be any more *neutral*, and they can actually be as discriminatory as, or even more than, human decision-makers [58]. Beyond benefiting from the general efforts towards fairer AI [25], this will also require HVAs developers to have a deeper understanding of discrimination in the medical industry and thoroughly evaluate future HVAs in the context of these ethical issues. With regards to financial issues, the cost of HVAs may also need to be considered.

6.3 Limitations and Future Work

This study has some limitations. First, our model is a significant step towards a holistic model to explain factors that influence trust in HVA. However, there are still many factors in all levels of designing an HVA that can impact on trust and should be taken into account (e.g., factors that were newly identified in qualitative results, financial causes, bias, and develop organisation). Future work should focus on this and continue to extend our model. It is crucial to delve deeper into the influential factors that affect this field of study. Considerations such as the systematic categorisation of functional factors, the improved capacity for rectifying communication errors and failures as evidenced in [40], and their subsequent effect on trust, demand attention. Additionally, privacy and security concerns associated with general voice assistants [103, 118, 124, 125, 157, 194], their application within the healthcare context, and their potential influence on trust also necessitate investigation. These represent significant opportunities for future work within this research domain. Second, trust may evolve as time goes by and users have more chances to interact with HVA. Therefore, longitudinal studies should also be considered in the future to consider factors that contribute to trust building over time, such as the type of relationship that could be formed between a user and an HVA, with recent studies having explored the relationships that users develop with general-purpose voice assistants [158]. Concurrently, as discussed in section 6.2, developers should continually update their efforts to ensure functional ease and user safety mechanisms. Given that attitudes, including privacy concerns and needs, can evolve over time [139], it's imperative to develop and offer services that provide enduring benefits to users. Third, we focused on users of HVAs as *patients*, but another interesting demographic to consider would be healthcare professionals, who may also utilize HVAs to facilitate their work, e.g. by allowing doctors to have voice interactions with electronic health records [98] or supporting the tasks of home health aides [21]. This would complement our study from the 'two sides' of potential HVA use.

7 CONCLUSION

Our work is the first one to examine different functional, personal and risk factors that affect trust in healthcare voice assistants and user intention to use these technologies. We find that anthropomorphism, perceived usefulness, effort expectancy, content credibility, and relative service quality, together with security risk, privacy risk and stance in technology substantially explain trust in HVA. In turn, trust in HVA moderately explains intention to use HVA. Based on these results, we derive some practical insights and recommendations to help design and develop next-generation HVAs that can promote user trust and encourage adoption.

ACKNOWLEDGMENTS

We thank CSCW's anonymous reviewers, Ruba Abu-Salma, and Hana Kopecka for their constructive feedback that helped improve the paper. This work is part of the EPSRC-funded Secure AI Assistants project (grant EP/T026723/1).

REFERENCES

[1] MM Abbasi and S Kashiyarndi. 2006. Clinical Decision Support Systems: A discussion on different methodologies used in Health Care. *Marlaedalen University Sweden* (2006).

[2] Noura Abdi, Kopo Ramokapane, and Jose Such. 2019. More than smart speakers: security and privacy perceptions of smart home personal assistants. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*. 451–466.

[3] Noura Abdi, Xiao Zhan, Kopo M Ramokapane, and Jose Such. 2021. Privacy Norms for Smart Home Personal Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[4] Sara Aeschlimann, Marco Bleiker, Michael Wechner, and Anja Gampe. 2020. Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior* 112 (2020), 106466.

[5] Ritu Agarwal and Jayesh Prasad. 1998. A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information systems research* 9, 2 (1998), 204–215. https://doi.org/10.1287/isre.9.2.204

[6] Rachith Aiyappa, Jisun An, Haewoon Kwak, and Yong-Yeol Ahn. 2023. Can we trust the evaluation on ChatGPT? *arXiv preprint arXiv:2303.12767* (2023).

[7] Jawaher Hamad Al Shamsi, Mostafa Al-Emran, and Khaled Shaalan. 2022. Understanding key drivers affecting students' use of artificial intelligence-based voice assistants. *Education and Information Technologies* (2022), 1–21.

[8] Emily Couvillon Alagha and Rachel Renee Helbing. 2019. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. *BMJ health & care informatics* 26, 1 (2019). https://doi.org/10.1136/bmjhci-2019-100075

[9] Areej AlHogail. 2018. Improving IoT technology adoption through improving consumer trust. *Technologies* 6, 3 (2018), 64. https://doi.org/10.3390/technologies6030064

[10] Mansour Naser Alraja, Murtaza Mohiuddin Junaid Farooque, and Basel Khashab. 2019. The effect of security, privacy, familiarity, and trust on users' attitudes toward the use of the IoT-based healthcare: the mediation role of risk perception. *IEEE Access* 7 (2019), 111341–111354.

[11] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput. Hum. Interact.* 26, 3 (2019), 28. https://doi.org/10.1145/3311956

[12] Alyssa Appelman and S Shyam Sundar. 2016. Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly* 93, 1 (2016), 59–79. https://doi.org/10.1177/1077699015606057

[13] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.

[14] Onur Asan, Alparslan Emrah Bayrak, Avishek Choudhury, et al. 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research* 22, 6 (2020), e15154. https://doi.org/doi:10.2196/15154

[15] Guclu Atinc, Marcia J Simmering, and Mark J Kroll. 2012. Control variable use and reporting in macro and micro management research. *Organizational Research Methods* 15, 1 (2012), 57–74.

[16] Richard P Bagozzi and Youjae Yi. 1988. On the evaluation of structural equation models. *Journal of the academy of marketing science* 16, 1 (1988), 74–94. https://doi.org/10.1007/BF02723327

[17] Phoebe E Bailey and Tarren Leon. 2019. A systematic review and meta-analysis of age-related differences in trust. *Psychology and aging* 34, 5 (2019), 674.

[18] Adam Baker, Yura Perov, Katherine Middleton, Janie Baxter, Daniel Mullarkey, Davinder Sangar, Mobasher Butt, Arnold DoRosario, and Saurabh Johri. 2020. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. *Frontiers in artificial intelligence* 3 (2020), 543405.

[19] Gaurav Bansal, Fatemeh Mariam Zahedi, and David Gefen. 2016. Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management* 53, 1 (2016), 1–21.

[20] Vince Bartle, Janice Lyu, Freesoul El Shabazz-Thompson, Yunmin Oh, Angela Anqi Chen, Yu-Jan Chang, Kenneth Holstein, and Nicola Dell. 2022. "A Second Voice": Investigating Opportunities and Challenges for Interactive Voice Assistants to Support Home Health Aides *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 45, 17 pages. https://doi.org/10.1145/3491102.3517683

[21] Vince Bartle, Janice Lyu, Freesoul El Shabazz-Thompson, Yunmin Oh, Angela Anqi Chen, Yu-Jan Chang, Kenneth Holstein, and Nicola Dell. 2022. "A Second Voice": Investigating Opportunities and Challenges for Interactive Voice Assistants to Support Home Health Aides. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[22] Izak Benbasat and Weiquan Wang. 2005. Trust in and adoption of online recommendation agents. *Journal of the association for information systems* 6, 3 (2005), 4.

[23] Urmil Bharti, Deepali Bajaj, Hunar Batra, Shreya Lalit, Shweta Lalit, and Aayushi Gangwani. 2020. Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In *2020 5th international conference on communication and electronics systems (ICCES)*. IEEE, 870–875.

[24] Timothy W Bickmore, Ha Trinh, Stefan Olafsson, Teresa K O'Leary, Reza Asadi, Nathaniel M Rickles, and Ricardo Cruz. 2018. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *Journal of*

medical Internet research 20, 9 (2018), e11510.

[25] Reuben Binns. 2018. Fairness in machine learning: Lessons from political philosophy. In *Conference on Fairness, Accountability and Transparency*. PMLR, 149–159.

[26] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012). https://doi.org/10.1037/13620-004

[27] Katarina Braune, Rayhan A Lal, Lenka Petruželková, Gary Scheiner, Per Winterdijk, Signe Schmidt, Linda Raimond, Korey K Hood, Michael C Riddell, Timothy C Skinner, et al. 2022. Open-source automated insulin delivery: international consensus statement and practical guidance for health-care professionals. *The Lancet Diabetes & Endocrinology* 10, 1 (2022), 58–74.

[28] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2022. An empirical study of older adult's voice assistant use for health information seeking. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 2 (2022), 1–32.

[29] Susana Buinhas, Ana Paula Cláudio, Maria Beatriz Carmo, João Balsa, Afonso Cavaco, Anabela Mendes, Isa Félix, Nuno Pimenta, and Mara Pereira Guerreiro. 2018. Virtual assistant to improve self-care of older people with type 2 diabetes: first prototype. In *International Workshop on Gerontechnology*. Springer, 236–248.

[30] Emily Buteau and Joonghwa Lee. 2021. Hey Alexa, why do we use voice assistants? The driving factors of voice assistant technology use. *Communication Research Reports* 38, 5 (2021), 336–345.

[31] Yakob Utama Chandra, Desi Maya Kristin, Joni Suhartono, Fina Shabrina Sutarto, and Minseol Sung. 2018. Analysis of determinant factors of user acceptance of mobile payment system in indonesia (a case study of go-pay mobile payment). In *2018 International Conference on Information Management and Technology (ICIMTech)*. IEEE, 454–459.

[32] Ching-Sheng Chang, Su-Yueh Chen, and Yi-Ting Lan. 2013. Service quality, trust, and patient satisfaction in interpersonal-based medical service encounters. *BMC health services research* 13, 1 (2013), 1–11.

[33] Shuchih Ernest Chang, Anne Yenching Liu, and Wei Cheng Shen. 2017. User trust in social networking services: A comparison of Facebook and LinkedIn. *Computers in Human Behavior* 69 (2017), 207–217.

[34] Cheng-Min Chao. 2019. Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model. *Frontiers in psychology* 10 (2019), 1652.

[35] Qian Qian Chen and Hyun Jung Park. 2021. How anthropomorphism affects trust in intelligent personal assistants. *Industrial Management & Data Systems* (2021).

[36] Amy Cheng, Vaishnavi Raghavaraju, Jayanth Kanugo, Yohanes P Handrianto, and Yi Shang. 2018. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–5.

[37] Oscar Hengxuan Chi, Shizhen Jia, Yafang Li, and Dogan Gursoy. 2021. Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior* 118 (2021), 106700. https://doi.org/10.1016/j.chb.2021.106700

[38] Eugene Cho. 2019. Hey Google, can I ask you something in private?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–9.

[39] Eugene Cho, S. Shyam Sundar, Saeed Abdullah, and Nasim Motalebi. 2020. *Will Deleting History Make Alexa More Trustworthy? Effects of Privacy and Content Customization on User Experience of Smart Speakers*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376551

[40] Janghee Cho and Emilee Rader. 2020. The role of conversational grounding in supporting symbiosis between people and digital assistants. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.

[41] Jinsook E Cho and Haiyan Hu. 2009. The effect of service quality on trust and commitment varying across generations. *International journal of consumer studies* 33, 4 (2009), 468–476.

[42] Avishek Choudhury and Hamid Shamszare. 2023. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *Journal of Medical Internet Research* 25 (2023), e47184.

[43] Ringle Christian M., Wende Sven, and Becker Jan-Michael. 2015. SmartPLS 3. Bönningstedt: SmartPLS. https://www.smartpls.com/

[44] Arlene E Chung, Ashley C Griffin, Dasha Selezneva, and David Gotz. 2018. Health and fitness apps for hands-free voice-activated assistants: content analysis. *JMIR mHealth and uHealth* 6, 9 (2018), e9705.

[45] Jacob Cohen. 1992. A power primer. *Psychological bulletin* 112, 1 (1992), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

[46] Tim Coughlan, Michael Brown, Richard Mortier, Robert J. Houghton, Murray Goulden, and Glyn Lawson. 2012. Exploring Acceptance and Consequences of the Internet of Things in the Home. In *2012 IEEE International Conference on Green Computing and Communications*. 148–155. https://doi.org/10.1109/GreenCom.2012.32

[47] Henry Kudzanai Dambanemuya and Nicholas Diakopoulos. 2021. Auditing the Information Quality of News-Related Queries on the Alexa Voice Assistant. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[48] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[49] Dimitri Dojchinovski, Andrej Ilievski, and Marjan Gusev. 2019. Interactive home healthcare system with integrated voice assistant. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 284–288.

[50] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In *Proceedings of the 25th international*

conference on intelligent user interfaces. 297–307.

[51] Aarthi Easwara Moorthy and Kim-Phuong L Vu. 2015. Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction* 31, 4 (2015), 307–335.

[52] Jide Edu, Jose Such, and Guillermo Suarez-Tangil. 2020. Smart home personal assistants: a security and privacy review. *ACM Computing Surveys (CSUR)* 53, 6 (2020), 1–36. https://doi.org/10.1145/3412383

[53] José Manuel Ortega Egea and María Victoria Román González. 2011. Explaining physicians' acceptance of EHCR systems: An extension of TAM with trust and risk factors. *Computers in Human Behavior* 27, 1 (2011), 319–332.

[54] Samer Ellahham, Nour Ellahham, and Mecit Can Emre Simsekler. 2020. Application of artificial intelligence in the health care safety context: opportunities and challenges. *American Journal of Medical Quality* 35, 4 (2020), 341–348.

[55] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

[56] Wenjuan Fan, Jingnan Liu, Shuwan Zhu, and Panos M Pardalos. 2020. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research* 294, 1 (2020), 567–592. https://doi.org/10.1007/s10479-018-2818-y

[57] Teresa Fernandes and Elisabete Oliveira. 2021. Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research* 122 (2021), 180–191.

[58] Xavier Ferrer, Tom van Nuenen, Jose Such, Mark Coté, and Natalia Criado. 2021. Bias and Discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine* 40, 2 (2021), 72–80.

[59] Geraldine Fitzpatrick and Gunnar Ellingsen. 2013. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)* 22 (2013), 609–665.

[60] Jonas Foehr and Claas Christian Germelmann. 2020. Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. *Journal of the Association for Consumer Research* 5, 2 (2020), 181–205.

[61] Claes Fornell and David F Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research* 18, 1 (1981), 39–50. https://doi.org/10.1177/002224378101800104

[62] Alan Fox. 1974. *Beyond contract: Work, power and trust relations*. Faber & Faber.

[63] Frontiers. 2021. The Trustworthiness of Voice Assistants in the Context of Healthcare Investigating the Effect of Perceived Expertise on the Trustworthiness of Voice Assistants, Providers, Data Receivers, and Automatic Speech Recognition. https://www.frontiersin.org/articles/10.3389/fcomp.2021.685250/full

[64] Lingling Gao and Xuesong Bai. 2014. A unified perspective on the factors influencing consumer acceptance of internet of things technology. *Asia Pacific Journal of Marketing and Logistics* 26, 2 (2014), 211–231. https://doi.org/10.1108/APJML-06-2013-0061

[65] David Gefen and Detmar W Straub. 2004. Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega* 32, 6 (2004), 407–424.

[66] Felix Gille, Anna Jobin, and Marcello Ienca. 2020. What we talk about when we talk about trust: Theory of trust for AI in healthcare. *Intelligence-Based Medicine* 1 (2020), 100001. https://doi.org/10.1016/j.ibmed.2020.100001

[67] Paul AM Gregory and Zubin Austin. 2021. How do patients develop trust in community pharmacists? *Research in Social and Administrative Pharmacy* 17, 5 (2021), 911–920. https://doi.org/10.1016/j.sapharm.2020.07.023

[68] Dogan Gursoy, Oscar Hengxuan Chi, Lu Lu, and Robin Nunkoo. 2019. Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management* 49 (2019), 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

[69] Dogan Gursoy, Giacomo Del Chiappa, and Yi Zhang. 2017. Preferences regarding external information sources: a conjoint analysis of visitors to Sardinia, Italy. *Journal of Travel & Tourism Marketing* 34, 6 (2017), 806–820. https://doi.org/10.1080/10548408.2016.1237405

[70] Melissa Gymrek, Amy L McGuire, David Golan, Eran Halperin, and Yaniv Erlich. 2013. Identifying personal genomes by surname inference. *Science* 339, 6117 (2013), 321–324.

[71] Taehyun Ha, Sangyeon Kim, Donghak Seo, and Sangwon Lee. 2020. Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation research part F: traffic psychology and behaviour* 73 (2020), 271–280.

[72] Marieke Haan, Yfke P Ongena, Saar Hommes, Thomas C Kwee, and Derya Yakar. 2019. A qualitative study to understand patient perspective on the use of artificial intelligence in radiology. *Journal of the American College of Radiology* 16, 10 (2019), 1416–1419.

[73] J Hair, W Black, B Babin, and R Anderson. 2009. Multivariate Data Analysis 7th Edition Pearson Prentice Hall: JOUR. (2009).

[74] Joe Hair, Carole L Hollingsworth, Adriane B Randolph, and Alain Yee Loong Chong. 2017. An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems* 117, 3 (2017), 442–458. https://doi.org/10.1108/IMDS-04-2016-0130

[75] Joseph F Hair, G Thomas M Hult, Christian Ringle, and Marko Sarstedt. 2014. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). (2014).

[76] Joe F Hair, Christian M Ringle, and Marko Sarstedt. 2011. PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice* 19, 2 (2011), 139–152.

[77] Joseph F Hair, Jeffrey J Risher, Marko Sarstedt, and Christian M Ringle. 2019. When to use and how to report the results of PLS-SEM. *European business review* (2019).

[78] Mark A Hall, Elizabeth Dugan, Beiyao Zheng, and Aneil K Mishra. 2001. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? *The milbank quarterly* 79, 4 (2001), 613–639. https://doi.org/10.1111/1468-0009.00223

[79] Sangyeal Han and Heetae Yang. 2018. Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management & Data Systems* 118, 3 (2018), 618–636. https://doi.org/10.1108/IMDS-05-2017-0214

[80] Christina N Harrington, Lyndsie Marie Koon, and Wendy A Rogers. 2020. Design of health information and communication technologies for older adults. In *Design for Health*. Elsevier, 341–363.

[81] Rajibul Hasan, Riad Shams, and Mizan Rahman. 2020. Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. *Journal of Business Research* 131 (2020), 591–597. https://doi.org/10.1016/j.jbusres.2020.12.012

[82] Allyson I. Hauptman, Wen Duan, and Nathan J. Mcneese. 2022. The Components of Trust for Collaborating With AI Colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) *(CSCW'22 Companion)*. Association for Computing Machinery, New York, NY, USA, 72–75. https://doi.org/10.1145/3500868.3559450

[83] David J Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods* 48, 1 (2016), 400–407.

[84] Erika C Hayden. 2013. Privacy loophole found in genetic databases. *Nature News* 17 (2013).

[85] Marina Jirotka, Rob Procter, Mark Hartswood, Roger Slack, Andrew Simpson, Catelijne Coopmans, Chris Hinds, and Alex Voss. 2005. Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work (CSCW)* 14 (2005), 369–398.

[86] Tanja B Jutzi, Eva I Krieghoff-Henning, Tim Holland-Letz, Jochen Sven Utikal, Axel Hauschild, Dirk Schadendorf, Wiebke Sondermann, Stefan Fröhling, Achim Hekler, Max Schmitt, et al. 2020. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Frontiers in medicine* 7 (2020), 233.

[87] Kadircan H Keskinbora. 2019. Medical ethics considerations on artificial intelligence. *Journal of clinical neuroscience* 64 (2019), 277–282.

[88] Yujin Kim, Jennifer Dykema, John Stevenson, Penny Black, and D Paul Moberg. 2019. Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review* 37, 2 (2019), 214–233. https://doi.org/10.1177/0894439317752406

[89] Bran Knowles and Vicki L. Hanson. 2018. Older Adults' Deployment of 'Distrust'. 25, 4, Article 21 (aug 2018), 25 pages. https://doi.org/10.1145/3196490

[90] Bran Knowles, Mark Rouncefield, Mike Harding, Nigel Davies, Lynne Blair, James Hannon, John Walden, and Ding Wang. 2015. Models and patterns of trust. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 328–338.

[91] Ahmet Baki Kocaballi, Juan C Quiroz, Dana Rezazadegan, Shlomo Berkovsky, Farah Magrabi, Enrico Coiera, and Liliana Laranjo. 2020. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. *Journal of medical Internet research* 22, 2 (2020), e15823. https://doi.org/doi:10.2196/15823

[92] Ned Kock and Gary Lynn. 2012. Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for information Systems* 13, 7 (2012).

[93] Yoon Jeon Koh and S Shyam Sundar. 2010. Effects of specialization in computers, web sites, and web agents on e-commerce trust. *International journal of human-computer studies* 68, 12 (2010), 899–912.

[94] Geir M Køien. 2011. Reflections on trust in devices: an informal survey of human trust in an internet-of-things context. *Wireless personal communications* 61, 3 (2011), 495–510. https://doi.org/10.1007/s11277-011-0386-4

[95] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly* 30, 4 (2006), 941–960. https://doi.org/10.2307/25148760

[96] Pascal Kowalczuk. 2018. Consumer acceptance of smart speakers: a mixed methods approach. *Journal of Research in Interactive Marketing* (2018).

[97] Songpol Kulviwat, Gordon C Bruner II, Anand Kumar, Suzanne A Nasco, and Terry Clark. 2007. Toward a unified theory of consumer acceptance technology. *Psychology & Marketing* 24, 12 (2007), 1059–1084.

[98] Yaa A Kumah-Crystal, Claude J Pirtle, Harrison M Whyte, Edward S Goode, Shilo H Anders, and Christoph U Lehmann. 2018. Electronic health record interactions through voice: a review. *Applied clinical informatics* 9, 3 (2018), 541.

[99] Yi Lai, Atreyi Kankanhalli, and Desmond Ong. 2021. Human-AI collaboration in healthcare: A review and research agenda. (2021).

[100] Airi Lampinen, Victoria Bellotti, Coye Cheshire, and Mary Gray. 2016. CSCW and theSharing Economy: The Future of Platforms as Sites of Work Collaboration and Trust. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 491–497.

[101] Raina Langevin, Mohammad Rafayet Ali, Taylan Sen, Christopher Snyder, Taylor Myers, E. Ray Dorsey, and Mohammed Ehsan Hoque. 2019. The PARK Framework for Automated Analysis of Parkinson's Disease Characteristics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 54 (jun 2019), 22 pages. https://doi.org/10.1145/3328925

[102] Bibb Latané. 1981. The psychology of social impact. *American psychologist* 36, 4 (1981), 343–356. https://doi.org/10.1037/0003-066X.36.4.343

[103] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–31.

[104] Chaiwoo Lee and Joseph F Coughlin. 2015. PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers. *Journal of Product Innovation Management* 32, 5 (2015), 747–759.

[105] DonHee Lee and Seong No Yoon. 2021. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health* 18, 1 (2021), 271. https://doi.org/10.3390/ijerph18010271

[106] Min Kyung Lee and Katherine Rich. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[107] Matthew KO Lee and Efraim Turban. 2001. A trust model for consumer internet shopping. *International Journal of electronic commerce* 6, 1 (2001), 75–91.

[108] One-Ki Daniel Lee, Ramakrishna Ayyagari, Farzaneh Nasirian, and Mohsen Ahmadian. 2021. Role of interaction quality and trust in use of AI-based voice-assistant systems. *Journal of Systems and Information Technology* 23, 2 (2021), 154–170. https://doi.org/10.1108/JSIT-07-2020-0132

[109] Stephan Lewandowsky, Michael Mundy, and Gerard Tan. 2000. The dynamics of trust: comparing humans to automation. *Journal of Experimental Psychology: Applied* 6, 2 (2000), 104.

[110] Yuting Liao, Jessica Vitak, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2019. Understanding the role of privacy and trust in intelligent personal assistant adoption. In *International Conference on Information*, Vol. 11420. Springer, 102–113. https://doi.org/10.1007/978-3-030-15742-5_9

[111] Hongxia Lin, Oscar Hengxuan Chi, and Dogan Gursoy. 2020. Antecedents of customers' acceptance of artificially intelligent robotic device use in hospitality services. *Journal of Hospitality Marketing & Management* 29, 5 (2020), 530–549. https://doi.org/10.1080/19368623.2020.1685053

[112] Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, et al. 2019. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health* 1, 6 (2019), e271–e297.

[113] Chen Lou and Shupei Yuan. 2019. Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising* 19, 1 (2019), 58–73.

[114] Ben Lowe, Yogesh K Dwivedi, and Steve d'Alessandro. 2019. Consumers and technology in a changing world. *European Journal of Marketing* 53, 6 (2019), 1038–1050.

[115] Yaobin Lu, Shuiqing Yang, Patrick YK Chau, and Yuzhi Cao. 2011. Dynamics between the trust transfer process and intention to use mobile payment services: A cross-environment perspective. *Information & management* 48, 8 (2011), 393–403.

[116] Niklas Luhmann. 2001. Familiarity, confidence, trust: Problems and alternatives. *Reseaux* 108, 4 (2001), 15–35.

[117] Poornima Madhavan, Douglas A Wiegmann, and Frank C Lacson. 2006. Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human factors* 48, 2 (2006), 241–256.

[118] Nathan Malkin, David Wagner, and Serge Egelman. 2022. Can Humans Detect Malicious Always-Listening Assistants? A Framework for Crowdsourcing Test Drives. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.

[119] Pedro M. Martins, João L. Vilaça, and Nuno S. Dias. 2021. A study about current digital assistants for healthcare and medical treatment monitoring. In *2021 IEEE 9th International Conference on Serious Games and Applications for Health(SeGAH)*. 1–7. https://doi.org/10.1109/SEGAH52098.2021.9551864

[120] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[121] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[122] D Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)* 2, 2 (2011), 1–25. https://doi.org/10.1145/1985347.1985353

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

[124] Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. 2019. Designing with gaze: Tama–a gaze activated smart-speaker. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[125] Nicole Meng, Dilara Keküllüoğlu, and Kami Vaniea. 2021. Owning and sharing: Privacy perceptions of smart speaker users. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

[126] Stephanie M Merritt and Daniel R Ilgen. 2008. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human factors* 50, 2 (2008), 194–210.

[127] Adam S Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA internal medicine* 176, 5 (2016), 619–625.

[128] Enid NH Montague, Brian M Kleiner, and Woodrow W Winchester III. 2009. Empirically understanding trust in medical technology. *International Journal of Industrial Ergonomics* 39, 4 (2009), 628–634. https://doi.org/10.1016/j.ergon.2009.01.004

[129] Sara Moussawi, Marios Koufaris, and Raquel Benbunan-Fich. 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets* 31, 2 (2021), 343–364.

[130] Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.

[131] Michael Muller, Katja Neureiter, Nervo Verdezoto, Alina Krischkowsky, Anna Maria Al Zubaidi-Polli, and Manfred Tscheligi. 2016. Collaborative appropriation: How couples, teams, groups and communities adapt and adopt technologies. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 473–480.

[132] Jinnatul Raihan Mumu, Paolo Saona, Md. Abdullah Al Mamun, and Md. Abul Kalam Azad. 2022. Is Trust Gender Biased? A Bibliometric Review of Trust in E-Commerce. *Journal of Internet Commerce* 21, 2 (2022), 217–245.

[133] Tom Nadarzynski, Oliver Miles, Aimee Cowie, and Damien Ridge. 2019. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital health* 5 (2019), 2055207619871808.

[134] Myura Nagendran, Yang Chen, Christopher A Lovejoy, Anthony C Gordon, Matthieu Komorowski, Hugh Harvey, Eric J Topol, John PA Ioannidis, Gary S Collins, and Mahiben Maruthappu. 2020. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *bmj* 368 (2020).

[135] Farzaneh Nasirian, Mohsen Ahmadian, and One-Ki Daniel Lee. 2017. AI-based voice assistant systems: Evaluating from the interaction and trust perspectives. (2017).

[136] Caroline A Nelson, Lourdes Maria Pérez-Chada, Andrew Creadore, Sara Jiayang Li, Kelly Lo, Priya Manjaly, Ashley Bahareh Pournamdari, Elizabeth Tkachenko, John S Barbieri, Justin M Ko, et al. 2020. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA dermatology* 156, 5 (2020), 501–512.

[137] Alicia L Nobles, Eric C Leas, Theodore L Caputi, Shu-Hong Zhu, Steffanie A Strathdee, and John W Ayers. 2020. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. *NPJ digital medicine* 3, 1 (2020), 1–3. https://doi.org/10.1038/s41746-019-0215-9

[138] Thomas P Novak and Donna L Hoffman. 2019. Relationship journeys in the internet of things: a new framework for understanding interactions between consumers and smart objects. *Journal of the Academy of Marketing Science* 47, 2 (2019), 216–237.

[139] Aisling A O'Kane, Helena M Mentis, and Eno Thereska. 2013. Non-static nature of patient consent: shifting privacy perspectives in health information sharing. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 553–562.

[140] Yfke P Ongena, Marieke Haan, Derya Yakar, and Thomas C Kwee. 2020. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. *European Radiology* 30, 2 (2020), 1033–1040.

[141] Shane O'Sullivan, Nathalie Nevejans, Colin Allen, Andrew Blyth, Simon Leonard, Ugo Pagallo, Katharina Holzinger, Andreas Holzinger, Mohammed Imran Sajid, and Hutan Ashrafian. 2019. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The international journal of medical robotics and computer assisted surgery* 15, 1 (2019), e1968.

[142] Leonard J Paas and Meike Morren. 2018. Please do not answer if you are reading this: Respondent attention in online panels. *Marketing Letters* 29, 1 (2018), 13–21.

[143] Raja Parasuraman and Christopher A Miller. 2004. Trust and etiquette in high-criticality automated systems. *Commun. ACM* 47, 4 (2004), 51–55.

[144] Sun Young Park, Yunan Chen, and Shriti Raj. 2017. Beyond health literacy: supporting patient-provider communication during an emergency visit. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2179–2192.

[145] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. 2019. Identifying challenges and opportunities in human-AI collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 506–510.

[146] Paul A Pavlou. 2003. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International journal of electronic commerce* 7, 3 (2003), 101–134.

[147] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* 46, 4 (2014), 1023–1031.

[148] Yuxin Peng, Pingping Yin, Zhaohua Deng, and Ruoxi Wang. 2020. Patient–Physician Interaction and Trust in Online Health Community: The role of perceived usefulness of health information and services. *International Journal of Environmental Research and Public Health* 17, 1 (2020), 139.

[149] Goda Perlaviciute and Linda Steg. 2014. Contextual and psychological factors shaping evaluations and acceptability of energy alternatives: integrated review and research agenda. *Renewable and Sustainable Energy Reviews* 35 (2014), 361–381.

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

[151] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[152] W Nicholson Price and I Glenn Cohen. 2019. Privacy in the age of medical big data. *Nature medicine* 25, 1 (2019), 37–43.

[153] Marianne Promberger and Jonathan Baron. 2006. Do patients trust computers? *Journal of Behavioral Decision Making* 19, 5 (2006), 455–468.

[154] Voicebot Research. 2022. *Healthcare Voice Assistant Consumer Adoption Report.* Technical Report. Voicebot Research. https://research.voicebot.ai/report-list/healthcare-voice-assistant-consumer-adoption-report-2022/

[155] Rebecca Schnall, Tracy Higgins, William Brown, Alex Carballo-Dieguez, and Suzanne Bakken. 2015. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. *Studies in health technology and informatics* 216 (2015), 467.

[156] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[157] William Seymour and Jose Such. 2023. Ignorance is Bliss? The Effect of Explanations on Perceptions of Voice Assistants. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24.

[158] William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–16.
[159] Emre Sezgin, Yungui Huang, Ujjwal Ramtekkar, and Simon Lin. 2020. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *NPJ Digital Medicine* 3, 1 (2020), 1–4.
[160] Emre Sezgin, Lisa K Militello, Yungui Huang, and Simon Lin. 2020. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. *Translational Behavioral Medicine* 10, 3 (2020), 606–628. https://doi.org/10.1093/tbm/ibz141
[161] Emre Sezgin, Garey Noritz, Alexander Elek, Kimberly Conkol, Steve Rust, Matthew Bailey, Robert Strouse, Aarti Chandawarkar, Victoria von Sadovszky, Simon Lin, et al. 2020. Capturing at-home health and care information for children with medical complexity using voice interactive technologies: multi-stakeholder viewpoint. *Journal of medical Internet research* 22, 2 (2020), e14202.
[162] Yeganeh Shahsavar, Avishek Choudhury, et al. 2023. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Human Factors* 10, 1 (2023), e47564.
[163] Varinder M Sharma and Andreas Klein. 2020. Consumer perceived value, involvement, trust, susceptibility to interpersonal influence, and intention to participate in online group buying. *Journal of Retailing and Consumer Services* 52 (2020), 101946. https://doi.org/10.1016/j.jretconser.2019.101946
[164] Faysal Hossain Shezan, Hang Hu, Gang Wang, and Yuan Tian. 2020. VerHealth: Vetting Medical Voice Applications through Policy Enforcement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 153 (dec 2020), 21 pages. https://doi.org/10.1145/3432233
[165] Osama Sohaib. 2021. Social networking services and social trust in social commerce: a PLS-SEM approach. *Journal of Global Information Management (JGIM)* 29, 2 (2021), 23–44.
[166] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication* 52, 2 (2005), 163–174.
[167] Jose Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3821–3832.
[168] Kil-Soo Suh, Hongki Kim, and Eung Kyo Suh. 2011. What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIs Quarterly* 35 (2011), 711–729. Issue 3. https://doi.org/10.2307/23042805
[169] Denok Sunarsi, Aris Baharuddin, Nur Afni, and Andi Anto Patak. 2019. The Effect of Service Quality and Price Accuracy on Consumer Confidence and Implications for Sales Increase. *PINISI Discretion Review* 3, 2 (2019), 101–110.
[170] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An integrative theory of intergroup conflict. *Organizational identity: A reader* 56, 65 (1979), 9780203505984–16.
[171] Tetsuya Tanioka, Tomoya Yokotani, Ryuichi Tanioka, Feni Betriana, Kazuyuki Matsumoto, Rozzano Locsin, Yueren Zhao, Kyoko Osaka, Misao Miyagawa, and Savina Schoenhofer. 2021. Development issues of healthcare robots: Compassionate communication for older adults with dementia. *International Journal of Environmental Research and Public Health* 18, 9 (2021), 4538.
[172] TechRepublic. 2020. Alexa Skills: Cheat Sheet. https://www.techrepublic.com/article/alexa-skills-cheat-sheet/
[173] R Thomson, DW Barclay, and CA Higgins. 1995. The partial least squares approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies: Special Issue on Research Methodology* 2, 2 (1995), 284–324.
[174] Eric J Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine* 25, 1 (2019), 44–56.
[175] Jenny Van Doorn, Martin Mende, Stephanie M Noble, John Hulland, Amy L Ostrom, Dhruv Grewal, and J Andrew Petersen. 2017. Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of service research* 20, 1 (2017), 43–58.
[176] Michelle ME van Pinxteren, Ruud WH Wetzels, Jessica Rüger, Mark Pluymaekers, and Martin Wetzels. 2019. Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing* (2019).
[177] Eleftheria Vaportzis, Maria Giatsi Clausen, and Alan J Gow. 2017. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. *Frontiers in psychology* 8 (2017), 1687.
[178] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.
[179] Viswanath Venkatesh and Fred D Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* 46, 2 (2000), 186–204.
[180] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* (2003), 425–478.
[181] Viswanath Venkatesh, James YL Thong, and Xin Xu. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly* (2012), 157–178.
[182] M Vimalkumar, Sujeet Kumar Sharma, Jang Bahadur Singh, and Yogesh K Dwivedi. 2021. 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior* 120 (2021), 106763. https://doi.org/10.1016/j.chb.2021.106763
[183] Sebastian Vollmer, Bilal A Mateen, Gergo Bohner, Franz J Király, Rayid Ghani, Pall Jonsson, Sarah Cumbers, Adrian Jonas, Katherine SL McAllister, Puja Myles, et al. 2018. Machine learning and AI research for patient benefit: 20 critical questions on transparency, replicability, ethics and effectiveness. *arXiv preprint arXiv:1812.10404* (2018).

[184] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–6.

[185] Weiyu Wang and Keng Siau. 2018. Living with Artificial Intelligence–Developing a Theory on Trust in Health Chatbots. In *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS*.

[186] Paul Ward. 2018. Trust and communication in a doctor-patient relationship: a literature review. *Arch Med* 3, 3 (2018), 36. https://doi.org/10.4172/2472-1654.100146

[187] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology* 52 (2014), 113–117.

[188] Maria Klara Wolters, Fiona Kelly, and Jonathan Kilgour. 2016. Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. *Health informatics journal* 22, 4 (2016), 854–866. https://doi.org/10.1177/1460458215593329

[189] Jen-Her Wu, Shu-Ching Wang, and Ho-Huang Tsai. 2010. Falling in love with online games: The uses and gratifications perspective. *Computers in Human Behavior* 26, 6 (2010), 1862–1871.

[190] Jie Xu, Kim Le, Annika Deitermann, and Enid Montague. 2014. How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology. *Applied ergonomics* 45, 6 (2014), 1495–1503.

[191] T Yan. 2008. Nondifferentiation. *Encyclopedia of survey research methods* 2 (2008), 520–521.

[192] Zheng Yan and Silke Holtmanns. 2008. Trust modeling and management: from social trust to digital trust. In *Computer security, privacy and politics: current issues, challenges and solutions*. IGI Global, 290–323.

[193] Samuel Yang, Jennifer Lee, Emre Sezgin, Jeffrey Bridge, and Simon Lin. 2021. Clinical Advice by Voice Assistants on Postpartum Depression: Cross-Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. *JMIR mHealth and uHealth* 9, 1 (2021), e24045. https://doi.org/10.2196/24045

[194] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. 2019. Privacy Perceptions and Designs of Bystanders in Smart Homes. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 59 (nov 2019), 24 pages. https://doi.org/10.1145/3359161

[195] Robert K Yin. 2013. Validity and generalization in future case study evaluations. *Evaluation* 19, 3 (2013), 321–332.

[196] Ryosuke Yokoi, Yoko Eguchi, Takanori Fujita, and Kazuya Nakayachi. 2021. Artificial intelligence is trusted less than a doctor in medical treatment decisions: Influence of perceived care and value similarity. *International Journal of Human–Computer Interaction* 37, 10 (2021), 981–990.

[197] Sangseok You and Lionel Robert. 2017. Emotional attachment, performance, and viability in teams collaborating with embodied physical action (EPA) robots. *You, S. and Robert, LP (2018). Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots, Journal of the Association for Information Systems* 19, 5 (2017), 377–407. https://doi.org/10.17705/1jais.00496

[198] Albert T Young, Dominic Amara, Abhishek Bhattacharya, and Maria L Wei. 2021. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *The Lancet Digital Health* 3, 9 (2021), e599–e611.

[199] Runting Zhong and Mengyao Ma. 2022. Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context. *BMC geriatrics* 22, 1 (2022), 1–13.

A QUESTIONNAIRE - PART ONE: SCENARIO DESCRIPTION

After the participant has chosen to consent to starting the questionnaire, a detailed scenario description is provided to help them understand the subject of the study (what we mean by HVAs):

'Imagine that you have an AI-based voice assistant at home which take care of your health, and you can seek help without going out to see a GP. You can issue commands directly to the voice assistant or have conservation with third-party skills that deployed on the voice assistants. It can take care of both your physical and mental health like a GP, such as: making a swift and effective diagnosis based on the symptoms that appear; assessing the likelihood of a certain illness; discussing and developing a treatment plan with you; helping you monitor and manage your chronic disease. Explain to you the test results such as blood test, x-ray diagnosis, etc. Combining with your own experience with voice assistants, please answer the following questions, we want your honest thoughts.'

B QUESTIONNAIRE - PART TWO: ITEM LIST

The second part of the survey is to let participants rate each of the items listed below using a 5-point likert scale that goes from from 'strongly disagree' (1) to 'strongly agree' (5).

Functional Factors

Anthropomorphism (A) - adapted from [150]:

A1. When I interacting with healthcare AI voice assistants, it would feel like someone is present in the room.

A2. I feel that interactions with healthcare AI voice assistants would be similar to those with a human.

A3. When communicating with healthcare AI voice assistants, I would feel like I am dealing with a real person.

A4. I think I would communicate with healthcare AI voice assistants in a similar way to how I would communicate with other people.

Effort Expectancy (EE) - adapted from [33, 37]:

EE1. I feel using healthcare AI voice assistants would take too much of my time.

EE2. I feel working with healthcare AI voice assistants would be difficult and I would not understand how to use them.

EE3. I feel it will take me too long to learn how to interact with healthcare AI voice assistants.

Perceived Usefulness (PU) - adapted from [39]:

PU1. I feel healthcare AI voice assistants would help me be more efficient when dealing with health-related issues.

PU2. I feel healthcare AI voice assistants would be useful when dealing with health-related issues.

PU3. I feel healthcare AI voice assistants could meets my needs.

PU4. I feel healthcare AI voice assistants would be able to do everything I would expect in order to take care of my health.

Perceived Content Credibility (PCC) - adapted from [39]:

Please indicate how well the following adjectives represent the healthcare AI voice assistant, from 1 = describes very poorly to 5 = describes very well:

PCC1. Accurate

PCC2. Complete

PCC3. Expert

PCC4. Consistent

PCC5. Authentic

Perceived Relative Service Quality (PRSQ) - adapted from [56]:

PRSQ1. I think the accuracy of information provided by a healthcare AI voice assistant would be higher than that of the average doctors.

PRSQ2. I think question answering and diagnosis speed of healthcare AI voice assistants would be faster than that of the average doctors.

PRSQ3. I think Healthcare AI voice assistants would provide a clearer and more understandable service delivery process.

Personal Factors

Stance in Technology (ST) - adapted from [37]:

ST1. I usually keep an eye on emerging products using AI technology, especially those that will be beneficial to my health.

ST2. I always try out emerging products using AI technology earlier compared to others especially if they will be beneficial to my health.

ST3. In general, I am willing to accept new emerging products using AI technology, especially if they will be beneficial to my health.

ST4. If I heard about an emerging product using AI technology, especially that is beneficial to my health, I would look for ways to use it.

Familiarity (F) - adapted from [37]:

F1. I know a lot about healthcare AI voice assistants.

F2. I have much knowledge about healthcare AI voice assistants.

F3. I am more familiar than the average person regarding healthcare AI voice assistants.

Technology Attachment (TA) - adapted from [37]:

TA1. I feel that AI technology is a part of my identity .

TA2. I identify strongly with the use of AI technology.

TA3. Using AI digital technology says a lot about who I am.

Social Influence (SI) - adapted from [9]:

SI1. Using healthcare AI voice assistants would be a status symbol in my social networks (friends, family, co-workers).

SI2. People in my social networks who would utilize healthcare AI voice assistants have more prestige than those who wouldn't.

SI3. People whose opinions I value would encourage me to utilize healthcare AI voice assistants.

SI4. People in my social networks who would utilize artificial intelligence such as healthcare AI voice assistants have a high profile.

Risk Factors

Security Risk (SR) - adapted from [39]:

SR1. I think it would be risky to interact with a Healthcare AI voice assistant.

SR2. I would be concerned if I had to deal with my health-related issues via a healthcare AI voice assistant.

SR3. There would be much uncertainty associated with my interactions with a healthcare AI voice assistant.

Privacy Risk (PR): - adapted from [39]:

PR1. I am concerned that a healthcare AI voice assistant would collect too much information about me.

PR2. I am concerned about who might access my personal information I had given to a healthcare AI voice assistant.

PR3. I feel that healthcare AI voice assistants would misuse my personal information.

PR4. There would be a large potential loss associated with providing personal information to healthcare AI voice assistant.

Perceived Substitution Risk (PSR) - adapted from [56]:

PSR1. I think that healthcare AI voice assistants are likely to replace doctors in the future.

PSR2. I think using healthcare AI voice assistants for a long time would make doctors dependent on them.

PSR3. I think the rise and development of healthcare AI voice assistants would likely lead to the unemployment of some doctors.

PSR4. I think using healthcare AI voice assistants for a long time would decrease doctors' own ability.

Trust and Intention to Use

Trust in HVAs (T) - adapted from [39]:

T1. I feel Healthcare AI voice assistants would be interested in my well-being.

T2. I feel like healthcare AI voice assistants would be truthful.

T3. I would characterize healthcare AI voice assistants as honest.

T4. I feel like a healthcare AI voice assistant would be sincere.

Intention to Use (IU) - adapted from [150]:

IT1. It is likely that I will use my voice assistant for Healthcare in the future.
IT2. I intend to use my voice assistant for healthcare frequently.
IT3. I expect to continue using my voice assistant for healthcare in the future.

C QUESTIONNAIRE - PART THREE: POST-SURVEY QUESTIONS

The last part of the survey contains both closed questions and open-ended questions that aims to understand more about how users perceive healthcare voice assistants. The responses to these questions were converted into a text format that could be analysed using thematic analysis methods.

Q1: Would you consider to use a healthcare voice assistant in the future?
- if No, then why you may NOT wish to use a healthcare AI assistant in the future?

Q2: I would trust the following organisations to make a healthcare voice assistant? not at all(1) - very much(5)
- The NHS (on another platform)
- Amazon
- Google
- Apple
- Please briefly describe any other organizations you trust to develop healthcare AI assistants (if any)

Q3: What aspects should be enhanced in terms of developing a trustworthy voice assistant in healthcare?

D THEMATIC ANALYSIS - THEMES AND CODES

Table 6 displays the themes and the corresponding codes investigated from participants' response to the post-survey questions.

Table 6. Themes and codes of thematic analysis.

Relation to the model	Themes	Codes
Retained	Security and Privacy	Reduce general privacy concern Data and usage transparency Ensure data storage security Obey regulations such as HIPAA Prevent data selling
Retained	Anthropomorphism	Provide natural conversation Empathy Compassionate Customize voice and care More human-like features
Retained	Other functional-related	Higher accuracy Higher usability Knowledgeable Timely Ease of use Help support avaliable Reliable information & advice
Excluded	Substitution Risk	Assist instead of replace Supervised by doctors Help contact doctors
Newly found	Financial Causes	Affordable to more people Special care to underprivileged Not designed for profit
Newly found	Discrimination Issues	Eliminate discrimination Treat patients equally
Newly found	Develop Organisation	NHS>Google≈Apple>Amazon Private tech companies Hospital & academic institutions

E ITEM LOADINGS

The final version of item loadings are shown in the Table 7. As can be seen from the table, the loadings are much larger than the preferred level of 0.7, which according to Hair et al. indicates that the construct explains more than 50 per cent of the item's variance, thus providing acceptable item reliability, except ST1 (0.479), SR2 (0.256), PR1 (0.410), A3 (0.569), PCC4 (0.357), PCC5 (0.365) and PSR2 (0.612), which were therefore removed and are not included in the table.

Table 7. Item Loading for Each Factor

Factor	Item	Loading	Factor	Item	Loading
Perceived Substitution Risk (PSR)	PSR1	0.906	Anthropomorphism (A)	A1	0.861
	PSR3	0.830		A2	0.872
	PSR4	0.817		A4	0.810
Perceived Usefulness (PU)	PU1	0.760	Social Influence (SI)	SI1	0.864
	PU2	0.760		SI2	0.713
	PU3	0.843		SI3	0.857
	PU4	0.760		SI4	0.885
Trust in HVA (T)	T1	0.787	Effort Expectancy (EE)	EE1	0.879
	T2	0.803		EE2	0.899
	T3	0.869		EE3	0.894
	T4	0.818	Perceived Relative Service Quality (PRSQ)	PRSQ1	0.871
Intention to Use (IU)	IU1	0.833		PRSQ2	0.845
	IU2	0.824		PRSQ3	0.822
	IU3	0.833	Security Risk (SR)	SR1	0.926
Technology Attachment (TA)	TA1	0.913		SR2	0.863
	TA2	0.909		SR3	0.914
	TA3	0.777	Privacy Risk (PR)	PR2	0.923
Stance in Technology (ST)	ST2	0.766		PR3	0.923
	ST3	0.889		PR4	0.907
	ST4	0.869	Perceived Content Credibility (PCC)	PCC1	0.838
Familiarity (F)	F1	0.828		PCC2	0.853
	F2	0.893		PCC3	0.772
	F3	0.869			

F POST-HOC ANALYSIS OF THE MODEL

As mentioned in Section 5.2.3, out of interest in whether the use of HVA and user-reported demographics also influenced trust in HVAs, we ran a second model which added the control variables to capture users' general demographics like *age* and *gender*, as well as whether they have *experience* in using voice assistant for healthcare purpose, and the voice assistant (*device type*) they normally use (e.g. Amazon Alexa, Google Assistant, etc.). Note that, from the demographics of the survey participants (see Table 2), 3 of the 300 participants reported that they were not willing to state their gender status, and in order to include the gender variable as a binary variable in the model, we evaluated the models after removing the data associated with these three participants. To summarise, the first model is the model *without* these three participants and *without* the control variables, and the second is the model *without* these three participants but *with* the control variables. The numbers in brackets in the table 8 are the results of the second model. Finally, the resulting model showed no significant differences in terms of the factors previously analyzed; those significantly and substantially influencing trust in HVA remained so, and those that were not significant remained so as well. In terms of the control variables, only ***gender*** was statistically significant, suggesting women to be more inclined to trust in HVAs (Sample Mean = 0.107, p <0.001).

Table 8. PLS-SEM Analysis Results: Hypotheses testing results (O: Original Sample Mean, M:Sample Mean, STDEV: Sandard Deviation). The numbers in brackets are results for models with control variables added.

	Hypotheses	O	M	STDEV	T statistics	p Value	Supported?
H1	Anthropomorphism ->Trust in HVA	0.076(0.073)	0.076(0.073)	0.038(0.036)	2.006(2.015)	0.045(0.044)	√(√)
H2	Effort Expectancy ->Trust in HVA	-0.059(-0.065)	-0.059(-0.065)	0.030(0.030)	1.977(2.147)	0.048(0.032)	√(√)
H3	Perceived Usefulness ->Trust in HVA	0.278(0.281)	0.278(0.282)	0.046(0.047)	6.011(5.956)	0.000(0.000)	√(√)
H4	Perceived Content Credibility ->Trust in HVA	0.159(0.156)	0.159(0.156)	0.034(0.035)	4.669(4.504)	0.000(0.000)	√(√)
H5	Perceived Relative Service Quality ->Trust in HVA	0.106(0.104)	0.105(0.103)	0.034(0.033)	3.125(3.108)	0.002(0.002)	√(√)
H6	Familiarity ->Trust in HVA	0.029(0.030)	0.029(0.031)	0.024(0.024)	1.2(1.215)	0.230(0.225)	x(x)
H7	Technology Attachment ->Trust in HVA	0.039(0.024)	0.041(0.025)	0.029(0.028)	1.376(0.864)	0.169(0.388)	x(x)
H8	Social Influence ->Trust in HVA	0.024(0.016)	0.025(0.015)	0.032(0.032)	0.761(0.505)	0.447(0.614)	x(x)
H9	Stance in Technology ->Trust in HVA	0.112(0.100)	0.11(0.099)	0.032(0.032)	3.549(3.082)	0.000(0.002)	√(√)
H10	Security Risk ->Trust in HVA	-0.138(-0.107)	-0.138(-0.108)	0.044(0.045)	3.104(2.399)	0.002(0.016)	√(√)
H11	Privacy Risk ->Trust in HVA	-0.175(-0.158)	-0.173(-0.156)	0.033(0.033)	5.224(4.840)	0.000(0.000)	√(√)
H12	Perceived Substitution Risk ->Trust in HVA	0.032(0.015)	0.03(0.015)	0.029(0.028)	1.087(0.516)	0.277(0.606)	x(x)
H13	Trust in HVA ->Intention to Use	0.738(0.738)	0.738(0.739)	0.031(0.030)	24.155(24.293)	0.000(0.000)	√(√)
	Gender ->Trust in HVA	(0.108)	(0.107)	(0.027)	(4.005)	**(0.000)**	**(√)**
	Age ->Trust in HVA	(0.009)	(0.009)	(0.023)	(0.375)	(0.708)	(x)
	Experience ->Trust in HVA	(0.002)	(0.003)	(0.026)	(0.093)	(0.926)	(x)
	Alexa ->Trust in HVA	(0.002)	(0.003)	(0.026)	(0.093)	(0.926)	(x)
	Google ->Trust in HVA	(0.026)	(0.052)	(7.595)	(0.003)	(0.997)	(x)
	Others ->Trust in HVA	(0.023)	(0.106)	(7.596)	(0.003)	(0.998)	(x)