Journal of the Academy of Marketing Science (2023) 51:843–866
https://doi.org/10.1007/s11747-022-00874-7

ORIGINAL EMPIRICAL RESEARCH

How artificiality and intelligence affect voice assistant evaluations

Abhijit Guha[1] · Timna Bressgott[2] · Dhruv Grewal[3,4] · Dominik Mahr[2] · Martin Wetzels[5] · Elisa Schweiger[6]

Received: 15 February 2021 / Accepted: 1 May 2022 / Published online: 28 June 2022

Abstract

Widespread, and growing, use of artificial intelligence (AI)–enabled voice assistants (VAs) creates a pressing need to understand what drives VA evaluations. This article proposes a new framework wherein perceptions of VA artificiality and VA intelligence are positioned as key drivers of VA evaluations. Building from work on signaling theory, AI, technology adoption, and voice technology, the authors conceptualize VA features as signals related to either artificiality or intelligence, which in turn affect VA evaluations. This study represents the first application of signaling theory when examining VA evaluations; also, it is the first work to position VA artificiality and intelligence (cf. other factors) as key drivers of VA evaluations. Further, the paper examines the role of several theory-driven and/ or practice-relevant moderators, relating to the effects of artificiality and intelligence on VA evaluations. The results of these investigations can help firms suitably design their VAs and suitably design segmentation strategies.

Keywords Voice assistants · Artificial intelligence · Signaling · Technology

The use of voice assistants (VAs), enabled by artificial intelligence (AI), is growing exponentially, greatly facilitated by their installations in various digital devices (e.g., Amazon Echo, Google Nest Hub, Apple Home Pod, most smartphones), as well as by their benefits relating to retailing and services. VAs are used globally, in a variety of domains ranging from homes to phones, to cars. Yet – as articles in the popular press indicate—consumers are not fully sold on VAs (e.g., Nguyen, 2021). So, it is important to identify and understand what factors drive consumers' VA evaluations, to advance both theory and practice.

We propose a model of VA evaluations, building from recent research on VAs (Table 1). However, this paper is also distinct from prior research, based on the points listed below. First, the primary mode of interaction between consumers and VAs is voice. In this sense, VAs are distinct from other technologies, such as websites or traditional apps, with which consumers interact primarily by clicking or typing. Second, building from prior work on VAs, we suggest that both social elements (e.g., perceived humanness) and functional elements (e.g., perceived usefulness) drive VA evaluations (Table 1; see Fernandes & Oliveira, 2021; McLean et al., 2021). Third, we draw from AI-related research, acknowledging explicitly that AI-enabled technology devices differ substantially from pre-AI versions. For example, while

Stephanie Noble & Martin Mende served as Guest Editors for this article.

✉ Elisa Schweiger
elisa.schweiger@kcl.ac.uk

Abhijit Guha
abhijit.guha@moore.sc.edu

Timna Bressgott
t.bressgott@maastrichtuniversity.nl

Dhruv Grewal
dgrewal@babson.edu

Dominik Mahr
d.mahr@maastrichtuniversity.nl

Martin Wetzels
martin.wetzels@edhec.edu

1 Department of Marketing, Darla Moore School of Business, University of South Carolina, Columbia, SC 29208, USA

2 Department of Marketing and Supply Chain Management, Maastricht University, Tongersestraat 53, 6211 LM Maastricht, The Netherlands

3 Department of Marketing, Babson College, Babson Park, MA 02457, USA

4 Fractional, University of Bath, Bath, UK

5 Department of Marketing, EDHEC Business School, 24 avenue Gustave Delory, CS 50411, 59057 Roubaix Cedex 1, France

6 Department of Marketing, King's College, Bush House, 30 Aldwych, London WC2B 4BG, UK

Table 1 Relevant studies relating to technology and voice assistants

(1) Source	(2) Drivers of VA evaluations Gap #1: Which VA features to focus on	(3) Process Mechanisms Gap #2: Process Mechanism is unclear	(4) Moderators Gap #3: Which Moderators to focus on	(5) Dependent Variables	(6) Insights
Dellaert et al. (2020)					Adopting AI-powered VA technologies (AIVAs), rather than engaging in traditional online purchase environments, may change consumer decision-making
Doyle et al. (2019)					Users regard the abilities of VAs as formal, fact-based, impersonal, and less authentic
Fernandes and Oliveira (2021)	Perceived ease of use, usefulness, humanness, social interactivity, social presence, trust, rapport		Preference for human interactions, experience	Intentions to use	Perceived social presence, perceived usefulness, trust, and rapport of the VA affect customers' desire to use VAs in the future. Both consumer experience with the digital VA and a preference for human interaction moderate these effects
Han and Yang (2018)	Task attraction, social attraction, physical attraction, security, privacy risk	Parasocial relationship		Satisfaction, continuance intention	Interpersonal attraction and security/privacy risks affect the adoption of VAs
Hernandez-Ortega and Ferreira (2021)	Experiences	Feeling of love	NA	Loyalty	Positive consumer experiences with AI VAs translate into feelings of love. These feelings of love in turn affect consumers' service loyalty
Mari and Algesheimer (2021)	Trust	Consideration size (number of options)	NA	Decision satisfaction	Trust in VA technologies increases consumers' satisfaction with their shopping decision. This effect is mediated by the consideration set size
McLean et al. (2021)	VA attributes (social presence, perceived intelligence, social attraction), technology attributes (perceived ease of use, perceived usefulness), situational attributes, (utilitarian beliefs, hedonic beliefs, distrust)		NA	Brand engagement	Social presence, perceived intelligence, and social attraction, as well as technology attributes and utilitarian benefits, positively influence consumer brand engagement. Trust concerns have a negative impact on customer brand engagement

Table 1 (continued)

(1) Source	(2) Drivers of VA evaluations Gap #1: Which VA features to focus on	(3) Process Mechanisms Gap #2: Process Mechanism is unclear	(4) Moderators Gap #3: Which Moderators to focus on	(5) Dependent Variables	(6) Insights
Moriuchi (2019)	Subjective norms	Ease of use, Perceived usefulness	Localization	Engagement, loyalty, attitudes	Subjective norms are antecedents of perceived ease of use and perceived usefulness (links to TAM), which positively influence engagement, attitudes, and loyalty. Localization moderates the relationship between perceived usefulness and engagement for non-transactional activities
Pitardi and Marriott (2021)	Perceived usefulness, perceived ease of use, enjoyment, social presence, social cognition, privacy	Attitude, trust	NA	Intentions to use	Perceived usefulness and perceived ease of use positively influence attitudes toward using VAs, whereas social presence and social cognition are antecedents of trust. Only attitude mediates the impact on intentions to use
Reis et al. (2017)					A comparison of multiple VAs reveals differences across functionalities
Schweitzer et al. (2019)	Relationship (role) with VA			Engagement, future usage intentions	Three primary relationship roles emerge from the interaction of voice-controlled smartphones with consumers (i.e., servant, partner, or master)
This article	**Artificiality-related (Natural speech, social cues, VA personality), Intelligence-related (accuracy, task range)**	**Perceptions of VA artificiality, perceptions of VA intelligence**	**Verbalizer cognitive style, perceived sacrifice, tech-savviness**	**VA evaluations (VA continued use intentions, VA purchase intentions)**	**Features that reduce perceptions of artificiality or increase perceptions of intelligence improve VA evaluations**

VA voice assistant, *AI* artificial intelligence, *NA* not applicable existing technology evaluation models predict preferences for increased perceived warmth and competence (e.g., Zhang & Zhang, 2021), consumers using AI-enabled devices start to feel unease when the perceived humanness or usefulness of the devices exceeds some threshold level (Kim et al., 2019; Liu et al., 2021). We thus propose that for VAs, which are inherently AI-enabled, perceived VA 'artificiality' and VA 'intelligence' (rather than warmth and competence) are better suited to be the mediator variables that mediate the effects of VA features on VA evaluations. Fourth, noting that product features that are both observable and controllable by the firm can function as signals (Spence, 1973), we propose that VA features represent signals of either VA artificiality or VA intelligence, which then drive VA evaluations. We know of no prior work that proposes the conceptualization of VA features as signals. The proposed model can guide firms in how to design VAs, as well as guide third-party providers in how to develop add-ins for existing VAs (e.g., new Alexa skills).

A review of recent VA and technology research (Table 1) highlights three important gaps that the current research is designed to address. First, a gap exists whereby prior research has not systematically examined, nor precisely determined, which specific VA features influence VA evaluations. In Table 1, column 2 (i) it is not fully clear why certain VA features – e.g., perceived ease of use—were selected, and (ii) the VA features selected e.g., perceived ease of use are more akin to perceived benefits, subjectively determined, and not a precisely defined VA feature. In Study 1, we text-mine more than 150,000 reviews of Amazon's Alexa device, seeking to systematically identify, and more precisely define, which features VA users emphasize when describing their VA usage experiences. These features then inform the key independent variables that we test in Study 2.

Second, a gap exists regarding the mechanisms via which VA features influence VA evaluations (Table 1, column 3). We propose a mechanism wherein VA features are signals of either perceptions of VA artificiality or VA intelligence, in turn, these perceptions of artificiality and intelligence affect VA evaluations. In so doing, we assert that VA intelligence provides a more overarching construct than ease of use or usefulness (Table 1; Moriuchi, 2019), and VA artificiality (or its lack) is a more comprehensive construct than feelings of love (Table 1; Hernandez-Ortega & Ferreira, 2021). Formally, we test whether perceptions of *reduced* artificiality and *increased* intelligence enhance VA evaluations.

Third, a gap exists regarding relevant moderators (Table 1, column 4). We draw from signaling theory to propose multiple moderators of the extent to which VA artificiality or VA intelligence influences evaluations; these moderators relate to receivers' differential motivation or ability to process signals.

Moving beyond qualitative investigations (Table 1; Doyle et al., 2019; Reis et al., 2017), we test the proposed model using a series of empirical examinations, using different methods, and present a host of robustness tests and post hoc tests. The results reveal the key mediating effects of VA artificiality and VA intelligence, with some evidence that their effects can backfire at extreme levels. This paper makes contributions to both theory and practice. To the best of our knowledge, we offer the first application of signaling theory to model VA evaluations. Conceptualizing VA features as signals, we demonstrate that these features affect perceptions of VA artificiality and VA intelligence, which then determine VA evaluations (e.g., continued use intentions). We also provide, to the best of our knowledge, the first effort to position perceived VA artificiality and VA intelligence as the central mediators in a model of VA evaluations. Factors such as artificiality and intelligence, relative to factors such as warmth or competence, are more tightly linked to VA features. Thus, researchers can use the proposed model to theorize how VA features—even those features yet to-be-introduced—affect VA evaluations. The signaling theory lens also suggests boundary conditions (i.e., moderators) that constrain the extent to which VA features affect consumers' evaluations; this article thus provides a guide to methods for identifying (other) moderators. Finally, building from the above, we derive a set of research questions to guide further research, pertaining to VAs, as well as (potentially) relating to other AI-enabled devices.

This paper also contributes to practice. By conceptualizing VA features as signals, firms can make more suitable design choices, depending on whether they seek to boost or suppress perceptions of artificiality or intelligence. Factors such as warmth and competence are relatively downstream from actual VA features, and thus offer weaker insights for how to design VA features (to increase evaluations). In contrast, VA artificiality and intelligence are – from a product design perspective—tightly linked to specific VA features, and thus allow for clearer guidance to VA designers about how best to design VA features. This paper also has implications for segmentation; drawing from signaling theory, it suggests consumer segments wherein certain VA features have more pronounced influences on VA evaluations.

Conceptual framework

In what follows, we develop a conceptual model of how consumers evaluate VAs. Prior work related to VAs offers various conceptualizations of how consumers evaluate them, including – for example—continued usage intentions (for existing VA users) and trial/purchase intentions (for potential VA users) (e.g., Fernandes & Oliveira, 2021; Pitardi & Marriott, 2021). Also, prior work has often suggested two primary drivers of VA evaluations. For example, Moriuchi (2019) proposes that ease of use and usefulness drive

VA evaluations, and Pitardi and Marriott (2021) propose that trust and attitude drive VA evaluations. Because VAs are powered by artificial intelligence (McLean & Osei-Frimpong, 2019), we propose two—novel—drivers of VA evaluations: perceived VA *artificiality* and perceived VA *intelligence*.

We conceptualize *VA artificiality* as the extent to which users perceive that VAs represent machines (i.e., are synthetic, not human). Consumers often interact with VAs as if they were interacting with other people (Pitardi & Marriott, 2021). Noting that the extent to which VA interactions mimic human–human interactions also positively influences downstream outcomes, such as rapport and usage intentions (Blut et al., 2021; Hassanein et al., 2009; Lu et al., 2016; McLean et al., 2020; Nasirian et al., 2017; Ye et al., 2019), we posit that diminished perceptions of VA artificiality (the extent to which the VA appears less synthetic, i.e., appears more human) are associated with more positive VA evaluations.

H1 A negative association exists between VA artificiality and VA evaluations.

We define *VA intelligence* according to how the VA responds to consumers' requests. Prior technology research (e.g., Blut et al., 2021) and studies of VAs (e.g., McLean et al., 2021) identify a positive association between enhanced perceptions of intelligence and VA evaluations. The VA trait of intelligence relates to users' perceptions of the VA's technology attributes and functional benefits, and we propose increased perceptions of VA intelligence are associated with better VA evaluations.

H2 A positive association exists between VA intelligence and VA evaluations.

We acknowledge that other papers have also posited that perceived intelligence has positive downstream consequences. However, there are some key differences which separate these papers from this paper. First, McLean et al. (2021) treat perceived intelligence as a triggering IV, not as a mediational pathway via which VA features impact continued use intentions. In contrast, we propose that VA features are the triggering IVs, and perceived intelligence is the mediating pathway via which VA features impact continued use intentions. Second, Blut et al. (2021) treat perceived intelligence as a mediational pathway, via which anthropomorphism impacts continued use intentions. In contrast, this paper does not link to anthropomorphism; instead, we propose that VA features are the triggering IVs.

As the key variables in our model, we propose that VA artificiality and VA intelligence mediate the impact of VA features on VA evaluations. In the following sections, we draw on signaling theory and (i) propose that various VA features signal about VA artificiality and VA intelligence (in turn impacting VA evaluations), and (ii) suggest factors that might moderate the impacts of VA artificiality and VA intelligence on VA evaluations.

Signaling theory: VA features as signals of artificiality and intelligence

By carefully designing or manipulating observable product features (Spence, 1973), firms (i.e., signal senders) can signal their type to consumers (i.e., signal receivers), and thus potentially influence their evaluations. For example, by offering a low-price guarantee, a retailer can signal to consumers that the prices it offers are low (relative to competitors or the market) (Biswas et al., 2006). We know of no prior studies that apply signaling theory to the context of VAs. However, signaling theory has been applied to other technology domains (e.g., Guo et al., 2020), so we adopt it as the underlying framework in this paper.

We define VA features, such as naturalness of speech or task range, as those features that are both observable (by consumers) and manipulatable (by firms), such that they qualify as signals. Then we can classify VA features as signals of either VA artificiality or VA intelligence. We propose that if the firm suitably manipulates VA features, then in turn it should affect consumers' perceptions of VA artificiality and VA intelligence, which in turn should affect VA evaluations. In Study 1, involving the text mining of more than 150,000 product reviews, we identify VA features that consumers cited most often in their evaluations of VAs: natural speech, social cues, task range, and accuracy. Here, we offer predictions about how these features may influence perceptions of artificiality and intelligence.

Natural speech Natural sounding speech may lead the listener to conclude that the speech is from a human. Thus, firms design VAs to have relatively natural sounding voices, seeking to enhance evaluations (e.g., "voice assistants, which designers aspired to make as 'natural' as possible, at first in their default middle-class female voice … produce voice and personae that would appear naturally… speech takes on features of natural conversation, such as pauses.."; Humphry & Chesher, 2021, pgs. 1975, 1979). Thus, we propose that the extent of naturalness in VAs' speech is a signal that reduces consumers' perceptions of VA artificiality.

H3 A negative association exists between the naturalness of VA speech and perceived VA artificiality.

Social cues Social cues, such as indicators that the VA has a specific gender or age, influence the extent to which users perceive the VA as humanlike and encourage them to refer to the VA using personalized words like "Alexa" or "she," rather than "it." Computational linguistics research into Reddit posts reveals that more than 70% of posts referred to Amazon's Alexa VA as "she," as did more than 80% of posts pertaining to Apple's Siri (Abercrombie et al., 2021). Even in the popular press, authors such as Ramos (2021) recognize that "many people refer to Siri, Alexa, and Cortana as 'she' and not 'it'." Consumers tend to experience feelings of connectedness toward non-human agents that are perceived as relatively human (Fernandes & Oliveira, 2021; van Pinxteren et al., 2019). We hence propose that social cues are signals that reduce consumers' perceptions of VA artificiality.

H4 A negative association exists between social cues and perceived VA artificiality.

Task range The number of tasks a VA can execute is its task range, which provides information about the usefulness of the VA. Such functional (cf. social) benefits influence VA evaluations (e.g., continued usage intentions) (Fernandes & Oliveira, 2021; also see McLean et al., 2021, who propose that evaluations may be higher if VAs can help execute multiple tasks). Therefore, we propose that an enhanced VA task range serves as a signal that prompts enhanced perceptions of VA intelligence.

H5 A positive association exists between VA task range and perceived VA intelligence.

VA accuracy Accuracy reflects how faithfully the VA executes the user's commands. Pitardi and Marriott (2021) propose and show that if a VA responds suitably, and if interactions with the VA go smoothly, it increases VA evaluations (e.g. continued use intentions). This research provides a basis for us to propose that enhanced VA accuracy offers a signal that increases perceptions of VA intelligence.

H6 A positive association exists between VA accuracy and perceived VA intelligence.

Moderating factors

In this section, relying on signaling theory, we propose and examine the role of several moderating factors that may affect the influence of VA artificiality and VA intelligence on VA evaluations.

Verbalizers Signal receivers must be able to understand the intended message of the signal, before any benefits can be realized (Kimery & McCord, 2008). Because VAs primarily communicate using voice, VA signals primarily are communicated by voice. Noting that verbalizers (cf. visualizers) are better able to comprehend text and voice cues (Koć-Januchta et al., 2017; Richardson, 1977), we predict that VA signals may be better comprehended by verbalizers. It follows that VA signals may have stronger impacts on consumers with a relative verbalizer cognitive style.

H7a The negative association between VA artificiality and VA evaluations is stronger among VA users who are relative verbalizers.

H7b The positive association between VA intelligence and VA evaluations is stronger among VA users who are relative verbalizers.

Tech-savviness Experts use cues differently than novices (Spence & Brucks, 1997; Wagner et al., 2001). That is, expert consumers are less likely to use signals to inform their evaluations (Biswas & Sherrell, 1993; Mattila & Wirtz, 2001), in domains like pricing (Gerstner, 1985; Rao & Monroe, 1988), and – importantly—in technology-related decision contexts (Park & Kim, 2008). More specifically, the role of such signals is less pronounced for consumers with higher levels of knowledge and/or expertise (Grewal & Compeau, 2007).

Those consumers with strong technological capabilities, who we refer to as being relatively tech-savvy (i.e., early adopters of new technologies, able to master new technologies quickly), may be able to judge how capable and intelligent a VA is on the basis of their existing technical knowledge, and so we propose that those who are tech-savvy are less likely to be swayed by signals.

H8a The negative association between VA artificiality and VA evaluations is weaker among VA users who are relatively tech-savvy.

H8b The positive association between VA intelligence and VA evaluations is weaker among VA users who are relatively tech-savvy.

Perceived sacrifice Some may argue that signals of technical capability exert stronger impacts on people who are more engaged (Wang et al., 2019). In general, VA users who score higher on perceived sacrifice—because they invest (or sacrifice) more resources (money, time, effort) to use VAs—should be relatively more engaged. In turn, these engaged VA users might react more strongly to signals of VA intelligence. This proposition is consistent with the theme in Cable and Turban (2001, pg. 145), who indicate that the impact of signals may be stronger for those who are relatively more vested in a product. Therefore, we predict:

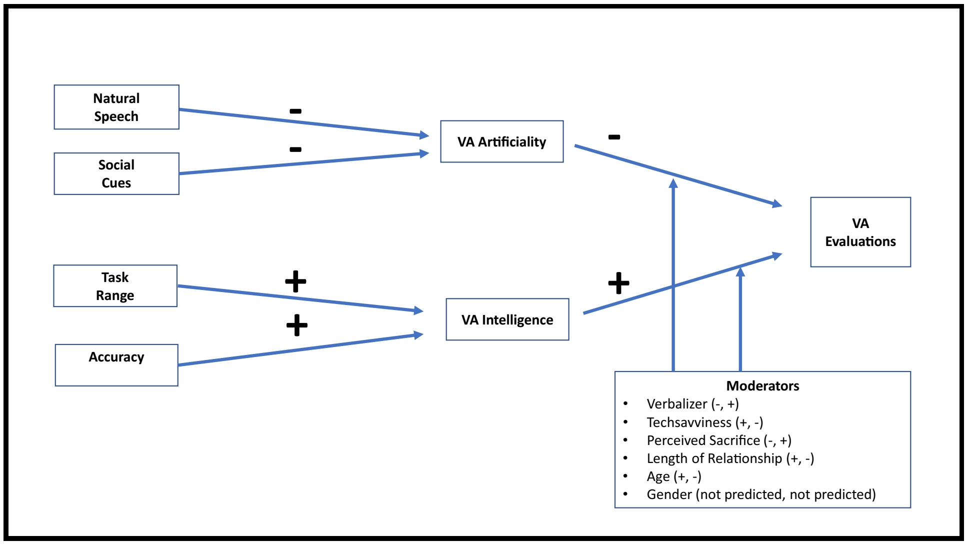


Fig. 1 Conceptual model

H9a The negative association between VA artificiality and VA evaluations is stronger among VA users who score higher on perceived sacrifice.

H9b The positive association between VA intelligence and VA evaluations is stronger among VA users who score higher on perceived sacrifice.

Length of ownership Prior work suggests that consumers make decisions in different ways, as the duration of ownership increases (Gounaris & Venetis, 2002). Specifically, the impact of signals and cues on judgments reduce as consumers' experience increases (Jin & Park, 2006). This point suggests that the impact of signals should reduce as the length of ownership increases.

H10a The negative association between VA artificiality and VA evaluations is weaker among VA users who have owned VAs for a longer period.

H10b The positive association between VA intelligence and VA evaluations is weaker among VA users who have owned VAs for a longer period.

Age Bennett and Hill (2012; pg. 202) suggest that younger consumers are "more naïve", due to "inexperience navigating the marketplace". In turn, such naivete may lead younger (vs. older) consumers to be relatively more impacted by signals like perceived VA artificiality and perceived VA intelligence.

H11a The negative association between VA artificiality and VA evaluations is weaker among older VA users.

H11b The positive association between VA intelligence and VA evaluations is weaker among older VA users.

Gender On the one hand, several industry reports have highlighted gender gaps in technology usage and the development of AI technology (e.g., OECD 2018). Therefore, using reasoning analogous to that used when making the tech-savviness prediction, male consumers may well be relatively less impacted by signals. On the other hand, in certain domains men do make relatively more use of technological cues (Devlin & Bernstein, 1995). Given this, we do not make any explicit hypotheses for gender effects.

The full model, including the predictions related to VA features, mediators, moderators, and VA evaluations, is shown in Fig. 1.

Study 1: Topic extraction of relevant VA features

To find and establish the VA features that initiate the process predicted by our conceptual model (i.e., features that signal VA artificiality and VA intelligence), we text-mined more than 150,000 consumer reviews posted on Amazon. Specifically, we applied an unsupervised Latent Dirichlet Allocation (LDA) to identify relationships among consumer reviews. With this probabilistic modeling approach, we classified sets of words according to unobserved groups (Blei et al., 2003; Milne et al., 2020), similar to LDA applications that extract topics from product reviews (Tirunillai & Tellis, 2014), reviews of tourist attractions (Taecharungroj & Mathayomchan, 2019), or loan requests (Netzer et al., 2019).

Data and data preparation

Using the web scraping tool WebHarvey, we collected consumer reviews relating to second-generation versions of the Amazon Echo and Amazon Echo Dot, from amazon.com, posted between October 24, 2016 (first review date), and December 10, 2018 (download date). In total, we collected 31,870 reviews for the Echo and 119,805 reviews for the Echo Dot, for a total of 151,675 reviews. We used Knime Analytics's (4.4.2) embedded text processing tools (Tursi & Silipo, 2018), in line with recommendations from Ordenes and Silipo (2021). Similar to previous analyses (Milne et al., 2020; Villarroel Ordenes et al., 2019), we removed stop words (e.g., "on," "and," "is"), reduced terms to their stems by applying the Kuhlen stemming algorithm (e.g., "buy" for "buying" and "buys"), erased punctuation, removed words with fewer than four characters, filtered out numerical terms (both numbers and terms that represent them), and converted all letters to lower case. Next, we excluded words that appeared fewer than five times in the text, to avoid any skewed effects due to outliers. Words with very high occurrences also can skew the analyses, so we removed context-specific, non-descriptive words, such as "Echo," "Amazon," and "generation." Then we identified the most frequent bigrams, such as "voice command" and "smart home," and included them in our analyses. From each review, which we define as separate documents, we extracted all remaining single words and bigrams with a bag-of-words approach, which presents these single words and bigrams as vectors, without preserving the word order. Finally, we excluded any vectors with fewer than 10 words (Ordenes & Silipo, 2021).

Identifying specific VA features with topic modeling

To extract topics from the documents (i.e., reviews), we implemented LDA using a simple parallel-threaded LDA algorithm based on the Mallet package (Tursi & Silipo, 2018). This algorithm implemented a sparse LDA sampling scheme, more efficient than traditional LDA approaches (Yao et al., 2009). As a generative model, this algorithm assumes that (i) words are interchangeable (bag of words), (ii) the order of documents is interchangeable, (iii) documents can belong to multiple topics (soft clustering), (iv) words can belong to multiple topics, and (v) every topic is represented by a multinomial distribution over a fixed word vocabulary (Tursi & Silipo, 2018).

Because LDA requires that the number of topics be specified before the analysis, we use a statistical approach to select the number of topics (Berger et al., 2020), across multiple iterations (Blei et al., 2003; Netzer et al., 2019). We varied the number of topics from 1 to 30. We kept the β prior parameter, which defines the prior weight of a word in a topic, constant ($\beta = 0.01$; Milne et al., 2020; Ordenes & Silipo, 2021). Then we included the number of topics (k) and an alpha parameter, equivalent to the prior weight of a topic in the document (Tursi & Silipo, 2018), as dynamic variables that can be imputed in each loop iteration (Griffiths & Steyvers, 2004; Ordenes & Silipo, 2021). The initial alpha was set to 50/k. To calculate the log-likelihood and perplexity (i.e., how well the model fits) of the entire data set, we determined the optimum number of topics, according to the elbow area, such that we found a range around which perplexity started to decrease. As Web Appendix A shows, the ideal number of latent topics is approximately seven.

These seven topics were natural speech, task range, social cues, accuracy, connection, smart home, and speakers. Figure 2 shows a bar chart for each of the seven topics, displaying the five most relevant terms for each topic, and the representativeness of each word for the topic. Words can be representative of more than one topic (i.e., "Alexa" appears in both Accuracy and Natural Speech topics). We ranked the terms from most to least representative, with the size of the horizontal bar indicating this representativeness. Larger horizontal bars indicate that the term is more frequently presented in reviews on that topic. For example, "connect" is highly representative of the connectivity topic, and we should expect to find that term fairly often in reviews related to the topic of connectivity. Returning to our overarching framework (Fig. 1), this topic mining analysis helped us focus our examination, because the first four topics reflect VA features that are likely to influence perceptions of VA artificiality and VA intelligence. The latter three topics relate to the VAs' hardware and set-up (i.e., connectivity, smart home, and speakers) and are less likely to influence perceptions of the devices' artificiality or intelligence; therefore, we do not study these further.

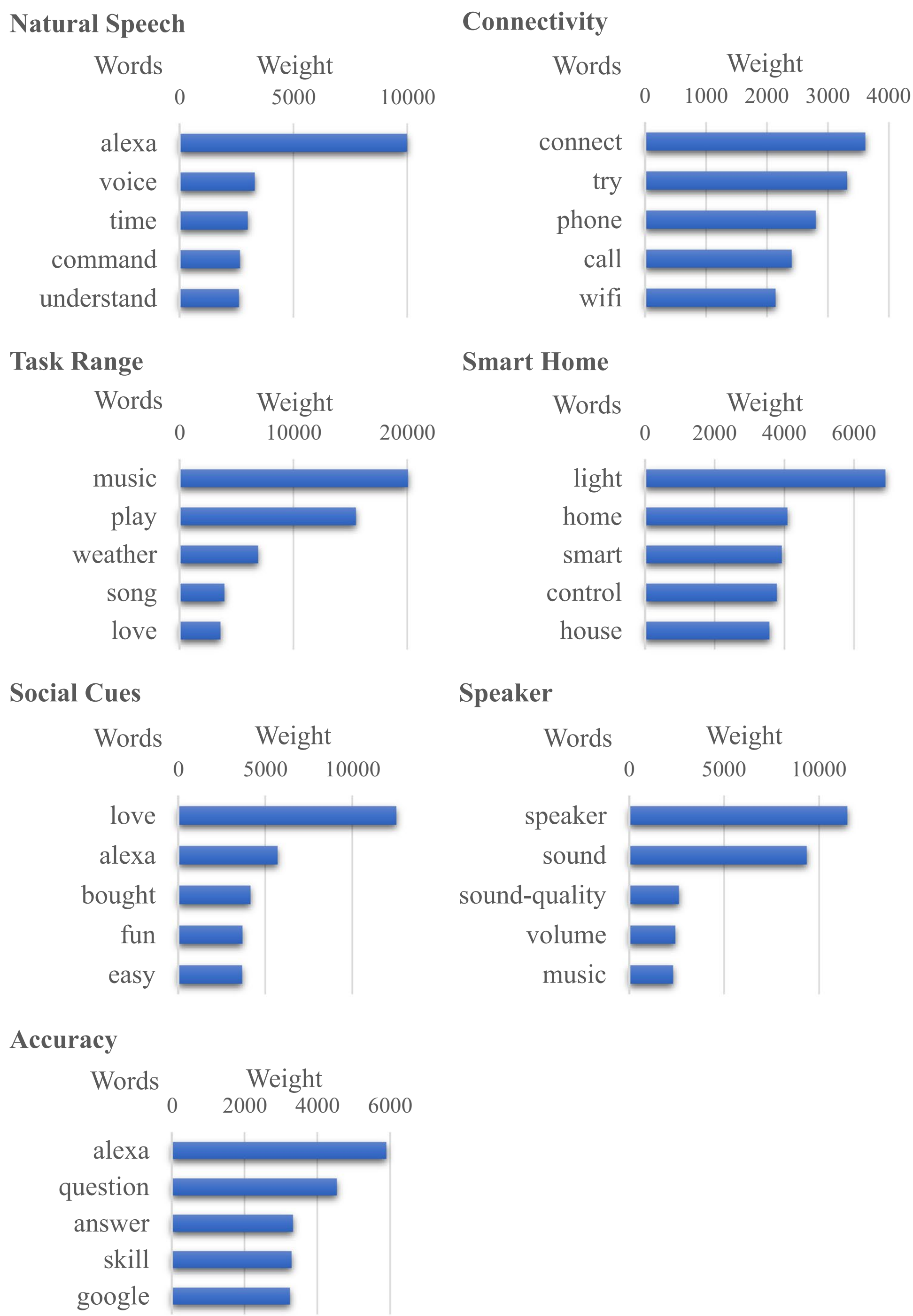
The four VA features (above) are fairly specific, linking tightly to actual VA attributes. For example, features like naturalness of speech and task range have rather specific definitions that a VA designer can execute on, whereas (say) ease of use (see Table 1) is relatively less specific (and is more of a benefit, and less of a feature), and so harder for a VA designer to execute on. We return to this point in the discussion section.

Study 2: Testing the model

Sample and data collection

We test the proposed model (Fig. 1) among respondents from the online Prolific platform. First, using a screening study, we identified participants likely to have an Amazon Echo (or related) device, by asking 3,000 respondents if they owned an Amazon Echo device, for how long, and if they would be willing to participate in a follow-up study.

Fig. 2 Study 1—Representativeness of terms for each topic



We identified 1,386 participants who indicated they had owned an Echo device for at least a month (see McLean et al., 2021), which helped ensure they had some experience with the VA (Pitardi & Marriott, 2021). Of these 1,386 respondents, 815 participated in the follow-up study, but 82 gave incorrect responses to the attention check question (i.e., indicated the number of cars in a picture incorrectly), leaving 733 respondents (48.8% women; $M_{Age} = 33.98$ years, $SD = 10.97$) whose responses we analyzed. In terms of experience with Amazon Echo, 8.3% of respondents had owned the device for 1–3 months, 11.7% for 3–6 months, 16.4% for 6–12 months, and the rest for more than a year. Furthermore, 48.2% respondents owned an Amazon Echo, 63.6% an Amazon Echo Dot, 13.8% an Amazon Echo Show, 1.4% an Amazon Echo Flex, and 2.2% a different Amazon Echo device (e.g., Amazon Echo Auto). The total exceeds 100%, because some participants owned more than one device.

Measurement scales

We used a mix of single-item and multi-item scales (see Appendix 1), adapted from extant literature. As a proxy for VA evaluations, we used continued usage intentions (Pitardi & Marriott, 2021). The mediator variables were perceptions of VA artificiality and VA intelligence; the independent variables were natural speech, social cues, task range, and accuracy. The moderator variables related to users' verbalizer cognitive style, tech-savviness, perceived sacrifice, and three demographic variables (age, length of ownership, and gender[1]). In addition, we controlled for ease of use,[2] as also elicited a latent factor to test for common method bias (CMB).

Testing for common method bias

Noting that elicited the independent and dependent variables from the same respondents, we report tests for CMB (Kock et al., 2021; Podsakoff et al., 2003). Upfront though, we note that we attempted to limit the CMB risk procedurally, by ensuring the clarity and conciseness of the items, as well as guaranteeing the anonymity of all responses (Podsakoff et al., 2012; Viswanathan and Kayande, 2012). Furthermore, the formats we used to elicit responses varied (e.g., we used a matrix table, graphic sliders, and a numerical slider), and we spatially separated the independent and dependent variable measures in the survey (Kock et al., 2021; Podsakoff et al., 2003).

We applied several statistical tests (Podsakoff et al., 2003). First, we applied Harman's single-factor test, using principal axis factoring (PAF), applied to the multi-item scales in our study. The extracted variance was just 27%, suggesting the relative absence of CMB (Fuller et al., 2016). Second, using an unmeasured latent variable approach (Schmid-Leiman transformation), on the basis of PAF, we include a method factor (Podsakoff et al., 2003; Yung et al., 1999). It accounts for only 29% of the variance, below the 50% threshold for CMB (Fuller et al., 2016). Furthermore, based on Liang et al. (2007), we applied the unmeasured latent method construct approach to our PLS path model to assess the potential threat of common method bias. This analysis reveals that the estimates in the structural model were not substantially affected by the inclusion of the unmeasured latent method factor. We also conducted a latent marker construct approach (cf., Richardson et al., 2009; Rönkkö & Ylitalo, 2011). Our analysis shows that the structural estimates were not impacted substantially by incorporating the latent marker construct (Web Appendix B). Finally, we calculated variance inflation factors (VIFs), generated by a full collinearity test (Kock, 2015). A model is relatively free of CMB if all factor-level VIFs are 3.3 or less; for our latent variables, inner VIFs ranged from 1.18 for verbalizer to 1.76 for ease of use, suggesting the absence of CMB (Kock, 2015).

Results

We used partial least squares structural equation modeling (PLS-SEM) to estimate the parameters in our measurement model (outer model) and structural model (inner model; Hair et al., 2017a, b). This composite-based SEM approach has limited distributional needs regarding the manifest variables and minimal computational requirements for the underlying algorithm (Hair et al., 2020; Rigdon, 2012; van Pinxteren et al., 2019). The use of PLS-SEM is appropriate, because it supports the prediction and explanation of endogenous variables in a theoretically grounded structural model (Hair et al., 2017a, b; Sarstedt et al., 2014). In addition, PLS-SEM allows for complex models, in terms of the number of variables and relationships, as well as modeling flexibility, limited requirements for the distributional assumptions of the variables and sample size, and convergence and stability of the results (Sarstedt et al., 2014). In turn, PLS-SEM generally achieves high levels of statistical power for hypotheses testing. We also note that PLS-SEM has been used in prior studies of VAs (e.g., Fernandes & Oliveira, 2021).

Measurement model We used the statistical package SmartPLS 3.3 for PLS-SEM, employing 10,000 bootstrap resamples to obtain robust standard errors and t-statistics for the parameters in our model (Hair et al., 2017a, b). Measurement models do not apply to single-item constructs, so we excluded natural speech, task range, accuracy, age, length of ownership, and gender measures from our reliability and validity assessments (Hair et al., 2017a, b). For internal reliability, we considered composite reliability, which takes the outer loadings of the indicator variables into account, and Cronbach's alpha, which is a more conservative measure (Hair et al., 2017a, b). Both measures indicated good internal reliability, and the values for all the multi-item constructs exceeded 0.7 (Hair et al., 2017a, b; Hulland, 1999). As evidence of convergent validity, the average variance extracted (AVE) values were all greater than 0.5 (Hair et al., 2017a, b) (see Table 2). To test for discriminant validity, we calculated the square root of the

[1] Note that we have no explicit prediction for the effect of gender.

[2] In Web Appendix C, we present results wherein we also controlled for anthropomorphism, normal speech rate, human presence, and social presence.

Table 2 Study 2—Measurement model of the main effect model

Factor	Item	Loading	AVE	Cronbach's α	CR
Continued Use (CU)	CU1	0.93	0.85	0.91	0.95
	CU2	0.90			
	CU3	0.95			
Artificiality (A)	A1	0.84	0.80	0.87	0.92
	A2	0.92			
	A3	0.92			
Social Cues (SC)	SC1	0.81	0.78	0.73	0.87
	SC2	0.95			
Intelligence (I)	I1	0.85	0.71	0.90	0.93
	I2	0.88			
	I3	0.85			
	I4	0.80			
	I5	0.84			
Perceived Sacrifice (PS)	PS1	0.70	0.73	0.82	0.89
	PS2	0.92			
	PS3	0.92			
Tech-savviness (TS)	TS1	0.82	0.69	0.79	0.87
	TS2	0.87			
	TS3	0.81			
Verbalizer	V1	0.98	0.94	0.93	0.97
	V2	0.96			
Ease of Use (EoU)	EoU1	0.87	0.78	0.90	0.93
	EoU2	0.86			
	EoU3	0.90			
	EoU4	0.90			

AVE average variance extracted, *CR* composite reliability. Natural speech, task range and accuracy are excluded (single-item scales)

AVE; none of these values exceeded the correlations between latent variables (Fornell & Larcker, 1981) (see Table 3). Because the Fornell–Larcker criterion arguably might not indicate discriminant validity accurately (Henseler et al., 2015), we also checked the heterotrait-monotrait (HTMT) ratio of correlations, that is, the ratio of within-trait to between-trait correlations, to identify true correlations among constructs. The HTMT values ranged between 0.006 and 0.756, below the conservative threshold of 0.85 (Hair et al., 2017a, b). Thus, the HTMT analysis corroborated AVE findings; the data set has adequate discriminant validity.

Structural models Model 1 examined the main effects of all variables (including main effects of moderator variables) without including interaction effects. To evaluate the structural model (Hair et al., 2017a, b), we relied on path coefficients and significance values (Hair et al., 2017a, b) (Table 4). We also calculated R^2 values for the latent constructs—VA artificiality (0.49), VA intelligence (0.26), and continued usage intentions (0.26)—which are medium to high (Chin, 1998; Hair et al., 2017a, b). The Stone-Geisser's Q^2 value indicated predictive relevance (Geisser, 1975); the calculated values, between 0.21 and 0.39, indicated medium to large out-of-sample predictive power, validating the predictive accuracy of our study (Chin, 2010; Hair et al., 2017a, b).

The structural model results are consistent with our conceptual model (see the path specifications in Table 4). More natural speech, and stronger social cues diminished perceptions of VA artificiality, whereas wider task range and increased accuracy enhanced perceptions of VA intelligence. In turn, stronger VA intelligence perceptions were *positively* associated with greater continued use intentions, but stronger VA artificiality perceptions were *negatively* associated with these intentions. Testing the indirect effects (Table 5), we found that more natural speech, stronger social cues, a wider task range, and greater accuracy exhibited positive indirect links to

Table 3 Study 2—Correlations and square root of the AVE

	Continued Use	Artificial	Natural Speech	Social Cues	Intelligence	Task Range	Accuracy	Perceived Sacrifice	TS	Verbalizer	Ease of Use	age	gender	Length of ownership
Continued Use	***0.92***													
Artificial	-0.28	***0.89***												
Natural Speech	0.25	-0.55	***NA***											
Social Cues	0.27	-0.65	0.49	***0.88***										
Intelligence	0.35	-0.48	0.38	0.36	***0.84***									
Task Range	0.27	-0.37	0.35	0.38	0.50	***NA***								
Accuracy	0.27	-0.37	0.40	0.32	0.54	0.45	***NA***							
Perceived Sacrifice	-0.14	-0.13	0.08	0.16	-0.04	0.01	-0.07	***0.86***						
Techsavviness (TS)	0.26	-0.24	0.21	0.19	0.24	0.17	0.24	-0.03	***0.83***					
Verbalizer	0.10	-0.20	0.16	0.21	0.13	0.12	0.15	0.25	0.19	***0.97***				
Ease of Use	0.40	-0.23	0.23	0.22	0.45	0.31	0.43	-0.35	0.41	0.12	***0.88***			
age	0.22	-0.10	0.19	0.12	0.03	0.07	0.07	-0.09	0.02	0.06	0.04	***NA***		
gender	0.06	0.04	-0.04	0.04	0.00	-0.01	-0.05	-0.08	-0.26	-0.05	0.04	0.04	***NA***	
length ownership	0.07	0.13	-0.05	-0.11	-0.09	-0.05	-0.08	-0.29	0.01	-0.18	0.06	0.15	0.10	***NA***

Values along the diagonal in boldface-italics represent the square roots of the AVE; *single-item scale or NA

continued use intentions, through the mediating pathways of VA artificiality perceptions and VA intelligence perceptions, in line with our conceptual model (Fig. 2). The main effects of the moderator variables indicate that more tech-savvy people, older people, and females showed higher VA continued use intentions; other moderators did not have significant main effects. With regards to our control variable, as may be expected, those who perceived greater ease of use expressed higher VA continued use intentions.

Next, we tested if perceived sacrifice, tech-savviness, verbalizer, age, length of ownership, and gender, moderate the influence of artificiality or intelligence on continued use. We test each moderator pair in a separate model (Table 4). We found no moderating effects for verbalizer cognitive style, tech-savviness, or gender. However, levels of perceived sacrifice (path coefficient = +0.11, see Table 4) and length of ownership (path coefficient = -0.10, see Table 4) moderated the impact of perceptions of intelligence on continued usage intentions, consistent with H_{9B} and H_{10B}. Finally, we found that the negative effects of artificiality and the positive effects of intelligence (path coefficients 0.10 and -0.09 respectively, see Table 4) were weaker amongst older consumers (H_{11A} and H_{11B}).

We produced Johnson-Neyman plots (Fig. 3), using latent variable scores from SmartPLS 3.3 (Hair et al., 2017a, b). Specifically, we used the R package interactions (Long, 2019), with a bootstrapped variance–covariance matrix of estimates (10,000 samples) obtained from the R package sandwich (Zeileis, 2004; Zeileis et al., 2020). The positive association between perceptions of VA intelligence and continued usage intentions was significant only when perceived sacrifice levels were greater than 2.04 (Plot A). The positive association between perceptions of VA intelligence and continued usage intentions was significant only when participants owned their device for less than 4.92 years (plot B). Finally, the positive (negative) associations between perceptions of VA intelligence (artificiality) and continued usage intentions was significant amongst those younger than 41.10 (42.13) years (plots C and D).

Robustness of results Other factors may influence VA evaluations, as suggested by prior research (Table 1). Therefore, we checked whether our findings sustained when we controlled for some of these alternative influences. Specifically, we tested four new models, in which we control for the impacts of anthropomorphism, normal speech rate, human presence, and social presence (see item details in Web Appendix C). These four variables

Table 4 Study 2—Structural models

Paths specified / Path Coefficients	1. Main Effect Model	2. Perceived Sacrifice Moderator	3. Tech-savviness Moderator	4. Verbalizer Moderator	5. Age Moderator	6. Length of Ownership Moderator	7. Gender Moderator
Model relationships							
Artificiality ➔ Continued Use	-.12**	-.12**	-.14***	-.13***	-.12**	-.12***	-.12**
Natural Speech ➔ Artificiality	-.30***	-.30***	-.30***	-.30***	-.30***	-.30***	-.30***
Social Cues ➔ Artificiality	-.50***	-.50***	-.50***	-.50***	-.50***	-.50***	-.50***
Intelligence ➔ Continued Use	.16***	.16***	.15**	.15***	.16***	.16***	.16***
Task Range ➔ Intelligence	.32***	.32***	.32***	.32***	.32***	.32***	.32***
Accuracy ➔ Intelligence	.40***	.40***	.40***	.40***	.40***	.40***	.40***
Moderator Variables							
Perceived Sacrifice ➔ Continued Use	-.03 ns	-.04 ns	-.03 ns	-.02 ns	-.02 ns	-.03 ns	-.03 ns
Tech-savviness ➔ Continued Use	.11**	.11**	.09*	.11**	.11**	.09*	.11**
Verbalizer ➔ Continued Use	.01 ns	.01 ns	.02 ns	.01 ns	.01 ns	.00 ns	.01 ns
Age ➔ Continued Use	.18***	.18***	.18***	.18***	.17***	.17***	.18***
Length of Ownership ➔ Continued Use	.04 ns	.04 ns	.04 ns	.04 ns	.05 ns	.05 ns	.04 ns
Gender ➔ Continued Use	.07*	.07*	.06 ns	.07 ns	.08*	.07*	.07*
Moderator X Artificiality ➔ Continued Use		.03 ns	.06 ns	.02 ns	.10**	-.04 ns	.02 ns
Moderator X Intelligence ➔ Continued Use		.11**	-.05 ns	-.04 ns	-.09*	-.10**	.01 ns
Control relationships							
Ease of Use ➔ Continued Use	.23***	.23***	.23***	.23***	.23***	.23***	.23***
Latent variable	R^2/ Q^2	R^2/ Q^2	R^2/ Q^2	R^2/ Q^2	R^2/ Q^2	R^2/ Q^2	R^2/ Q^2
Artificial	.49 / .39	.49 / .39	.49 / .39	.49 / .39	.49 / .39	.49 / .39	.49 / .39
Intelligence	.26 / .26	.37 / .26	.37 / .26	.37 / .26	.37 / .26	.37 / .26	.37 / .26
Continued Use	.26 / .21	.27 / .22	.27 / .22	.26 / .21	.28 / .23	.27 / .22	.26 / .21

***$p < .001$, **$p < .01$, *$p < .05$, ns. non-significant

Table 5 Study 2—Indirect effects

Routes	1. Main Effect Model	2. Perceived Sacrifice Moderators	3. Tech-savviness Moderators	4. Verbalizer Moderators	5. Age Moderators	6. Length of Ownership Moderators	7. Gender Moderators
Natural Speech ➔ Artificiality ➔ Continued Use	.04**	.04**	.04***	.04**	.04***	.04**	.04**
Social Cues ➔ Artificiality ➔ Continued Use	.06***	.07***	.07***	.06**	.07***	.06**	.06**
Task Range ➔ Intelligence ➔ Continued Use	.05**	.05**	.05**	.05**	.05**	.05**	.05**
Accuracy ➔ Intelligence ➔ Continued Use	.06***	.06***	.06**	.06***	.06***	.06***	.06***

***$p < .001$, **$p < .01$, *$p < .05$, ns. non-significant

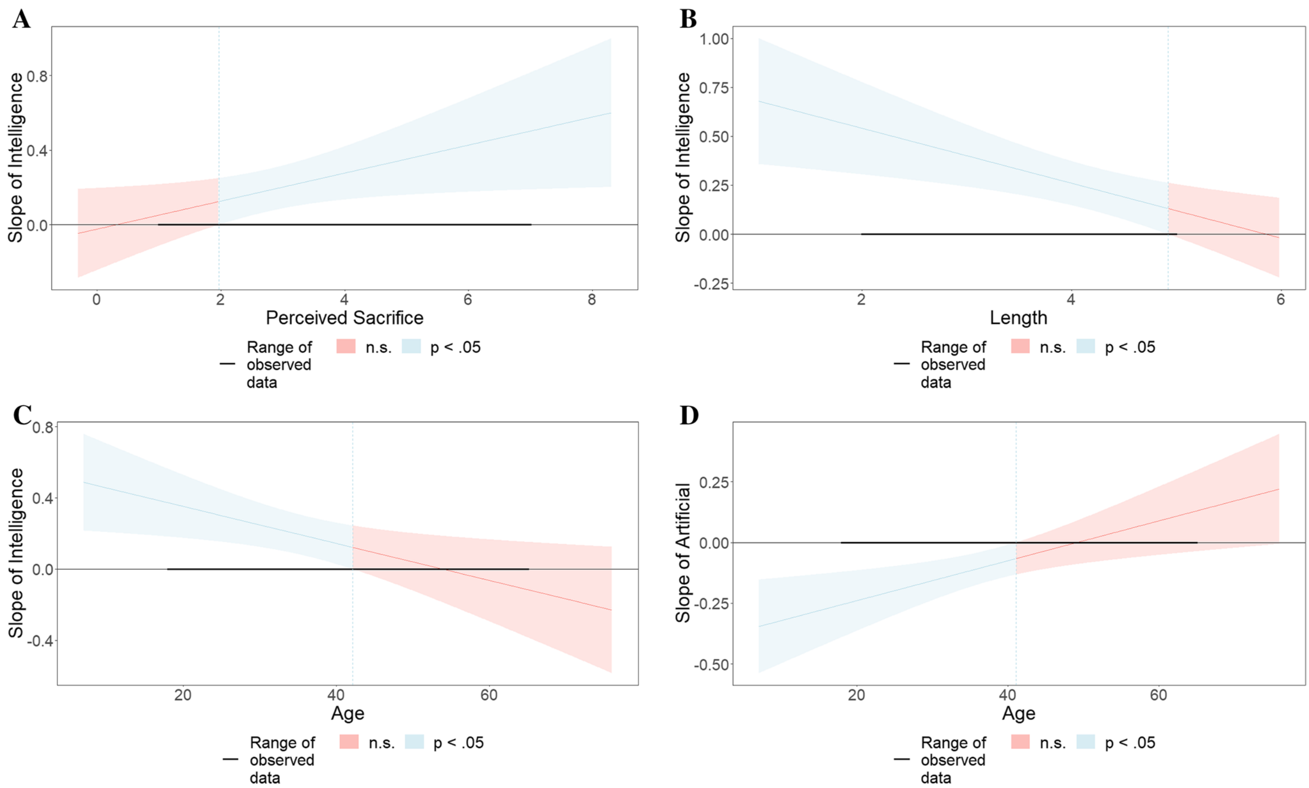


Fig. 3 Moderation plots—Study 2. **A** Intelligence x Perceived Sacrifice. When PS values ≥ 2.04,, the effects of VA Intelligence on VA Continued Use Intentions is significant ($p < .05$). **B** Intelligence x Length. When Length of relationship values ≤ 4.92, the effects of VA Intelligence on VA Continued Use Intentions is significant ($p < .05$). **C** Intelligence x Age. When age values ≤ 42.13, the effects of VA Intelligence on VA Continued Use Intentions is significant ($p < .05$). **D** Artificial x Age. When age values ≤ 41.10, the effects of VA Artificiality on VA Continued Use Intentions is significant ($p < .05$)

might affect VA evaluations, according to prior literature (Table 1). We found similar indirect and direct effects, even when controlling for these variables; none of these four variables had significant effects on continued usage intentions (see Web Appendix C).

Nonlinear effects By way of an exploratory analysis, we explored possible quadratic effects of VA artificiality and VA intelligence, as well as the artificiality × intelligence interaction. In the main effect model, we ran separate tests for the quadratic effects of artificiality, the quadratic effects of intelligence, both quadratic effects together, and the artificiality × intelligence effect (see Table 6).

We noted nonlinear effects. First, the quadratic effect of artificiality on continued usage intentions was negative and significant (path coefficient = -0.06), suggesting boundaries to the general negative association between artificiality and VA evaluations. Second, the quadratic effect of intelligence was not significant. Third, the artificiality × intelligence interaction was significantly positive (path coefficient = + 0.08); higher levels of intelligence were associated with increased VA evaluations, but this effect was not significant when VA artificiality levels dropped below 3.97 (see floodlight analysis—per Spiller et al., 2013 – in Fig. 4). We discuss these quadratic and interaction effects in the discussion section.

Table 7 contains an overview of the hypotheses and results.

Study 3: Re-testing the role of VA artificiality and intelligence

Study 3 has two objectives. First, we test for the effects of VA artificiality and VA intelligence on VA evaluations, formally controlling for factors like warmth and competence. Second, prior examinations have primarily involved those who own a specific VA; would the effects

Table 6 Study 2—Non-linear effects

Paths specified / Path Coefficients	Main Effect Model	Artificial Quadratic Effects	Intelligence Quadratic Effects	A & I Quadratic Effects	A x I Interaction
Model relationships					
Artificiality ➔ Continued Use	-.12**	-.15**	-.12***	-.14***	-.15**
Natural Speech ➔ Artificiality	-.30***	-.30***	-.30***	-.30***	-.30***
Social Cues ➔ Artificiality	-.50***	-.50***	-.50***	-.50***	-.50***
Intelligence ➔ Continued Use	.16***	.15***	.15***	.16***	.12***
Task Range ➔ Intelligence	.32***	.32***	.32***	.32***	.32***
Accuracy ➔ Intelligence	.40***	.40***	.40***	.40***	.40***
Control Relationships					
Perceived Sacrifice ➔ Continued Use	-.03 ns	-.02 ns	-.02 ns	-.02 ns	-.02 ns
Tech-savviness ➔ Continued Use	.11**	.10**	.11**	.10**	.10**
Verbalizer ➔ Continued Use	.01 ns	.02 ns	.01 ns	.02 ns	.01 ns
Age ➔ Continued Use	.18***	.18***	.18***	.18***	.18***
Length of Ownership ➔ Continued Use	.04 ns	.05 ns	.04 ns	.05 ns	.05 ns
Gender ➔ Continued Use	.07*	.06	.07 ns	.06 ns	.06 ns
Ease of Use ➔ Continued Use	.23***	.24***	.24***	.24***	.24***
Robustness Checks					
Artificiality Quadratic Effect ➔ Continued Use		-.06*		-.06*	
Intelligence Quadratic Effect ➔ Continued Use			-.01 ns	.00 ns	
Artificiality x Intelligence ➔ Continued Use					.08*
Indirect Effects					
Natural Speech ➔ Artificiality ➔ Continued Use	.04**	.04**	.04**	.04**	.05***
Accuracy ➔ Intelligence ➔ Continued Use	.06***	.06***	.06**	.06***	.05**
Task Range ➔ Intelligence ➔ Continued Use	.05**	.05**	.05**	.05**	.04**
Social Cues ➔ Artificiality ➔ Continued Use	.06***	.07**	.06***	.06***	.08***

***$p < .001$, **$p < .01$, *$p < .05$, ns. non-significant

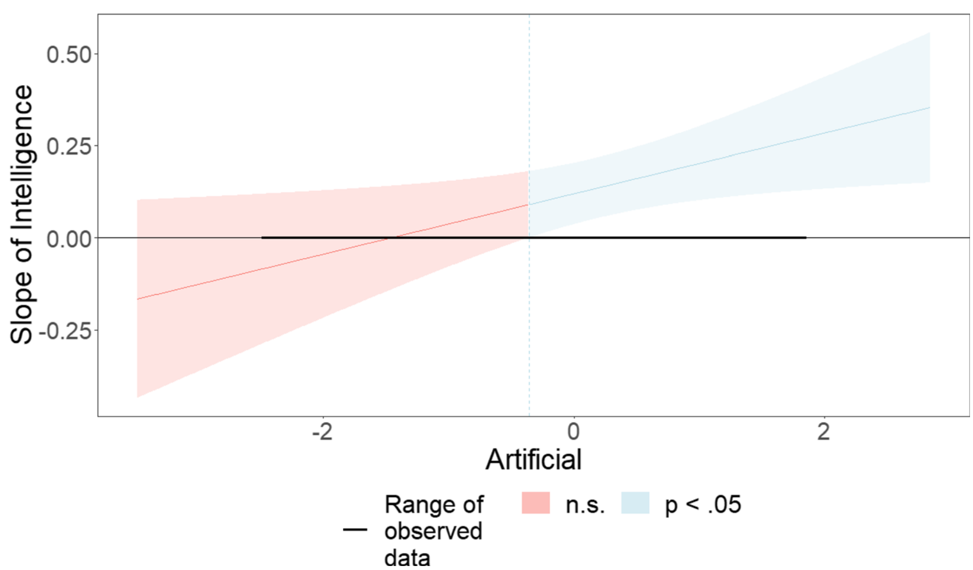


Fig. 4 Moderation Plots – Study 2. Intelligence x Artificial. When VA Artificiality values ≥ 3.97, the effects of VA Intelligence on VA Continued Use Intentions is significant ($p < .05$)

Table 7 Summary of results

Hypothesis		Support
1	A negative association exists between VA artificiality and VA evaluations	Supported (B = -.12**) Table 4, Model 1
2	A positive association exists between VA intelligence and VA evaluations	Supported (B = .16***) Table 4, Model 1
3	A negative association exists between naturalness of VA speech and perceived VA artificiality	Supported (B = -.30***) Table 4, Model 1
4	A negative association exists between social cues and perceived VA artificiality	Supported (B = -.50***) Table 4, Model 1
5	A positive association exists between VA task range and perceived VA intelligence	Supported (B = .32***) Table 4, Model 1
6	A positive association exists between VA accuracy and perceived VA intelligence	Supported (B = .40***) Table 4, Model 1
7a	The negative association between VA artificiality and VA evaluations is stronger among VA users who are relative verbalizers	non-significant Table 4, Model 4
7b	The positive association between VA intelligence and VA evaluations is stronger among VA users who are relative verbalizers	non-significant Table 4, Model 4
8a	The negative association between VA artificiality and VA evaluations is weaker among VA users who are relatively tech-savvy	non-significant Table 4, Model 3
8b	The positive association between VA intelligence and VA evaluations is weaker among VA users who are relative tech-savvy	non-significant Table 4, Model 3
9a	The negative association between VA artificiality and VA evaluations is stronger among VA users who score higher on perceived sacrifice	non-significant Table 4, Model 2
9b	The positive association between VA intelligence and VA evaluations is stronger among VA users who score higher on perceived sacrifice	Supported (B = .11**) Table 4, Model 2
10a	The negative association between VA artificiality and VA evaluations is weaker among VA users who have owned VAs for a longer period	non-significant Table 4, Model 6
10b	The positive association between VA intelligence and VA evaluations is weaker among VA users who have owned VAs for a longer period	Supported (B = -.10**) Table 4, Model 6
11a	The negative association between VA artificiality and VA evaluations is weaker among older VA users	Supported (B = .10**) Table 4, Model 5
11b	The positive association between VA intelligence and VA evaluations is weaker among older VA users	Supported (B = -.09**) Table 4, Model 5

***$p < .001$, **$p < .01$, *$p < .05$, ns. non-significant

sustain amongst those who do not own the VA (but, may be exposed to information about the VA)?

Participants ($N = 496$; 30.6% women; $M_{Age} = 32.44$ years, $SD = 11.51$; recruited from Prolific) indicated whether they owned an Amazon Echo device. If participants did not own an Amazon Echo ($n = 182$), then they viewed a 60-s clip of people interacting with Alexa. Thereafter participants completed the (i) artificiality and intelligence items (as in Study 2), (ii) a two-item warmth scale and a two-item competence scale, and (iii) a three-item purchase intention scale pertaining to the Amazon Echo. Warmth and competence items were answered on a seven-point Likert scale, from 'not at all descriptive' to 'very descriptive' (Aaker et al., 2012). Re. warmth, participants were asked to indicate the extent to which they found the terms "warm" and "friendly" described the VA. Re. competence, participants were asked to indicate the extent to which they found the terms "competent" and "capable" described the VA (Aaker et al., 2012). Finally, re. purchase intentions, participants responded on a seven-point Likert scale, from 'very low' to 'very high,' and responded to the following three questions (a) The likelihood that I would buy an Amazon Echo is…, (b) The probability that I would consider buying an Amazon Echo is … and (c) My willingness to buy an Amazon Echo is….

For participants who already owned an Amazon Echo ($n = 314$), we asked them to think about their own Echo device when responding to the artificiality and intelligence items (from Study 2), the warmth and competence items (described above), and the continued usage intentions scale (from Study 2). In addition, in all cases ($N = 496$), we elicited demographics, and a latent factor to test for CMB.

Measurement models

We used a latent marker construct approach (cf., Richardson et al., 2009; Rönkkö & Ylitalo, 2011) on

Table 8 Study 3—Measurement model

Factor	Item	Loading	AVE	Cronbach's α	CR
Artificiality (A)	A1	0.972	0.700	0.873	0.873
	A2	0.711			
	A3	0.806			
Intelligence (I)	I1	0.912	0.674	0.911	0.911
	I2	0.869			
	I3	0.749			
	I4	0.807			
	I5	0.755			
Warmth	W1	0.844	0.775	0.871	0.873
	W2	0.914			
Competence	C1	0.943	0.875	0.933	0.933
	C2	0.928			

AVE average variance extracted, *CR* composite reliability

Table 9 Study 3—Descriptive statistics, correlations, and square root of the AVE

	Mean (SD)	1	2	3	4
1. Artificiality	4.65 (1.35)	*0.836*			
2. Intelligence	5.41 (0.98)	-0.385	*0.821*		
3. Warmth	4.33 (1.36)	-0.489	0.564	*0.88*	
4. Competence	5.31 (1.19)	-0.282	0.795	0.559	*0.936*

Values along the diagonal in italics represent the square roots of the AVE

the (a) purchase intention model and (b) continued use model, employing PLS, to assess the potential threat of CMB. Our analysis (Tables 8, 9 and 10) shows that neither the structural estimates for continued use nor purchase attention were substantially impacted by incorporating the latent marker construct.

With a confirmatory factor analysis, we evaluated the measurement model in terms of its ability to represent the artificiality, intelligence, warmth, and competence of the participants ($N = 496$), according to its internal reliability, discriminant validity, and convergent validity (Hair et al., 2020), using PLS-SEM with 10,000 bootstrapping samples. The composite reliability and Cronbach's alpha values were greater than 0.87 for all measures, and the outer loadings for the multi-item constructs exceeded 0.71, indicating good internal reliability (Hair et al., 2017a, b) (Tables 8, 9 and 10). The AVEs were above 0.7 for all measures, which indicated good convergent validity (Hair et al., 2017a, b) (Tables 8, 9 and 10). To check for discriminant validity, we used both the Fornell–Larcker criterion and the HTMT ratio, with values between 0.28 and 0.80 (Hair et al., 2017a, b). Thus, we affirmed the internal reliability, discriminant validity, and convergent validity of the four measures. In summary, the VA artificiality and VA intelligence measures

Table 10 Study 3—Structural models, with and without latent markers

Paths specified	Standardized coefficients without Latent Marker Variable	t-value bootstrap	Standardized coefficients with Latent Marker Variable
(a) Not owning an Alexa Echo ($N = 182$)			
Artificiality ➔ Purchase Intention	-0.240	2.841**	-0.248
Intelligence ➔ Purchase Intention	0.232	2.449*	0.196
Warmth ➔ Purchase Intention	0.178	1.860 ns	0.159
Competence ➔ Purchase Intention	0.052	0.543 ns	0.046
(b) Owning an Alexa Echo ($N = 314$)			
Artificiality ➔ Continued Use	0.030	0.621 ns	0.031
Intelligence ➔ Continued Use	0.486	6.908***	0.487
Warmth ➔ Continued Use	0.182	2.950**	0.182
Competence ➔ Continued Use	0.019	0.267 ns	0.019
Latent variable	R^2	Q^2	
(a) Purchase Intention	0.295	0.262	
(b) Continued Use	0.356	0.312	

***$p < .001$,**$p < .01$, *$p < .05$, ns. non-significant

Italics: control variables

were separate, and – in our data—distinct from warmth and competence.

Structural model

Among participants who did not own an Echo device, when we controlled for warmth and competence perceptions, VA artificiality (path coefficient = -0.24, $p < 0.01$) and VA intelligence (path coefficient = + 0.23, $p < 0.05$) were significant predictors of VA purchase intentions (Tables 8, 9 and 10). For participants who already owned an Echo device, VA intelligence (path coefficient = + 0.49, $p < 0.01$) was a significant predictor of continued usage intentions, whereas VA artificiality was not (Tables 8, 9 and 10). Overall, the path coefficient signs were consistent with those from Study 2. Further, the effects (largely) sustained irrespective of whether the respondents formally owned a VA, or were merely exposed to information about a VA.

Discussion

The purpose of this study was to show that (i) in our data, VA artificiality, VA intelligence, warmth and competence are all sufficiently differentiated, and (ii) the effects of VA artificiality and VA intelligence sustain, despite controlling for warmth and competence. The latter point is especially important, noting that – in our data – there is strong correlation between artificiality-warmth and intelligence-competence (Table 9). This probably reflects that today the tasks VAs execute are simple enough such that VA intelligence maps onto VA competence. But in the future tasks that VAs execute, the linkage may be less strong. This point is best illustrated using the example of Stitch Fix (see below).

Stitch Fix uses AI to process consumers' requests, and suitably suggest clothing options. In many cases, the AI does a fine job. But in some cases, the AI does not, and it is for these reasons that Stitch Fix also has a human stylist overview/ modify the suggestions made by the AI (Davenport, 2021). One consumer had requested something to wear at "at a wedding where my ex will also be at". The AI suggested standard clothes one wears at a wedding, whereas the human stylist understood the subtext of the consumer's request and suggested clothes of a different type. In this example, the AI's response was "intelligent" but perhaps not "competent". In a world wherein VAs are asked to take on more complex roles, one may well imagine a consumer making similar requests of a VA (as above) and getting responses that are "intelligent" but not "competent".

General discussion

To develop a model for VA evaluations (Fig. 1), we build on extant theory pertaining to VAs, AI, technology adoption, and signaling. We conceptualize VA features as signals of VA artificiality or VA intelligence, which in turn affect VA evaluations, and these effects are moderated by various signal receiver characteristics. Study 1, based on text-mining of more than 150,000 consumer reviews, identified key VA features that may function as signals. Study 2 (together with post hoc analyses and tests for nonlinear effects) affirmed the validity of the proposed model. In particular, more natural speech and more social cues signal a lower level of VA artificiality, whereas a wider VA task range and greater VA accuracy signal greater VA intelligence. In turn, lower levels of perceived VA artificiality and higher levels of perceived VA intelligence increase consumers' VA evaluations, with these effects moderated by users' perceived sacrifice, length of VA ownership and age (nuanced details presented in Table 4). We confirmed that VA artificiality and VA intelligence are distinct from perceived warmth and competence, and that the effects of VA artificiality and VA intelligence on VA evaluations sustained even when we control for these alternative, potentially influential factors. Also, the effects sustained irrespective of whether respondents formally owned the VA or were merely exposed to information about the VA (Study 3). Finally, we offered initial evidence that there exist quadratic and interaction effects (Table 6).

Contributions

We propose various theoretical contributions pertaining to the important and substantial domain of VAs. VA use is substantial and growing, likely to increase even more as VAs are used more widely in other domains, such as cars. To the best of our knowledge, this article represents the first use of signaling theory in a model of VA evaluations, and the first use of signaling theory to work on AI. To the extent that VA features may be conceptualized as signals, we reveal a specific pathway by which VA features affect consumers' evaluations (e.g., continued use intentions), through their impacts on VA artificiality and VA intelligence. This application of signaling theory also suggests

Table 11 Future research agenda

Research Question	Subtopics and Questions
1	Do other mediators (beyond artificiality and intelligence) affect VA evaluations? Can VA features impact both perceptions of artificiality and intelligence? Are artificiality and intelligence orthogonal and independent? If not, how do they relate, and what form does their interdependence take?
2	Do reduced artificiality perceptions become eerie, and so stop enhancing VA evaluations when they reach a very low point, in line with the uncanny valley effect (Mori, 1970)? Because VAs lack embodiment, the eeriness threshold might differ, but the phenomenon appears similar. Research should investigate explicitly whether and how these effects arise for VAs In what conditions do effects identified in research into robots apply to VAs?
3	Can VA features backfire, such as by undermining consumers' sense of autonomy, with negative implications for evaluations (André et al., 2018; Grewal et al., 2021), such that consumers may select nonpreferred options to reaffirm their autonomy (Davenport et al., 2020)? Which trait and state factors determine whether (and how much) consumers value autonomy in VA settings? For example, consumers' culture and perceptions of VAs as servants versus partners may be relevant moderators; perceived autonomy may be more important in individualistic cultures or when users perceive VAs as servants
4	Are there other moderators that are specific to voice technology?
5	Do the findings apply to VAs other than Amazon's Echo (Alexa), which is the focus of this study? Across platforms that host VAs, such as cellphones, tablets, personal computers, and cars, does it matter whether the platform host matches the developer? Are VA evaluations contingent on the type of VA, the platform, or their match?
6	How do new add-ins (e.g., Alexa skills) introduced by third-party developers affect VA evaluations? Does integration with third-party platforms affect VA evaluations?

suitable moderators (boundary conditions) that influence the extent to which VA features affect consumers' evaluations. As technology advances, and new or improved models of VAs emerge, and new use cases emerge, researchers can apply the lens of signaling theory to identify and examine other moderators that may be relevant to VA evaluations.

Also, to the best of our knowledge, this study is the first to position VA artificiality and VA intelligence as mediators in a model of VA evaluations, rather than alternatives like (i) warmth and competence, and (ii) love, ease of use and usefulness, from Table 1. In a conceptual sense, artificiality and intelligence are more closely linked to VA features, than the alternative mediators listed above. Here again, as technology advances and new VA features emerge, researchers can use these theoretical lenses to elaborate on how VA features (including yet-to-be introduced features) affect VA evaluations. We also find and emphasize that the benefits of artificiality and intelligence for VA evaluations may not be wholly linear. Prior models involving mediators mostly consider linear effects of mediators, so establishing the role of nonlinearities represents a theoretical contribution.

Beyond contributions to theory, we propose contributions to practice, related to the design of VAs and consumer segmentation efforts. By conceptualizing features as signals, we propose that firms should make design choices and trade-offs according to whether they seek to boost or suppress perceptions of artificiality or intelligence. For example, a firm that serves expert investors might want to signal intelligence, so any feature enhancements it develops (e.g., add-ins, Alexa skills) should prioritize that signal. A firm serving novice investors instead might attempt to manipulate (more specifically, reduce) perceptions of artificiality. Such design trade-offs arguably can make the VA more appealing while also reducing the firm's overall development costs.

Furthermore, firms can a priori predict the impact of introducing or modifying specific VA features on subsequent evaluations. First, the VA features we identify are fairly specific, linking tightly to actual VA attributes. Thus, natural speech and task range map back to actual VA attributes, unlike factors like ease of use etc. (see Table 1), which are more akin to benefits than features. Second, unlike variables such as warmth and competence (Belanche et al., 2021; van Doorn et al., 2017), which are relatively distant from actual product features and offer fewer insights for how to design VA features to increase evaluations, VA artificiality and VA intelligence are closely linked to specific VA features and thus reveal more obvious pathways to VA evaluations. Also, VA product designers may be better able to suitably design VA features if asked to design features that affect perceptions of VA artificiality and VA evaluations (than if asked to design features that affect perceptions of warmth and competence).

Finally, we note implications for segmentation. For example, VA features that signal intelligence have stronger impacts among relatively new VA owners, VA owners who have incurred much sacrifice (in terms of money, time etc.) and younger consumers. Also, VA features that signal artificiality have stronger impacts among younger consumers.

Limitations and further research

We relied on surveys to test our conceptual model and find consistent results; these results showed good external validity. It would be helpful to complement these studies with experimental studies that manipulate levels of VA artificiality and VA intelligence and test the same conceptual model. Other limitations relate to (i) the relatively high correlation between intelligence and competence, although we note that that the two measures are conceptually distinct, and that the effects of VA intelligence sustain even when controlling for competence, and (ii) use of single-item scales for some constructs.

We proposed that VA features effect VA evaluations via perceptions of artificiality and intelligence. Further research might seek mediators we may have missed. Furthermore, we treat artificiality and intelligence as orthogonal variables, which is not accurate (see nonlinear effects, Study 2 – the benefits of perceived intelligence increase as perceptions of perceived artificiality increase – values $\geq$ 3.97). Also, some VA features could have a positive impact on VA evaluations through artificiality but a negative impact through intelligence, or vice versa, such that the net effect is uncertain. These points suggest important questions for continued research (Table 11, RQ1).

Increased humanness perceptions (i.e., reduced artificiality perceptions) may be eerie, in line with the uncanny valley effect (see work on robots—Mende et al., 2019; Mori, 1970). Because VAs lack embodiment, this eeriness threshold might differ vis-à-vis work on robots, but the basic phenomenon may sustain. Research should investigate explicitly whether and how these effects arise for VAs, to advance both the theory surrounding the uncanny valley effect and its application to VAs. A related, broader research direction might outline and specify the conditions in which effects related to robots apply to VAs (Table 11, RQ2).

Our findings suggest potential benefits of targeting relatively less tech-savvy users with VA features that signal VA intelligence, because use of such VAs might reduce their search costs. However, such efforts also could undermine consumers' sense of autonomy, with negative implications for their evaluations (André et al., 2018; Grewal et al., 2021), such that consumers may select nonpreferred options to reaffirm their autonomy (Davenport et al., 2020, 2021). Researchers may consider which factors determine whether (and how much) consumers value autonomy in VA settings. For example, consumers' culture and perceptions of VAs as servants versus partners may be relevant moderators; perceived autonomy tends to be more important in individualistic cultures and when users perceive VAs as servants (Table 11, RQ3).

Continued work might examine other moderators that are specific to voice technology. We found no significant moderating effects of verbalizer cognitive style; other individual differences might influence the extent to which voice-related VA features affect VA evaluations though (Table 11, RQ4).

We primarily focus on Amazon's Echo (and, by implication, Alexa), but there are many other popular VAs, such as Apple's Siri. In addition, there are a wide variety of platforms that host VAs, including cellphones, tablets, personal computers, and cars. In some cases, the platform host matches the developer of the VA (e.g., Siri on Apple iPhones), but not always (e.g., Amazon's Alexa on Apple iPhones). Might VA evaluations be contingent on the type of VA, the platform, and/or their match (Table 11, RQ5)?

Additional work might examine how new add-ins (e.g., Alexa skills) introduced by third-party developers affect VA evaluations. When more such third-party–provided skills are available, the underlying VA might appear more intelligent, which affects VA evaluations, even though those skills are not (directly) under the control of the firm that designs the VA. Similarly, integration with third-party platforms could affect VA evaluations. If this integration increases perceptions of VA intelligence, even if the platforms involve external, third parties, then collaborative approaches might make sense, in that they could prompt direct integration benefits (e.g., revenue sharing), as well as indirect benefits through enhanced perceptions of VA intelligence. (Table 11, RQ6). Examining these research questions can substantially enhance understanding of VAs and VA evaluations, as well as the theoretical lenses that are appropriate for predicting VA evaluations.

Appendix 1: Study 2—Measures

Measure	Question	Scale Points	Source
Continued Use	It is likely that I will use my Amazon Echo device in the future	7-Point Likert (Disagree—Agree)	Adapted from Pitardi and Marriott (2021)
	I intend to use my Amazon Echo device frequently	7-Point Likert (Disagree—Agree)	
	I expect to continue to use my Amazon Echo device in the future	7-Point Likert (Disagree—Agree)	
Artificiality	In my opinion, Amazon Echo appears fake	7-Point Likert (Natural—Fake)	Adapted from Moshkina (2011) and Moshkina (2012)
	In my opinion, Amazon Echo appears artificial	7-Point Likert (Lifelike—Artificial)	
	In my opinion, Amazon Echo appears machinelike	7-Point Likert (Humanlike—Machinelike)	
Natural Speech	In my opinion, my Amazon Echo device sounds natural	11-Point Likert (Disagree—Agree)	n.a
Social Cues	How would you describe 'Alexa"?	7-Point Likert (I would describe Alexa by saying "it is"—I would describe Alexa by saying "she is")	Adapted from Stroessner and Benitez (2019)
	How lifelike is 'Alexa'?	7-Point Likert (Not at all lifelike—Extremely lifelike)	
Intelligence	In my opinion, Amazon Echo appears competent	7-Point Likert (Strongly Disagree—Strongly Agree)	Adapted from McLean et al. (2021)
	In my opinion, Amazon Echo appears knowledgeable	7-Point Likert (Strongly Disagree—Strongly Agree)	
	In my opinion, Amazon Echo provides relevant information	7-Point Likert (Strongly Disagree—Strongly Agree)	
	In my opinion, Amazon Echo is intelligent	7-Point Likert (Strongly Disagree—Strongly Agree)	
	In my opinion, Amazon Echo provides accurate information	7-Point Likert (Strongly Disagree—Strongly Agree)	
Task Range	Amazon Echo can perform a wide variety of tasks	7-Point Likert (Disagree—Agree)	
Accuracy	Amazon Echo executes commands accurately	11-Point Likert (Disagree—Agree)	Atkinson and Rosenthal (2014)
Perceived Sacrifice	The price charge to use Amazon Echo is low / high	7-Point Likert (Low—High)	adapted from Cronin et al. (2000)
	The time required to use Amazon Echo is low / high	7-Point Likert (Low—High)	
	The effort I must make to use Amazon Echo is low / high	7-Point Likert (Low—High)	
Tech-savviness	I am constantly being sought after by people for advice on new digital technology	7-Point Likert (Disagree—Agree)	Ng (2012)
	I am typically one of the first to use new digital technology when it appears	7-Point Likert (Disagree—Agree)	
	I am able to use multiple digital technologies with ease	7-Point Likert (Disagree—Agree)	
Verbalizer	I prefer to learn verbally	7-Point Likert (Strongly Disagree—Strongly Agree)	Adapted from Mayer and Massa (2003)
	I am a verbal learner	7-Point Likert (Strongly Disagree—Strongly Agree)	
Ease of Use	Learning to work with Amazon Echo was easy for me	7-Point Likert (Strongly Disagree—Strongly Agree)	Adapted from McLean et al. (2021) and Davis (1989)
	I find it easy to get Amazon Echo to do what I want it to do	7-Point Likert (Strongly Disagree—Strongly Agree)	
	I find Amazon Echo easy to use	7-Point Likert (Strongly Disagree—Strongly Agree)	
	It is easy for me to become skilful at using Amazon Echo	7-Point Likert (Strongly Disagree—Strongly Agree)	

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11747-022-00874-7.

Authors' Contribution The first and sixth authors collected and analyzed data. All authors contributed to the writing of the paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aaker, J. L., Garbinsky, E. N., & Vohs, K. D. (2012). Cultivating admiration in brands: Warmth, competence, and landing in the "golden quadrant." *Journal of Consumer Psychology, 22*(2), 191–194.

Abercrombie, G., Curry, A. C., Pandya, M., & Rieser, V. (2021). Alexa, Google, Siri: What are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants. Retrieved from https://arxiv.org/abs/2106.02578. Accessed 13 June 2022.

André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., Huber, J., van Boven, L., Weber, B., & Yang, H. (2018). Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions, 5*(1–2), 28–37.

Atkinson, L., & Rosenthal, S. (2014). Signaling the green sell: The influence of eco-label source, argument specificity, and product involvement on consumer trust. *Journal of Advertising, 43*(1), 33–45.

Belanche, D., Casaló, L. V., Schepers, J., & Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The Humanness-Value-Loyalty model. *Psychology & Marketing, 38*(12), 2357–2376.

Bennett, A. M., & Hill, R. P. (2012). The universality of warmth and competence: A response to brands as intentional agents. *Journal of Consumer Psychology, 22*(2), 199–204.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing, 84*(1), 1–25.

Biswas, A., Dutta, S., & Pullig, C. (2006). Low price guarantees as signals of lowest price: The moderating role of perceived price dispersion. *Journal of Retailing, 82*(3), 245–257.

Biswas, A., & Sherrell, D. L. (1993). The influence of product knowledge and brand name on internal price standards and confidence. *Psychology & Marketing, 10*(1), 31–46.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science, 49*, 632–658.

Cable, D. M., & Turban, D. B. (2001). Establishing the dimensions, sources and value of job seekers' employer knowledge during recruitment. In *Research in Personnel and Human Resources Management*. Emerald Group Publishing Limited.

Chin, W. W. (1998). The partial least squares approach to structural equation modelling. *Modern Methods for Business Research, 295*(2), 295–336.

Chin, W. W. (2010). How to write up and report PLS analyses. In *Handbook of partial least squares* (pp. 655–690). Springer, Berlin, Heidelberg.

Cronin Jr, J. J., Brady, M. K., & Hult, G. T. M. (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing, 76*(2), 193–218.

Davenport, Tom (2021). The Future of Work Now: AI-Assisted Clothing Stylists At Stitch Fix. Forbes, May 12. https://www.forbes.com/sites/tomdavenport/2021/03/12/the-future-of-work-now-ai-assisted-clothing-stylists-at-stitch-fix/?sh=28d7bb153590. Accessed 13 June 2022.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42.

Davenport, T., Guha, A., & Grewal, D. (2021) How to design an AI marketing strategy. *Harvard Business Review,* July-August, 42–47.

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

Dellaert, B. G., Shu, S. B., Arentze, T. A., Baker, T., Diehl, K., Donkers, B., ... & Steffel, M. (2020). Consumer decisions with artificially intelligent voice assistants. *Marketing Letters*, *31*(4), 335-347.

Devlin, A. S., & Bernstein, J. (1995). Interactive wayfinding: Use of cues by men and women. *Journal of Environmental Psychology, 15*(1), 23–38.

Doyle, P. R., Edwards, J., Dumbleton, O., Clark, L., & Cowan, B. R. (2019). Mapping perceptions of humanness in speech-based intelligent personal assistant interaction. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, 1–12.

Fernandes, T., & Oliveira, E. (2021). Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants' adoption. *Journal of Business Research, 122*, 180–191.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research, 69*(8), 3192–3198.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association, 70*(350), 320–328.

Gerstner, E. (1985). Do higher prices signal higher quality? *Journal of Marketing Research, 22*(2), 209–215.

Gounaris, S. P., & Venetis, K. (2002). Trust in industrial service relationships: Behavioral consequences, antecedents and the moderating effect of the duration of the relationship. *Journal of Services Marketing, 16*(7), 636–655.

Grewal, D. & Compeau, L. (2007). Consumer responses to price and its contextual information cues: A synthesis of past research, a conceptual framework, and avenues for further research. *Review of Marketing Research*, Volume 3, ed. Malhotra, N. M.E. Sharpe, 109–131

Grewal, D., Guha, A., Satornino, C. B., & Schweiger, E. B. (2021). Artificial intelligence: The light and the darkness. *Journal of Business Research, 136*, 229–236.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5228–5235.

Guo, X., Deng, H., Zhang, S., & Chen, G. (2020). Signals of competence and warmth on e-commerce platforms. *Data and Information Management, 4*(2), 81–93.

Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017a). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017b). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science, 45*(5), 616–632.

Hair, J. F., Jr., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101–110.

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management & Data Systems, 118*(3), 618–636.

Hassanein, K., Head, M., & Ju, C. (2009). A cross-cultural comparison of the impact of social presence on website trust, usefulness and enjoyment. *International Journal of Electronic Business, 7*(6), 625–641.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.

Hernandez-Ortega, B., & Ferreira, I. (2021). How smart experiences build service loyalty: The importance of consumer love for smart voice assistants. *Psychology & Marketing, 38*(7), 1122–1139.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal, 20*(2), 195–204.

Humphry, J., & Chesher, C. (2021). Preparing for smart voice assistants: Cultural histories and media innovations. *New Media & Society, 23*(7), 1971–1988.

Jin, B. & Park, J. (2006). The moderating effect of online purchase experience on the evaluation of online store attributes and the subsequent impact on market response outcomes. *NA - Advances in Consumer Research.* Volume 33, eds. Pechmann, C. & Price, L. Duluth, MN: Association for Consumer Research, 203–211

Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. (2019). Eliza in the uncanny valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing Letters, 30*(1), 1–12.

Kimery, K. M., & McCord, M. (2008). Seals on retail web sites: A signalling theory perspective on third-party assurances. In *Web Technologies for Commerce and Services Online,* IGI Global. 111–134.

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study. *Computers in Human Behavior, 68*, 170–179.

Kock, F., Berbekova, A., & Assaf, A. G. (2021). Understanding and managing the threat of common method bias: Detection, prevention and control. *Tourism Management, 86*, 104330.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10.

Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly, 31*(1), 59–87.

Liu, S. X., Shen, Q., & Hancock, J. (2021). Can a social robot be too warm or too competent? Older Chinese adults' perceptions of social robots and vulnerabilities. *Computers in Human Behavior, 125*, 106942.

Long, J. A. (2019*). Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.0*. from https://cran.r-project.org/package=interactions. Accessed 13 June 2022

Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior, 56*, 225–237.

Mari, A., & Algesheimer, R. (2021, January). The role of trusting beliefs in voice assistants during voice shopping. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, 4073.

Mattila, A., & Wirtz, J. (2001). The moderating role of expertise in consumer evaluations of credence goods. *International Quarterly Journal of Marketing, 1*(4), 281–292.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior, 99*, 28–37.

McLean, G., Osei-Frimpong, K., & Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement? Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research, 124*, 312–328.

McLean, G., Osei-Frimpong, K., Wilson, A., & Pitardi, V. (2020). How live chat assistants drive travel consumers' attitudes, trust and purchase intentions. *International Journal of Contemporary Hospitality Management, 32*, 1795–1812.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research, 56*(4), 535–556.

Milne, G. R., Villarroel Ordenes, F., & Kaplan, B. (2020). Mindful consumption: Three consumer segment views. *Australasian Marketing Journal, 28*(1), 3–10.

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33–35.

Moriuchi, E. (2019). Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing, 36*(5), 489–501.

Moshkina, L. (2011). An integrative framework of time-varying affective robotic behavior, Doctoral dissertation, Georgia Institute of Technology.

Moshkina, L. (2012). Reusable semantic differential scales for measuring social response to robots. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. ACM, 89–94.

Nasirian, F., Ahmadian, M., & Lee, O. (2017). AI-Based voice assistant systems: Evaluating from the interaction and trust perspectives. *Twenty-third Americas Conference on Information Systems*, Boston.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research, 56*(6), 960–980.

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education, 59*, 1065–1078.

Nguyen, J. (2021). Turns out, no one wants to talk to Amazon's Alexa. *Mashable*, December 23. Retrieved from https://mashable.com/article/amazon-alexa-usage-drop. Accessed 13 June 2022.

Ordenes, F. V., & Silipo, R. (2021). Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications. *Journal of Business Research, 137*, 393–410.

Organisation for Economic Co-operation and Development (OECD). (2018). *Bridging the digital gender divide: Include, upskill, innovate*. OECD.

Park, D. H., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications, 7*(4), 399–410.

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology & Marketing, 38*(4), 626–642.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Ramos, D. (2021). Voice Assistants: How artificial intelligence assistants are changing our lives every day. Smartsheet.com, July 26. Retrieved from https://www.smartsheet.com/voice-assistants-artificial-intelligence. Accessed 13 June 2022.

Rao, A. R., & Monroe, K. B. (1988). The moderating effect of prior knowledge on cue utilization in product evaluations. *Journal of Consumer Research, 15*(2), 253–264.

Reis, A. M., Paulino, D., Paredes, H., & Barroso, J. (2017). Using intelligent personal assistants to strengthen the elderlies' social bonds. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction,* 593–602.

Richardson, A. (1977). Verbalizer-Visualizer: A cognitive style dimension. *Journal of Mental Imagery, 1*, 109–126.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance. *Organizational Research Methods, 12*(4), 762–800.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning, 45*(5–6), 341–358.

Rönkkö, M., & Ylitalo, J. (2011). PLS marker variable approach to diagnosing and controlling for method variance. *Proceedings of the International Conference on Information Systems*, ICIS 2011, Shanghai, China, December 4–7, 2011. Association for Information Systems 2011, ISBN 978–0–615–55907–0. https://aisel.aisnet.org/icis2011/proceedings/researchmethods/8. Accessed 13 June 2022

Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the emancipation of PLS-SEM: A commentary on Rigdon (2012). *Long Range Planning, 47*(3), 154–160.

Schweitzer, F., Belk, R., Jordan, W., & Ortner, M. (2019). Servant, friend or master? The relationships users build with voice-controlled smart devices. *Journal of Marketing Management, 35*(7–8), 693–715.

Spence, A. M. (1973). Job market signalling. *Quarterly Journal of Economics, 87*, 355–374.

Spence, M. T., & Brucks, M. (1997). The moderating effects of problem characteristics on experts' and novices' judgments. *Journal of Marketing Research, 34*(2), 233–247.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research, 50*(2), 277–288.

Stroessner, S. J., & Benitez, J. (2019). The social perception of humanoid and non-humanoid robots: Effects of gendered and machine-like features. *International Journal of Social Robotics, 11*(2), 305–315.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management, 75*, 550–568.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479.

Tursi, V., & Silipo, R. (2018). *From Words to Wisdom, an Introduction to Text Mining with Knime*. KNIME Press.

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research, 20*(1), 43–58.

van Pinxteren, M. M., Wetzels, R. W., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: Implications for services marketing. *Journal of Services Marketing, 33*(4), 507–518.

Villarroel Ordenes, F., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., & Wetzels, M. (2019). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research, 45*(5), 988–1012.

Viswanathan, M., & Kayande, U. (2012). Commentary on "common method bias in marketing: Causes, mechanisms, and procedural remedies." *Journal of Retailing, 88*(4), 556–562.

Wagner, J. A., Klein, N. M., & Keith, J. E. (2001). Selling strategies: The effects of suggesting a decision structure to novice and expert buyers. *Journal of the Academy of Marketing Science, 29*(3), 290–307.

Wang, L., Fan, L., & Bae, S. (2019). How to persuade an online gamer to give up cheating? Uniting elaboration likelihood model and signalling theory. *Computers in Human Behavior, 96*, 149–162.

Yao, L., Mimno, D., & McCallum, A. (2009). Efficient methods for topic model inference on streaming document collections. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Ye, S., Ying, T., Zhou, L., & Wang, T. (2019). Enhancing customer trust in peer-to-peer accommodation: A "soft" strategy via social presence. *International Journal of Hospitality Management, 79*, 1–10.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113–128.

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Research Report Series, 10*, 1–17.

Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software, 95*(1), 1–36.

Zhang, X., & Zhang, R. (2021, August). Impact of physicians' competence and warmth on chronic patients' intention to use online health communities. In *Healthcare* (Vol. 9, No. 8, p. 957). Multidisciplinary Digital Publishing Institute.