Pontifícia Universidade Católica
do Rio de Janeiro

Isabela Canellas da Motta

Exploring proposals to align users' mental models and improve interactions with Voice Assistants (VAs)

Masters Dissertation

Dissertation presented to the Graduate Program in Design of PUC-Rio in partial fulfillment of the requirements for the degree of Masters in Design.

Advisor: Prof. Manuela Quaresma

Rio de Janeiro
February 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Isabela Canellas da Motta

Exploring proposals to align users' mental models and improve interactions with Voice Assistants (VAs)

Masters Dissertation

Dissertation presented to the Graduate Program in Design of PUC-Rio in partial fulfillment of the requirements for the degree of Masters in the Graduate Program in Design of the Department of Arts and Design in the Center of Theology and Humanities of PUC-Rio. Approved by the Examination Board signed below.

Prof. Maria Manuela Rupp Quaresma
Advisor
Department of Arts and Design – PUC-Rio

Prof. Marcelo Fernandes Pereira
Department of Arts and Design – PUC-Rio

Heloisa Caroline de Souza Pereira Candello
IBM Research

Prof. Monah Winograd
Sectorial Coordinator of Graduate Studies and Research of the Center of Theology and Humanities – PUC-Rio

Rio de Janeiro, February 11th 2022

Bibliographic data

Motta, Isabela Canellas da

Exploring proposals to align users' mental models and improve interactions with Voice Assistants (VAs) / Isabela Canellas da Motta ; advisor: Manuela Quaresma. – 2022.
203 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design, 2022.
Inclui bibliografia

1. Artes e Design – Teses. 2. Assistentes de voz. 3. Modelos mentais. 4. Interação humano-computador. 5. Usabilidade. 6. Delphi. I. Quaresma, Maria Manuela Rupp. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Artes e Design. III. Título.

CDD: 700

Acknowledgements

To CAPES and FAPERJ, for funding for this research.

To my advisor, Prof. Manuela Quaresma, for her essential guidance since under-graduate research.

To LEUI – Laboratory of Ergodesign and Usability of Interfaces, for supporting this research, and to all its members, for the continuous encouragement.

To my parents, Henrique and Zoê, for their unconditional support.

To all classmates and professors who guided and supported this research.

To all my friends, especially João Victor, who helped me through two years of the coronavirus pandemic.

To all of this research's volunteers, without whom this dissertation would not have been possible.

Abstract

Motta, Isabela Canellas; Quaresma, Manuela (Advisor). **Exploring proposals to align users' mental models and improve interactions with Voice Assistants (VAs).** Rio de Janeiro, 2022, 203p. Masters Dissertation – Department of Arts and Design, Pontifical Catholic University of Rio de Janeiro.

Voice Assistants (VAs) bring various benefits for users and are increasingly popular, but some barriers for VA adoption and usage still prevail, such as users' attitudes, privacy concerns, and negative perceptions towards these systems. An approach to mitigating such obstacles and leveraging voice interactions may be understanding users' mental models of VAs, since studies indicate that users' understandings of VAs are unaligned with these systems' actual capabilities. Thus, considering the importance of a correct mental model for interactions, exploring influential factors causing misperceptions and solutions to deal with this issue may be paramount. The objective of this research was to identify leading causes of users' misperceptions and offer design recommendations for aligning users' mental models of VAs with these systems' real capacities. In order to achieve this goal, we conducted a systematic literature review (SLR), exploratory interviews with experts, and a questionnaire-based three-round Delphi study. The results indicate that design aspects such as VAs' high humanness and the lack of outputs' transparency are influential for mental models. Despite the indication that these drivers lead to users' misperceptions, removing VAs' humanness and excessively displaying information about VAs might not be an immediate solution. In turn, developers should assess the context and task domains in which the VA will be used to guide design decisions. Moreover, developers should understand the users' profiles and backgrounds to adjust interactions, as users' characteristics are influential for how they perceive the product. Finally, developing teams should have a correct and homogeneous understanding of VAs and possess the necessary knowledge to employ solutions properly. This latter requirement is challenging since VAs' novelty might demand professionals to master new skills and tools.

Keywords

Voice Assistants; Conversational Agents; Voice Interfaces; Mental Models; Human-Computer Interaction; Usability; Delphi.

Resumo

Motta, Isabela Canellas; Quaresma, Manuela. **Explorando propostas para alinhar os modelos mentais de usuários e melhorar as interações com Assistentes de Voz.** Rio de Janeiro, 2022, 203p. Dissertação de Mestrado – Departamento de Artes e Design, Pontifícia Universidade Católica do Rio de Janeiro.

Assistentes de Voz (AVs) trazem diversos benefícios para os usuários e estão se tornando progressivamente populares, mas algumas barreiras para adoção de AVs ainda persistem, como atitudes dos usuários, preocupações com privacidade e percepções negativas desses sistemas. Uma abordagem para mitigar os obstáculos e melhorar as interações pode ser entender os modelos mentais dos usuários de AVs, uma vez que estudos indicam que o entendimento dos usuários não é alinhado com as reais capacidades desses sistemas. Assim, considerando a importância de um modelo mental correto para as interações, explorar fatores geradores de percepções inadequadas e soluções para lidar com tal questão pode ser essencial. O objetivo desta pesquisa foi identificar fatores influentes para as percepções inadequadas de usuários e oferecer recomendações de design para alinhar os modelos mentais de usuários com as reais capacidades desses sistemas. Para alcançar esse objetivo, nós conduzimos uma revisão sistemática de literatura, entrevistas exploratórias com experts e um estudo Delphi de três rodadas com base em questionários. Os resultados indicam que os aspectos de design como a humanização dos AVs e a transparência em respostas do sistema são influentes para os modelos mentais. Apesar desses fatores terem sido indicados como causas para incorreções em modelos mentais, remover a humanização dos AVs e apresentar informações excessivas pode não ser uma solução imediada. Indica-se que designers devem avaliar o contexto de uso e os domínios de tarefa em que os AVs serão usados para guiar as soluções de design. Além disso, os designers devem entender os perfis e *backgrounds* dos usuários para ajustar as interações uma vez que as características dos usuários são influentes para sua percepção do produto. Finalmente, o time de desenvolvimento deve ter um entendimento correto e homogêneo do AVs, e deve possuir o conhecimento

necessário para aplicar soluções corretamente. Esse último requisito é desafiador porque os AVs são produtos relativamente novos e podem demandar que os profissionais dominem novas habilidades e ferramentas.

Palavras-chave

Assistentes de Voz; Agentes Conversacionais; Interfaces de Voz; Modelos Mentais; Interação Humano-Computador; Usabilidade; Delphi.

Table of contents

List of figures

List of tables

List of panels

1
Introduction

Voice Assistants (VAs) are artificial intelligence (AI)-powered virtual assistants that can perform a range of tasks in a system, which users interact through a voice interface that may be supported by a visual display (WEST; KRAUT; HAN EI, 2019). These systems have been developed by multiple technology-related enterprises such as Apple (Siri), Amazon (Alexa), Microsoft (Cortana), Google (Google Assistant), and Samsung (Bixby), and run on several devices such as earphones, smart speakers, and smartphones.

As exemplified by Amazon's Alexa, which was able to perform over 70.000 skills in the USA by 2020 (STATISTA, 2020), available features are rapidly growing in number, ranging from tasks such as weather forecast to home automation. VAs were estimated to be in use in over four billion devices by 2020 (MOAR; ESCHERICH, 2020), with 20% of all population in western countries reporting using VAs several times a day in 2021 (VAILSHERY, 2022). Forecasting indicates that VAs are expected to reach 8.4 billion units by 2024 (VAILSHERY, 2021), and that the voice recognition technology market will be worth 30 billion U.S dollars by 2026 (VAILSHERY, 2022). The projections for VAs indicate that interfaces for human-computer interaction (HCI) are in the midst of a paradigm shift from visual interfaces to hands-free, voice-based interactions (WEST; KRAUT; HAN EI, 2019).

Before exposing the matters around VAs, it is important to outline a few relevant characteristics that set them apart from other systems. Firstly, VAs must not be confused with other types of Conversational Agents (CAs) or Voice User Interfaces (VUIs). For example, since VAs apply Natural Language Processing (NLP) algorithms (PEARL, 2016) – a type of AI – users are usually not limited to fixed queries and may formulate commands in different ways to request the same action. Such feature makes them different from most Interactive Voice Response (IVR) systems – commonly used for Customer Service –, which can usually deal

only with previously established commands or manual input (PEARL, 2016). Moreover, contrarily to *chatbots*, in which the primary interaction channel is written text, users interact with VAs through speech. Finally, differently from *virtual agents*, VAs are not accompanied by any projection of a human (or human-like) virtual image that illustrates a visible entity (WEST; KRAUT; HAN EI, 2019). Hence, in this study, we will refer to VAs as a type of VUI that presents the following characteristics:

1) It applies AI to the processing of commands.
2) Its primary interaction channel is speech, not written text. However, it may present some visual information on a screen or the device.
3) It does not project a human-like virtual image to represent an entity.

This chapter will present this research's outline, including our research problem, question, primary and secondary objectives, and the research's method and techniques. At this chapter's end, we also present the structure of this dissertation.

1.1. Research problem

As with any VUI, VAs enable users' interaction through a voice command from the user to the system (i.e., voice input) and a voice response from the system to the user (i.e., voice output) (BHOWMIK, 2015). The use of voice makes VAs notably different from other interaction modalities (e.g., haptic, visual) due to the speech's intuitiveness. Since humans evolved to be able to understand speech (NASS; BRAVE, 2005) and speaking is a constant action in users' routines, voice interaction is natural and intuitive (MEEKER, 2016; PEARL, 2016).

Despite the VAs' benefits and the projections suggesting growth in VA usage, studies indicate that barriers to the adoption of these systems still prevail. In the first place, surveys have shown that users consider VAs as not relevant or not very useful (MOTTA; QUARESMA, 2019; ROBART, 2017), and such perceptions of low usefulness have been reported to negatively impact VA adoption (MCLEAN; OSEI-FRIMPONG, 2019). Additionally, both market and scientific publications have shown that users frequently report facing errors throughout interactions (MAUÉS, 2019; WHITE-SMITH *et al.*, 2019), and such technical issues

have been related to low satisfaction measures (PURINGTON *et al.*, 2017). To recover from failures, users apply strategies that might hamper the interaction's naturality, such as repeating requests, adjusting a command's structure, wording, or information amounts, changing pronunciation, and speaking louder (BENETEAU *et al.*, 2019; GARG; SENGUPTA, 2020; LOVATO; PIPER; WARTELLA, 2019; PORCHERON *et al.*, 2018; PORCHERON; FISCHER; SHARPLES, 2017; YAROSH *et al.*, 2018). Finally, users' attitudes – that is, their "tendencies of approach or avoidance" (OSGOOD, 1957, p. 189) – towards voice interaction affect VA usage (MORIUCHI, 2019). Particularly, several studies have indicated that users are concerned about the privacy of their data, creating negative attitudes towards VAs (BURBACH *et al.*, 2019; DE BARCELOS SILVA *et al.*, 2020; HOY, 2018; MCLEAN; OSEI-FRIMPONG, 2019; PITARDI; MARRIOTT, 2021).

While there still exist limitations in speech recognition technology (especially in NLP technology; PEARL, 2016) that might partially account for the issues above, users' mental models of VAs might also play a significant role in the quality of interactions. Mental models are a type of conceptual model that represents how a product or system works (NORMAN, 2013), comprising a set of expectations about a system's components, functioning, and proper usage (WICKENS; LEE; LIU; BECKER, 2014). These models are essential for users since they are closely related to how people perform tasks and dictate performance levels (WILSON; RUTHERFORD, 1989).

Although users' mental models are essential for the quality of interactions, studies indicate that users' mental models of VAs do not match these systems' actual capabilities (CHO; LEE; LEE, 2019; LUGER; SELLEN, 2016). Overall, users have unrealistic mental models of VAs, displaying a lack of understanding concerning VAs functioning and high expectations for system features, intelligence, and conversational capabilities (see chapter 3 for a complete review on users' mental models of VAs).

The users' misperceptions might be relevant to the beforementioned adoption and usage barriers. For example, in a previous study, we showed that users do not utilize some tasks due to the unawareness of their availability (MOTTA; QUARESMA, 2021), which could account for the belief that VAs have low usefulness. The possible unawareness of available features might also lead to errors

since some failures may be caused by requests to perform activities out of the VAs' scope. Moreover, users' difficulty in recovering from failures might be related to their low comprehension of the reasons behind errors, as studies suggest that users apply different error-recovery strategies based on their understanding of error sources (KIM; JEONG; LEE, 2019; MOTTA; QUARESMA, 2022; MYERS *et al.*, 2018; PORCHERON *et al.*, 2018; PORCHERON; FISCHER; SHARPLES, 2017). Likewise, negative attitudes towards VAs and privacy concerns might be caused by the lack of understanding of VAs' functioning, as users are unaware of privacy controls and privacy-related information such as data collection, storage, and sharing (AMMARI *et al.*, 2019; COWAN *et al.*, 2017; JAVED; SETHI; JADOUN, 2019; WEBER; LUDWIG, 2020).

1.2. Research question and objectives

Considering the issues presented above, aligning users' mental models of VAs with these systems' actual capabilities is paramount for VA adoption, and improving VAs' design characteristics might be essential for such an alignment. As explained by Norman (2013), users develop their mental models by relying on the *system image* (perceivable physical cues of the product itself, past experiences, advertisements, manuals, etc.), making design aspects vital to developing correct mental models. Hence, this research's primary goal is to identify leading causes of users' misperceptions and offer design recommendations for aligning users' mental models of VAs with these systems' real capacities. This research poses the following research question: "*How can VAs be improved to mitigate gaps between users' mental models and the VAs' actual capabilities?*". To support such an investigation, we aimed to understand 1) the causes for users' misperceptions of VAs and 2) solutions to deal with the issue. Moreover, the research aims to fulfill the following secondary goals (SG):

1. To gather and understand the characteristics of human-human communication;
2. To comprehend the concept of mental models and their impacting factors;
3. To identify the state-of-the-art of users' mental models of VAs;

4. To explore how developers and researchers of conversational interfaces understand users' mental models of VAs;
5. To identify the main causes that lead to misalignments in users' mental models of VAs;
6. To identify adequate solutions to deal with the issue of users' mental models of VAs.

We highlight that we employ the word "developer" in this work to refer to any professional working in VA development, and not only those involved in programming. Developers may include interaction and conversational designers, UX writers, programmers, data analysts, and many others working in VA projects.

1.3. Research method and techniques

We applied an exploratory method with a mixed-methods approach to address this research's questions and objectives, combining qualitative and quantitative techniques. We conducted a literature review on human conversational practices, human cognition, and mental models to address specific goals 1 and 2. Then, we systematically reviewed the literature to identify the state-of-the-art of users' mental models of VAs (SG3). Based on the literature review, we identified the need to explore how developers and researchers of conversational interfaces understand users' mental models of VAs (SG4), and therefore, we conducted exploratory semi-structured interviews with these professionals. The interviews also served as a preparation for the following research technique: a three-round Delphi study with professionals involved in VA development or research. This questionnaire-based study aimed to identify the leading causes of misalignments in users' mental models (SG5) and gather solutions to deal with this issue (SG6). The complete description of this research's methodology is presented in chapter 4.

1.4. Relevance of the research

Offering recommendations for improving VAs and aligning users' mental models with these systems' actual capabilities can bring several benefits. Firstly, a

variety of studies in the literature have explored issues in users' perceptions and understanding of these interfaces (see chapter 3). However, few publications offer concrete solutions to deal with such a matter, and, although some have assessed paths to solving the problem, most of these studies evaluated other voice interfaces and not VAs specifically (KIM; JEONG; LEE, 2019; KIRSCHTHALER; PORCHERON; FISCHER, 2020; MYERS, Chelsea M, 2019). This specificity is relevant since VAs are not limited to fixed commands and operate in a much larger domain than voice interfaces that serve a specific purpose (e.g., voice-based calendar, receipt assistant, driving assistant).

In addition, to the extent of our knowledge, the literature still lacks a study that aims to find solutions to the users' mental model issue by gathering professionals involved in the research or development of such interfaces. Since they are closely involved in VAs' development (or the research that aids it), they can propose solutions while keeping development constraints in mind. Furthermore, the professionals who took part in this study came from different backgrounds, and such a variety of opinions can contribute to the proposed solutions.

Understanding how users think and conceptualize tasks and tools is vital for good interface design of any product (HACKOS; REDISH, 1998). Hence, design solutions for improving VAs must consider how to align users' expectations with the system's features and functioning. A set of recommendations is necessary to support the interaction designers' work, especially for novel products such as VAs, which use not only new technology (artificial intelligence and machine learning), but also a communication channel rather than the visual (speech). By improving the quality of interactions with VAs, these systems' adoption and usage may also be leveraged. Such an effort is essential since VAs can bring various economic and social benefits.

Firstly, VAs offer several economic advantages. According to Liu (2021), the transaction value of eCommerce purchases made through VAs has reached 4.6 billion US dollars worldwide in 2021, and a 400% growth is expected to happen until 2023, raising such statistic to 19.4 billion US dollars. For online shopping, specifically, VAs are expected to generate a great profit, since 73% of VAs' owners have already shopped through the assistants and 31% stated that an interaction with a VA influenced them to buy a product (INVOCA, 2018). From 2014 to 2016, the use of VAs also caused an increase of 33% in the number of consumers calling

sellers after conducting previous research or considering a purchase. Such calls are essential since it helps the consumer to decide for the purchase and generate more than a trillion dollars in the US (INVOCA, 2018). Additionally, VAs produce profit for a vast chain of actors that are a part of its development, implementation, and integration with other services. These include developers, cloud storage services, IT companies that develop API integrations, hardware manufacturers, and resellers (MOAR, 2019).

VAs also have great potential for accessibility and user experience. They benefit from the voice interaction's characteristics, such as the possibility of monitoring complex and specific information without requiring the visual channel, being physically distant from the device, and speech's intuitiveness. Furthermore, voice interfaces are ideal for products with small or nonexistent displays, freeing them from the need for a visual apparatus, thus, potentially favoring the Internet of Things type of products (MEEKER, 2016; PEARL, 2016).

Accessibility is also a crucial advantage for voice interfaces. The majority of visual displays available on the market are not accessible for all types of users. Touch-screen smartphones with visual interfaces may present challenges for users with visual impairments. Moreover, considering that 6.8% of the Brazilian population was illiterate by 2018 (IBGE, 2019), the majority of textual information may be excluding for such users. Thus, VAs may present benefits since messages are presented through voice responses. Similarly, the complexity of some interactions on visual interfaces may pose challenges for elderly users or people with cognitive impairment (BALASURIYA *et al.*, 2018; PRADHAN; LAZAR; FINDLATER, 2020). As interactions are intuitive and straightforward (BHOWMIK, 2015), voice interfaces may simplify interactions for specific tasks (MEEKER, 2016), which is suitable for accessibility and positive for user experience overall.

1.5. Document structure

This document is structured as follows. **Chapter 2** outlines users' communication with VAs by explaining how voice interactions pattern after human-human communication. We present relevant concepts and practices applied by humans in their daily conversations, such as Grice's cooperative principle and mechanisms of

conversation exposed in the field of conversation analysis (e.g., turn-taking, adjacency pairs, and repair). Furthermore, to understand how language production is linked to cognition – which is important for comprehending the concept of mental models – we expose how humans process information, solve problems, and make decisions. Throughout the chapter, we present various examples to illustrate how the concepts relate to users' interactions with VAs.

After establishing how users communicate with VAs, we further explore users' mental models of VAs in **Chapter 3.** The chapter starts by presenting the concept of mental models and then presents the results of a Systematic Literature Review (SLR) on users' mental models of VAs. The chapter's objective is to provide an overview of the state-of-the-art concerning this issue. The review showed that users are not aware of relevant information for data privacy, do not correctly understand VAs' general functioning and actions, and have trouble understanding error sources and recovering from failures. Likewise, users have unrealistic expectations for these systems' intelligence and technical, social, and conversational capabilities. Consequently, users face hardships throughout interactions and get frustrated, which leads to the underutilization or complete abandonment of the Voice Assistant. At the end of the chapter, we present a diagram describing relevant factors impacting users' perceptions.

Following, in **Chapter 4**, we report the research method. Such description includes the Systematic Literature Review (SLR; section 4.1.), exploratory interviews with experts (section 4.2.), and a Delphi study (section 4.3.). As mentioned, the SLR's goal was to understand the literature's state-of-the-art on users' mental models of VAs. The exploratory interviews with experts had the purpose of surveying how professionals involved in the research and development of VAs perceive the issue of users' mental models. Additionally, the Delphi study – a three-phased questionnaire – aimed to support the communication of such professionals in order to identify causes and solutions for the mental model issue.

Thereafter, we proceed to present the research's results. **Chapter 5** reports the results of the interviews with the experts. In **Chapter 6,** we present the results of the Delphi study, dividing the chapter into sections according to the study's phases. We discuss the interviews' and Delphi's results in light of the literature in **chapter 7** and present the research's conclusions in **chapter 8.** Finally, we present the **bibliography** and this document's **appendixes.**

2
Users' communication with Voice Assistants (VAs)

The voice interfaces (PEARL, 2016) that support users' interactions with Voice Assistants (VAs), as any type of Human-Machine Interface (HMI), intermediate the exchange of information between a human and a system (KROEMER; GRANDJEAN, 1997). However, speech interaction makes voice interfaces notably different from other interfaces such as Graphic User Interfaces (GUI) or gestural interfaces. Since the early days of humanity, the human brain evolved in several ways to process and understand speech, leading to a successful and stable perception of voice by humans (NASS; BRAVE, 2005). Thus, one of the main benefits of voice interaction is the speech's intuitiveness, as conversations are natural to human beings and are continuously present in users' routines (AMAZON, 2020; MEEKER, 2016; PEARL, 2016).

The advantage of speech for user-VA interaction is possible because people tend to easily and naturally apply social rules to Human-Computer Interaction (HCI). Nass and colleagues conducted a series of experiments (see NASS; MOON, 2000; NASS; STEUER; TAUBER, 1994) which demonstrated that users mindlessly attribute human characteristics to computers, for example, by expressing politeness, addressing gender stereotypes, and applying notions of "self" and "other" (NASS; STEUER; TAUBER, 1994). According to Nass and Moon (2000),

> "Mindless behavior (...) occurs as a result of conscious attention to a subset of contextual cues (Langer, 1992). These cues trigger various scripts, labels, and expectations, which in turn focus attention on certain information while diverting attention away from other information. Rather than actively constructing categories and distinctions based on all relevant features of the situation, individuals responding mindlessly prematurely commit to overly-simplistic scripts drawn in the past" (NASS; MOON, 2000, p. 83)

The authors argue that users express these types of behavior when the system: 1) has words as outputs, 2) provides answers based on users' inputs, and 3) performs roles traditionally performed by humans. Since a VA provides *spoken*

responses to *users' commands* and acts as an *assistant*, it is possible to argue that users may mindlessly apply social rules to voice interactions. Additionally, humans are unable to suppress natural reactions to speech (NASS; BRAVE, 2005), reinforcing that user-VA interaction may present similarities to natural, human-human communication. Therefore, as suggested by Nass, Steuer, and Tauber (1994), investigating social communication principles and understanding its implications for human-computer interaction is paramount to design user interfaces.

This chapter aims to present theories on how conversations are structured and the underlying cognitive processes behind human communication. We expose conversational theories that have been previously demonstrated to apply to users' interactions with conversational interfaces, such as Grice's Cooperative Principe (GOOGLE, 2017) and the field of conversation analysis (MOORE; ARAR, 2019). To illustrate how users' interactions with VAs fit into such theories, we provide hypothetical interactional scenarios. We created such examples based on observations of user-VA interactions in previous studies (see Motta and Quaresma (2021, 2022) for a complete description).

2.1.
Human communication and conversations

Conversations, written or spoken, are human's primary channel for communicating in a society. People engage in conversations for various reasons, from emotional exchanges to practical, goal-oriented everyday activities. Although conversations may feel like an organic action learned since early childhood, human-human communication is not random. Conversely, human conversations are highly methodic (HUTCHBY; WOOFFITT, 1998) and tend to be characterized by the mutual recognition of a common purpose or direction among participants, therefore being a cooperative effort (GRICE, 1991).

2.1.1.
Grice's Cooperative Principle

According to Grice (1991), people tend to be cooperative with their conversational partner and apply a set of rules to do so, except for cases in which the

speaker intentionally violates them (e.g., irony, lies, metaphors). The author defines this tendency as the Cooperative Principle: "make your conversational contribution on such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (GRICE, 1991, p. 26). To adhere to the Cooperative Principle, talk exchanges usually regard a range of maxims, comprised of four categories:

1. *Quantity* - Provide the right amount of information necessary to convey a message: nor too little, nor too much.
2. *Quality* - Do not say information considered to be false or which lacks evidence to be supported.
3. *Relation* - Provide relevant information to the conversation at hand.
4. *Manner* - Express a message in a comprehensible manner by avoiding obscurity of expression, ambiguity, prolixity, and unorderly exposition.

In accordance with the before-mentioned proneness of people to apply social norms to HCI, it is possible to observe situations in which users are cooperative and follow the maxims to communicate with VAs (or, at least, try to). Although some users may use VAs as a source of entertainment (i.e., jokes), these systems are mostly used to perform practical tasks and are valued for the rapidness of their interaction (PEARL, 2016). Thus, it could even be counterproductive to violate the conversational maxims**,** as making commands excessively long would slow down interactions, and providing the VA with false, irrelevant, or incomprehensible information will likely result in an error.

For example, if a user wants to set up a reminder to call someone, they may obey the maxims of *quantity* and say, "remind me to call Joey at 8 am", rather than adding unnecessary information such as, "remind me to call Joey, who is my boyfriend, at 8 am since I know this is the time he wakes up". Likewise, if a user wants to listen to The Beatles' "Yesterday," it is unlikely that they would intentionally violate a maxim of *quality* and say, "Play 'Yesterday' by Beyoncé." Similarly, if, while scheduling an appointment to the calendar, the VA asks, "when is the appointment?", users will probably follow the maxim of *relation* and answer, "November third at nine o'clock" instead of deliberately saying something irrelevant such as "Pizza." In the last example, if a user wants to follow the maxims of *manner*, they may even try to mitigate ambiguity by saying, "November third at *9 pm*".

It is necessary to point out that Grice (1991) also explains that participants of human-human conversations unintentionally violate a maxim on occasions. Similarly, even though VA users may (mindlessly) *intend* to be cooperative and follow Grice's maxims, they may not always *succeed* in doing so. Consider figure 2.1:

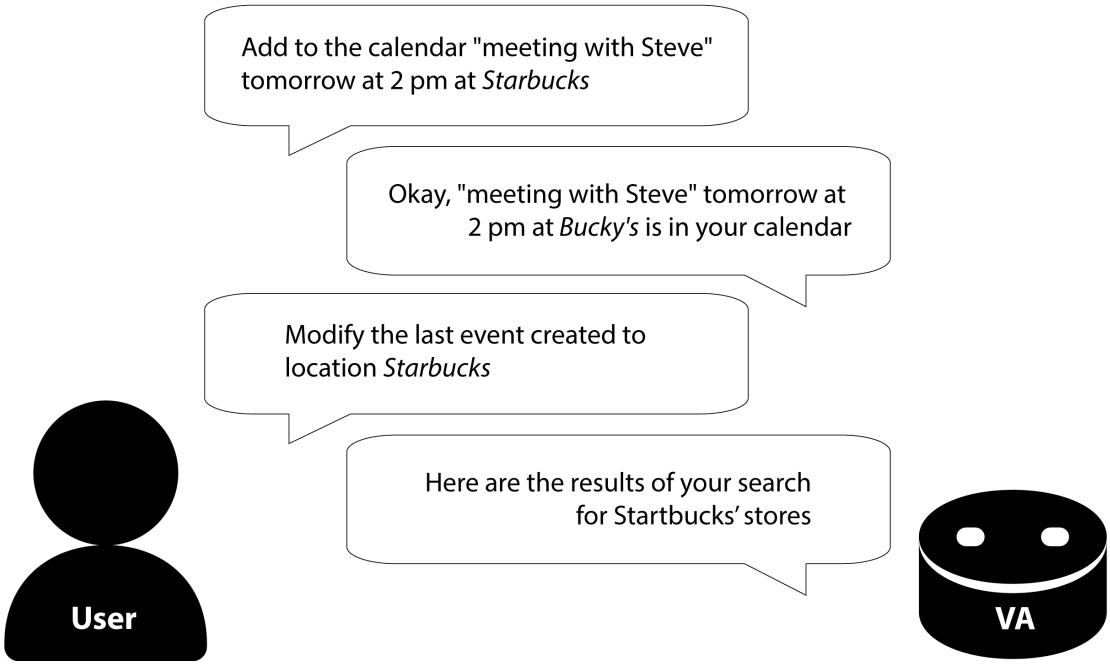


Figure 2.1 – User-VA dialogue 1. Source: the authors.

In this situation, the user adds a meeting to the calendar, but the VA misrecognized the appointment's location. Then, they try to edit the event by specifying "*the last event created,"* arguably trying to provide enough information to the VA and avoid ambiguity. However, instead of editing the event, the VA performs a web search for locations, maybe because it could not correctly "interpret" the user's reference to a past interaction (*last event*). Thus, the user was unsuccessful in evaluating the right *quantity* of information to complete the task and the correct *manner* of speaking to the VA. In turn, the VA failed to present a relevant response to the user correctly. Situations such as this suggest that there might be a gap between behaviors believed to be cooperative by users and VAs.

The examples provided in this section illustrate how users may mindlessly (try to) follow Grice's Cooperative Principle and maxims when communicating with VAs. Nevertheless, these conversational rules are more related to the message's content than to the structure of a conversation. The following section will present the structure of conversations according to the Conversation Analysis filed

and show why user-VA interaction may follow similar patterns to human conversations.

2.1.2.
Conversation Analysis

The field of Conversation Analysis, originated from sociology, is defined by Hutchby and Wooffitt (1998, p. 13) as "(...) the study of talk. More particularly, it is the systematic analysis of the talk produced in everyday situations of human interaction: talk-in interactions". The field considers the utterances spoken throughout a conversation (or a talk-in interaction) as a means for speakers to accomplish a goal when interacting with others. Additionally, conversation production is methodic, resulting in systematic and deeply ordered talk exchanges (HUTCHBY; WOOFFITT, 1998). The following subsections present some of the key concepts, initially defined by Sacks and colleagues (SACKS; SCHEGLOFF; JEFFERSON, 1974; SCHEGLOFF; JEFFERSON; SACKS, 1977; SCHEGLOFF; SACKS, 1973), that characterize how people structure conversations.

2.1.2.1.
Turn-taking

The first important notion about conversations is their organization in turns, in a process defined as turn-taking. Sacks, Schegloff, and Jefferson (1974) explain that turn-taking is a type of organization used to order a series of social activities (e.g., games, customer service, traffic), including talk exchanges.

The authors provide a systematic model for describing turn-taking, showing that, in conversations, speaker change is recurrent and one speaker talks at a time, although there are brief occurrences in which multiple participants may speak simultaneously. Furthermore, the transitions between speakers' turns happen with as little gap (silence) or overlap (interruption) as possible, and both turn order and turn size are not fixed. Likewise, the length and content of conversations, the number of participants, and the turn distribution are not predefined. To select a conversation's next speaker, participants use turn-allocation techniques. Moreover, talks are not necessarily continuous and are composed of several turn-constructional units.

Finally, speakers may employ repair mechanisms when errors or violations occur (see subsection 2.1.2.3 for further repair definitions; SACKS; SCHEGLOFF; JEFFERSON, 1974).

Considering this definition of turn-taking, it is possible to identify similarities of human-human conversation to user-VA interaction. Firstly, voice interaction is essentially organized in turns: a back-and-forth style conversation between user and VA. Also, the system must, ideally, not interrupt the user (overlap), although the opposite is not true. Nevertheless, as pointed out by Moore & Arar (2019), not all features of talk-in interactions can be observed in the interaction with virtual agents. Since VAs need to process users' inputs, there is a gap between turns, and VAs' utterances (or range of utterances) are specified in advance by its developers. Additionally, the number of participants is unlikely to exceed two (VA and one user), as virtual agents still struggle with multi-party interactions (MOORE; ARAR, 2019).

2.1.2.2. Adjacency pairs

To identify the aspects that indicate the opening and closing of a conversation, Schegloff and Sacks (1973) presented the concept of adjacency pairs. According to the authors, adjacency pairs are a set of two utterances in which the first pair part requires a second, complementary part uttered by a different speaker. The second utterance does not need to follow the first pair part instantly to characterize an adjacency pair, but it must happen after the first utterance. Therefore, a relevant characteristic of adjacency pairs is their *sequential implicativeness*, that is, when speaker A produces the first part of an adjacency pair, they "projects for the sequentially following turn [speaker B's answer] the relevance of a determinate range of occurrences (be they utterance types, activities, speaker selections etc.)" (SCHEGLOFF; SACKS, 1973, p. 296). Question-answer, greeting-greeting, and offer-acceptance/ refusal combos are examples of adjacency pairs.

Due to its sequential implicativeness, Schegloff and Sacks (1973) argue that an essential feature of the adjacency pairs is that they allow speakers to present their comprehension and sense-making of the conversation. Hence, B's response (second part) to A's utterance (first part) supports B in showing whether they understood

A's intention and aids A in identifying whether B comprehended and accepted their utterance. This characteristic of adjacency pairs may present significant implications for voice interface design.

As voice interactions are composed of user input and system output, they are a set of adjacency pairs by nature. In some interactions, the first pair part is the user command, and the second is the VA response (figure 2.2):

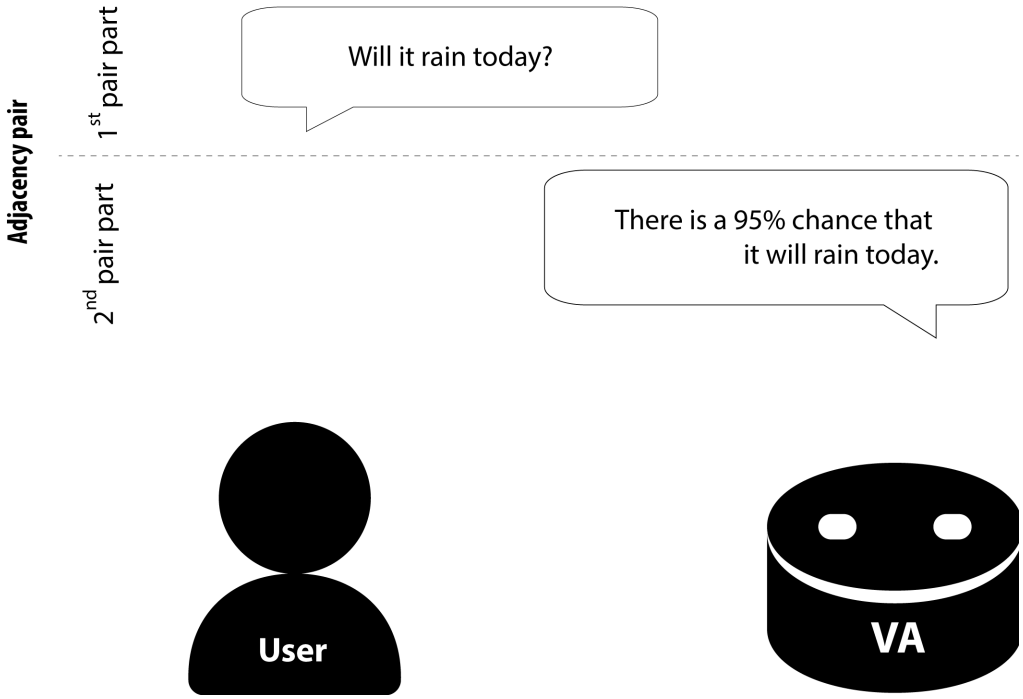


Figure 2.2 – User-VA dialogue 2 (Adjacency pair). Source: the authors.

Nevertheless, VAs responses may also initiate a new adjacency pair by posing a question to the user in order to perform a request (figure 2.3):

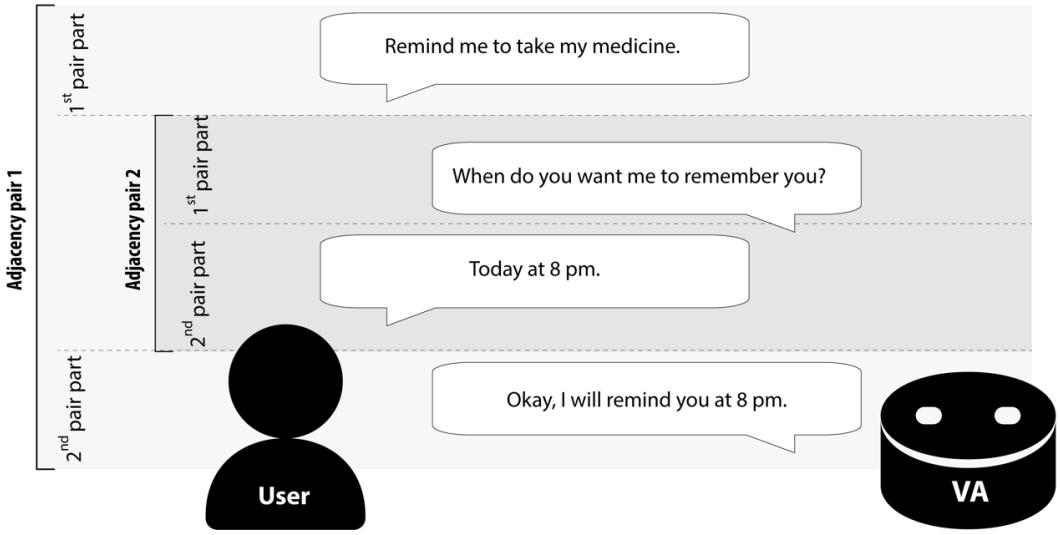


Figure 2.3 – User-VA dialogue 3 (Adjacency pair). Source: the authors.

As parts of an adjacency pair, VA responses have the intrinsic role of displaying the system's "understanding" of users' commands. To better visualize this, consider the first example (figure 2.2) presented above. The VA response appropriately answers the user's question, pointing to three implications: 1) the user's input was correctly captured, 2) the VA was able to understand the user's intention, and 3) the VA can perform the required action. As for the second situation (figure 2.3), the system output implicates 1 and 2, but by asking for additional details, it suggests that the information provided by the user was insufficient. Nevertheless, by creating a new adjacency pair, the VA indicates that further user action is needed and even directs the user to an appropriate type of answer (benefiting from sequential implicativeness).

Therefore, the nature of voice interaction as a group of adjacency pairs is influential for VA design since system responses become a means for users to comprehend VAs' interpretation of their commands. As will be discussed throughout this chapter, such a characteristic may be vital to mitigating errors throughout interactions.

2.1.2.3. Repair

Throughout daily conversations, issues constantly happen. People may violate the Cooperative Principle, mispronounce words, or even interrupt their

conversational partner. The comic below, inspired by a popular video, depicts a funny misunderstanding in a conversation between God and an angel (figure 2.4):

Figure 2.4 - A case of conversation repair. Adapted from a popular video[1] (PETERSEN, 2020). Illustration by Luiza Dias.

The dialogue above is merely a joke, but it creates humor by illustrating the type of conversational failure that requires what Schegloff, Jefferson, and Sacks (1977) define as *repair***.** According to the authors, when speakers are faced with issues in conversations, they try to fix them by correcting trouble sources. Although it was a little too late for the dinosaurs, a simple utterance such as "what?" is sufficient to initiate repair.

Schegloff, Jefferson, and Sacks (1977) consider repair as an action that generates a reaction, and therefore it can be divided into *initiation* and *outcome* (be it successful or not). Initiation concerns the conversational party who starts the repair action, which may be *self-initiated* or *other-initiated*. Differently, the outcome is related to the speaker that finishes the repair action (SCHEGLOFF; JEFFERSON; SACKS, 1977). Speakers may achieve outcomes through *self-repair*, conducted by

[1] https://www.tiktok.com/@lizemopetey/video/6870511001536646405 (PETERSEN, 2020)

speakers themselves, or *other-repair*, conducted by the conversational partner. Thus, the speaker who initiates repair is not necessarily the same who finishes it.

Schegloff, Jefferson, and Sacks (1977) comprise repair attempts into the following categories:

1. *Self-initiated self-repair*: Speaker A signalizes an issue in his own turn and corrects it themselves.
2. *Other-initiated self-repair*: Speaker B signalizes an issue in speaker A's turn, but speaker A is the one who corrects it.
3. *Self-initiated other-repair*: Speaker A signalizes an issue in his own turn, but speaker B is the one who corrects it.
4. *Other-initiated other-repair*: Speaker B signalizes speaker an issue in speaker A's turn and corrects it.
5. *Self-initiated failure*: Speaker A tries to initiate repair of an issue in his own turn, but both A and B cannot repair the conversation.
6. *Other-initiated failure*: Speaker B tries to initiate repair of an issue in speaker A's turn, but both A and B cannot repair the conversation.

As discussed before, users' interactions with VAs are not always successful and eventually require repair. The literature provides evidence that users of voice interfaces apply strategies to deal with query misrecognition, indicating attempts to repair the conversation (for a review, see chapter 3, section 3.2.3). Some examples of repairs in user-VA interactions are shown below (figure 2.5; for the sake of differentiating between human-human and user-VA conversations, repairs will be addressed as "user"- and "VA"-initiation and repair):

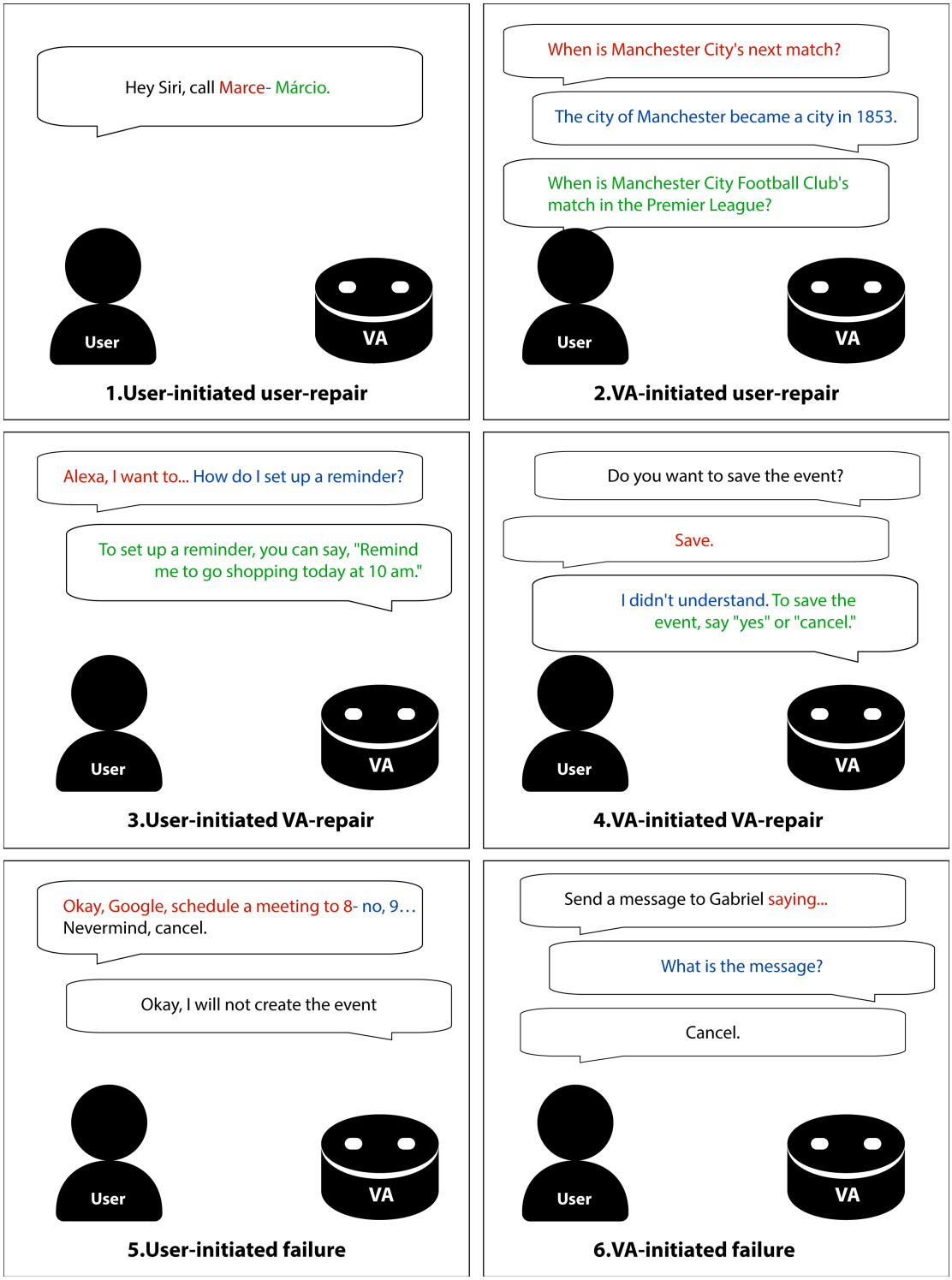


Figure 2.5 – Types of user-VA conversation repair. Red = issue; Blue = repair initiation; Green = repair. Source: the authors.

As can be observed in the dialogues above, both users and VAs may initiate and conduct the repair. However, instances in which users repair the interaction may be more frequent since, in most cases, VAs are not "aware" of their own mistakes (otherwise, they would not have performed that action in the first place). Even

in situations in which the assistant can *identify* an error (e.g., example 4), it might not be capable of *correcting* the user, as VAs can perform a wide range of tasks, and predicting what a user wants to say might be troublesome.

Hence, the before-mentioned role of system responses as parts of adjacency pairs may be essential to repairing interactions, as they communicate VAs' understanding of users' commands. In VA-initiated user-repairs (example 2), the system response indicates the *need for correction* and provides a *cue* for handling the issue, highlighting that appropriate system responses may be vital for successful interactions between users and VAs.

This section presented theories of how human-human conversations occur. It was shown that talk exchanges are a cooperative effort among speakers, achieved through a series of conversational maxims. Moreover, an essential aspect of conversations is their sequential order, that is, talk-in interactions are constituted of turns that are linked to make definitive sequences. These sequences can demonstrate speakers' understanding of the last turn, both in terms of a message's content (by inference) and the turn's completion (HUTCHBY; WOOFFITT, 1998). Finally, when issues happen, people are naturally prone to initiate repair.

Nevertheless, to fully comprehend users' communications with VAs, it is necessary to investigate their conversational behavior and the underlying cognitive processes that lead to such actions. The next section aims to present key concepts to language production and user-VA interaction: information processing, problem-solving, and decision-making.

2.2.
Conversations and Cognition

The way people talk, which follow the structures and conventions presented before, is a product of the human mind's cognitive processes. The first of these actions is called language processing: how people deal with the verbal messages (be it oral or written) they receive. According to Massaro (1975), language processing follows some stages between abstraction and meaning attribution to a received stimulus (i.e., speech). These stages are part of a more comprehensive process defined as information processing.

2.2.1.
Information processing

The cognitive action of information processing is responsible for the perception and meaning attribution of all stimuli that people receive. According to Wickens et al. (2014), there are three stages to information-processing: perception, transformation of information, and response selection. To illustrate this cognitive process, consider the simple dialog above between a user, Steve, and Siri:

Figure 2.6 – User-VA dialogue 4. Source: the authors.

As for any information humans receive, interactions with VA also go through information processing. Firstly, information is gathered by the *sensory systems* (i.e., ears) and perceived through a meaningful interpretation based on prior knowledge. The *perception* stage comprises three usually simultaneous perceptual processes: bottom-up feature analysis, unitization, and top-down processing. Wickens et al. (2014) explain that information is firstly analyzed through a bottom-up process, which relies exclusively on the stimulus itself (e.g., phonemes of spoken messages) rather than on past experiences. Nevertheless, as the stimuli occur jointly and are familiar to humans due to previous experiences, information becomes unitized (e.g., words, phrases), making the perceptual processing faster and more automatic. Finally, top-down processing performs inferences about an event based on

expectations derived from past experiences kept in long-term memory (e.g., the message).

In the example above, after Steve received Siri's response by his auditory system, his perception stage began with the identification of the phonological characteristics of the auditory message (i.e., disconnected phonemes: yo͞o SHo͝od tāk an' əmˈbrelə). Then, in the unitization phase, the perceived sounds started coming together to form words and phrases (e.g., əmˈbrelə stands for "umbrella"). As for the final perception stage, Steve relied on past experiences with the same type of stimuli (i.e., conversations, interactions) to make inferences about the message. Hence, knowing what "you", "should", "take", "an", "umbrella" means, Steve inferred that Siri is talking about the act of carrying the object umbrella.

Following perception, information is manipulated in working memory**,** which creates more permanent representations of information and retrieves familiar information from long-term memory. This stage is when people actually think or interpret information. For language processing, this stage attributes meaning for messages, makes inferences about the implication of these messages, and aids the decision of response approaches (WICKENS; LEE; LIU; BECKER, 2014).

It is necessary to note that two components are essential to information processing and, specifically, for information transformation: working memory and long-term memory. Wickens et al. (2014) define *working memory* as a type of memory storage responsible for temporarily holding *limited* amounts of information to be rehearsed or cognitively transformed. Contrarily, *long-term memory* is a mechanism to hold information for longer periods and retrieve such information (retrieval). Wickens et al. (2014) define the process of storing information in long-term memory as learning, in which working-memory is responsible for forming meaningful associations to be stored. Long-term memory's structure is associative: semantic networks keep related pieces of information in organized sections. One way information can be organized in long-term memory is through *mental models* (see chapter 3 for a complete definition).

As for Steve's interaction, he interprets what has been perceived in his working memory while retrieving information from long-term memory to aid his interpretation. Although Siri did not directly answer his question by saying "yes, it will rain", it implied this information by advising him to take an umbrella. Based on Steve's previous interactions and Siri's response, he uncovers two implications: 1)

Siri understood his command and was able to present a correct answer, and 2) It will rain.

Therefore, the information transformation stage is closely related to the system responses' role as a source of users' understanding of the system (as previously mentioned for the concept of sequential implicativeness; section 2.1.2.2). Users interpret messages and make implications about the system based on their previous knowledge and interactions, and then select a response to such an action.

Finally, the last stage of information processing is *response selection and execution*, which usually produces new information to be perceived, creating a feedback loop. In the case above, Steve decided not to answer Siri, as it completed the task. However, had Siri posed a new question (e.g., For which city?), Steve would have to answer the assistant, which would, in turn, provide more information for Steve to process.

Figure 2.7, retrieved from Wickens et al. (2014), summarizes the information processing action:

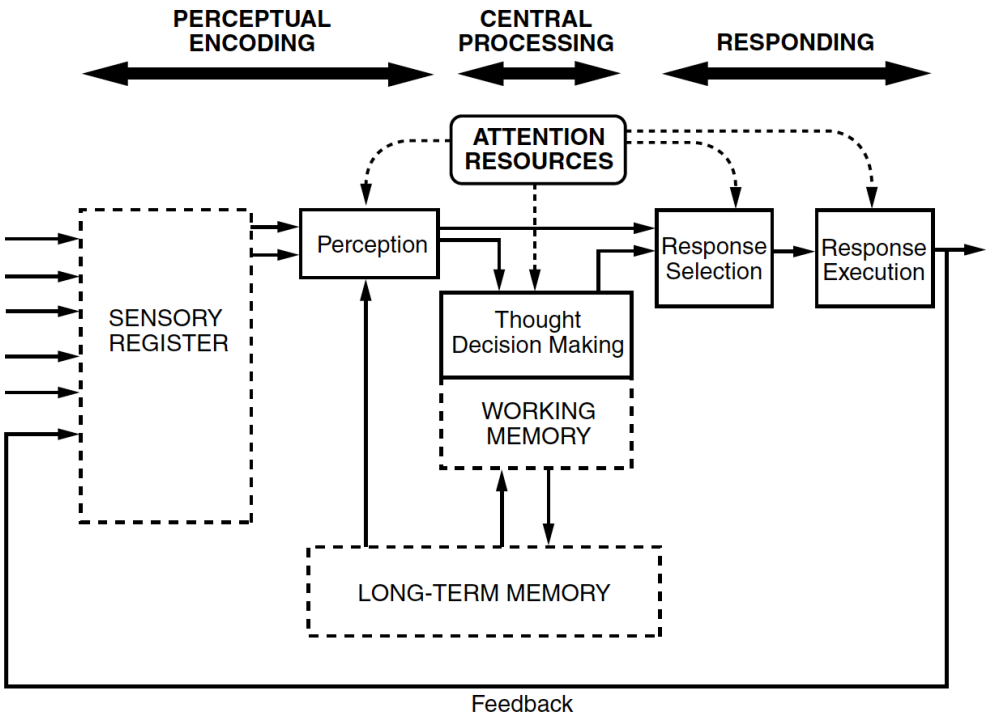


Figure 2.7 - A model of human information processing. Retrived from Wickens et al. (2014, p. 102).

As shown above, information processing is relevant for studying user-VA interactions since it is intrinsically related to language processing and response selection. Furthermore, information processing affects other cognitive activities that

impact conversation repairs, such as problem solving, troubleshooting, and decision making (HUTCHBY; WOOFFITT, 1998). The next section aims to present such concepts and their implications for VA design.

2.2.2.
Problem solving, troubleshooting, and decision making

Problem solving and *troubleshooting* are cognitive phenomena in which a human applies a series of often unconscious cognitive operations to go from an "initial state" to a "goal state" (WICKENS; LEE; LIU; BECKER, 2014), and which path between these stages are often unknown (LETHO; NAH; YI, 2012). Wickens et al. (2014) argue that troubleshooting differs from problem solving due to its focus on identifying a problem rather than finding a solution. Thus, although it is not always necessary to diagnose a problem to solve it (e.g., adding several ingredients to a soup until it tastes good), troubleshooting is usually a part of problem solving. Furthermore, while troubleshooting involves the performance of diagnosing tests to identify an issue (e.g., tasting the soup), problem solving often requires solutions to be employed (e.g., adding salt).

Closely related to problem solving, *decision making* is a process with a relatively long timeframe that involves selecting one among several options based on some knowledge about this alternative. Moreover, the outcome of the choice to be made is uncertain (WICKENS; LEE; LIU; BECKER, 2014). According to Letho et al. (2012), problem solving and decision making are overlapping processes since decision making can be considered problem solving and vice-versa. The authors argue that both consist in "a state of not having a selection toward a state with a selection" and "choosing a path out of potential paths or generating alternatives" (LETHO; NAH; YI, 2012, P. 230).

Considering these similarities, it can be argued that both issuing commands and repairing interactions are decision-making/ problem-solving activities. Users want to achieve a goal (e.g., search for information online) and, to do so, they need to choose how to speak their commands and find solutions to eventual failures. Therefore, interactions with VAs are as intrinsically related to information processing as these cognitive actions. Figure 2.8 illustrates how decision making/ problem solving overlaps with information processing.

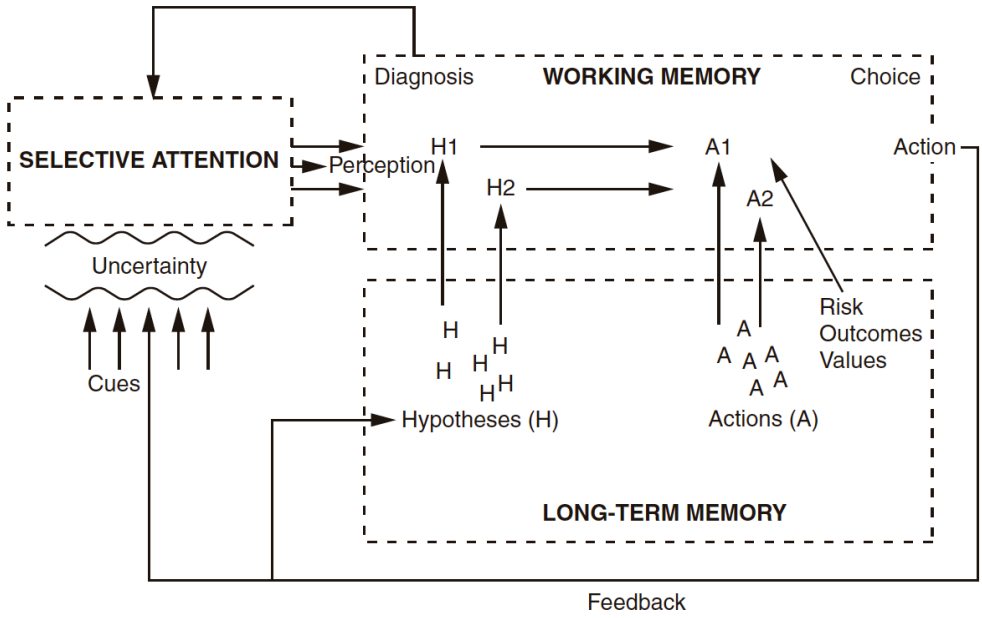


Figure 2.8 - Information processing and decision making. Retrieved from Wickens et al. (2014, p. 143).

Returning to the last section's example, Siri could have posed another question for the user, such as "For which city?". In this case, Steve would have to go through a decision-making process to answer.

Wickens et al. (2014) explain that external cues are received and perceived into working memory and then used to create hypotheses or diagnoses that attribute meaning to such cues concerning a system's current or future states. This process is supported by long-term memory. In the example, Steve could hypothesize that Siri could not complete the task due to lack of information (h1) or misrecognized his command and performed another task (h2), which are common issues during his interactions with VAs.

After that, the likelihood of these hypotheses' correctness is evaluated in working memory, and possible plans or actions are retrieved from long-term memory (WICKENS; LEE; LIU; BECKER, 2014). To do so, the user may consider, for example, the visual feedback written on his iPhone screen, showing that Siri correctly captured and recognized his command. Therefore, h2 could be considered less likely, and the user may proceed to evaluate actions based on h1.

The possible outcomes for such actions are also assessed to decide which path to take (WICKENS; LEE; LIU; BECKER, 2014). Due to information stored in his long-term memory, Steve probably knows that weather variates in different

locations. Hence, he might decide that Siri needs his location to provide correct weather data and proceeds to answer "Rio de Janeiro."

After a decision is made, new cues from the exterior (Siri) are provided as feedback, updating the operator's situation assessment and aiding the decision to conduct further actions. That is, when Siri completes the task by providing weather information for Rio de Janeiro city, Steve knows he made the correct decision.

As argued throughout this chapter, similar characteristics of human communication can be observed in users' interactions with voice assistants. This tendency means that users try to be cooperative and apply conversational rules to their commands, and also try to repair mistakes as they happen. Thus, as users are experts in engaging in conversations, interactions should be intuitive and with few errors. Notwithstanding, the literature shows that errors are recurrent in user-VA interaction.

A survey conducted by Ipsos (WHITE-SMITH *et al.*, 2019) showed that 23% of the participants agreed that the recognition of their commands by VAs are not very good. Accordingly, Maués (2019) gathered users' opinions on personified virtual assistants through focus groups and identified that the participants had a strong perception that their assistants had problems interpreting and fulfilling commands. Such errors throughout interactions may cause frustration for users (KISELEVA *et al.*, 2016b; LOPATOVSKA *et al.*, 2019; PURINGTON *et al.*, 2017) and ultimately impact VA adoption (BURBACH *et al.*, 2019; MCLEAN; OSEI-FRIMPONG, 2019; MORIUCHI, 2019).

The questions that arise from these findings are: If users actively attempt to provide good commands, why do so many errors occur in interactions with VAs? Also, if users are willing to repair issues throughout interactions, why are they frequently unable to recover from system errors? The answers to these questions may lay in two essential components presented in this chapter: 1) users' understanding of VAs, stored in their long-term memory and 2) VA responses, a part of the adjacency pair mechanism used for sense-making in conversations. Users' perceptions of VAs seem to be influential for their decisions on how to issue their commands while respecting the Cooperative Principle. Likewise, VA responses are sources of information for users to perceive, interpret, and make implications, and thus may affect problem-solving actions such as conversational repairs.

Considering these possibilities, the next chapter will present the concept of mental models - users' understandings of how a system works - and discuss its implications for VA design. Moreover, the state of the art of users' mental models of VAs will be presented, and potential issues around the topic will be discussed.

3
Mental Models and Voice Assistants

The previous chapter presented human communication tendencies and how they are related to user-VA interaction, as well as the underlying cognitive processes behind conversations. As discussed, information stored on users' long-term memory, collected from past interactions, are essential for users to understand how to use VAs and repair failed interactions. This chapter aims to define *mental models*, users' cognitive organization of such past experiences, and discuss the design implications of these models. In addition, this chapter intends to present the current state of the art of users' mental models of VAs, identified through a systematic literature review (SLR; see chapter 4 for the detailed methodology).

3.1.
Mental models in Human-Computer Interaction

Mental models are a type of conceptual model people create to represent how a product or system works, although it may not necessarily match the actual system functioning (NORMAN, 2013). More specifically, mental models comprise a set of expectations about a system's components, functioning, and proper usage (WICKENS; LEE; LIU; BECKER, 2014). Correct mental models help users predict a system's behavior and comprehend its actions in unforeseen situations. Otherwise, poor models may lead to a decreased understanding of the system's functioning, which might be problematic for unexpected interactions (NORMAN, 2013).

Wilson and Rutherford (1989) explain that users create their mental models "based on previous experience as well as current observation, which provides most (if not all) of their subsequent system understanding and consequently dictates the level of task performance." (WILSON; RUTHERFORD, 1989, p. 3). Similarly, Norman (2013) argues that users rely on available information to form their mental

models: perceivable physical cues of the product itself, past experiences, advertisements, manuals, and others. The combination of these factors is defined by the author as the *system image* and highlights the role of a product's design for users to develop correct mental models.

Norman (2013) proposed the diagram illustrated in figure 3.1 to show that designers, users, and the system image comprise three vertices of a triangular relationship. Designers have their own conceptual models of an object (a) but, as the product is no longer with them, it becomes an isolated vertex (b), represented only by its physical structure (i.e., the system image). While designers wish that users' conceptual models (c) are equal to their own, it is usually impossible for users to communicate with designers. Therefore, users are left to make inferences based on their interactions with the product itself, highly relying on the system image.

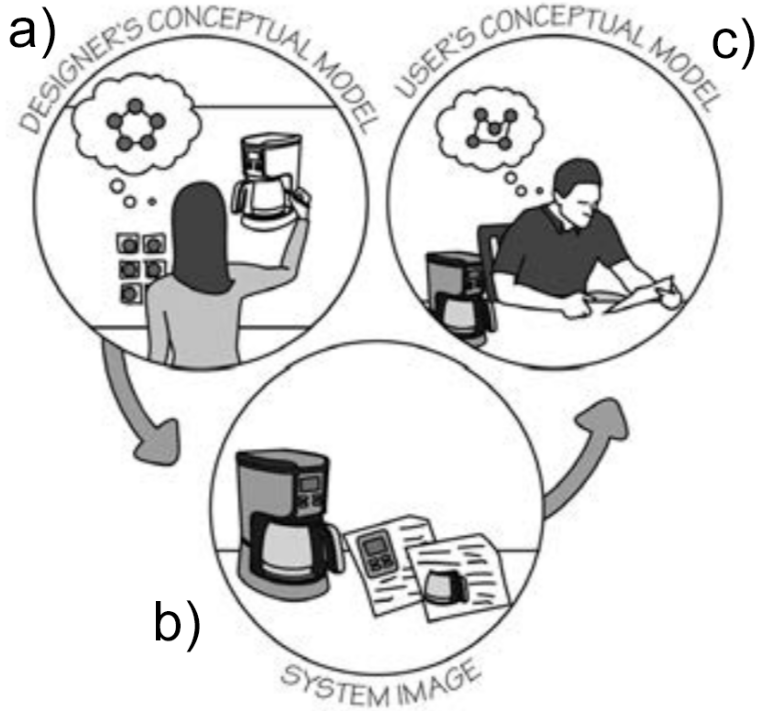


Figure 3.1. – Norman's diagram of conceptual models. Adapted from Norman (2013, p. 68).

The concept of system image may be related to the last chapter's discussions concerning the role of VA responses for users' interactions. As for VAs, the system image may differ for different devices. A smart speaker such as Amazon Echo has its system image composed not only of the device's hardware but also from Alexa's voice, system outputs, documentation, and Alexa's app, which works as visual

support. Differently, VAs on smartphones such as Siri may also have their system image composed of information on the device's screen, as illustrated by figure 3.2.

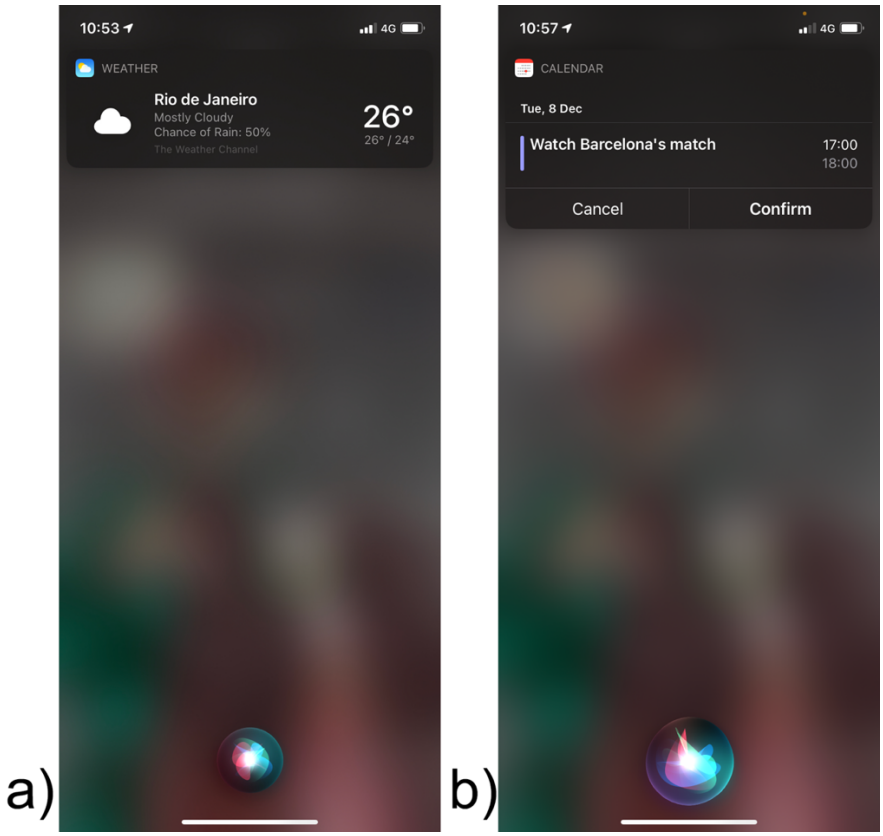


Figure 3.2. - Siri's interface. Check the weather (a) and to Add an appointment to the calendar (b).

Mental models are closely related to how people perform tasks. Norman (2013) argues that when users interact with a product,

> they face two gulfs: the Gulf of Execution, where they try to figure out how it operates, and the Gulf of Evaluation, where they try to figure out what happened. […] The Gulf of Evaluation reflects the amount of effort that the person must make to interpret the physical state of the device and to determine how well the expectations and intentions have been met (NORMAN, 2013, p. 77).

According to the author, a designer's job is bridging such gulfs, that is, making them small by providing clear and objective information about the system operation and states. Both gulfs are bridged through the use of conceptual models, as well as signifiers and constraints (Gulf of Execution) and feedback (Gulf of Evaluation).

To illustrate this, a hypothetical scenario of a users' interaction with Alexa will be considered. The user wants to set the alarm clock for the following morning

and decides to ask Alexa. They know, from previously collected information (system image), that Alexa can perform such a task (mental model). The user is also aware that they need to say the wake-word "Alexa" before issuing their request (mental model), as learned from the device's initial setup (signifier/ gulf of execution). Therefore, they say, "Alexa, wake me up tomorrow at 7 am", but Alexa gives no response (lack of feedback). From past interactions, they know that the Echo turns on a blue light when Alexa is capturing user input (mental model), and therefore concludes that they were not heard and need to issue their command again (gulf of evaluation).

The previous example shows the importance of both mental models and design aspects that compose a system's image. If the user did not have a correct mental model of the system operation (i.e., blue lights indicate turn-taking), they would not have been able to recover from Alexa's error (i.e., failure to capture the command). At the same time, have design aspects been absent (i.e., no lighting mechanism has been implemented), the user's evaluation of the system state would not have been possible. Therefore, design aspects must form a system image that leads users to develop correct mental models of a product, which bridges both the gulf of execution and evaluation.

3.2. Users' mental models of Voice Assistants

Designing products that align users' perceptions with actual system capacity is essential for a product's usability. To help developers match users' expectations with VAs' capacity, it is first necessary to understand *how* these mental models can be described and identify potential misinformation sources. Although some studies have pointed out issues around users' mental models of VAs (CHO; LEE; LEE, 2019; LUGER; SELLEN, 2016), to the extent of our knowledge, no systematic review has been conducted to identify patterns and discrepancies on this topic across the literature. Thus, a systematic literature review (SLR) was conducted to identify the state of the art of users' mental models of VAs.

The SLR aimed to answer the following research questions (RQs):

- **RQ1:** How can users' mental models of VAs be described in terms of system features, functioning, and way to use?

- **RQ2:** How does previous experience with VAs affect users' mental models? We created this RQ since the literature indicates that users' mental models may vary according to previous system experience (NORMAN, 2013; WICKENS; LEE; LIU; BECKER, 2014).

We collected 557 primary studies and accepted 57 based on inclusion and exclusion criteria and a quality assessment checklist (see chapter 5 for the methodology). This section reports the findings from the full-text revision of these 57 primary studies, divided into five findings' groups. For the first research question, we created three major groups following Wickens et al.'s (2014) definition of mental models: 1) users' expectations for VAs' features, 2) users' understanding of VAs' functioning, and 3) users' perceptions of VAs' usage. Nevertheless, as we observed during the analysis that the "learning" topic was found for all three categories, we added the extra category "4) users' learning practices". Group 5 comprised contributions to the second RQ. At the end of this chapter, we summarize the results by presenting a framework illustrating the impacting factors for users' mental models of VAs according to the SLR.

3.2.1. Users' expectations of VAs' features

Group 1 comprises users' expectations of features supported by VAs and tasks they can or wish to accomplish with these interfaces. The first emerging topic was *the relationship between physical spaces and users' expectations for VAs' features*. Through a survey, Cambre et al. (2020) asked users to imagine future VA usage scenarios and observed that participants pictured assistants as bearing separate identities and capable of moving freely through devices to achieve tasks. Similar results were found by Lee, Cho, and Lee (2020), who assessed the effects of a smart speaker's physical presence on users' mental models. The authors observed that, in the presence of visible devices, users perceived VAs as separate entities and attributed different roles and expertise according to their placement.

This relationship may encompass not only users' perceptions of VAs' roles but also their usage behavior. Abdolrahmani et al. (2020) interviewed legally blind users and identified that they employ different devices to perform varied task types based on each device's perceived strengths. Li, Rau, and Huang (2019) observed

through a Wizard-of-Oz experiment that users believe VAs should provide more services in living spaces than in workspaces, and such understanding ultimately influenced their willingness to disclose information to VAs. Likewise, Lopatovska et al. (2019) reported that users who placed smart speakers in kitchens or living rooms had different task usage patterns than users who placed their devices in bedrooms. Accordingly, users' surroundings and a command's nature seem to affect their usage behavior (KENDALL; CHAUDHURI; BHALLA, 2020). Porcheron et al. (2018) argue that users are often surrounded by third parties and need to account for social settings while interacting. Issues related to background noise and, especially, a sense of social embarrassment to interact in front of others have been shown to undermine interactions (BALASURIYA *et al.*, 2018; COWAN *et al.*, 2017; LOPATOVSKA; OROPEZA, 2018; PARK; LIM, 2020; TRAJKOVA; MARTIN-HAMMOND, 2020).

The second emerging topic is that *users' utilization of features is limited to a set of tasks.* Commonly used features reported in the literature are playing music, checking the weather, and setting timers, alarms, and reminders (AMMARI *et al.*, 2019; CHO; LEE; LEE, 2019; GARG; SENGUPTA, 2020; HUXOHL *et al.*, 2019; LOPATOVSKA; OROPEZA, 2018; LUGER; SELLEN, 2016; OH; CHUNG; JU, 2020; PRADHAN; MEHTA; FINDLATER, 2018; PRIDMORE *et al.*, 2019; TRAJKOVA; MARTIN-HAMMOND, 2020; WEBER; LUDWIG, 2020; YANG; AURISICCHIO; BAXTER, 2019). Users also utilize VAs for looking up information such as recommendations on places to eat or visit, recipes, information about sports and culture, and for learning-related activities. Another frequently mentioned task in the literature is controlling Internet of Things (IoT) devices such as lights and thermostats (AMMARI *et al.*, 2019; CHO; LEE; LEE, 2019; GARG; SENGUPTA, 2020; HUXOHL *et al.*, 2019; KENDALL; CHAUDHURI; BHALLA, 2020; LOPATOVSKA *et al.*, 2020; LOVATO; PIPER; WARTELLA, 2019; PRADHAN; MEHTA; FINDLATER, 2018; PRIDMORE *et al.*, 2019; WEBER; LUDWIG, 2020; YANG; AURISICCHIO; BAXTER, 2019). Moreover, VAs are used for entertainment purposes such as telling jokes, playing games, and exploring the VAs' personality, particularly by children and people with intellectual disabilities (BALASURIYA *et al.*, 2018; CHO; LEE; LEE, 2019; FESTERLING; SIRAJ, 2020; GARG; SENGUPTA, 2020; KENDALL; CHAUDHURI; BHALLA, 2020; LOVATO; PIPER; WARTELLA, 2019; PRIDMORE *et al.*, 2019;

TRAJKOVA; MARTIN-HAMMOND, 2020; WEBER; LUDWIG, 2020). Additionally, studies focusing on people with disabilities identified other uses for VAs, such as productivity-related tasks (e.g., writing emails, managing a calendar) and safety improvements (AMMARI *et al.*, 2019; PRADHAN; MEHTA; FINDLATER, 2018).

Although the literature indicates the underutilization of VA tasks, *users perceive the features supported by VAs to be insufficient to their needs.* Primary studies showed that users consider available tasks insufficient or useless, letting down their initial expectations and leading to frustration, abandonment of the VA, underutilization of features, and perceptions of VAs as toys (BENETEAU *et al.*, 2020; CHO; LEE; LEE, 2019; OH; CHUNG; JU, 2020; PRADHAN; MEHTA; FINDLATER, 2018; TRAJKOVA; MARTIN-HAMMOND, 2020; WEBER; LUDWIG, 2020). VAs have also been judged limited compared to humans, not trustworthy even for simple tasks due to inconsistent performance, and to lack accessibility (COWAN *et al.*, 2017; PRADHAN; MEHTA; FINDLATER, 2018).

In line with the previous topic, we observed that *users wished VAs could do more*. These requests included controlling mental and physical health states for elders and people with disabilities, enabling macros and logic-based commands, improving hardware, increasing home safety, and managing family schedules (CLARK; PANTIDI; *et al.*, 2019; HUXOHL *et al.*, 2019; LOPATOVSKA *et al.*, 2020; PARK; LIM, 2020; TRAJKOVA; MARTIN-HAMMOND, 2020; WEBER; LUDWIG, 2020). The most prominent of users' aspirations was integration with IoT devices, including offline modes. Users desire a smart home full of IoT devices, valuing the VA as an agent that brings together all products without requiring access to multiple third-party apps (AMMARI *et al.*, 2019; CAMBRE *et al.*, 2020; CHO; LEE; LEE, 2019; HUXOHL *et al.*, 2019; LOPATOVSKA *et al.*, 2020; OH; CHUNG; JU, 2020; PRADHAN; MEHTA; FINDLATER, 2018; WEBER; LUDWIG, 2020).

Moreover, the SLR indicated *users' desire for developing personal relationships with VAs.* This topic comprised two distinct aspirations. The first is a wish for customization and personalization and integration with other products and services (ABDOLRAHMANI *et al.*, 2020; COWAN *et al.*, 2017; HUXOHL *et al.*, 2019; LOPATOVSKA *et al.*, 2020; WEBER; LUDWIG, 2020). The second is to develop an emotional relationship with the VA, a theme in which findings were

heterogeneous across primary studies. Cho, Lee, and Cho (2019) conducted a long-term study by observing participants' interactions and gathering users' opinions about Alexa. Users aspired to have a unique, distinguishable agent who could engage in emotional exchanges and develop relationships with users. Similarly, Park and Lim (2020) asked families to propose an imaginary VA and identified that the hypothetical assistants had an essential role in family dynamics: aiding family conversations, supporting resting rituals, managing family memories, and repairing emotional conflicts. Other studies strengthen the perceptions of VAs as a family member, companion, or at least as a social entity (CAMBRE *et al.*, 2020; GARG; SENGUPTA, 2020; OH; CHUNG; JU, 2020; PRADHAN; MEHTA; FINDLATER, 2018; TRAJKOVA; MARTIN-HAMMOND, 2020; XU; WARSCHAUER, 2020b).

Oppositely, some primary studies concluded that users have more instrumental expectations for VAs, recognizing it as knowledgeable for informational purposes but unable to display warmth (DOYLE *et al.*, 2019; FESTERLING; SIRAJ, 2020; GARG; SENGUPTA, 2020). Studies have also shown that users do not want to bond with VAs probably due to the assistants' limitation as dialogue partners and the understanding that these systems are subservient tools to users (CLARK; PANTIDI; *et al.*, 2019; FESTERLING; SIRAJ, 2020).

3.2.2.
Users' understanding of VAs' functioning

Group 2 presents users' understanding of VAs' interactional and technical functioning. As with the previously discussed theme, primary studies showed discrepancies in *users' understanding of VAs' conversational capabilities.* Some studies suggested that users thought of VAs as a person and initially expected to interact in a human-like manner. We observed frequent mentions to desires of engaging in more natural conversations, which encompassed accurate speech recognition (including accents and foreign languages), the ability to maintain a conversation for several turns, and proper speaker recognition. (CHO; LEE; LEE, 2019; HUXOHL *et al.*, 2019; LOPATOVSKA *et al.*, 2020; LOVATO; PIPER; WARTELLA, 2019; LUGER; SELLEN, 2016; OH; CHUNG; JU, 2020; WEBER; LUDWIG, 2020; XU; WARSCHAUER, 2020b). Contrariwise, other publications reported that users

considered interactions to be command-based, unauthentic conversations, missing the capacity to maintain conversation flows (CLARK; PANTIDI; *et al.*, 2019; COWAN *et al.*, 2017; DUBIEL *et al.*, 2020; HUXOHL *et al.*, 2019; PORCHERON *et al.*, 2018).

We also observed *conflicting perceptions regarding VAs' humanness levels*, even among participants from the same study. Xu and Warschauer (2020a) uncovered children's perceptions and did not observe a consensus among participants regarding whether VAs' capabilities arose from human technology (i.e., men made) or natural intelligence (i.e., an aware entity). Likewise, Guzman (2019) identified through interviews that while some users perceived VAs as embedded in a device (e.g., a smartphone's feature), others thought of these systems as an independent entity, that is, the interaction's source. Bonfert et al. (2018) developed a VA that rebuked users' impolite behaviors and collected assumptions about the interface, pointing out that while some users considered virtual agents an emotionless machine, others believed they deserved politeness and should be treated like humans. Furthermore, the literature suggested that users are divided regarding anthropomorphized agents' likability (WEBER; LUDWIG, 2020; YAROSH *et al.*, 2018).

Although the literature does not fully address the reasons for the inconsistencies in users' perceptions, we observed that *VA design influences such perceptions*. Firstly, social cues (e.g., name) and humoristic interactions (e.g., jokes) may elicit the image of an intelligent, human-like agent and impact users' perception of VAs' embodiment in a device (CAMBRE *et al.*, 2020; GUZMAN, 2019; LUGER; SELLEN, 2016). Inconsistently, funny VA responses have been reported to be considered both funny and welcome and fake (DOYLE *et al.*, 2019; LOPATOVSKA, 2020). As for likability, Kuzminykh et al. (2020) identified that users perceive VAs as distant or approachable depending on characteristics such as the capability to provide information, interaction style (i.e., responsiveness to prompts, jokes), voice, and companionship. The authors found that approachability was related to system trust and expectations for warmth. Displaying empathy and engaging in reciprocal self-disclosure have also been demonstrated to affect an agent's likability (CHIN; MOLEFI; YI, 2020; LI; RAU, 2019).

Doyle et al. (2019) identified further impacting constructs for perceived humanness: type and level of knowledge (also indicated by Lovato, Piper, and Wartella, 2019), interpersonal connections, linguistic content, conversational

interactivity, dialogue performance, sense of identity, voice's expressiveness and clarity, and behavioral affordances. Particularly, the literature indicates that anthropomorphic perceptions can be yielded solely by speech and that VAs' voices impact several perceptions (CHO; LEE; LEE, 2019; LOVATO; PIPER; WARTELLA, 2019; XU; WARSCHAUER, 2020b). Cowan et al. (2015) assessed computational voices with different anthropomorphic levels and found that a human-like voice was perceived as the most capable, competent, flexible, and trustworthy among the conditions. Other studies reinforce that, compared to synthetic voices, anthropomorphized voices increase an agent's trustworthiness and lead to augmented involvement and behavioral intentions to make purchases. Nuanced voice features, such as accents, may also implicate the VAs' personality (CHÉRIF; LEMOINE, 2019; COWAN *et al.*, 2017; DUBIEL *et al.*, 2020).

Although users have been demonstrated to have high hopes for VAs' functioning, *such perceptions may differ from actual system capabilities, leading to frustration and abandonment when expectations are not met*. Cho, Lee, and Cho (2019) reported that users initially wished to engage in conversations with Alexa but were disappointed with its intelligence and ability. Similarly, Luger and Sellen (2016) argue that VAs' playful interactions may act as affordances that convey system capabilities that are beyond reality, ultimately leading to frustration. Differently, studies have found that successful results do not guarantee user satisfaction, as adequate outcomes may be rated as unsatisfactory and vice-versa (BALASURIYA *et al.*, 2018; KISELEVA *et al.*, 2016a; LOPATOVSKA *et al.*, 2019). For example, the literature suggests that children (FESTERLING; SIRAJ, 2020) and users with intellectual disabilities (BALASURIYA *et al.*, 2018) may attribute more value to the act of interacting alone than to system outcomes. Hence, anthropomorphism should not be disregarded.

Our analysis indicated another emerging topic related to VA design: *the role of VAs' feedback*. Lopatovska et al. (2020) gathered users' recommendations for VAs through focus groups and showed that users consider VAs to have low transparency, leading to a lack of understanding of the VAs' functioning and frustration. Consistently, Weber and Ludwig (2020) showed that users recommend that VAs have feedback for interactions. As for feedback type, studies showed that participants consider audio-only feedback confusing when compared with visual feedback, and believe that animations and colored lights are appropriate for indicating

conversational turns (BALASURIYA *et al.*, 2018; DOYLE *et al.*, 2019; HUXOHL *et al.*, 2019; LEE; CHO; LEE, 2020).

This category's final theme revolved around the VAs' technical operation and its implications for *privacy concerns.* The literature indicated that users are concerned about the privacy of their data regarding surveillance and data protection, especially for sensitive information (AMMARI *et al.*, 2019; COWAN *et al.*, 2017; HUXOHL *et al.*, 2019; JAVED; SETHI; JADOUN, 2019; LAU; ZIMMERMAN; SCHAUB, 2018; PRIDMORE *et al.*, 2019; WEBER; LUDWIG, 2020). However, studies point to the existence of a trade-off between privacy and convenience that encouraged some users to utilize VAs despite being worried about privacy. Other reasons for trusting VAs included the belief that handling huge data amounts would be unfeasible for companies, relationships of trust with developers, and a sense that data collection is necessary for user profiling (LAU; ZIMMERMAN; SCHAUB, 2018; PRIDMORE *et al.*, 2019). Notwithstanding, most users are unaware of privacy-related information, such as knowledge about data collection, storage, and sharing, and the possibility of viewing, editing, or deleting history logs (AMMARI *et al.*, 2019; COWAN *et al.*, 2017; JAVED; SETHI; JADOUN, 2019; WEBER; LUDWIG, 2020). While unawareness impacts users' trust in VAs (COWAN *et al.*, 2017), Lau, Zimmerman, and Schaub (2018) suggested that current privacy controls are viewed as effortful, confusing, or insufficient, and preventive features (e.g., incognito mode) are preferred over such retroactive controls.

3.2.3.
Users' perceptions of VAs' usage

This section presents contributions that address people's understanding of how to speak to VAs. Firstly, we observed that *people tend to apply habits of human-human conversations to user-VA interactions.* These include non-verbal communication, implicit contextual references (e.g., saying "this phone" when referring to VAs' device), indirect requests (e.g., saying "alright" to request a recipe's next step), and imprecise temporal expressions (e.g., "schedule appointment *later today*") (FOURNEY; DUMAIS, 2016; RONG *et al.*, 2017; VTYURINA; FOURNEY, 2018; XU; WARSCHAUER, 2020a). However, such behavior may be an unconscious action caused by the naturality of speech. Cowan and colleagues

(2015, 2019) suggest that partner models influence users' speech irrespectively of the interlocutor (e.g., VA or human). The authors found that both variations in lexical choices caused by an accented speech (COWAN *et al.*, 2019) and employment of syntactic alignment (i.e., use of a grammatical structure that matches the conversational partner; COWAN *et al.*, 2015) can be observed in people's communication regardless of the interlocutor (i.e., human or computer). As demonstrated by Wu et al. (2020), a particular device type (smartphone or smart speaker) is also not determinant for lexical choices in user-VA interactions.

Despite people's seemingly natural tendency to translate human-human dialogue trends to user-VA interaction, we observed that *people tend to adapt their speech to interact with VAs.* Such adaptations include pronouncing words more accurately, removing words, using specific terms, changing accents, and speaking louder or more clearly. As VAs fail to recognize contextual references to past interactions and physical locations, users also remove these types of utterances from their commands (AMMARI *et al.*, 2019; BALASURIYA *et al.*, 2018; COWAN *et al.*, 2017; DOYLE *et al.*, 2019; LUGER; SELLEN, 2016; PRADHAN; MEHTA; FINDLATER, 2018). The need for such adaptations is likely caused by VAs' speech recognition failures and limitations and may cognitively strain users (CHO; LEE; LEE, 2019). Similarly, *users of voice interfaces apply strategies to deal with errors.* When faced with failures, users try to repair interactions by repeating requests, adjusting a command's structure, wording, or information amounts, changing pronunciation, and speaking louder (BENETEAU *et al.*, 2019; GARG; SENGUPTA, 2020; LOVATO; PIPER; WARTELLA, 2019; PORCHERON *et al.*, 2018; PORCHERON; FISCHER; SHARPLES, 2017; YAROSH *et al.*, 2018).

However, users do not randomly employ such tactics. Instead*, system responses may be significant for error repair*. Myers et al. (2018) showed a relationship between obstacle types and error handling tactics applied by participants using a voice-based calendar: while misrecognition errors were highly related to hyper articulation, other types of mistakes (unfamiliar intent, failed feedback, system error) caused users to adopt a range of frustration-related strategies. Similarly, Porcheron and colleagues observed users' interactions with conversational agents and showed that VAs' responses were indicators for failures, serving as resources for users to reason and interpret outputs, identify errors, and reformulate commands (PORCHERON *et al.*, 2018; PORCHERON; FISCHER; SHARPLES, 2017).

Reinforcing this idea, Kirschthaler, Porcheron, and Fischer (2020) demonstrated that increasing a voice interface's discoverability (e.g., providing instructions) led to better performance and usability. Further evidence reinforces the idea that VA responses may affect users' attitudes and queries (BONFERT *et al.*, 2018; CHIN; MOLEFI; YI, 2020; XU; WARSCHAUER, 2020a).

3.2.4.
Users' learning practices

As explained previously, emerging topics around learning were recurrent in the primary studies. To present this topic, we considered the four learning methods suggested by Beneteau et al. (2020), who conducted a long-term study to understand how families learn about a Alexa. Firstly, *the VA may present information about itself* (e.g., "I can set up a timer") or *the user may ask the VA about it* (e.g., "What can you do?"). The literature suggests that active user requests for information often aim at learning the VAs' personality, opinions, and capabilities, but VAs offer little support for learning, causing users to struggle for mastering the system (HUXOHL *et al.*, 2019; LOPATOVSKA, 2020; PRADHAN; MEHTA; FINDLATER, 2018; WEBER; LUDWIG, 2020; YAROSH *et al.*, 2018). Another technique was to *learn about the system from third parties* (i.e., friends, family, advertisements) (BENETEAU *et al.*, 2020), but no other publication approached this method.

Finally, the most cited practice in the literature was *the trial-and-error approach*. Users try to learn about the VAs' functioning, features, and usage through experimenting: issuing commands to understand the VAs' capabilities and the appropriate syntax, rhythm, intonation, accents, and tactics they should apply (BENETEAU *et al.*, 2020; LUGER; SELLEN, 2016; TRAJKOVA; MARTIN-HAMMOND, 2020). As argued by Myers, Furqan, and Zhu (2019), who observed participants' interactions with a voice-based calendar, users who explored the voice interface's limits were more effective in learning about supported intents and utterances types. Nevertheless, employing such an approach was often effortful, and not all users were willing to engage in such a method (BENETEAU *et al.*, 2020; LUGER; SELLEN, 2016; TRAJKOVA; MARTIN-HAMMOND, 2020; WEBER; LUDWIG, 2020).

3.2.5.
Users' VA previous experience and mental models

In the last group, we identified that *users' previous experience with VAs affects interaction performance*. Primary studies indicate that heavy and expert users are more efficient and self-reportedly more successful in their interactions since they are used to VAs' communication requirements and actively search for VAs' capacities when the system provides good results (LOPATOVSKA *et al.*, 2019; LUGER; SELLEN, 2016; MYERS; FURQAN; ZHU, 2019). Similarly, Myers et al. (2019) modelled participants' behavior when interacting with a voice calendar and observed three types of users: proficient, explorers, and strugglers. Although the authors did not control for VA experience, proficient users, who relied less on the visual menu and produced less utterances and misfires, had better performance metrics. In a similar stream, Lau, Zimmerman, and Schaub (2018) identified that light users are more likely to be unaware of privacy controls.

Notwithstanding, conflicting literature's findings make it *unclear if these experiences affect other perceptions*. Lopatovska and Willians (2018) identified no differences in anthropomorphizing behaviors between heavy and light users. As for perceived ease of use, Jung, Kim, and Ha (2020) observed that user types were affected by different types of obstacles: while heavy users were more sensitive to failures and desired more control over the interaction, light users preferred more guided interactions. Contrarily, Chen and Wang (2018) assessed the effects of prior experience with a conversational agent on users' perceived usability of Siri and found that inexperienced users perceived the system to be less usable than experienced participants.

Furthermore, users' mental models might be influenced not only by previous experience with voice interaction but also by *users' technical backgrounds.* Chen & Wang (2018) observed that technical knowledge acted as a moderating factor between perceived usability and VA experience, as technical users with no VA experience had their perceived usability restored after experiencing issues. Similarly, Luger and Sellen (2016) found that users who considered themselves more technically knowledgeable had lower initial expectations for VAs' capabilities and intelligence and were more forgiving of errors. Age, which is hypothesized to be related to technical comprehension of machines (e.g., young adults vs. elders,

young vs. older children), might also influence VAs' perceived social skills and humanness (OH; CHUNG; JU, 2020; XU; WARSCHAUER, 2020b). Nevertheless, the *influence of users' technical background is also contradictory among studies*. Myers, Furqan, and Zhu (2019) did not observe any effect of programming experience on performance metrics other than verbosity levels. Similarly, Javed, Sethi, and Jadoun (2019) identified no correlation between technical background and knowledge about privacy-related information.

3.2.6. Framework

Resulting from the literature analysis, figure 3.3 illustrates a framework describing factors related to users' mental models of VAs, including design aspects and users' conceptualizations and behaviors. We remark that our representation does not aim to establish a conclusive framework to describe users' mental models of VAs, nor to assert all factors relevant to this concept. Rather, our contribution lies in identifying emergent themes in the literature and their observed relationships. In addition to indicating potential gaps and significant research topics, the framework intends to organize the analyzed literature according to a specific definition of mental models since most primary studies did not explicitly address such a concept but present relevant findings for its uncovering. Thus, the framework was developed as an attempt to illustrate the state of the art in the literature by clustering similar findings in broad categories and highlighting observed relationships among them, thus visually representing the last sections' results.

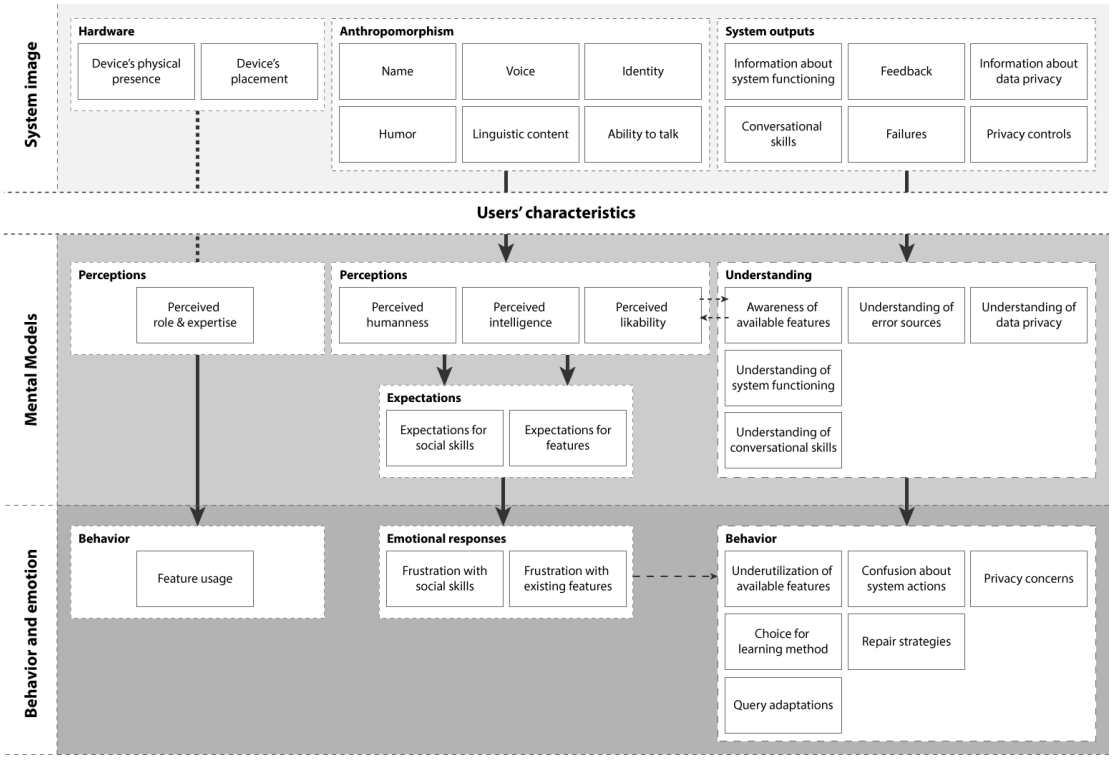


Figure 3.3. – Framework illustrating the SLR's results.

On the framework's first layer, we positioned VA design aspects that represent the system image: factors related to *hardware*, characteristics related to *anthropomorphism*, and types of *system outputs* that explicitly or implicitly present information about the system or the interaction for users. Consistently with Norman (2013), the system image seemed to affect users' mental models represented in the second layer.

The second level illustrates the dimensions considered relevant to users' mental models. We divided users' models as follows. "Perceptions" are related to the VAs' social nature and more subjective variables such as expertise, humanness, intelligence, and likability. "Understanding" addressed topics related to the system's actual technical capabilities and functioning. "Expectations" describe how users' wish VAs would behave.

The influence of the system image on users' perceptions appears to be moderated by varying users' characteristics, including age, previous experience with VAs, and technical background. Finally, we added the third layer, "behavior and emotion", since the analysis suggests that users' mental models impact their behavior and feelings towards interactions, as studies have indicated (KIERAS; BOVAIR, 1984; NORMAN, 2013).

This chapter exposed the concept of mental models in HCI and described of the state-of-the-art in the literature on users' mental models of VAs. As discussed, mental models are vital for how users utilize a product and are closely related to the levels of task performance. Nevertheless, as indicated by the SLR, users have high expectations for VAs' intelligence and capabilities, provoked mainly by anthropomorphic features that induce high levels of anthropomorphism on VAs. Moreover, the literature suggests that users have an inaccurate comprehension of VAs functioning, including information on the VA operation, data handling, feedback, failures, available features, error sources, and speech recognition thresholds. Such unawareness of VAs' functional aspects suggests that the way information is presented on VAs' outputs (i.e., feedback, instructions, cues, etc.) might be inappropriate.

Due to the unaligned mental models described above, users underutilize or abandon features, report confusion regarding actions and error sources, express privacy concerns, and employ effortful and inefficient learning methods such as trial-and-error approaches. Such behavior may be related to relevant barriers for VA adoption, such as negative attitudes towards VAs, privacy concerns, and perceptions that VAs have low usefulness and ease of use (BURBACH *et al.*, 2019; MAUÉS, 2019; MCLEAN; OSEI-FRIMPONG, 2019; MORIUCHI, 2019; MOTTA; QUARESMA, 2019; PITARDI; MARRIOTT, 2021; PURINGTON *et al.*, 2017; ROBART, 2017; WHITE-SMITH *et al.*, 2019).

A previously suggested solution (LUGER; SELLEN, 2016) might be mitigating anthropomorphic features – such as the interaction style and VAs' identities – to decrease perceived levels of intelligence and humanness and align users' expectations with the system's actual capacity. Nonetheless, studies have pointed out that the VAs' personality traits are a reason for VA adoption and are particularly valued by user niches such as children and people with intellectual disabilities (BALASURIYA *et al.*, 2018; FESTERLING; SIRAJ, 2020; LUGER; SELLEN, 2016; MOTTA; QUARESMA, 2019). Moreover, even subtle human aspects (e.g., voice, speech) may elicit unaligned expectations, as exposed in chapter 2. Thus, other tactics should be applied to manage people's perceptions of VAs.

An alternative may lay in increasing the transparency in VAs outputs. For example, as argued by Porcheron et al. (2018) and demonstrated by Kim, Jeong, and Lee (2019), system responses are important sources for users to diagnose and recover from failures, and low transparency regarding error sources may affect error

handling. Hence, information should be adequately available for users, for example, by offering initial learning guidance, clarifying error sources, and instructing users to handle errors.

Considering the influential factors for users' mental models exposed throughout this chapter, especially VAs' humanness and outputs, we aimed to explore solutions to deal with the matter. The following chapter will describe the methodological aspects of this research and the subsequent techniques applied to achieve this research's main objective.

4
Methodology

This research's main objective was to identify leading causes of users' misperceptions and offer design recommendations for aligning users' mental models of VAs with these systems' real capacities. We aimed to understand how VAs can be improved to mitigate gaps between users' mental models and the VAs' actual capabilities. Considering such goals, we applied an exploratory method with a mixed-method approach that combined both quantitative and qualitative data collection.

The first step to solving misalignments in users' perceptions is correctly understanding how their mental models can be described and which factors are influential to users' understandings. Therefore, we conducted a systematic literature review (SLR) on users' mental models of VAs to comprehend the state-of-the-art on the subject. After identifying relevant factors for users' misperceptions and surveying potential solutions, we aimed to validate some assumptions before focusing on a specific proposal. For this purpose, we conducted exploratory interviews with experts in conversational interfaces, aiming to understand their opinions on the mental model matter. Finally, to address our main research objective, we conducted a questionnaire-based, three-round Delphi study with professionals experienced in the research or development of conversational interfaces. This chapter presents the planning, procedures, and analysis of these techniques.

4.1.
Systematic literature review (SLR)

In order to understand the state-of-the-art concerning users' mental models of VAs, we conducted a systematic literature review (SLR). The SLR aimed to answer two research questions:

- **RQ1:** How can users' mental models of VAs be described in terms of system features, functioning, and way to use?

RQ2: How does previous experience with VAs affect users' mental models?

4.1.1. Search engines and search terms

We chose four databases that store a great number of articles in the field of Human-Computer Interaction (HCI) and Human Factors and Ergonomics to collect the primary studies: ACM Digital Library, SAGE, ScienceDirect, and Scopus.

The search strings comprised two types of terms: interface keywords and model keywords, assembled by an "AND" boolean. "Interface" keywords are synonyms for VAs and voice interfaces ("*voice assistants*", "*speech interfaces*", "*voice user interfaces*", "*virtual assistants*", "*intelligent personal assistants*", "*conversational agents*", "*conversational interfaces*", "*digital home assistants*", "*home virtual assistants*"). "Model" keywords are a set of terms that represent users' mental models or behavior ("*mental models*", "*strategy*", "*query adaptations*", "*expectations*", "*system understanding*", "*system perception*", "*tactics*").

We searched for "model" keywords in full text, but to avoid retrieving a large number of publications unrelated to voice interfaces, we limited the search of "interface" keywords to abstracts, titles, and keywords. For the same reason, we added the string [AND NOT "*chatbots*" OR "*chatterbots*" OR "*embodied conversational agents*" OR "*robots*"] to our "abstract, title, keyword" search. We chose the search terms from previously known literature and refined the strings through pilot testing.

4.1.2. Inclusion and exclusion (I/E) criteria

We established three inclusion criteria and two exclusion criteria for accepting primary studies.

- Publications should have been peer-reviewed full papers published in conference proceedings or journals between 2011 (iOS's Siri release year; HOY, 2018) and September 2020, when this review was conducted.
- Publications should have approached users' mental models (i.e., understanding, perceptions, and expectations for the system, following Wickens et al.'s

(2014) definition) and/or behavior when interacting with voice interfaces. This criterion was left intentionally broad since most primary studies did not directly mention (nor measure) the concept of mental models but assessed relevant variables to this study's goal.

- Publications should have employed a user study as a research method (e.g., survey, interview, experiment). Although there is a large corpus of valuable theoretical studies, we established that empirical testing would be essential to explore users' actual mental models of their interactions with voice interfaces.
- Publications must not have investigated embodied interfaces, in which the interaction with the voice interface is not the primary task (e.g., in-vehicle interfaces, virtual assistants embedded to learning platforms, image-based virtual assistants), to avoid intervening variables (CLARK; DOYLE; *et al.*, 2019).
- Publications must not have compared voice interaction to other interaction channels since this dynamic could affect users' perceptions (i.e., focus on comparison due to being more aware of differences between the two modalities). Moreover, since this research is focused on VAs specifically, comparing interaction modalities was out of the SLR's scope.

4.1.3. Quality Assessment

We measured the quality of accepted primary studies through a checklist in which we assigned scores based on the publications' abstracts (minimum score = 12; i.e., 60% of 20, the maximum score). The checklist's items concerned both general paper quality (questions 1 and 2; retrieved from Kitchenham, 2007) and paper suitability to answer the SLR questions (questions 3 to 6):

- Q1: How clear is the main research question? (score: 1 = implicit; 2 = explicit, but confusing; 3 = explicit and clear, but general; 4 = explicit, clear, and specific)
- Q2: How adequate is the study to address the research question? (score: 1 = inadequate; 2 = partially adequate; 3 = perfect fit)

- Q3: What type of study was conducted? (score: 1 = online questionnaire; 2 = in-depth interviews; 3 = experiment/ usability testing/ observational data)
- Q4: What type of voice interface was evaluated? (score: 1 = voice interface supported by visual display; 2 = purely auditory voice interface; 3 = voice assistant supported by visual interface; 4 = purely auditory voice assistant)
- Q5: What type of participants took part in the study? (score: 1 = a specific audience, such as children or elderly; 2 = different groups of audiences to draw comparisons, i.e., parents and children; 3 = general audiences)
- Q6: Does the study address the research questions for this SLR? (score: 1 = potentially; 2 = implicitly; 3 = explicitly)

For Q3, we assigned those specific scores since our main research question is qualitative, and we believe that primary studies that collected observational or self-reported qualitative data would be more fitting to our objective. Similarly, in Q4, we assigned higher scores for studies that have examined commercial VAs (or, at least, a simulation) rather than other voice interfaces since this study's object is VAs. We also prioritized studies that evaluated voice-only interfaces (e.g., Alexa on Amazon Echo) since graphic interfaces could be intervening variables. For the same reason, studies that focused on a specific type of participant received a lower quality score.

4.1.4. Procedure

One researcher conducted the SLR by applying the search string to the search engines and downloading the resulting citations to the software Mendeley. The researcher evaluated the primary studies' abstract and metadata to check for the I/E criteria. Suitable papers also passed through an abstract-based quality assessment according to the aforementioned checklist. Nevertheless, as all primary studies that passed the I/E criteria scored above the minimum quality score, no primary study was excluded from the review.

We recorded all accepted papers' information on a table containing: authors' names, publication year, source, and a summary of the study's research question, method, and key findings. Thereafter, one researcher conducted a full-text analysis of the primary studies and clustered them according to the type of contribution it

provided for our research questions. RQ1 was divided into the first three contribution groups (groups 1-3, see below) based on Wickens et al.'s (2014) definition of mental models. Nevertheless, during analysis, we observed that "learning" was found across all three groups, and therefore we created the category "users' learning practices" (group 4). Additionally, group 5 comprised contributions to RQ2.

- **Group 1:** users' expectations of VAs' features.
- **Group 2:** users' understanding of VAs' functioning.
- **Group 3**: users' perceptions of VAs' usage.
- **Group 4**: users' learning practices.
- **Group 5:** the impact of users' previous experience on users' mental models of VAs.

Through full-text analysis, one researcher identified relevant findings and their respective group and transposed them into a table. For each group, contents were registered according to the papers' authors, year, quote, subject, and researcher's comments fields (table 4.1). We then used the "subject" field to spot patterns across studies and identify emerging topics to address the RQs.

Table 4.1. – Example of analysis for group 1.

Author(s)	Year	Quote	Subject	Comments
Doyle et al.	2019	"IPAs were described as largely devoid of the ability to show warmth compared to a human partner." (p. 5)	User-VA relationship	

4.1.5.
Primary studies' gathering and sample profile

We collected 773 primary studies from four databases (figure 4.1.), but we excluded papers for which full-text access was unavailable for our institution (i.e., CAPES access on Scopus; n = 88). After eliminating 128 duplicates, 557 primary studies had their abstract analyzed, and 57 passed the I/E criteria, being the sample for this study (n = 57).

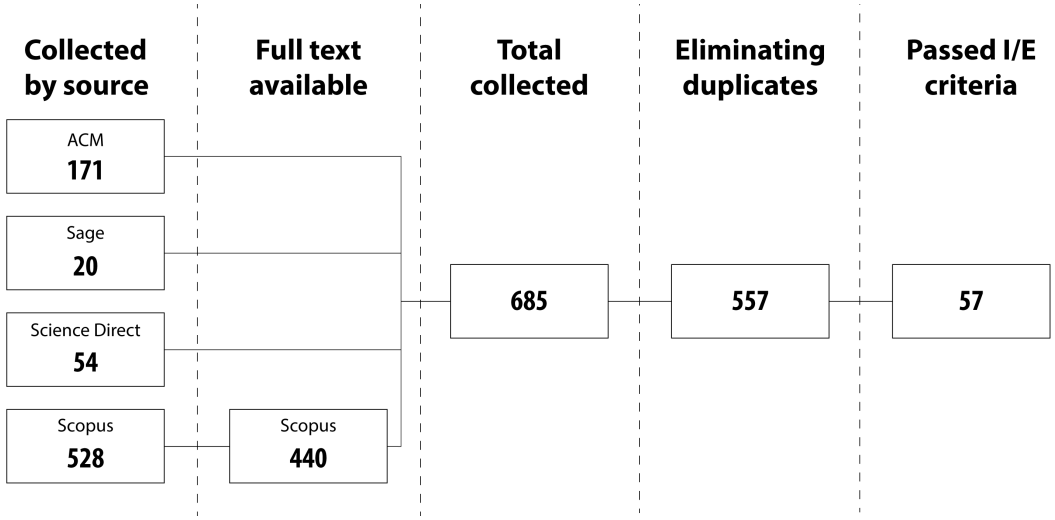


Figure 4.1. – SLR's procedure.

Appendix 1 contains a table showing the accepted primary studies, along with their contribution group, employed research technique, evaluated interface – VA or Simulation (Sim) – and participant type.

4.2. Exploratory interviews

The SLR led to an understanding of the state-of-the-art in the literature concerning users' mental models of VAs. Specifically, the review indicated two relevant aspects related to how VAs are designed (i.e., system image) that impact how users perceive VAs: 1) anthropomorphic features embedded in the VA design, and 2) system outputs provided to users. On the one hand, while an important driver for adoption, the VAs' humanness causes misperceptions about the system's actual functioning, capabilities, and limitations. On the other hand, the outputs offered to users are not sufficient to align users' mental models with reality. Therefore, the SLR's results indicated that increasing the transparency about how VAs works in their outputs by providing more information and explanations might correct users' perceptions about these systems.

Although the literature has brought upon valuable insights, the studies were focused on examining users' interactions alone. Such restriction was expected – and induced by one of the I/E criteria – since the SLR's goal was to understand how users' mental models of VAs can be described and which factors impact them.

Nevertheless, as presented in section 3.1, the system image results from the work of those who design a product: the developers. Therefore, VA developers must be aware of the causes affecting users' mental models and employ appropriate solutions to the design of VAs.

Considering the above, we conducted exploratory interviews with professionals experienced with the research and/or development of conversational interfaces. The interviews' goal was to comprehend whether such experts' opinions on the impacting factors for users' unaligned mental models are in line with the SLR's results. Such an assessment was necessary before moving on to the research's following technique (i.e., Delphi study; section 4.3) in order to survey these professionals' vocabulary and validate assumptions that would guide the Delphi's preparation. Specifically, we aimed to gather their perceptions on the role of the two factors mentioned above (1 – anthropomorphism and 2- system outputs), as well as possibly surveying other relevant causes and solutions to deal with the issue. Moreover, while increasing the transparency in VAs' outputs might have seemed like an adequate path, we needed to assess how desirable and feasible it would be to implement such a solution. Since these professionals work closely in developing conversational interfaces, they are prone to being aware of restraints to this proposal.

Considering the above, we conducted exploratory interviews with professionals experienced with the research and/or development of conversational interfaces. The interviews aimed to answer three research questions:

RQ1: Which causes do these professionals attribute to users' unaligned mental models of VAs?

RQ2: Which solutions do these professionals believe to be effective and adequate to solve the issue of users' unaligned mental models of VAs?

RQ3: What are these professionals' opinions on the feasibility and desirability of increasing the transparency in VAs' outputs?

To answer the research questions, we chose the technique of semi-structured interviews, which does not require that questions have a fixed order or number. Thus, the moderator could adjust the questions according to the participant's responses while following a series of previously established topics (GIL, 2008). Such an aspect was advantageous for this study since participants could express themselves freely while approaching the matters at hand.

4.2.1.
Participants

Participants were recruited through email and social media (LinkedIn), following the method of snowball sampling and convenience sampling. We recruited both researchers from academia and developers working in projects of conversational interfaces. To participate, researchers should have held a Doctoral degree, while developers should have at least three years of experience in the field of conversational design. All participants should have been currently or recently involved in studies or projects about conversational interfaces. To diversify the visions of such professionals as much as possible, we recruited participants from different backgrounds (see panel 4.1.). In total, eight professionals participated in the interviews.

Panel 4.1. – Participants' characteristics

Gender	Male = 5 Female = 3
Country	Brazil = 7 UK = 1
Place of work	Industry = 6 Academia = 1 Both = 1
Background	Computer science = 2 Communication studies = 2 Design = 1 Librarianship = 1 Multimedia production = 1 Languages = 1
Position	UX/VUI designer = 2 Consultant = 1 Voice Product Specialist = 1 UX Writer = 1 Conversational Design Lead = 1 Research scientist = 1 Lecturer = 1

4.2.2.
Procedure

Firstly, we scheduled the interviews and sent a free and informed consent term to the participants through email (see Appendix 2-3). Once in the interview, after greeting participants and superficially presenting the study's objective, the interviewer followed the script below:

1- **Explain the concept of mental models.** During a pilot interview, the interviewee mentioned that participants might be acquainted with different definitions of the concept of mental models. Thus, in order to align all participants' understanding of the concept, the interviewer explained the definition of mental models being considered in the study (based on Norman, 2013 and Wickens *et al.*, 2014): "a type of conceptual model created by users to represent how a product or system works, including a series of expectations about its components, functioning, and proper usage".

2- **Explain the issue of users' mental models of VAs.** The interviewer explained that indications suggest that users' mental models of VAs are unaligned with these systems' actual capabilities, resulting in abandonment and underutilization.

3- **Ask about causes for unaligned mental models and solutions to deal with such an issue.** The interviewer asked participants to express their opinions on what leads users to form mental models unaligned with reality. The participants also had to provide suggestions on how to solve this problem. The interviewer encouraged the participants to provide examples of real situations based on their work experience whenever possible.

4- **Explain and ask participants to comment on the assumption that VAs' anthropomorphic features and outputs influence users' mental models.** The interviewer explained the two main causes that, as mentioned above, were believed to strongly influence users' perceptions of VAs. Thereafter, the interviewer asked the professionals to give their opinion about such an assumption.

5- **Explain and ask participants to comment on the assumption that increasing the VAs' transparency would improve users' mental models.** The interviewer presented the assumption that increasing the system

transparency while maintaining similar levels of humanness on VAs would align users' mental models with the system's actual capabilities. Participants should comment on whether they believed such a solution would be feasible and desirable.

6- **Final greetings.**

As stated before, the interview followed the semi-structured approach, and therefore the script was flexible and varied depending on the participants' responses. The interviews were conducted remotely through the platform Zoom. The procedure lasted an average of 45 minutes and were recorded in both audio and video for analysis. The interviews were conducted between June and July 2021.

4.2.3. Analysis

To analyze the data, we conducted a thematic analysis following a top-down approach. In the first place, after transcribing the interviews, we identified how the professionals' responses fit into the following categories (according to the RQs):

1. **Causes for users' unaligned mental models of VAs.**
2. **Solutions to deal with users' unaligned mental models of VAs.**
3. **Professionals' opinions on the viability and desirability of increasing the transparency in VAs' outputs.**

Then, we identified emerging topics for each category by transposing passages to the platform Miro and elaborating an affinity diagram (BARNUM, 2011).

4.3. Delphi study

The exploratory interviews pointed out relevant factors for users' construction of their mental models of VAs. These included not only VAs' anthropomorphic features and outputs but also other aspects such as business and marketing limitations. Similarly, although most participants recognized the potential for increasing the VAs' transparency, the professionals also suggested other paths to correct users' perceptions. Thus, although we had initially planned to focus on the transparency topic in this research's following phase (based on the SLR's results), the interviews'

findings suggested it was still necessary to explore other causes and solutions to deal with the mental model issue. To achieve this outcome, we conducted a three-round Delphi study using online questionnaires as means of data collection.

4.3.1.
The Delphi method

According to Linstone and Turoff (1975, p. 3): "Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem." The method aims to provide a procedure for a group of people knowledgeable on a subject to reach a consensus of opinions on a topic of interest (FISH; BUSBY, 2005). The Delphi method benefits from the belief that reaching consensus through a collective process is valuable to find solutions to problems (LINSTONE; TUROFF, 1975 apud FISH; BUSBY, 2005).

The structured communication made possible by the method is based on four important aspects: "some feedback of individual contributions of information and knowledge; some assessment of the group judgement or view; some opportunity for individuals to revise views; and some degree of anonymity for the individual responses." (LINSTONE; TURROF, 1975, p. 3). Therefore, the Delphi method involves different rounds of anonymized, remote data collection, usually employing questionnaires as tools.

Linstone and Turoff (1975) highlight a variety of scenarios in which an issue could benefit from the collective communication driven by the Delphi method. These include (but are not limited to):

- Gathering unknown or unavailable data.
- Planning academic programs or urban development.
- Defining pros and cons of policies.
- Dealing with matters that require collective and subjective judgments.
- Dealing with matters that involve individuals from different backgrounds and expertise, or who disagree so severely that anonymity is necessary.

4.3.2.
Research questions

We applied the Delphi method due to the complexity of the issue of users' mental models of VAs and the need to identify solutions to the problem. Such a method could support experienced professionals in communicating and reaching a consensus on recommendations to align users' perceptions with VAs' actual capabilities. We aimed to answer two research questions with the Delphi study:

- **RQ1:** What are the opinions of professionals experienced with research and/or development of conversational interfaces concerning the causes for users' unaligned mental models of VAs?
- **RQ2:** Which solutions do professionals experienced with research and/or development of conversational interfaces suggest for improving users' mental models of VAs?

4.3.3.
Participants and recruitment

Fish and Busby (2005) explain that selecting a Delphi study's participants is critical since their knowledge on the topic of interest is vital for the quality of the study's outcome. Participants should be involved in the matter being studied, have information and knowledge to share, and be interested in participating (DELBECQ; VAN DE VEN; GUSTAFSON, 1975).

In this study, we recruited professionals with experience in researching or developing conversational interfaces. For developers, we conducted an unsystematic search on the social media LinkedIn by using terms such as "Voice User Interface (VUI)", "UX Writing", "Conversational interface", "Chatbot", "Voice Assistants" etc. For researchers, we started the selection by identifying authors who had published in the field of conversational interfaces. Then, we used the Google Scholar and Research Gate platforms to identify those who had other publications and interests in the field. We also recruited participants through indications given by the experts interviewed previously and professionals who had already agreed to take part in the Delphi. Hence, we had a combination of snowball sampling and purposive sampling.

The area of conversational interfaces is a relatively new field of work, and therefore it would be challenging to find professionals who have been working exclusively on projects of this kind for longer periods. Therefore, we did not impose a minimum time spam in work experience. We believed that even professionals who are entering the field could contribute to the study. For a similar reason, we did not limit our sample to professionals experienced with VAs only (e.g., Siri, Alexa) and allowed the participation of developers of chatbots, voice bots, and Interactive Voice Response (IVR) systems. We considered that these participants could contribute to the study since such conversational interfaces share enough similarities with VAs to cause analogs issues on users' perceptions (see chapter 2). Finally, we allowed participants from different backgrounds since VA development requires varied profiles of professionals.

After selecting eligible participants, we started the recruitment by email, Research Gate, and LinkedIn. As suggested by Delbecq, Van de Ven, and Gustafson (1975), the invitation contained a superficial explanation of the research goal, obligations, estimated completion time, and instructions for the study's procedure. We recruited participants continuously over September and November 2021 to reach the ideal sample of 15 to 30 participants (DELBECQ; VAN DE VEN; GUSTAFSON, 1975). We invited 90 professionals, and 22 agreed to participate.

4.3.4.
Study's format, procedure, and materials

The Delphi method consists of the following rounds (figure 4.2): 1) initial exploration, 2) consensus and disagreements, and 3) review and final considerations. Each stage may be completed by using the questionnaire tool as means of data collection (LINSTONE; TURROF, 1975). Thus, this study followed the format of a three-round online questionnaire.

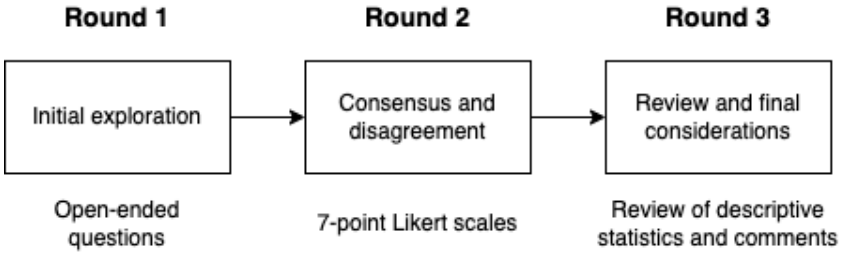


Figure 4.2. – Summary of the Delphi study's rounds.

4.3.4.1. Round one

Before starting round one, we selected eligible participants and invited them by email and social media. Such recruitment did not happen at once for all of the 90 invited professionals since we needed to gradually search for new participants while waiting for the others to respond to the invitation. Once the professionals agreed to participate, we sent the link to the first questionnaire and a free and informed consent term (Appendix 4-5), with which the respondents agreed by clicking on a button inside of the questionnaire. A total of 22 professionals agreed to participate and answered the first stage.

The first questionnaire was created with the tool Google Forms. Following the literature recommendations (DELBECQ; VAN DE VEN; GUSTAFSON, 1975), the first questionnaire (see Appendix 6-7) started with overall instructions about the study's first phase. Then, it presented the issue to be discussed: users' mental models of VAs. The passage exposed a definition of mental models to align all participants' understanding of the concept (as discussed in section 4.2.2.) and presented the issue of users' unaligned mental models of VAs. Thereafter, it showed two open-ended questions:

1. In your opinion, what are the causes that lead users to form mental models that are unaligned with Voice Assistants' real capabilities? Please, state at least three causes and, for each one, explain why it is relevant.
2. In your opinion, what solutions could solve the issue of users' incorrect mental models of Voice Assistants? Please, present at least three solutions and, for each one, explain why it is appropriate.

We asked participants to provide at least three causes/ solutions since the subsequent phases of the study were highly dependent on the quality and quantity of the first phase's answers. Finally, we provided an open field for optional comments and asked questions about social-demographic data. The results of the first round were analyzed (see section 4.3.6. for the analysis' description), and we summarized the participants' answers in statements to be used in round two.

4.3.4.2. Round two

In the second round, we sent the second questionnaire's link along with instructions and a 15-day deadline through email to all participants at the same time. When needed, we also sent reminders highlighting the deadline. A total of 18 participants completed the second questionnaire.

The second questionnaire (see Appendix 8-9) – which we created on the Eval&Go platform due to the better visualization it provides – aimed to enable participants to judge the group's opinions provided on the first round. Firstly, the questionnaire presented instructions for the second round, and participants could read the explanation on the issue of users' mental models of VAs again. Then, we presented the statements summarizing the group's responses to the first phase and asked the professionals to provide their level of agreement with each phrase on a 7-point Likert scale. We also provided open fields for optional comments.

The questionnaire presented a total of 35 statements (see chapter 6 for the statement list), divided into two parts: statements representing the causes that lead to misalignments in users' mental models of VAs (16 statements) and solutions to deal with the issue (19 statements):

1) PART 1: *"Below, we present statements that summarize the groups' opinions about the **CAUSES** that lead users of Voice Assistants to form mental models that are unaligned with these systems' actual capabilities. Please state your **level of agreement** with each statement below on a 1 to 7 scale. You may also leave **optional comments** on the statements."*

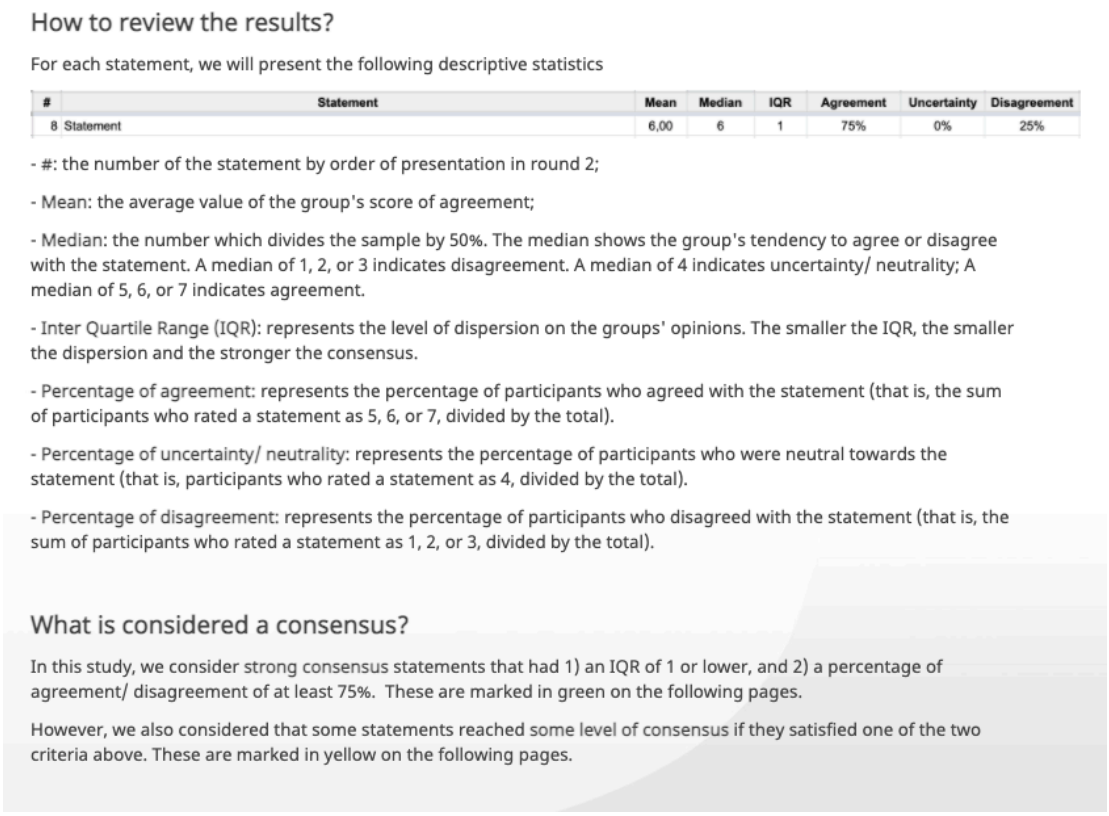
2) PART 2: "*Below, we present statements that summarize the groups' opinions on the **SOLUTIONS** to align users' mental models with the Voice Assistants' actual capabilities. Please state your **level of agreement** with each statement below on a 1 to 7 scale. You may also leave **optional comments** on the statements."*

The statements passed through pilot testing with a VA developer and a plain language specialist to ensure that they conveyed their intended meaning. During such evaluation, a reviewer considered the following phrase confusing: "*Voice Assistants update silently, not explaining to users about the updates in their skills*." Therefore, we modified the statement to clarify its meaning: "*Voice Assistants do not tell users when they update, nor explain the updates in their skills."* The pilot testing did not point to any other issues.

4.3.4.3. Round three

Finally, the third and last round of the Delphi study was to report the results of the second round to the participants (Appendix 10-11). The objective of this round was to give them the opportunity to review their judgments after viewing the group's opinion, besides conveying the feeling of closure to the participants (DELBECQ; VAN DE VEN; GUSTAFSON, 1975). The questionnaire was sent by email, but, as reviewing the results was an optional step, only 5 participants filled the form.

The first pages of the questionnaire presented an introduction explaining its aim and instructions on how to read the results. To direct the participants, we provided an image and written instructions (figure 4.3):



How to review the results?

For each statement, we will present the following descriptive statistics

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
8	Statement	6,00	6	1	75%	0%	25%

- #: the number of the statement by order of presentation in round 2;

- Mean: the average value of the group's score of agreement;

- Median: the number which divides the sample by 50%. The median shows the group's tendency to agree or disagree with the statement. A median of 1, 2, or 3 indicates disagreement. A median of 4 indicates uncertainty/ neutrality; A median of 5, 6, or 7 indicates agreement.

- Inter Quartile Range (IQR): represents the level of dispersion on the groups' opinions. The smaller the IQR, the smaller the dispersion and the stronger the consensus.

- Percentage of agreement: represents the percentage of participants who agreed with the statement (that is, the sum of participants who rated a statement as 5, 6, or 7, divided by the total).

- Percentage of uncertainty/ neutrality: represents the percentage of participants who were neutral towards the statement (that is, participants who rated a statement as 4, divided by the total).

- Percentage of disagreement: represents the percentage of participants who disagreed with the statement (that is, the sum of participants who rated a statement as 1, 2, or 3, divided by the total).

What is considered a consensus?

In this study, we consider strong consensus statements that had 1) an IQR of 1 or lower, and 2) a percentage of agreement/ disagreement of at least 75%. These are marked in green on the following pages.

However, we also considered that some statements reached some level of consensus if they satisfied one of the two criteria above. These are marked in yellow on the following pages.

Figure 4.3. – Instructions on how to review round three's results

The main part of the questionnaire displayed tables showing the statements alongside a set of descriptive statistics (as exemplified by figure 4.4 below). We divided the results into two parts: 1) statements related to the **causes** leading to issues in users' mental models of VAs, and 2) the **solutions** to deal with this matter. For each part, we presented a table with statements that reached a strong consensus,

statements that reached some level of consensus, and statements that did not reach consensus. Following each table, we provided an open field for participants to write comments. The questionnaire ended with a field for the professionals to leave their names and another open field for free comments on the study.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
18	Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).	3,83	4	2	39%	22%	39%
21	The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.	5,50	6	2,75	72%	22%	6%
33	Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).	4,61	4	4	44%	22%	33%

Figure 4.4. – Example statements displayed to participants.

4.3.5. Analysis

The study's analysis followed three steps to accommodate all the rounds.

4.3.5.1. Round one

Due to the qualitative nature of the data collected in the first questionnaire, we conducted a thematic analysis using an affinity diagram (BARNUM, 2011). The analysis consisted in the five stages, following Delbecq, Van de Ven, and Gustafson's (1975) recommendations:

1- We transposed all the participants' responses into two tables (one for each question). Since we asked the participants to provide at least three topics for each question, we separated each topic into different cells while maintaining its identification by participant number (e.g., P1, P2, etc.).

2- For each topic, we identified its general theme and attributed a code that represented such theme.

3- After coding each topic, we reviewed all of the codes to identify similarities among them, and, from this process, we created categories.

4- Once we defined all categories, we pasted all answers in the software Miro by creating a "note" for each table cell (figure 4.5). We positioned all "notes" according to the major categories and then elaborated the affinity diagram using a bottom-up approach to find emerging topics and similar contents among the "notes". This process was necessary since we needed

an adequate level of granularity in order to generate the statements for the following round. If the statements were too broad, they would not reflect the group's true opinion, but if we made every "note" a statement, the questionnaire would become too long, endangering its response rate.

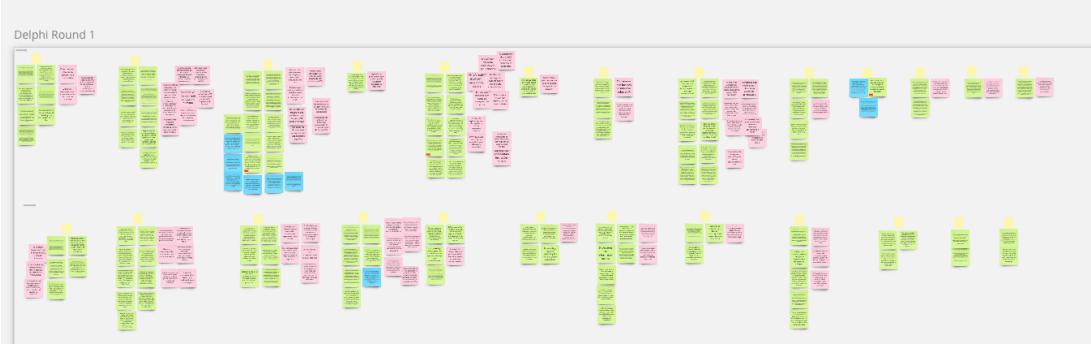


Figure 4.5. – Affinity diagram created on the Miro platform.

5- From the emerging topics identified in the previous step, we created the statements for round two. These phrases were validated and refined through expert review.

4.3.5.2. Round two

The second questionnaire required both quantitative and qualitative analysis due to the nature of the data collected. The quantitative analysis was important to determine which statements would be a consensus among the group. However, as Giannarou and Zervas (2014) have shown, there are several ways to analyze the quantitative data of a Delphi questionnaire and identify consensus. Since we had participants answer a 7-point Likert scale for this study, we adhered to Fish and Busby's (2005) recommendation of measuring consensus using the Interquartile Range (IQR). Nevertheless, Giannarou and Zervas (2014) pointed that many studies have used a combination of the IQR with supplementary metrics such as median, mean, standard deviation, or the percentage of participants agreeing/ disagreeing.

After tabulating the data, we used the software Excel to analyze the data and calculate the following descriptive statistics:

- **Mean:** shows the average value of the participants' level of agreement with a statement.
- **Median:** shows the sample tendency towards a value on the scale (i.e., one of the 7 points). A median of 7, 6, or 5 indicates a tendency to agree with a

statement. A median of 4 indicates a tendency of neutrality. A median of 3, 2, or 1 indicates a tendency of disagreement.

- **Interquartile Range (IQR):** according to Fish and Busby (2005, p. 247):

> Interquartile ranges are calculated by taking half the difference between the "upper quartile," or the point in the distribution below which 75% of the cases lie (the 75th percentile), and the "lower quartile," the point below which 25% of the cases lie (the 25th percentile). This type of statistic provides information about the range of scores that lie in the middle 50% of the cases, and in doing so provides information about the consensus of response on a particular item.

Considering this definition, we measured the IQR for each statement to identify the degree of consensus among the group. A small IQR means that the "middle 50%" of the responses are less dispersed across the scale's values, meaning that the participants provided mostly similar answers. The IQR value metric is beneficial since it is not easily affected by outliers (FISH; BUSBY, 2005).

- **Percentages of agreement, neutrality, and disagreement:** represents the number of respondents who tended to agree, disagree, and stay neutral towards a statement. The sum of scores of 7, 6, and 5 indicate agreement; the score of 4 indicate neutrality; the sum scores of 1, 2, and 3 indicate disagreement (table 5.2).

Table 4.2. – Example of the percentage's calculation

Agreement			**Neutrality**	**Disagreement**		
Completely agree (7)	Agree (6)	Partially agree (5)	Nor agree, nor disagree (4)	Partially disagree (3)	Disagree (2)	Completely disagree (1)
10%	60%	5%	5%	10%	5%	5%
Agreement: **75%**			Neutrality: **5%**	Disagreement**: 20%**		

To determine whether the statements have reached a consensus among the group, we established two criteria:

1) An IQR of 1 or lower, as recommended by Fish and Busby (2005).
2) A percentage of agreement, neutrality or disagreement of 75% or higher. We determined the threshold of 75% since the literature has exemplified that 70% to 80% thresholds are common for indicating consensus (GIANNAROU; ZERVAS, 2014).

We determined that statements that met both criteria have reached a strong consensus among the group. However, we also considered that statements that met only one of the criteria had reached a mild consensus.

As for the qualitative data analysis, we performed top-down thematic analysis. We transposed the responses to a digital board on Miro according to the statements to which they corresponded. Thereafter, we identified similarities to determine emerging topics of interest.

5.3.6.3. Round three

The third questionnaire aimed to let the group review the results and provide comments arguing in favor or against the results. Thus, the data collected was qualitative, and we conducted the same top-down analysis described for round two's qualitative data.

5
Exploratory interviews with experts

Aiming to understand the opinions of experts of conversational interfaces on the impactful factors for users' mental models of VAs and potential solutions to deal with the issue, we conducted exploratory semi-structured interviews. In total, eight experts participated in the study. This chapter presents the interviews' results in three sections according to the three research questions we aimed to answer. In section 5.1, we present the professionals' opinions on the causes of misalignments on users' mental models of VAs. In section 5.2, we report the experts' opinions on solutions to deal with such issues. Finally, section 5.3[2] exposes the participants' views on the feasibility and desirability of increasing the transparency in VAs' outputs as a proposal to solve the mental model matter.

5.1.
Causes leading to issues in users' mental models of VAs

The experts indicated which factors they believed to impact users' mental models of VAs and lead to misaligned perceptions about these systems. Some of these causes were in line with the SLR's results, while others were yet to be reported. As explained in the methodology chapter (section 4.2), we let participants talk freely first and then asked their opinions on two specific causes: VAs' anthropomorphic features and VAs' outputs (i.e., VA responses to users' requests, including feedback, instructions, and cues for interactions).

In the first place, as shown by the literature, the participants considered that VAs have a high level of humanness, affecting how users perceive the product. Some participants mentioned this cause before we directly asked for their

[2] The results exposed in section 5.3 have been accepted for presentation in the conference "Ergodesign & USIHC 2022", due to take place virtually between March 7th-10th. The paper, entitled "*Exploring challenges and opportunities to increase the transparency of Voice Assistants (VAs)"* will be published in the event's proceedings in March 2022.

opinions, while others did not initially mention it. However, all participants agreed that VAs' anthropomorphic features influence users' mental models. They mentioned that such anthropomorphism leads users to believe that they may interact with VAs in a similar manner to human-human conversations. "*I think that directing people to this understanding, 'look, you can talk as if you were talking to a friend', that is a mistake*" (P3). According to the professionals, users' perceptions gradually change into a more realistic perception as errors and failures happen.

The interviews outcomes were also in line with previous studies concerning the aspects that induce such humanness. The experts explained that design characteristics such as a human-like voice (i.e., recorded voice instead of text-to-speech), conversational style (e.g., less robotic and more natural), and vocabulary might incite high humanity in VAs. One participant mentioned that a less robotic voice might induce lower expectations concerning the system's capabilities: "*When we use a TTS [Text-To-Speech], people eventually get it (...) It is clearer that, okay, it is a technology, it is a machine here, so it may be a little more limited*" (P6).

Unexpectedly, the experts argued that attributing such high levels of humanness to conversational interfaces may even cause users to doubt if the system is actually a machine or a human being. Some participants reported examples of users who confused a conversational agent for a human, making them feel tricked and embarrassed: "*Some people know that they are interacting with a machine, so sometimes they even feel a little tricked [when the interface tries to mimic a human]. (...) But it happens, some people interact and take a little bit longer to realize that it was a machine, and when they find out, they get super embarrassed*" (P2). Such confusion was also reported to affect the task performance: "*Sometimes, some less attentive people did not get it from the start that it was a recording [and not a human]. Sometimes, they told their whole story to the digital agent, and it was pointless because it could not understand. We just needed them to tell us if they were able to pay [a debt] or not*" (P6). For one of the experts, increasing the VAs' humanness to the point that users are not even sure whether it is a machine or a human anymore is also an ethical issue: "*I think it is not ethical, omitting to the user that they are talking to a machine and not a human. They have the right to know with whom they are talking to.*" (P1).

In a similar stream with the previous topic, the experts commented on how such anthropomorphism might cause users to expect human-level conversations from VAs, confusing them on how to interact. The participants argued that as daily human communication leads to a well-established idea of how to engage in conversations, users believe that should mimic such conversational practices when faced with highly human interfaces. "*If I am talking to a human being, [I know] how to interact with a human being, we learn it through experience during our lives. If the machine presents itself for me as something as close as possible to a human, I understand that, to interact with it, I should interact more or less in the same way I interact with humans*" (P3). Nonetheless, when the interface fails to live up to such expectations due to its limitations, users cannot interact and feel frustrated. "*The idea that comes to our mind is 'I am going to have a conversation.' So, I will say something, the machine will totally understand what I say, and it will answer me just like a human. But, sometimes, even the humans themselves do not answer in the way you have imagined*." (P4). *"[The VA] is sold as conversational, with a more natural language. [But] sometimes, the way people talk, the naturality, the tool is not ready yet to this amount of information*" (P6).

Although the adverse effects of VAs' humanness on users' perceptions have been reported in the literature, the interviews indicated that applying such anthropomorphic aspects is not driven by design decisions alone. The experts mentioned that business partners, managers, or clients frequently intervened in projects of conversational interfaces, pledging that a more human-like virtual agent would make interactions more natural and leverage both user experience and user acceptance. "*They wanted us to humanize the applications, and 'humanizing' comes from that, from understanding that the closest the system acted as a human, the more fluid [the interaction] would be. And the better the [users'] reactions and their acceptance of the system would be."* (P3). However, two experts also mentioned that these stakeholders usually do not possess the technical knowledge to apply solutions adequately: "*It might be the client, the businessperson, sometimes it is an executive who does not even know anything [about the project and VAs] but wants to have a say in it. (...) [People who] do not have a single clue of what they are talking about but give their opinions*." (P5).

Another emerging topic on the exploratory interviews related to findings in the literature is VAs' outputs. In a similar manner to VAs' humanness, while not

all experts mentioned aspects related to VAs' outputs at first, we observed a general agreement with the relevance of this factor. The experts argued that issues in users' perceptions are caused by the lack of information provided since conversational interfaces eventually do not present what the system is capable of doing nor how to utilize it. One participant argued that VAs do not propose an initial preparation for users: "*If you initiate the interaction with these systems without more or less preparing the users to such interactions, be it with an initial explanation, be it with a more subtle way of showing that they are talking to a system that has specific interaction manners, you are already distorting their perceptions that [the system] has these possibilities*". (P3). Since the interface does not clarify what it can do, users are left to explore the interface by themselves, utilizing a trial-and-error approach, as reported in section 3.2.4 (users' learning practices). "*Only by interacting, making mistakes, does the user realizes that it [the VA] is not what they initially expected*" (P3). Furthermore, one participant argued that the VAs' error-handling mechanisms are usually limited and fail to provide relevant information to users, who cannot ask questions to investigate error sources: *"[When you talk to a human] you can do this sort of interrogation. But when you talk to a conversational interface, it says, 'sorry, I didn't get that', and you are not really able to [ask] 'What didn't you get?'*" (P8).

The last of the interviews' emerging themes that showed similarities with the SLR's results was the influence of users' characteristics on their perceptions about the VAs. As suggested in the literature, such characteristics included both their previous experiences with other voice interfaces and their educational backgrounds or interests. One participant mentioned that some users feel tense from the start of a voice interaction because of previous bad experiences with voice interfaces: "*People feel tense, they do not know what is behind the technology to support the interaction and are already used [to the speech recognition technology being unprecise]. 'Damn it, I am going to talk to a robot, they never understand me*'" (P1). Based on their research, another expert argued that users show differences in behavior depending on their level of experience and interest in technology. "[*People interested in technology] understand the technology's functioning, the environment, and the possibilities. (...) Sometimes, they even [interact] to test [the VA]. (...) [People who are less interested in technology] will use only the most basic features. They may even obey what the interface is trying to teach, repeat the exact same words*" (P6). Finally, one professional argued that the issue of users'

misperceptions of VAs is natural since these systems are a relatively new technology and there is always a stage of strangeness: "*I think it is natural because it is definitely not new, but it is still in the beginning, recognizing how to deal with all of this expectation*" (P7).

Finally, the following emerging topics are causes which – just as the users' characteristics – are not embedded on the VA itself but belong to their usage context: VAs' marketing and users' perceptions of machines in general. According to the experts, the marketing of VAs is unrealistic, provoking incorrect perceptions on users, especially for the VAs' conversational capabilities. In their views, companies attribute terms such as human-like or conversational to the VA to create the image of a fluid interaction, similar to human communication. "*The marketing itself contributes to a distorted view of the product or what it is capable of doing*" (P3). "*A long time ago, eight years or so, Apple was saying 'you can talk to Siri how you would talk to your friend, and it works.' Of course, that is not true now, and it certainly was not true then*." (P8). One participant also argued that the term "artificial intelligence" itself might cause misperceptions: "*I think the term artificial intelligence complicates a lot because the word 'intelligence' brings the imaginary that [the VA] is, in fact, intelligent*" (P5). Similarly, some of the professionals believed that perceptions of increased intelligence and capacity might be due not only to VAs' marketing but also to people's general impression that machines are flawless and highly capable. "*Sometimes we attribute a level of perfection to machines, and we do not do that when we talk to other people*" (P4). "*The image of a VA goes back to the image of the 60s' robots.*" (P7).

Considering the results above, we may address the first RQ for this research technique: "*Which causes do these professionals attribute to users' unaligned mental models of VAs*?". The following points highlight the key findings:

- The VAs are highly humanized, tailoring users' perceptions to a level that might confuse them on how to interact and impact performance.
- The VAs' humanness is eventually pushed on the VA's design by stakeholders outside the developing team.
- The VA presents little information about its functioning and has poor error-handling support.

- Users' backgrounds and interests are relevant to their perceptions and behaviors towards interactions with VAs.
- Marketing and users' perceptions of machines in general might set unrealistic expectations on users.

5.2. Solutions to leverage users' mental models of VAs

This section presents the participants' opinions on which solutions could be implemented to deal with issues in users' mental models of VAs. As with the previous section, we will expose both the participants' unbiased thoughts and their opinions on the two topics we suggested (i.e., VAs' humanness and outputs).

Firstly, while the experts considered that VAs' anthropomorphic features were a relevant driver for users' misperceptions, their opinions on paths to deal with this aspect were not homogeneous. On the one hand, the professionals recognized the value of having users interact with a human-like agent to make interactions more natural. "*Saying that it is a (chat)bot makes the person tense, like, 'oh, I will need to choose which words to use to be understood' (...) But, if [the chatbot] is laid-back, if it talks in a conversational manner, humanized and informally (...), it gives the chance for people to relax about their answers too*." (P1). One participant commented that humanizing the way VAs speak may also benefit the presentation of information for users, bridging the gap between users and machines: "*When you talk about 'explainability', I think that the anthropomorphism helps a little. You can use a language that is maybe more accessible, that is not a machine's language*." (P2). Likewise, one expert pointed out that such anthropomorphic aspects are important for the VAs' image: "*Obviously, you are not going to sell off a virtual assistant that is sulky, boring, and with a weird voice that no one wants to hear*." (P4). Moreover, as suggested at the end of chapter 3, a participant commented that eradicating the VAs' humanization would be challenging due to the strong association between human beings and conversations: "*Taking off the humanization is nearly impossible because, even if the voice is robotic, we are already bringing a human element. There are no other beings that speak. Only humans speak. (...) Maybe we can reduce [the humanization], but not remove it*" (P5).

On the other hand, the participants pondered how to implement the correct levels of humanness on VAs, arguing that exaggerating some features may be more harmful than beneficial. Many experts mentioned that developers must favor the interactions' objectiveness and efficiency over anthropomorphic characteristics such as a playful conversational style and funny prompts (e.g., jokes). "*It is better that they [the user] understands that [the VA] is a machine and can interact in a more precise manner, making fewer mistakes and save their time. That creates the satisfaction that the clients [stakeholders] frequently believe that lies in the humanization"* (P3). In a similar line of thought, one participant suggested that developers should choose the right moments to present playful prompts to users: "*You need to know the right moment to be human in the interface (...) maybe in the beginning or at the end (...), not during the interaction.*" (P5). Comments in that sense highlighted that the users' objectives in interacting should be investigated and prioritized: "*What do you expect? If you expect [to perform a task], is it really worth it for me to make [the VA] so humanized that it confuses the person? Maybe it is better if it is a little bit more objective, right*?" (P2). "*Does the public want a companion at these moments [when performing a task]? Does the public want a 'Google' at these moments? It seems to me that the public wants a 'Google'*" (P7).

The second emerging topic was related to how VAs' outputs are designed. According to the participants, a VA should present more information about itself and its functioning to users, including explicit or implicit clarifications that the conversational agent is not a human. Two experts reported cases in which they employed prompts instructing users how to interact in order to imply that the agent was a machine: *"[The agent] explained to the person, 'we are going to interact a little differently now [...] instead of typing, you are going to answer by talking to me, okay?'*" (P1). The instructions on how to interact were also mentioned by other professionals, who argued that conversational interfaces should explain what the system can do or how it works to align users' understanding of its functioning. The professionals suggested that more objective and instructive interfaces lead to a better task performance: "*Now that we are working with microcredit, the chatbot's first prompt is to present its scope (...) The more directional, the more proactive, the better it works*" (P2). Nevertheless, one participant argued that deciding how much information is enough for users to understand the system is a challenging task for developers. "*How much do people need to know for their mental models? (...)*

Voice interfaces are relatively new, so we do not really know how to explain them particularly well and what explanations we need to give to people. (...) Perhaps, we do not need to go into so many details, or perhaps, we do." (P8).

Similarly, the participants also argued that designing proper error-handling strategies is paramount to helping users recover from eventual mistakes. "*When an error happens, you need to show the user the context of why it happened. This helps the person to answer the question the second time, to reformulate their response without not repeating the same mistake.*" (P6). However, one expert was concerned that directly teaching the users exactly what to speak might harm the interaction: "*This makes the text poor and very robotic. Sometimes it works because the person will know exactly what to say, but it is possible to leave it implicit in the text*" (P6).

Besides dealing with humanness levels and presenting information to users, when we asked the experts which solutions they applied in previous works, some mentioned that it is necessary to mind general design guidelines and conversational design's best practices. "*There are some basic rules with design principles that should be applied in any type of interaction*" (P3). These recommendations guided decisions from the conversation's structure, such as the type of question to be applied (e.g., open-ended or multiple choice), to more detailed, content-level design: "*When we talk about content, I think a lot about bias, to which level we can bring a neutral language*" (P5).

Likewise, the professionals recommended that considering the users' context and the system types is crucial to implementing solutions. Developers must consider who the users are and in which situations, for what purposes, and how they will use the VA. "*We need to consider the context in which that will be approached. The audiences, the users, their personas, the technology.*" (P1). Some of the experts also argued that, due to such dependency on the context, it would be hard to provide general, absolute solutions to the mental model matter: "*There is no formula to elaborate a script for a voice assistant. (...) You would have to create best practices, and, when we talk about language, it is really hard to create best practices. (...) Best practices need to support any type of audiences, market, content, and that is really hard. (...) Each content is a different content. It is not like math."* (P7). Consistently, many experts reported that, although they try to foresee potential user action that could lead to failures, in order to fix errors and apply solutions, they usually need to examine the interface and deal with each trouble separately. "*I think*

that there is this first analysis of what is the error (...), and then we propose a solution." (P6).

Based on the findings above, we may answer the RQ2: "*Which solutions do these professionals believe to be effective and adequate to solve the issue of users' unaligned mental models of VAs?*". The highlights are summarized below:

- The VAs' anthropomorphic features bring benefits but should be used carefully not to get in the way of the main tasks.
- Prioritizing the interaction's objectiveness, VAs should present information to users on what they can do and how to recover from errors.
- Developers should mind best practices when implementing solutions, considering the users, their objectives, and their contexts.
- Due to the dependency on the context, some solutions might need to be designed on a case-by-case basis.

5.3. The feasibility and desirability of increasing VAs' transparency

This section specifically approaches the participants' opinions on whether increasing the transparency in VAs' outputs would be an adequate solution, which was our third RQ. As explained before, we asked them about this topic directly because the SLR's results suggested that more transparent outputs could be beneficial to aligning users' mental models.

The participants' responses revolved around two main themes: 1) the feasibility and 2) the desirability of increasing the transparency in VAs' outputs. Concerning the proposal's feasibility, the participants mentioned technological, design, and business aspects. As for desirability, the experts' responses approached usability issues and how adequate such a solution would be to the users' objectives.

In the first place, the professionals' opinions on technological restraints to increasing VAs' outputs were not homogeneous. While some experts mentioned that the solution seems technically feasible, others were dubious and pessimistic towards the subject. One participant mentioned that, although there have been great technological improvements in speech recognition technology over the past years, errors are prone to happen. According to them, to help users recover from failures

in voice interactions, developers should understand how humans repair their communication during real conversations: "*It is important to look how people solve problems and misunderstandings (...) They understand what happened, and then they fix it*" (P8).

However, the professional explained that VAs require multiple and varied systems and technology to function, leading to several potential failure sources. Therefore, the system cannot always diagnose why an error happened nor explain how to recover. The participant exemplified such a situation with other AI-based systems: "*An algorithm for image recognition does not explain why it could not recognize an image, it only says that it could not recognize it*" (P8). Hence, it might be difficult for VAs to communicate error sources or how to handle failures – essential instructions to the system's transparency – since the system itself might not diagnose the issue in all situations.

Besides the technological aspects, some experts commented on the project's design limitations that might hinder the feasibility of increasing VAs' transparency. The professionals reported that designing conversational interfaces is highly connected to constructing conversational flows: "*We still think about the basic model of interaction, in the 'tree' with the main user journeys. (...) It's a limited flow*" (P6). "*When we [the user] ask a question [to the VA], we are dealing with a flowchart*" (P1). According to these participants, predicting scenarios and including high transparency outputs in previously established flows may be a challenge. "*The users' head is a world of its own. They interact with the same tools in distinct ways. The way they speak the same command is different*" (P6). Such a hardship becomes even more complex for VAs, generalist interfaces that can perform a broad set of tasks. Contrarily to other conversational interfaces, such as Customer Service voice bots, VAs are not limited to a specific domain of features (e.g., vehicular insurance).

In a similar stream, another obstacle related to VAs' generalist nature is the role of the context for the design of voice interfaces: "*It is highly contextual, and it is a system with machine learning, and the context influences the conversation part a lot. (...) Just imagine the differences in the domains of finances, arts, and for people with low income. Look at the difference in explainability needed."* (P2). As exposed in the previous section, the use context is essential to guide design decisions, and thus issues in predicting usage context may also limit the development of transparent outputs.

The participants also expressed concerns about limitations derived from the development team's skills and eventual demands from business partners. An expert reported that designing VAs requires a large team of professionals, making the interface a combination of these designers' perceptions: "*There are two [mental] models: the user's and the system's. A 'system's model' might sound weird because you go to the designer. But it is not only the designer because the conversational system is so more complex, so many people work in it, so many curators*" (P2). Similarly, a participant mentioned that the development team should possess the correct knowledge on designing such interfaces "*The only obstacle comes from the team's lack of knowledge. The person needs to know how to structure the output's content using linguistic best practices*." (P5). Finally, several participants mentioned the difficulty in communicating with managers and business partners, who commonly do not value the benefits of increasing transparency in VAs' outputs. Such an issue was also pointed as a cause for excessive humanization of the VAs in section 5.1. The real example given by participant 1 illustrates this theme: "*The person called and talked to an IVR [Interactive Voice Response] system (...) So, the person was aware that they were talking to a robot because it had DTMF[3] interaction technology (...) Then, the IVR system said: 'okay, now I am transferring you to one of our attendants', which did not make it clear that it would be another robot. (...) They [the client] were against [the system] presenting itself [as a robot]. (...) They [customers] expected to be answered by a human, but it was a robot instead. (...) I argued a lot with them, but of course, I lost because they were a multinational [company]."* (P1).

We also asked participants whether they believed increasing the transparency in VAs' outputs was desirable from a business, user, and design point of view. Most participants considered the solution advantageous, particularly for the VAs' usability. In general, those who judged the proposal as desirable commented that leveraging the transparency could result in fewer errors and make users more confident and secure. "*Just let the system say what it can understand instead of leaving the person to find out alone and fail*" (P5). Some experts also considered it beneficial for users to understand the machine's limits: "*Makes sense being honest with the user (...) People understand that, okay, it is a machine, it will not do everything*

[3] Interactions that utilize DTMF technology (Dual Tone Multi Frequency) are those that require the user to press a button on the phone to select a command (e.g., to talk to an attendant, press 1).

(...) I think it is interesting for the user because they know how far they could go within the system. (...) Of course, there is the issue of expectation and reality, but the reality is not always frustrating" (P4).

Furthermore, the experts mentioned that increasing the VAs' transparency may support not only the users but also the other stakeholders involved in the system's project itself: "*It is worth it for the comprehension of everybody who is in the process. Be it the user, be it the people who design it, because it [understanding the system's limitations] is also a little complicated for us, too*" (P4). Making it clear for users which features are available can also support identifying gaps in the product: "*In the end, it turns out cheaper for the company because then you already have a mapping of what the users are missing*" (P4).

Oppositely, other professionals expressed concerns that an excessive increase in transparency could negatively impact the VAs' usability. The experts reported that presenting too many explanations on the system functioning might make interactions slower, tedious, confusing, or even hamper the voice interaction's easiness: "*A device that keeps explaining to you all the time how you should interact is boring, slow, effortful. People usually do not tolerate that much (...) [the interaction] stops exploring one of these system's positive aspects – the easiness in learning – since people already have an idea of how that works."* (P3). Likewise, one participant questioned whether designing an interface that presents so much information about its functioning is actually in line with users' goals while performing tasks: "*When I say that I want to play a song, I want to play the song. (...) I do not need to say, 'Oh look, I found the song 'X', and I am going to play it for you'. Okay, it is a kind of confirmation, cool, but I do not want it. I want it to play the song. And, if you are not able to play it, then, I want to know why immediately"* (P7).

Below, we report the key-findings to address RQ3: "*What are these professionals' opinions on the viability and desirability of increasing the transparency in VAs' outputs*?":

- There may be technological restraints to transparency, especially for implementing more transparent error-handling mechanisms.
- The flow-based nature of conversational design and the difficulty of predicting usage contexts for generalist interfaces might pose challenges to designing transparency in VAs' outputs.

- The developing team should possess a homogeneous understanding of the system and have the required skills to appropriately design transparent outputs.
- While increasing transparency is desirable to mitigate errors and install trust in users, it should be used carefully not to make interactions slow and boring.

This chapter presented the opinions of experts on conversational interfaces on factors causing misperceptions on VA users and solutions to address the issue. As reported, we identified several similarities between the professionals' views and the SLR's findings. However, while the participants generally agreed with the causes assumed to be relevant for users' mental models (i.e., anthropomorphism and VAs' outputs), they also pointed out other drivers such as marketing and the influence from stakeholders. Likewise, the experts indicated important considerations to employing solutions, such as the role of the usage context and the voice interface operational domain. Finally, although increasing the transparency in VAs' outputs was considered desirable overall, the experts also mentioned several technological, technical, and business restraints to such a solution.

Given the results, we considered it necessary to further investigate how professionals experienced with conversational interfaces view the mental model issue. Instead of focusing only on how to increase the VAs' transparency, as initially planned, we believed that a broader exploration could lead to uncovering more factors impacting users' perceptions, as well as surveying other appropriate solutions. Nonetheless, as we mentioned throughout this chapter, the interviewed experts did not always have convergent opinions on the discussed topics. Thus, to support a collective communication that could gather a high number of professionals and aid a consensus among them, we conducted a Delphi study, which will be described in the next chapter.

6
Delphi study

The interviews' results indicated the need to further explore the opinions of professionals experienced with the research or development of conversational interfaces. However, surveying the professionals individually could obscure potential disagreements among the participants, who might have opposing views – as shown in the previous chapter's results. On the other hand, gathering such professionals to a collective discussion (e.g., a focus group) could be challenging due to issues in managing disagreements and scheduling. Therefore, we applied the Delphi technique with professionals who have worked researching or developing conversational interfaces to examine their opinions on the matter of users' mental models of VAs. The three-round questionnaire allowed us to identify consensus in the participants' views, benefitting the exploration of relevant factors for users' mental models and adequate solutions to deal with misperceptions.

This chapter's first section (section 6.1) will present the participants' characteristics. Section 6.2 will report the results of the study's first phase. Finally, we will show the second and third rounds' findings in section 6.3[4].

6.1.
Sample's characteristics

Table 6.1 shows the characteristics of the 22 professionals who answered the first questionnaire. The first column shows the characteristic type, the second column shows the options inside each category, and the final column displays how many professionals fell into each option.

[4] The results exposed in chapter 6 have been accepted for presentation in the conference "Human-Computer Interaction International 2022 (HCII 2022)", due to take place virtually between June 26th and July 1st, 2022. The paper, entitled "*Exploring the opinions of experts in conversational design: A study on users' mental models of Voice Assistants (VAs)"* will be published in the event's proceedings in July 2022.

Table 6.1 – Sample's characteristics.

Characteristic	Categories	Number
Country of work	Brazil	16
	Ireland	2
	United States	2
	Sweden	1
	Austria	1
Place of work	University	12
	Enterprise	8
	Both	2
How long has it been since last worked with a conversational interface	Currently involved	11
	Less than a year ago	7
	Three years ago	4
Years of work experience	Less than a year	2
	Between 1 and 2 years	6
	Between 2 and 3 years	7
	Between 3 and 5 years	3
	For more than 5 years	4
Highest level of education	High School	1
	Graduation	6
	Masters	6
	Doctorate	4
	Postdoctorate	5
Field of graduation	Computer science	8
	Communication studies	2
	Information systems	3
	Artificial Intelligence	1
	Painting	1
	Design	1
	Informatics	1
	Pharmacy	1
	Marketing	1
	Eletronic Engeneering	1
	Cognitive Science/ Psychology	1
Job position	Researcher	13
	Developer	7

	Design lead	1
	Designer	1
	CTO	1
	UX Designer	1
	Professor	4

As shown above, most participants were from Brazil, where we conducted this study. The sample was relatively balanced in "place of work", with professionals from universities and companies. Furthermore, most participants were currently or recently involved in projects of conversational interfaces. Although two participants had a somehow short work experience in the field, we did not exclude them from the sample. As mentioned in chapter 5, we considered that even professionals with short experience could contribute to the study. These participants specifically reported holding a Graduate degree (masters and doctorate). Similarly, one participant reported that he had not finished his undergraduate studies but had five years of experience in the field. Since this participant was indicated by another professional, and considering their experience, we also did not exclude them.

As for the professionals' educational backgrounds, we identified that they come from areas related to computer science. We also observed varied graduation fields such as pharmacy and painting. Finally, most participants declared themselves to be either researchers or developers. However, many professionals did not specify in which part of development they work. The sum of the "job position" cells does not result it 22 as some professionals reported more than one job position.

6.2.
First round's results

In the study's first round, we exposed the issue of users' mental models of VAs through a brief text and asked the professionals (n = 22) to answer two open-ended questions. For the analysis, we categorized the responses and identified emerging topics in each category. From this process, we produced 35 statements that were used in the second round. It is important to highlight that due to the many categories and emerging themes identified, it was necessary to summarize some

topics together when turning them into statements to avoid an extensive questionnaire on the second round.

In the two following sections, we will present the categories, their emerging themes (if applicable), examples of participants' responses, and their resulting statement(s).

6.2.1.
Causes for issues in users' mental models

Round one's first open-ended question was: "*In your opinion, what are the causes that lead users to form mental models that are unaligned with Voice Assistants' real capabilities? Please, state at least three causes and, for each one, explain why it is relevant.*" From the 68 responses provided, we identified twelve categories of factors mentioned by the participants to influence users' mental models, from which 16 statements were created. Table 6.2 shows the categories along with the number of participants who cited them.

Table 6.2. – Categories of causes considered to misalign users' mental models with reality.

Category	Number of participants
Users' limited understanding of technology	13
Users' previous experiences with voice interfaces	1
Users' learning practices	3
Users' privacy concerns	2
Users' expectations for human-like conversations	6
VAs' anthropomorphic features	8
VAs' transparency	8
VAs' high complexity	1
Differences among VAs	1
Unrealistic marketing	8
Influences from science fiction	5
Lack of user research	1

The first four categories illustrated in table 6.2 are somehow related to the users' characteristics, interests, or behavior. "*Users' limited understanding of*

technology" comprised responses that highlighted the influence of low comprehension of the technology on users' mental models.

The first theme in this category was that people who face hardships dealing with technology in general, specifically AI-powered systems, are not used to VAs and may have trouble interacting. "*Lack of knowledge of what Artificial Intelligence is: due to the AI application's success that has been reported in the past years, I see that many people sympathize with the area and wish to use the technology, but still get confused about what is AI and its functioning*" (P8).

Some participants specifically mentioned users' lack of understanding of the limits in VAs' speech recognition and awareness of a social context. "*Lack of technological knowledge about the tool: voice recognition is a reality, but still exist limitations such as perfect semantic recognition and the natural capacity to formulate answers that are not embedded in the system*" (P19).

Finally, the professionals also mentioned that such a lack of knowledge leads to frustrating user-VA interactions. "*The impossibility for users to recognize the assistants' cognitive limits. The users, when engaged in dialogues with the assistant, end up being frustrated since they tend to require what [tasks] the assistant have trouble in comprehending and answering."* (P18).

We summarized the comments in this category to create the following statement: **Users do not know the Voice Assistants and the Artificial Intelligence technical limitations and require the Assistant to perform tasks and recognize commands beyond its capabilities.** Although the participants mentioned the frustration caused by such low technology comprehension, we did not include it in the statement since such a feeling is a consequence of a misperception and not a leading cause.

Only one professional mentioned the following category, "*Users' previous experiences with voice interfaces"*. The professional argued that: "*There is also a negative expectation about the VAs by some people who, for example, have had previous contact with conversational interfaces such as extremely limited Interactive Voice Response (IVR) [systems*]." (P11). For this category, we created the following statement: **Bad previous experiences with other voice interfaces create negative expectations for the Voice Assistants on users.**

"*Users' learning practices*" is a category mentioned by three participants and revolved around users' lack of dedication and interest in learning about the VA

before purchasing or starting using it. "*Users' interest in reading about the subject: many VA users install the assistant and straight-up ask it about everything. They do not dedicate some time to read about the subject and try to utilize it in the best way possible.*" (P6). One participant also mentioned that such a lack of interest in learning might originate from users' perceptions of low usefulness for VAs. "*VAs are systems that, for most users, don't solve any real problems. If you use it to control IoTs [Internet of Things products] or set an alarm, you'd probably have a pretty accurate idea of the VA model. If you don't need it, why invest time into understanding it*?" (P15). Deriving from such comments, we generated one statement: **Users do not look for information about the Voice Assistant before buying it, especially for tasks that are deemed unnecessary.**

The fourth category of relevant factors for users' mental models was "*Users' privacy concerns"*, which was reported by two participants. In summary, the professionals argued that users are afraid to interact due to concerns about how the VA will manage their data or that others could hear their interactions, impeding them from constructing a solid mental model. "*There is the matter that using the voice implies in caution about comfort and privacy. While a user may perform actions following (and improving) their own mental models [of visual interfaces], through voice, the user must feel comfortable with privacy to perform an action through such a modality since they cannot have control over who will hear their requests*." (P5). The statement created for this category was the following: **Privacy concerns hinder users from interacting with the Assistants for long enough to construct correct mental models.**

The following five cause types for misperceptions in users' models were related to VAs' design. "*Users' expectations for human-like conversations",* cited by six professionals, exposed how the use of speech in VAs leads users to expect sophisticated and natural conversations from the system. Generally, the responses mentioned at least one of three topics: 1) people are used to having conversations with other humans; 2) users have high expectations for VAs' conversational capabilities since they compare VAs and humans; and 3) VAs' actual conversational capabilities do not match those of humans, causing frustration. Participant 17's comment is an example that comprises all themes: *"Voice is often associated with social interactions, so some users may assume that conversational interactions with voice assistants are more 'human-like' than they actually are. This is relevant*

because my perception is that sometimes users expect voice assistants to interact with them in a way that they are currently not capable, which can lead to frustration and disappointment when the voice assistant does not interact as expected." (P17).

The previous category led to one statement: **Users construct their mental models of conversations through human interactions, but Voice Assistants have lower conversational skills, letting down expectations and causing frustration.**

In line with the influence of speech for users' models, "*VAs' anthropomorphic features*" were considered an impacting factor by eight participants. We identified two main emerging topics for this category. Firstly, the professionals argued that the VAs' humanness might lead users to believe that VAs are more capable than reality, suggesting that they have human-level skills. "*Cognitive anthropomorphism. We tend to think that the systems have the same cognitive abilities as humans. This can lead the user to assume that the machine presents the same language processing capability, and consequently, there is no need to use specific words or speak slower*" (P3).

The participants also included a set of characteristics that may humanize the VAs, which we considered a separate emerging topic. The features reported to cause anthropomorphism were voice, name, gender, metaphors, and humorist prompts. For example, P14 argued: "*Poor mental models develop through improper setting of expectations. (...) Expectations set through metaphor - calling a system an "assistant" implies it might behave like a human assistant*." (P14).

To conceive a statement for this category in a concise manner, we decided to combine both topics into one phrase: **Characteristics that induce anthropomorphism (e.g., voice, name, gender, metaphors, humor) cause users to expect Voice Assistants would be as capable as a human.**

The second category related to VAs' design is "*VAs' transparency",* from which we observed four emerging topics and developed four distinct statements. This category led to a higher number of phrases because the emerging topics had relevant nuances, and including all specificities in a single statement could make it confusing.

Firstly, we identified comments arguing that VAs generally present little information about various aspects of their functioning and utilization. These included: available features, ways of usage, command processing, decision-making,

and error-recovery mechanisms. For example, one participant answered that: "*The voice assistant does not tell why it came to a distinct conclusion, i.e. why it has recognized an input (ASR; [Automatic Speech Recognition]), how it has mapped the input to an understanding/intend (NLU; [Natural Language Understanding]) and how this understanding has triggered a distinct response (DM [Dialogue Manager] =>NLG [Natural Language Generation]). So, in a sense, we miss error transparency*." (P16). We synthesized several information types reported by the participants to create the following statement: **Voice Assistants do not explain to users about their skills, how they should be utilized, how they process commands, how they make decisions, or how to recover from failures.**

Furthermore, some professionals specifically pointed out that VAs do not explain their limitations for specific tasks or actions to users, lacking clarifications on the differences between actions and their requirements. "*Inconsistent Conversational Context: Some queries can become multi-turn and refer to the previous query, others can't. There is nothing in the query to easily identify when or why this is the case*." (P12). Thus, we developed a statement for this topic: **Voice Assistants do not explain their limitations for certain actions, such as recognizing the conversational context.**

Another matter explained by one participant was the lack of transparency in VAs' updates: "*One other issue is that these systems update silently and their capabilities change without any easy way for the user to know when a mental model of use should be updated.".* (P12). We rephrased such response into a statement: **Voice Assistants do not tell users when they update, nor explain the updates in their skills.**

For this category's last topic, we identified that some experts argued that VAs do not present instructions during users' initial interactions. According to the responses, this absence affects users' understanding and expectations towards the system. "*Lack of expectation setting - because most voice assistants are ready for interaction without any instructions or training, users are left wondering 'what can I say' without any expectation of the system's capabilities."* (P14). The statement that summarized this theme follows: **Voice Assistants do not present initial instructions to users, leaving them without knowing what to expect from the product.**

The next two categories were mentioned by only one participant each and are also linked to design influences on users' perceptions. Firstly, "*VA's high complexity*" encompassed a single response: "*Many of these interfaces are overly complex, demanding a high cognitive load from users*" (P4). We rewrote the comment to create a statement representing this category: **Voice Assistants are too complex and demand too much of the users' cognition.** Similarly, only one answer fell into the category "*Differences between VAs":* "*The comparison between two or more assistants at the functionality level. There are possibilities that you may find in Alexa that you may not find in Siri or Google. Maybe this 'addiction' that one assistant allows you to perform a functionality might project that other assistant should behave in the same way*" (P7). As with the previous category, we changed the comment to generate a statement: **Voice Assistants from different brands have different skills, leading to the belief that an Assistant might have the same skills as the others.**

The three final categories are drivers caused by external actors to the interaction itself (i.e., external to the user and the VA). Eight participants reported that "*unrealistic marketing*" leads to issues in users' mental models of VAs. This category comprised two emerging topics, resulting in two phrases. The first topic emerged from comments arguing that VAs' developing companies do not present enough content to explain how conversational interfaces and AI works. "*With voice assistants, users buy a product whose actual capabilities they do not know, as those capabilities are not visible. If one buys a piece of software, there is a way to check its features before purchasing. Voice assistants do not offer this discrete set of features. (...). It's like buying the proverbial pig in a poke.*" (P16). We synthesized this topic's responses as: **Brands present little institutional content about the Assistants and Artificial Intelligence, leading users to buy the product without being aware of its capabilities.**

The second emerging topic was the unrealistic nature of VAs' marketing. The professionals argued that the developing companies usually advertise the system by choosing simple and flawless use cases. Such marketing also tends to overplay the VAs' intelligence and capacity. "*Mismatch between how systems are marketed (ie, social capabilities are oversold and underdelivered) and the nature of current interactions (ie, limited to question-response). Similarly, this raises expectations about how people can interact with these systems*." (P13). This topic

resulted in another statement: **Marketing raises users' expectations by exaggerating the Voice Assistant's social skills and intelligence, showing use cases that are too simple and fluid.**

In a similar stream, five participants indicated "*influences of science fiction"* on users' perceptions. Overall, the responses reported that science-fiction media picture AI systems as highly intelligent and capable of features that current VAs do not match. "*The perception that the vast majority holds towards virtual assistants is of an extremely complex and futuristic, as in [science] fiction movies*." (P4). We considered all the comments to create one statement to this category: **In Science Fiction, systems powered with Artificial Intelligence are pictured as futuristic, intelligent, sensitive, talkative, and capable, creating unaligned perceptions about current Assistants.**

Finally, the last type of impacting factor for users' models was "*the lack of user research*" conducted by developers. Only one participant approached this theme, arguing that "*I believe [this factor] to be of great importance for the creation of conversational flows: not conducting studies about UX [user experience] with the assistant (...) It is important to define a real focus to the skill and work in creating that flow, considering particularly the user experience."* (P7). We rephrased the professional's comment to develop a statement: **Developers conduct little research about user experience, which is paramount to creating conversational flows.**

Considering the results, we present below the 16 statements resulting from the first question of the Delphi's first round (panel 6.1).

Panel 6.1 – Statements generated from round one's first question.

PART 1 - Causes for misalignments in users' mental models
Users do not know the Voice Assistants and the Artificial Intelligence tech-nical limitations and require the Assistant to perform tasks and recognize commands beyond its capabilities.
Bad previous experiences with other voice interfaces create negative expectations for the Voice Assistants on users.
Users do not look for information about the Voice Assistant before buying it, especially for tasks that are deemed unnecessary.
Privacy concerns hinder users from interacting with the Assistants for long enough to construct correct mental models.

Users construct their mental models of conversations through human interactions, but Voice Assistants have lower conversational skills, letting down expectations and causing frustration.
Characteristics that induce anthropomorphism (e.g., voice, name, gender, metaphors, humor) cause users to expect Voice Assistants would be as capable as a human.
Voice Assistants do not explain to users about their skills, how they should be utilized, how they process commands, how they make decisions, or how to recover from failures.
Voice Assistants do not explain their limitations for certain actions, such as recognizing the conversational context.
Voice Assistants do not tell users when they update, nor explain the updates in their skills.
Voice Assistants do not present initial instructions to users, leaving them without knowing what to expect from the product.
Voice Assistants are too complex and demand too much of the users' cognition.
Voice Assistants from different brands have different skills, leading to the belief that an Assistant might have the same skills as the others.
Brands present little institutional content about the Assistants and Artificial Intelligence, leading users to buy the product without being aware of its capabilities.
Marketing raises users' expectations by exaggerating the Voice Assistant's social skills and intelligence, showing use cases that are too simple and fluid.
In Science Fiction, systems powered with Artificial Intelligence are pictured as futuristic, intelligent, sensitive, talkative, and capable, creating unaligned perceptions about current Assistants.
Developers conduct little research about user experience, which is paramount to creating conversational flows.

6.2.2.
Solutions to deal with misalignments in users' mental models

As for the first round's second open-ended question, we displayed a query to the participants: "*In your opinion, what solutions could solve the issue of users' incorrect mental models of Voice Assistants? Please, present at least three solutions and, for each one, explain why it is appropriate*." The 59 answers to this query resulted in twelve categories of potential solutions to be applied to VAs. Table 6.3 shows the categories and their respective citation count.

Table 6.3 – Categories of solutions and the number of participants who cited them.

Category of solution	Number of participants
Changes in users' behavior	7
Increase VAs' transparency	10
Highlight user-VA colaboration	2
Improve error-handling mechanisms	3
Mitigate VAs' anthropomorphism	4
Offer supplementary content and tutorials	5
Handle privacy concerns	3
Change marketing strategies	3
Improve the developers' skills	1
Conduct research and understand usage contexts	6
Apply best practices	3
Improve speech recognition	3

As shown above, seven professionals provided comments arguing that "*changes in users' behavior"* would be a fitting solution to align their mental models. We identified three main themes for this category. Firstly, some participants suggested that users should look after information on VAs to improve the quality of interactions: "*Read first, use later: if this happens, they [the users] will certainly have a mental model that is more central on the real capacity of the acquired product*" (P6). From this topic, we created a statement: **Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).**

In the second place, two professionals suggested that users should be trained on how to interact with VAs: "*As for the language understanding, maybe we need to train users in the kind of language patterns voice assistants do understand*." (P16). Such a topic resulted in a phrase for the second questionnaire: **Users should receive training on how to use Voice Assistants, including supported language patterns.** This category also had a third theme, which only one participant mentioned. They argued that, similar to other technology, people will learn how to use VAs over time: "*Just like we have with digital education to access the internet and other digital services, I believe that soon the moment will arrive when we will learn how to interact with conversational services.*" (P2). Since this comment was very

distinct from the others, we developed a statement: **No solution should be applied to Voice Assistants since users will naturally learn how to interact.**

The category with the greater number of citations was "*increase VAs' transparency"*, mentioned by ten participants. The first theme that emerged from responses in this category was the need for explicit explanations from the VA about how it functions and how broad its scope is. "*Explanation methods of the assistant's decisions and action should be explained so that the comprehension of the limits and scope may be apprehended by the user*" (P18). In summary: **Voice Assistants should provide examples and explanations about their skills' scope and action execution, decision making, and learning processes.**

Likewise, some professionals proposed that these explanations should be embedded throughout interactions to guide the users to a more solid understanding of the system: "*Embedding conversational gambits that clarify context as a human would, and storing the results, would correct the mental model over time. (i.e 'Kitchen Hob not found, what light would you like to turn on?' -- 'Kitchen Top')"* (P12). We synthesized comments in this sense into a statement: **Voice Assistants should present usage tips throughout interactions, including mechanisms to clarify the conversation context.** The main nuance between the two phrases in this category is that the prior focuses on explanations of the VAs' scope and functioning, while the latter suggests presenting information on how to utilize the VA.

The following category is also related to changes in VA design: three participants suggested that VAs' should "*highlight user-VA collaboration"*. The two responses in this category argued that VAs could learn from users, improving the VA generally and aligning it with the users' preferences: "*Highlight the importance of collaboration: the VA's learning depends on the user's interaction. Such collaboration allows the customization and improvement of suggestions, according to the user's preferences*." (P8). The combined comments resulted in a statement: **The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.**

Three participants indicated that, to align users' mental models with VAs' actual skills, developers should "*improve error-handling mechanisms"*. The main difference between this category's two themes is the channel in which such mechanisms would be presented. One participant proposed creating an independent interface for users to review failures: "*Possibly having an error-correction interface*

on another modality, as a web page or app, allowing users to look at the unsuccessful queries and understand where they were recognition errors (...) and where they were out of scope and the idea that the system can perform such a query should be removed." (P12). We summarized this comment into a single phrase: **There should be a platform that shows failed past interactions to help users understand the reasons for errors and the system's scope.** In addition, the other professionals argued that error-recovery strategies should be embedded in interactions: "*Exposing more data on where it doesn't understand would also help train the user to understand the computational model. Instead of just saying 'I didn't understand, try saying it another way', it could expose what was outside the model with 'Asking 'closest restaurant to' is unsupported, try a different type of question*." (P12). The resulting statement is as follows: **Developers should create error recovery mechanisms (ex: inform what was misunderstood and how to reformulate the command).**

Interestingly, while eight professionals have pointed to VAs' humanness as a cause for misperceptions, only four participants suggested to "*mitigate VAs' anthropomorphism".* We identified two solutions in this category's comments: 1) remove or diminish features such as voice, name, gender, and metaphors to mitigate anthropomorphism, and 2) clarify to users that the VA is not a human. The following response comprises these two themes: "*I saw cases in which (...) the chatbot itself was programmed to inform [the user] that it was a robot (...) Another strategy was not to use a person's name nor attribute gender.*" (P21). This category resulted in two statements: 1) **The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.,** and 2) **Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).**

Furthermore, five participants argued that companies should "*offer supplementary content and tutorials*" to users. A group of experts suggested that users' initial interactions should be supported by tutorials and instructions to aid users' learning and improve their mental models. "*Companies making mini-tutorials: I believe this would help people a lot when they acquired their first virtual assistant*" (P6). Such comments served as a basis for a phrase: **Voice Assistants should present initial instructions, tutorials, and information about new supported actions and new ways to formulate commands.**

On the other hand, two participants argued that developers and companies should offer content explaining how the technology, specifically, AI, works. "*Demystification of the technology behind the conversational interface through evangelist contents, in different levels of detail to reach all audiences*." (P9). Considering their responses, we developed another statement: **Manufacturers and professionals from the Artificial Intelligence field should offer information about such technology to the population in an accessible manner (e.g., institutional material).**

The following solution types was suggested by the professionals was to "*handle privacy concerns"* from users. The participants argued that users' privacy concerns might step in the way of other solutions, and therefore it is necessary to make it clear how users' data are handled. "*Clearer understanding of privacy policies related to voice interactions. This can help users better understand if they want to use their voice assistant to perform certain interactions*." (P17). Consequently, the following phrase was developed: **The Voice Assistant should explain to users about the privacy of their data to help them decide which tasks to perform.**

Besides the solution categories that proposed changes to the VA design, we also observed indications that developers and companies should undergo modifications to improve users' perceptions. In the first place, three participants argued that it is essential to "*change marketing strategies".* In line with the findings of the previous section, showing that VAs' marketing is considered unrealistic, the professionals reported that VAs' advertisements should present to these system's true capacities: "*With more appropriate branding companies can signpost to people they are interacting with a machine, setting appropriate expectations and improving transparency simultaneously*." (P13). This category led to a single statement: **Marketing on Voice Assistants should stick to these systems' actual capabilities, presenting common and realistic use cases.**

Similarly, the participants argued that the development team should "*improve the developers' skills"* and *"conduct research and understand usage contexts"*. One professional explained that properly designing the conversational flow is key to improving VAs, but conversational designers are not always aware of technical limitations: "*I think that an alignment of technical requirements could happen so that the [conversational] flow may be constructed (...) someone from the team should be up to date and aligning both sectors, the conversational and the*

development" (P7). We rephrased the professional's response: **Developers of conversational flows should receive training on the Assistans' technical limitations so they can produce appropriate flows.**

Likewise, six experts answered that developers should understand the users and their contexts when designing solutions. Such comprehension included users' objectives, characteristics, interactional behavior, and mental models. "*VAs may be used only thorough voice or through text, with a screen or with other interface types. They may be used for home automation in a silent and calm ambient, may be used in chaotic traffic, may be used in a dangerous and noisy street, may be used from afar (...) or close to the user (...). The mental models will change, and the research and evolution should be continuous*" (P11). We comprised the comments in this category in a statement: **Developers should understand users (e.g., profiles, goals, contexts, behavior, semantics, mental models) to create solutions that address their needs and context.** Expressly, some experts also indicated the need to research human-human communication practices to adapt user-VA interactions. Comments from this topic relate to the last section's results, which pointed that users tend to construct their mental model of conversations from human interactions. "*The last one is tough, but much more research needs to go into understanding how supplementary information is exchanged during dialogue between humans, and how this might be adapted (not mimicked!) for a human-machine dialogue context*." (P13). Such a theme resulted in a phrase: **Developers should research mechanisms in human conversations and adapt them to interactions with Voice Assistants.**

"*Apply best practices*" of usability and conversational design was another category focused on the developers' work in designing VAs. "*There are some generic key points that might be employed, for example, apply best practices in the first use (users' onboarding), reinforcing what was understood by the system, [applying] only one question per query (...), expose possible options for [questions which have] limited options etc*" (P11). We synthesized this category in a single statement: **Developers should always apply best practices of usability and voice interaction when designing Voice Assistants.**

Finally, three participants suggested that improving speech recognition would aid users in improving their mental models. Responses from this category seemed to propose dealing with users' misperceptions by improving the VA to meet

their expectations. "*Improvements in the recognition of synonyms and similar commands to improve the recognition of the users' intents*" (P1). Therefore, the last phrase was created: **Developers should improve speech recognition technology (e.g., synonyms, intents, words in other languages, accents, localization, and different voice types and users).**

Below, panel 6.2 presents the 19 statements resulting from the analysis reported above:

Panel 6.2 – Statements generated from round one's second question.

PART 2 - Solutions to deal with the mental model issue
Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).
Users should receive training on how to use Voice Assistants, including supported language patterns.
No solution should be applied to Voice Assistants since users will naturally learn how to interact.
Voice Assistants should provide examples and explanations about their skills' scope and action execution, decision making, and learning processes.
Voice Assistants should present usage tips throughout interactions, including mechanisms to clarify the conversation context.
The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.
There should be a platform that shows failed past interactions to help users understand the reasons for errors and the system's scope.
Developers should create error recovery mechanisms (ex: inform what was misunderstood and how to reformulate the command).
The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.
Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).
Voice Assistants should present initial instructions, tutorials, and information about new supported actions and new ways to formulate commands.
Manufacturers and professionals from the Artificial Intelligence field should offer information about such technology to the population in an accessible manner (e.g., institutional material).
The Voice Assistant should explain to users about the privacy of their data to help them decide which tasks to perform.
Marketing on Voice Assistants should stick to these systems' actual capabilities, presenting common and realistic use cases.
Developers of conversational flows should receive training on the Assistans' technical limitations so they can produce appropriate flows.

Developers should understand users (e.g., profiles, goals, contexts, behavior, semantics, mental models) to create solutions that address their needs and context.
Developers should research mechanisms in human conversations and adapt them to interactions with Voice Assistants.
Developers should always apply best practices of usability and voice interaction when designing Voice Assistants.
Developers should improve speech recognition technology (e.g., synonyms, intents, words in other languages, accents, localization, and different voice types and users).

This section presented the categories of responses given by participants concerning causes for users' misperceptions and solutions to address this issue. The 35 statements resulting from such analysis were used in round two, as presented below.

6.3. Second and third round's results

In the Delphi's second questionnaire, we presented the statements derived from the first round along with a 7-point Likert scale, asking participants to rate each phrase according to their level of agreement. The professionals could also leave optional comments. Although only eight participants have left comments, a total of 18 participants evaluated the statements, being the sample for this round (n = 18). We will present the results by exposing, for each statement, the following descriptive statistics: mean, median, Interquartile Range (IQR), and percentages of agreement, disagreement, and neutrality. The sections will also present some open comments from the participants, which support the interpretation of the quantitative data.

As for the study's third round, we presented the before-mentioned statistics for each statement and provided open fields for optional comments. Five professionals answered the form, but only two left comments on the results (n = 2). Such a low response rate was somehow expected since leaving comments was optional. Considering the third round's small sample, we decided to add these comments to this section instead of creating a separate part.

Each section will start by presenting the statements for which the experts reached a strong consensus (i.e., IQR $\leq$ 1 AND agreement/disagreement percentage $\geq$ 75%). Then, we proceed to statements that reached only mild consensus (i.e., IQR

≤ 1 OR agreement/disagreement percentage ≥ 75%). Finally, we display the phrases that caused incongruencies among the group's views.

6.3.1.
Causes for issues in users' mental models

The first part of the questionnaire prompted the participants the following instruction: "*Below, we present statements that summarize the groups' opinions about the **CAUSES** that lead users of Voice Assistants to form mental models that are unaligned with these systems' actual capabilities. Please state your **level of agreement** with each statement below on a 1 to 7 scale. You may also leave **optional comments** on the statements"*. In total, the professionals reached a strong consensus on five statements (table 6.4), mild consensus on four statements (table 6.5), and could not reach a consensus for the remaining seven phrases (table 6.6).

Table 6.4 – First part's statements for which the professionals reached a strong consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
S1	Bad previous experiences with other voice interfaces create negative expectations for the Voice Assistants on users.	6,22	6	1	100%	0%	0%
S2	Characteristics that induce anthropomorphism (e.g., voice, name, gender, metaphors, humor) cause users to expect Voice Assistants would be as capable as a human.	5,61	6	1	89%	6%	6%
S5	Users construct their mental models of conversations through human interactions, but Voice Assistants have lower conversational skills, letting down expectations and causing frustration.	6,11	6	1	89%	11%	0%
S9	Voice Assistants are too complex and demand too much of the users' cognition.	2,67	2	1	11%	11%	78%
S14	In Science Fiction, systems powered with Artificial Intelligence are pictured as futuristic, intelligent, sensitive, talkative, and capable, creating unaligned perceptions about current Assistants.	6,22	6,5	1	89%	11%	0%

The results from round two show that all participants strongly accorded that previous bad experiences with other voice interfaces influence users' mental models of VAs (S1). Such a phrase was part one's only statement that reached a 100% level of agreement. One professional highlighted in the comments that such an affirmation only applies to real-life interfaces, not mediatic representations that could lead to an opposite effect.

Both statements 2 and 5 – which are somehow related to the impact of VAs' humanness on users' perceptions – caused a strong consensus. These were the only drivers directly related to VA design that reached a strong consensus in the first part. On this topic, Participant 17 commented that "*From my experiences, users that anthropomorphize VAs can expect more human-like interactions which are not yet available in many VA devices and given the limited visual interaction, some may find it confusing to interact with VAs without more training and knowledge of what to do*." (P17).

Oppositely, statement 9 was the only phrase leading to a strong consensus of disagreement among all of the questionnaire's statements (median = 2 / disagree). That is, it was unanimous that a high level of complexity in VAs does not lead to issues in users' understandings. One professional highlighted that VAs are not complex, nor require too much cognitive load. However, another participant argued that VAs complexity could be an issue for novice users: "*As I see it, this scenario [high complexity of VAs] could be applied to novice users, without previous experience. Something like the beginning of the learning curve*" (P5).

In addition, the group mutually agreed that media picturing AI-powered systems as futuristic and intelligent (S14) causes misalignments in users' models. No professional further commented on this affirmation.

Table 6.5 – First part's statements for which the professionals reached mild consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
S3	Users do not know the Voice Assistants' technical limitations and require the Assistant to perform tasks and recognize commands beyond its capabilities.	6,17	6,5	1,75	94%	6%	0%
S13	Marketing raises users' expectations by exaggerating the Voice Assistant's social skills and intelligence, showing use cases that are too simple and fluid.	5,78	6	1,75	78%	17%	6%
S8	Voice Assistants do not explain their limitations for certain actions, such as recognizing the conversational context.	5,61	6	2	83%	6%	11%
S10	Voice Assistants do not explain to users about their skills, how they should be utilized, how they process commands, how they make decisions, or how to recover from failures.	5,72	6	2	89%	6%	6%

The medians and agreement percentages indicate that the group recognized the impact of the statements in table 6.5 on users' mental models. Nonetheless, the

IQR value of > 1 suggests that the participants could not reach a consensus on how much these factors are relevant. The group reached a mild consensus for statement 3 but did not provide further arguments. On the other hand, the professional who disagreed with statement 13 mentioned that: "*I think the marketing of what can be done is fine and matches what the devices can do. I think that the challenge is user expectations and desires of what could be*." (P17).

Similar to the phrases in table 6.4, only two statements (S8 and S10) focused on VA design, originating from the previously mentioned "VAs' transparency" category. In statement 8, one professional argued that providing detailed information such as limitations could jeopardize the experience and make the VA more complex. Similarly, participant 1 also questioned the need for presenting certain information: "*I agree that [the VAs] do not explain about its capabilities, but I do not understand the 'processing of commands'. Why would they need to explain how they process their commands or make decisions?*" (P1).

As mentioned before, we will also present the third round's results in this section. Only one participant left comments on the phrases above. They mentioned that VAs are still early concepts, still in development, making marketing and support strategies precocious. According to them, this idea is reinforced by the fact that misunderstandings about the VAs are not limited to users only, extending even to developers. They also mentioned that VAs are, in fact, complex when compared with other products due to the large scope made possible by free speech.

Table 6.6 – First part's statements for which the professionals did not reach consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
12	Brands present little institutional content about the Assistants and Artificial Intelligence, leading users to buy the product without being aware of its capabilities.	4,56	5	1,75	61%	11%	28%
16	Users do not look for information about the Voice Assistant before buying it, especially for tasks that are deemed unnecessary.	5,06	5,5	2	56%	33%	11%
6	Voice Assistants do not tell users when they update, nor explain the updates in their skills.	4,94	6	2,5	72%	0%	28%
4	Privacy concerns hinder users from interacting with the Assistants for long enough to construct correct mental models.	3,83	3	2,75	33%	11%	56%
15	Voice Assistants from different brands have different skills, leading to the belief that an Assistant might have the same skills as the others.	5,17	5	2,75	67%	22%	11%

11	Developers conduct little research about user experience, which is paramount to creating conversational flows.	4,22	4,5	3	50%	11%	39%
7	Voice Assistants do not present initial instructions to users, leaving them without knowing what to expect from the product.	5,00	5,5	3,75	56%	17%	28%

Table 6.6 presents the phrases that did not lead to any consensus among the group. While the sample tended to agree with statement 12 (median = 5/ partially agree), a high percentage of the group disagreed with such an idea (28%). Participant 17 explained that brands do advertise VAs properly, but some users unexpectedly receive VAs as gifts from other people, and therefore had no previous awareness of their functioning. As for statement 16, two participants argued against the affirmation by mentioning that it would not make sense to read about skills considered unnecessary, and not reading manuals is standard behavior for any product. Both professionals also considered that such a statement unfairly blames the user for issues with VAs rather than the product's design: "*I think it is complex to "blame" the user. The user should not have to look after information on how to use, the communication should be intuitive and guided. Maybe it is a cause (okay, users do not inform themselves), but it is not, in my opinion, a fair cause*" (P1).

The professionals tended to agree with statement 6, but 28% disagreed with the affirmation. Participant 5 argued that "*updates concerning the scope or the assistant's functionalities do not seem to me as something reasonable to notify the user*". Despite a slight tendency to disagree with statement 4 (median = 3/ partially disagree), the percentages show that the group's opinions were split for this cause. Overall, three participants commented that, while people with strong privacy concerns simply avoid using VAs, such concerns are not significant for how users interact: "*I don't think that users are really adapting/reducing their interactions with voice assistants due to privacy concerns. From my experience, people who have privacy concerns do not use voice assistants at all (...). Others, although they are aware of and know about privacy risks (...), seem to not really change their interaction behavior." (P16).*

Statement 15 also provoked a tendency of agreement (media = 5/ partially agree) that could not reach a consensus. Only one participant attributed a possible confusion between VAs from different brands to their humanness: "*The distinction between brands is clear (such as smartphones). However, the conversational*

interfaces' humanization makes this distinction increasingly invisible." (P5). The phrase closest to neutrality (S11; median = 4,5) also divided the group. In their comments, two participants reported that conversational design is a new area, still lacking tools or skilled professionals to deal with user experience appropriately: "*I think developers are aware of the importance of user experience, however, the expertise required for CUI design might not always be available since it is a reemerging area*." (P17). Finally, statement 7 had the higher IQR, indicating the highest dispersion in the professionals' opinions. No comments were provided for this statement.

In round three, the participants left two comments about the statement in table 6.6. Participant 5 argued in favor of the VAs' proactivity, explaining that no other products could find appropriate moments to proactively provide information for users. Nevertheless, using Alexa and Amazon's newsletter as an example, they argued that such content is still limited to English, posing barriers to non-English speakers. Furthermore, participant 2 considered the lack of consensus good. "*I believe there are differences in roles, lives, and knowledge among the other participants in this study. Some aspects of this market may change a lot according to with whom you are talking to"* (P2).

6.3.2. Solutions to deal with misalignments in users' mental models

The questionnaire's second part presented the following instructions to the professionals: "*Below, we present statements that summarize the groups' opinions on the **SOLUTIONS** to align users' mental models with the Voice Assistants' actual capabilities. Please state your **level of agreement** with each statement below on a 1 to 7 scale. You may also leave **optional comments** on the statements."* The participants reached a strong consensus for eight statements (table 6.7) and mild consensus for eight phrases (table 6.8). Only three statements did not lead to consensus (table 6.9).

Table 6.7 – Second part's statements for which the professionals reached a strong consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement

24	Voice Assistants should present initial instructions, tutorials, and information about new supported actions and new ways to formulate commands.	6,00	6,5	1	83%	6%	11%
25	Marketing on Voice Assistants should stick to these systems' actual capabilities, presenting common and realistic use cases.	6,28	6	1	100%	0%	0%
26	Developers should improve speech recognition technology (e.g., synonyms, intents, words in other languages, accents, localization, and different voice types and users).	6,11	6,5	1	89%	0%	11%
27	Developers should create error recovery mechanisms (ex: inform what was misunderstood and how to reformulate the command).	6,44	7	1	94%	6%	0%
29	Developers of conversational flows should receive training on the Assistants' technical limitations so they can produce appropriate flows.	6,11	7	1	83%	11%	6%
30	Developers should understand users (e.g., profiles, goals, contexts, behavior, semantics, mental models) to create solutions that address their needs and context.	6,33	7	1	94%	6%	0%
31	Developers should research mechanisms in human conversations and adapt them to interactions with Voice Assistants.	6,11	6	1	94%	6%	0%
32	Developers should always apply best practices of usability and voice interaction when designing Voice Assistants.	6,28	7	1	89%	11%	0%

As presented in table 6.7, six out of the eight statements that reached a strong consensus were solutions directed to developers. On this topic, one participant stated that the word “developer” should encompass all the professionals who are involved in VAs’ development, and not only software developers. We also identified that the solutions reaching a solid consensus in part two had more extreme mean and median values than phrases in part one. These statistics might mean that the group had stronger opinions towards the solutions.

Interestingly, presenting information, tutorials, and instructions (S24) was unanimously considered an adequate solution to align users’ perceptions (median = 6,5), even though the professionals have expressed concerns about presenting too much information in their previous comments. Likewise, while marketing did not induce a strong consensus as a cause for issues in users’ perceptions, statement 25 was the only solution to cause a 100% level of agreement. The group did not report any comments for these statements.

The professionals also agreed that developers should focus on improving speech recognition (S26; median = 6,5) and error-handling mechanisms (27; median = 7/ strongly agree), but no comments were provided. The group tended to strongly agree that developers of conversational flows should receive technical

training (S29), with one participant extending this recommendation for developers responsible for error-handling strategies. No comments were reported for statement 30, with which the group tended to strongly agree (median = 7).

As for statement 31, the group unanimously agreed (median = 6), but two participants left concerns. Participant 16 mentioned that VAs already count on qualified professionals from the linguistics or natural language processing (NLP) fields to study human communication. Participant 17 also argued that command-based interactions might be beneficial in some cases: "*For my own experiences, some users appreciate the command interactions for their convenience in certain tasks. I think it could be beneficial to explore for more complex tasks, but I think it would depend on the task the VA is being used to facilitate*." (P17). Similarly, applying best practices (S32) led to a strong consensus of agreement (median = 7/ strongly agree), but participant 11 regarded: "*Yes, but also [it is necessary to] study the current solutions and evolve them when necessary, test, and experiment other paths. It [VAs] is still a very new scenario, with iterations every day. A best practice for one scenario may not work for others*" (P11).

The only professional to comment on this group of statements in round three was participant 5. They mentioned that the solutions in table 6.7, if applied, will surely be beneficial for the yet-to-be-explored world of conversational interfaces.

Table 6.8 – Second part's statements for which the professionals reached mild consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
34	Users should receive training on how to use Voice Assistants, including supported language patterns.	4,00	4	1	33%	44%	22%
20	Voice Assistants should present usage tips throughout interactions, including mechanisms to clarify the conversation context.	6,00	6,5	1,75	89%	6%	6%
35	No solution should be applied to Voice Assistants since users will naturally learn how to interact.	2,72	2,5	1,75	17%	6%	78%
17	Voice Assistants should make it clear for users that they are talking to a machine and not a human.	5,56	6,5	2	78%	11%	11%
19	Voice Assistants should provide examples and explanations about their skills' scope and action execution, decision making, and learning processes.	6,06	6,5	2	94%	6%	0%
22	The Voice Assistant should explain to users about the privacy of their data to help them decide which tasks to perform.	5,67	6	2	78%	17%	6%
23	Manufacturers and professionals from the Artificial Intelligence field should offer information about such technology to the population in an accessible manner (e.g., institutional material).	5,72	6	2	78%	17%	6%

28	There should be a platform that shows failed past interactions to help users understand the reasons for errors and the system's scope.	5,50	5,5	2	78%	17%	6%

Table 6.8 shows another eight statements that reached mild consensus. Statement 34 was the only consensus of the group's opinions tending to neutrality. Although the agreement/disagreement percentages were balanced, the IQR value of 1 and the median of 4 suggest that the participants collectively tended to stay neutral towards training users. Four participants commented on this topic, arguing that users usually do not read manuals or instructions and that such a solution could diminish the naturality of user-VA interactions. The participants proposed that information should be provided to users subtly throughout interactions. "*On the one hand, we want to humanize/ make interactions natural, but on the other hand, we want to instruct users. There is a controversy. (...) I doubt users would spend time searching and reading information about how to use the VA. Here we could insert the [VAs'] proactivity*." (P5).

In line with the previous suggestion, the group reached a mild consensus in agreeing that usage tips could be included throughout interactions (S20). However, one participant was concerned that adding too many tips could make interactions tedious and slow. "*I think there might be many ways to make mental models more accurate, but I don't think all of them ought to be simultaneously applied as it would eliminate the fun and convenience that leads people to interact in the first place*" (P14).

The only statement with which the group tended to disagree was that no solution should be employed (S35). Such a result was somehow expected since the group acknowledged many impactful factors for issues in users' perceptions (section 6.2), and therefore not applying any solution might not solve the problem. Only one participant commented this phrase: "*I do think there should be an improvement which is why I somewhat disagree with the idea that users will learn how to interact, but again I think it comes down to the conceptual and mental models aligning, and I believe this will be at the task level.*" (P17).

The participants reached a mild consensus in agreeing with statements 17 and 19 (median = 6,5), but no comments were provided. The results for statement 19 were aligned with the statistics described for statement 10 in section 6.2, meaning that the group kept their mild consensus for the adequacy of a cause (statement

10) and a solution (statement 19). Statements 22 and 23, with which the group tended to agree, had the same median (6/ agree) and agreement percentage (78%), and only one participant commented on each phrase. Participant 17 explained that privacy issues still prevent VA adoption (S22), so it is useful to clarify this topic. Moreover, participant 11 agreed with statement 23 but argued that such a solution alone would not solve the issue of users' mental models.

As for creating a platform showing failed past interactions (S28), although the group accorded slightly that it would be a proper solution, two participants expressed concerns. Participant 5 believed that users would not consult such an interface, and since user-VA interactions are quick, there is room to accommodate the trial-and-error approach: "*It is not interesting that the user finds out a failure through trial-and-error and ends up giving up. However, I imagine that such cases are similar to a human conversation in which (...) they try to clarify what was said*" (P5). Participant 11 also argued that solutions should not place the responsibility of change on the user, but should modify the product itself: "*Again, it seems to me that we are suggesting educating the users to understand our project instead of evolving the project to have adequate usability to the users*".

Again, only one professional left a comment on round three. They repeated that the field of conversational design is emergent, and the solutions represented in statements 17, 19, 20, and 22 could be particularly beneficial for users.

Table 6.9 – Second part's statements for which the professionals did not reach consensus.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
18	Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).	3,83	4	2	39%	22%	39%
21	The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.	5,50	6	2,75	72%	22%	6%
33	Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).	4,61	4	4	44%	22%	33%

Finally, table 6.9 displays the statements that did not lead the group to a consensus. The participants had split opinions on statement 18, tending to stay neutral. We identified that the agreement and disagreement percentages are the same (39%), reinforcing that the group could not reach a consensus. Three participants

left comments on this topic. One of them, who tended to agree, argued that: "*the developers should humanize the assistant to make it more 'approachable', particularly for a modality with such immediatism and rich social presence*" (P5). On the other hand, other professionals also mentioned that, for some tasks, humanizing the VA might not be beneficial or safe. Participant 16 argued that it might be too late to abolish the anthropomorphic features. "*I believe we are past this point of avoiding human characteristics. Maybe we should focus on making people understand that although the voice assistant has human characteristics, its capabilities are different - there are things the voice assistant can do better than humans, and others which it cannot do at all.*" (P16). Curiously, while the group could not reach a consensus on removing anthropomorphic characteristics off VAs, they had previously accorded that VAs' humanness is a cause for issues in mental models (section 6.2).

Although 72% of the participants tended to agree with statement 21, the IQR value of 2,75 suggests a high dispersion on the group's view. Two participants explained that users might not want to teach the VA for all tasks and that manufacturers might not allow such collaboration. Finally, the participants tended to stay neutral towards statement 33, resulting in the highest dispersion metric among the phrases (IQR = 4). No comments were provided for this phrase.

The only professional who commented on these phrases in round three, participant 5, reinforced their previous opinions. They argued that anthropomorphic characteristics are essential for VAs to connect with users and make interactions natural. The professional also proposed that user-system collaboration could result from a long user-VA relationship, being attractive for people interested in buying a VA.

In the Delphi study, we identified that not every cause to issues in users' misperceptions led the group to suggest a specific solution. For example, while we observed a category of "VAs' anthropomorphic features" as a driver for misperceptions and an equivalent solution of "mitigate anthropomorphism", other categories did not provoke such parallelism. Factors such as differences between VAs, influences from science fiction media, and the VAs' complexity were not approached directly in the improvement proposals. Likewise, the analysis showed solutions that did not seem to directly address any of the reported causes. Some examples are the

categories of highlighting user-VA collaboration and improving speech recognition technology.

Moreover, the participants tended to accord more on solutions to deal with users' mental models than on the drivers causing such misperceptions. On the one hand, sixteen solutions led to some consensus (strong consensus = 8; mild consensus = 8), and only three did not provoke any accordance. On the other hand, only nine causes made the group agree to some level (strong consensus = 5; mild consensus = 4), and the remaining seven factors did not lead to consensus. These results indicate that the group was more confident in evaluating solutions than determining causes to issues in users' mental models.

Similarly, we observed different results for associated causes and solutions, such as the before-mentioned anthropomorphism topic. The group reached a strong consensus on the influence of VAs' humanness on users' perceptions, but did not collectively agree that developers should avoid human-like characteristics. Although this incongruency was more salient for the humanness theme, we also identified such a pattern for other categories, such as VAs' transparency, users' privacy concerns, influences of marketing, and lack of user research.

The previous chapters reported the results of exploratory interviews and the Delphi study. In the next chapter, we will discuss the results of both techniques in light of the literature and address this research's goals and questions.

7
Discussion

This research aimed to offer recommendations to align users' mental models of VAs with these systems' real capabilities. For this purpose, we conducted literature reviews, exploratory interviews, and a Delphi study applying a three-phased questionnaire. In this chapter, we will discuss our findings in light of the literature to address the research's main goal and question (*how can VAs be improved to mitigate gaps between users' mental models and the VAs' actual capabilities?).*

Overall, the themes discussed in the interviews and in the Delphi study were similar, presenting partially coinciding results with the systematic literature review (SLR). We observed that the Delphi study led to more topics around users' mental models than the interviews, which can be attributed to a greater number of participants, who also had a longer deadline to prepare their answers. However, in both techniques, we identified that the participants did not accord on all emerging themes, as shown throughout this chapter. In addition, this study's main difference from the SLR's findings were discussions on peripheral factors to VAs' design, such as the influence of marketing and science fiction, the weight of stakeholders, and the significance of the development team's skills. This incongruence might be explained by the fact that we purposely limited the SLR's scope to publications reporting user studies, while our studies involved developers and researchers. Below, we will discuss the main topics identified in the results.

In the first place, the most controversial subject was the influence of VAs' anthropomorphic characteristics on users' perceptions and how to manage such features. Similar to the SLR's findings, the participants pointed to names, natural voices, metaphors, gender, and a humoristic conversational style as aspects that induce anthropomorphism. According to both techniques' findings, there is a general agreement that VAs' humanness leads users to expect human-level skills from these systems, including conversational capabilities. As a consequence, the experts reported that users frequently hoped to interact with VAs in a human-like manner,

applying conversational styles that are not supported by the interface. These results also echo the SLR's findings. Specifically, in the interviews, the gap between users' behavior and the VAs' limits was reported to affect task performance and cause frustration on users. Moreover, in the Delphi study, the statements reflecting such notions were among the only five phrases to cause a strong consensus on participants in part one.

Unexpectedly, some interviewed professionals reported cases of users confusing a conversational interface for a human, making the realization that an agent is a machine embarrassing and frustrating. These examples were all originated from phone-based conversational agents for customer service. Unlike VAs – that have a dedicated device (e.g., Alexa and the Amazon Echo) or are embedded in a smartphone (e.g., Siri and the iPhone)– users interact with such agents through phone calls, and this specificity might account for this for the confusion. Nonetheless, the number of skills performed by VAs is increasing steadily, reaching over 70,000 in 2020 for Alexa alone (STATISTA, 2020). Among these skills, shopping applications are common, and VAs might undergo similar issues of the phone-based agents if customer support skills are highly anthropomorphized. Thus, the concern that users might confuse VAs for humans may not be dismissed and, as pointed by one professional, may have ethical implications. The concern about ethics is in line the literature on principles for AI systems, as shown by Ruiz and Quaresma (2021). The authors reviewed a range of guidelines for AI systems and observed that, to follow the principle of human dignity, these systems should indicate to users whether they are talking to a machine or another human.

Despite the professionals' general agreement on the influence of VAs' humanness on users' misperceptions, the participants in both techniques were divided on the adequacy of removing the VAs' anthropomorphic features. On the one hand, some professionals argued that humanizing the VAs makes interactions more natural for users, reducing potential tension in talking to robotic-like agents. Similarly, interviewed participants commented that humanizing VAs might benefit presenting information about the system itself or other complex information. Such a path might make information more accessible to the products' target audiences since the machine's vocabulary can be adjusted to match the users' knowledge or linguistic practices. This view is somehow consistent with the prevailing belief that the naturality

of speech benefits voice interaction in terms of accessibility and user experience (BHOWMIK, 2015; PEARL, 2016).

Arguments favoring anthropomorphism also indicated that the aspect contributes to VAs' image and popularity, which the literature reinforces. Studies have pointed that VAs' humanness is a driver for adoption and is specifically valued by niches such as children and people with cognitive disabilities (BALASURIYA *et al.*, 2018; FESTERLING; SIRAJ, 2020; GARG; SENGUPTA, 2020). Moreover, a study by Purington et al. (PURINGTON *et al.*, 2017) showed a relationship between users' tendency to anthropomorphize VAs and satisfaction, suggesting that users may be found of these interfaces' humanness generally.

On the other hand, experts who disagreed with humanizing VAs argued that more straightforward interactions might lead to fewer errors and support task performance. Such opinions echo some studies in the literature which suggest reducing VAs' humanness (DOYLE *et al.*, 2019; LUGER; SELLEN, 2016). The participants also mentioned that it is necessary to favor the users' goals and task types when determining VAs' humanness levels. According to the findings, developers should identify what users wish to accomplish and when (or whether) it is adequate to insert features such as playful prompts or overly human voices and conversational styles. In the Delphi, the group even mildly accorded that VAs should clarify that the system is not human, following the ethical guidelines exposed previously.

Furthermore, participants in both techniques suggested that "we are past the point" of removing the humanness off VAs, and even mitigating anthropomorphic characteristics might not prevent users' from humanizing the interface. This view is consistent with the work of Nass and colleagues, who indicated that users' mindlessly attribute human perceptions to computers when the system performs a human's role (e.g., assistant) and provide word-based answers to the users' inputs (NASS; MOON, 2000; NASS; STEUER; TAUBER, 1994).

Another broadly discussed theme caused friction of opinions among both techniques' participants: VAs' outputs and their transparency levels. During the interviews and the Delphi's first round, several transparency-related factors were linked to users' misunderstandings about VAs. Notably, various participants reported that the system fails to present relevant information for users' mental models. Echoing the SLR's results, such information included available features, ways of usage, technical functioning (i.e., command processing and decision making),

initial instructions to support learning, explanations on error sources and recovery paths, and clarifications on system updates.

Although we initially surveyed various information-presentation issues, the Delphi's second round indicated that the professionals did not reach a strong consensus on the effects of these matters on users' mental models. Among the four statements reflecting the transparency category, none led to a strong consensus, two reached a mild consensus, and two did not reach any accordance. These results might mean that while transparency is a general issue to users' perceptions, the group could not agree on what information specifically is troublesome. The possibility is reinforced by comments left by some professionals, arguing that information such as "limitations for certain actions", "command processing", and "updates in skills" are not relevant for users.

The lack of accordance on how much information to display was also identified in results concerning the solutions' adequacy. Interestingly, all phrases related to transparency, error-handling, and supplementary contents led to some consensus among the group (strong consensus = 2; mild consensus = 5), contradicting the results of the cause-type statements. For example, although the professionals did not consider that the absence of initial instructions and update notifications caused issues, the group had a strong consensus that applying such solutions could leverage users' mental models. Likewise, while the group only mildly accorded with the cause-type statement approaching error-handling, the creation of error recovery mechanisms was unanimously considered a paramount solution (median = strongly agree). Meanwhile, concerns that interactions could become slow, unnatural, tedious, or that users would not want to consult supplementary content prevented some phrases from reaching a strong consensus. The statements affected by these concerns on the Delphi were: embedding usage tips in interactions, training users, presenting examples and explanations on VAs' scope, and creating an error-visualization platform. Therefore, we observed divergencies in the professionals' opinions on what to present to users, probably driven by the uncertainty on which information is truly relevant and the threshold between too little or excessive content.

Similar concerns and heterogeneity of views were observed in the exploratory interviews. The experts commented that increasing transparency is a desirable proposal if it leads to fewer errors and installs confidence in users. Nevertheless, the professionals highlighted that presenting too much information might

jeopardize the interactions' quickness and easiness. These usability regards are aligned with literature indicating that, although increasing transparency leverages task performance, providing instructions excessively might displease users and make interactions slower (KIRSCHTHALER; PORCHERON; FISCHER, 2020; MOTTA; QUARESMA, 2022). Developers also recommend that instructions, confirmations, acknowledgments, and other outputs aiming to direct the interaction should be used carefully to avoid repetition (GOOGLE, 2017b).

As discussed, we observed an essential factor to be considered when designing both VAs' levels of transparency and humanness: the users' context. In the Delphi, the group reached a strong consensus that developers should understand users and tailor solutions to address their needs. Such comprehension included users' profiles, goals, usage contexts, behavior, semantics, and current mental models. Although no notices were made for this statement specifically, the participants left various comments addressing the subject throughout the Delphi's second round. As mentioned above, some professionals were against presenting too much information, arguing that such a proposal places the responsibility of learning about the VA on users and not on the product's design (e.g., training users, letting them learn naturally over time, or telling them to inform themselves better). Similarly, participants commented that "users do not look for information about VAs" – a phrase that did not lead to any consensus – puts the blame on users for misalignments in their mental models. The professionals argued that this view is inappropriate since the product and its developers are to blame for any issues in users' mental models. This recommendation is in line with interaction design literature suggesting that designers should adequate products to users' needs, goals, and conceptual models (HACKOS; REDISH, 1998).

Strengthening the proposition above, comments in the Delphi and in the interviews suggest that it might be essential to consider the conversational agents' operational domain or task types to determine VAs' transparency and humanness levels. As the professionals explained, domains that involve more complex information might require more transparency and explicability from the VA than more simple activities (e.g., investment skills *versus* music players). As shown by this dissertation's authors in a previous study (MOTTA; QUARESMA, 2021), VAs' tasks have different usability metrics regarding error number, error types, task completeness, and user satisfaction. Such indications reinforce the idea that some

activities may be easier to achieve than others, and developers need to carefully examine the correct information amounts required for users to complete the tasks. This recommendation may be extended to VAs' humanness since tasks such as games and playful interactions are utilized by user niches that value VAs' anthropomorphism (e.g., children; FESTERLING; SIRAJ, 2020).

In a similar stream, the Delphi's results indicate that users' previous experiences and characteristics influence their mental models. The group strongly accorded (100% agreement) that previous bad experiences with other conversational interfaces create negative expectations for VAs on users. The interviewed professionals also reported identifying differences in behavior from novice and heavy users of VAs and technology in general, reporting a more exploratory attitude from expert users. This tendency is in line with the SLR's results, which showed that users' age, technical backgrounds (e.g., people who work or study technology-related areas), and level of experience with voice interaction affect their behavior and perceptions. As shown, users with high experience with VAs and technology are more successful in interactions and better understand VAs' functioning (LAU; ZIMMERMAN; SCHAUB, 2018; LOPATOVSKA *et al.*, 2019; LUGER; SELLEN, 2016; MYERS; FURQAN; ZHU, 2019). Thus, as suggested by some professionals, information amounts to be displayed could be defined – among other parameters – by the users' level of experience with VAs and technology. Less experienced users might need more guidance and information to interact (CHEN; WANG, 2018).

Furthermore, we observed that science fiction might influence users' perceptions, a possibility with which the group unanimously agreed. This kind of media portraits robots as futuristic and highly intelligent, sociable, sensitive, talkative, and capable, which possibly induces users to expect such capacities from VAs. Although science fiction was not observed as impactful for users' mental models in the SLR, Cambre *et al.*'s (2020) forecasting study might reinforce its influence. The authors asked users to create hypothetical usage scenarios for 30 years in the future, resulting in utopic (e.g., AI systems becoming humans' friends) and dystopic (e.g., the world ruled by dictator machines) plots, usually observed in sci-fi media. Considering this trend, factors external to the VAs design might affect users' perceptions, including users' experiences with technology and VAs and other actors such as science fiction.

Additionally, the device type that holds the conversational agent might influence users' mental models, as shown in the SLR's results. Although this theme was not broadly discussed nor accorded by this research's participants, the possibility raised on the Delphi's first round that users expect VAs from different brands to present the same features might be related to the literature. Studies exposed that the device's type, physical presence, and placement influences users' perception of the VAs' roles and available features (ABDOLRAHMANI *et al.*, 2020; CHO; LEE; LEE, 2019; LI; RAU; HUANG, 2019; LOPATOVSKA *et al.*, 2019). Nevertheless, as few publications approached the influences of VAs' hardware on users' perceptions, this theme requires further investigation.

Independent of the parameters that should be used to decide the adequate information amount, the interviewed professionals also showed skepticism on the feasibility of increasing VAs' transparency. For instance, technological limitations in the AI technology supporting VAs might impede the interface from correctly diagnosing and explaining failures. On this topic, the literature about eXplainable Artificial Intelligence (XAIs) has already indicated that many AI systems – especially those with better performance – are black boxes, unable to explain how they make decisions (HOLZINGER, 2018). Furthermore, the experts reported that designing transparency in conversational interfaces would require a solid previous understanding of the conversational flow, but since VAs are generalist interfaces, it is challenging to foresee usage contexts. Likewise, studies show that, for humans, producing and understanding explanations is highly dependent on a context (MILLER, 2019), making it necessary for XAI systems to consider the context and relevant information to present outputs (FERREIRA; MONTEIRO, 2020). Nonetheless, VAs do not understand contextual information (AMMARI *et al.*, 2019; LUGER; SELLEN, 2016), resulting in additional obstacles to designing transparency.

Finally, another barrier to designing not only transparent outputs but also VAs, in general, might be the development team's qualifications. The professionals mentioned that developers should possess the necessary knowledge and skills to design interactions. Applying usability and voice interaction best practices were considered paramount, and the professionals strongly agreed that researching human linguistic practices and adapting interactions is an adequate solution to deal with users' misperceptions. Once again, these solutions did not seem to address a

specific issue since "lack of user research from developers" was not unanimously considered a driver for problems in users' mental models. The group possibly tended to stay neutral towards such a cause since many participants reported working in academia rather than in development teams, being unsure of market practices.

Whereas employing adequate tools and skills seem significant to developing VAs that induce correct mental model on users, the literature suggests that some fields involved in VA development might not follow such recommendations. For example, Souza and Quaresma (2019) reviewed market reports to understand the profile of User Experience (UX) designers. The authors observed that a significant share of these professionals has no graduate degree in Design or related areas and often search for knowledge in a self-taught manner. Additionally, many technology-related companies employ agile development methodologies. As shown by Da Costa Brito and Quaresma (2019), these workflows only partially support important User-Centered Design (UCD) guidelines, which are frequently deferred in favor of deadlines. These indications call into question whether VA development comprises the required parameters exposed by the participants to employ appropriate solutions.

Similarly, some professionals also commented that voice interaction and VAs are an emerging field of work and that new unexplored skills might be essential to design interactions. As a participant on the Delphi proposed, it might be necessary to assess the suitability of existing best practices and make adaptations when needed. Such tendency may also apply to development and evaluations tools, as indicated in the literature. For example, Zwakman, Pal, and Arpnikanondt (2021) evaluated the suitability of the System Usability Scale (SUS) questionnaire to assess the usability of VAs. The authors reported that the SUS – initially planned to evaluate Graphical User Interfaces (GUIs) – had drawbacks to assessing voice interactions' learnability and user-friendliness. In turn, the authors developed a Voice Usability Scale (VUS), made to address VAs interaction's easiness, affective aspects (i.e., user satisfaction), and recognizability & visibility (related to transparency). Thus, developers may need to acquire different knowledge from existing GUI guidelines and develop new tools to accommodate VAs' specificities.

Nonetheless, VAs' novelty might be challenging for developers to understand the technology's functioning and limitations. As explained by one interviewed professional, as AI-based voice interfaces are a relatively new trend, developers

themselves might not fully understand all of these systems' aspects. The Delphi's participants' previously mentioned difficulty in determining which information to display for users might be indicative of this hypothesis. Since there are many characteristics to uncover and understand, prioritizing content might be challenging. Accordingly, the group strongly accorded that designers of conversational flows should be trained on technical aspects of VAs' functioning to solve the users' mental model matter.

The need to understand VAs may not be restricted to developers only, extending to managers, businesspeople, and other stakeholders involved in the VAs' projects. The interviewed experts reported difficulties communicating with managers and businesspeople commissioning conversational interfaces. They explained that these actors usually do not possess much knowledge on the conversational agents' limitations nor experience in conversational design. However, their requests are still influential for design decisions, such as the agents' level of humanness. These actors' lack of knowledge may also account for VAs' unrealistic marketing, a highly mentioned topic by this research's participants. In the Delphi, the group mildly accorded that marketing raises users' expectations by exaggerating VAs' skills and advertising overly simple use cases. Consistently, the group reached a 100% agreement that marketing should be loyal to VAs' actual capabilities.

The significance of solutions proposing changes for developers – especially acquiring knowledge on designing voice interactions and understanding the VAs' functioning – aligns with Norman's (2013) description of the relationships between designers and the product. According to the author, the designer parts from their own understanding of the product to build the system image, which is fundamental for users' mental models. Following such a definition, it is coherent that developers' understandings of VAs are considered paramount since misperceptions could result in a biased system image, hampering users' models. Likewise, developers must possess the required skills and tools to translate their perceptions appropriately into a solid system image.

Nevertheless, as an interviewed expert suggested, conversational agents such as VAs are complex systems, usually put together by large development teams who might possess incongruent mental models. The results of the Delphi study, which gathered professionals from different graduation backgrounds, development roles, and institution types (i.e., academia or companies), reinforce this possibility.

In the second round, the group accorded and tended to have more assertive evaluations for solutions than for causes of users' misperceptions (i.e., more extreme mean and median values), reaching a strong consensus for only five causes. Moreover, although the solutions on which the group strongly accorded are valuable, we did not identify a strong cause-solution parallelism. That is, among the eight solution-type statements reaching strong accordance, none seem to directly tackle the five cause-type phrases that also led to a strong consensus. Consistently, aside from the need for error-handling mechanisms, all solutions leading to strong consensus seemed to suggest generic actions, such as improving speech recognition technologies, applying usability best practices, and improving the developers' skills.

Hence, the lower accordance on influential issues, the lack of cause-solution parallelism, and the solutions' generic nature reinforce the possibility that a diverse group of professionals might have divergent views. Such variations in the VA developers' views could bring both benefit and harm. On the one hand, the professionals' views should be carefully assembled so the team can accord on which issues to address and how to solve them. As Ackoff (1974 *apud* MORAES, 1997) argues, designers should be more attentive in selecting the wrong problem rather than the wrong solution to the right problem. However, as we mentioned above, the group had a lower tendency in agreeing to issues leading to users' mental models. On the other hand, incongruent points of view may result in identifying a higher number of trouble sources and surveying numerous proposals to address them, as observed in the Delphi's first round. As discussed throughout this chapter, such proposals are highly valuable for designing VAs' that induce correct mental models on users.

8
Conclusion

Voice Assistants (VAs) bring several benefits to users and are increasingly popular, but some barriers to these systems' usage and adoption still prevail. Such obstacles are related to how users perceive and understand VAs, that is, their mental models. The literature indicated that users' mental models are unaligned with these systems' capabilities, lacking an adequate understanding of VAs' functioning and comprising high expectations for VAs' features, intelligence, and conversational capabilities. Such an issue not only leads to frustration and abandonment of VAs but could also account for adoption barriers such as negative attitudes, privacy concerns, and perceptions of low usefulness and ease of use.

Considering these consequences and the role of an appropriate mental model for task performance levels, it was necessary to investigate how to improve VAs to mitigate gaps between users' mental models and the VAs' real skills. Thus, the objective of this research was to identify leading causes of users' misperceptions and offer design recommendations for aligning users' mental models of VAs with these systems' real capacities. To achieve this goal, firstly, we systematically reviewed the literature to understand the state of the art of users' mental models since comprehending how users currently perceive VAs was the first step to identify influential factors and solutions to misperceptions. Then, we interviewed experts in conversational interfaces to understand their opinions on users' mental models of VAs. Finally, we conducted a questionnaire-based, three-round Delphi study with professionals experienced in the research or development of conversational interfaces. The latter technique aimed to identify leading causes of misalignments in users' mental models of VAs and survey solutions to deal with such an issue.

This research's findings indicate that anthropomorphic features strongly influence users' perceptions, creating exaggerated expectations for VAs' intelligence and conversational skills. Highly humanized conversational agents could also confuse users on how to interact, hampering task performance and even creating ethical

concerns. However, whereas this research's participants agreed that VAs' humanness led to misperceptions, removing the VAs' humanization was not a unanimous solution. Both the literature and our studies suggest that anthropomorphism benefits VA adoption and is valued by some user niches. Therefore, developers should ponder the advantages and drawbacks of anthropomorphism when designing VAs.

In a similar manner, the literature pointed that VAs might lack transparency in their outputs to explain to users about their functioning and features, a notion with which our participants generally accorded. Nevertheless, we observed that the professionals disagreed on which information is missing specifically and what pieces of information should be prioritized throughout interactions. Whereas increasing the VAs' transparency through cues, instructions, feedback, and tutorials was considered an adequate solution to mitigate errors and align users' understandings, we observed regards that interactions could become slow and tedious. Hence, similarly to VAs' levels of humanness, developers should determine the right information amount to display.

To aid design decision on both topics mentioned above, this research's participants' overall agreement was that VA developers should conduct research to understand the usage context and establish the user's requirements. We observed in all techniques that users' backgrounds and characteristics are influential for their mental models, including their interests in technology, age, and educational background. For example, it is possible that infrequent or novice users might need more assistance on interacting, suggesting the need to adapt VAs' transparency levels to the users' profiles. Likewise, users' goals, task aspects, and the conversational agents' usage domains are relevant to comprehending the interaction's requirements. Varied activities might require the support of different information and explainability levels, and could benefit (or not) from the VAs' humanization. Thus, developers should have a solid understanding of these context-related aspects when applying solutions so that VAs' humanness or transparency levels do not hamper task performance or impede users from reaching their goals.

Despite the recommendations mentioned above, there might still exist restraints to the feasibility of the solutions. In the first place, technological limitations might hinder increased VAs' transparency levels and error-handling mechanisms, as AI systems might not be able to diagnose errors and suggest recovery paths correctly. Similarly, since VAs are generalist interfaces that can perform a wide range

of tasks (e.g., skills), foreseeing usage contexts might be challenging. These predictions are crucial to the beforementioned assessment of users' needs and are also necessary to design conversational flows with increased transparency.

Furthermore, the interviews and Delphi's findings suggested that VA developers must possess the proper knowledge and skills to apply usability and voice interaction best practices. Nonetheless, due to VAs' relative novelty, it is possible that misunderstandings concerning VAs' functioning and limitations also exist among developers. For a similar reason, some participants indicated that currently available best practices might not meet the requirements for VA development, and existing knowledge should be adapted when necessary. Therefore, professionals should search for new required skills and alter or create new development and testing tools when needed.

Such a demand may also extend for other stakeholders involved in VA development since these actors might have an influence on design and marketing decisions despite their possible low understanding of these systems' thresholds. Such an issue might originate from incipient development processes and methodologies since the technology market might still be adapting to VAs' novelty. Stakeholders should be aware of the VAs' limitations and benefits to identify market gaps that could benefit from voice interaction rather than apply conversational agents based on the desire for innovation. It might also be advantageous that companies attempt to conduct tests and adapt their methodologies and processes to improve VA development.

Finally, the Delphi results suggest that team members from different backgrounds might have varied perceptions, and it might be challenging to align all of the developers' mental models of VAs. Considering that a homogeneous understanding among the development team may be essential to designing a consistent system image for VAs, it might be needed to identify paths and processes to support these professionals' communication. While varied views are vital to identifying problems and survey solutions, a congruency of perceptions might be valuable when dealing with a specific trouble source that requires a focused solution.

Considering the conclusions above, we propose the following recommendations to align users' mental models with VAs' actual capabilities:

- Developers should consider VAs usage context when designing solutions, including users' profile, interests, and goals, VAs' usage domains and task characteristics, and device specificities (see chapters 3, 5, and 6);
- Developers should adequate VAs' levels of humanness and transparency based on the usage context (as specified above), granting that these features will not hamper the interaction's usability or create obstacles for task performance (see chapters 3, 5, and 6);
- Developers must possess the required skills to design VAs and search for new knowledge and adapt development and evaluation tools when needed (see chapters 5 and 6);
- Both developers and stakeholders should search for information to adequately understand VAs' functioning and limitations (see chapters 5 and 6);
- New methodologies and processes might be beneficial to VA development, aiding the teams' communication with stakeholders and supporting a more homogeneous understanding of the VAs' project at hand. Such a solution might lead to the better identification of issues and problem solving, possibly inducing improvements in the system image (see chapter 6).

8.1. Limitations

This research provided insights and recommendations to align users' mental models of VAs, but some limitations existed. Firstly, we only conducted three rounds in the Delphi study, and we did not allow the participants to review their quantitative evaluations of the statements. Adding subsequent questionnaires until more statements reached a consensus could have strengthened the study's findings, although such an attempt might have failed considering the low number of respondents in the final round.

Secondly, in the interviews and the SLR, we observed topics that were not suggested in the Delphi's first round, such as the influence of stakeholders outside of development teams and the impacts of device type or placement on users' mental models. These causes could have been included as statements for the second round. Alternatively, we could have conducted more interviews, using the two open-ended questions of the Delphi's first round as a script, and then analyzed the results to

summarize the interviewees' comments into statements for the second round. This method – modified Delphi (LINSTONE; TURROF, 1975) – could have improved the statements to be evaluated since the interviews allowed the experts to speak in more detail, and the moderator could interrogate them into explaining real-life scenarios from their work experience. However, the modified Delphi would have required initial planning that we did not expect since the interviews had an exploratory purpose, aiming to prepare the subsequent Delphi.

Furthermore, as discussed in the previous chapter, although valuable, some of the statements reaching the highest level of consensus among the Delphi group were relatively generic. As reported above, the usage context is highly influential to designing solutions, and therefore we attribute this result to the lack of a specific scenario describing a specific issue in users' mental models of VAs (e.g., error-handling for appointment scheduling with Alexa). On the one hand, the absence of a particular usage context benefited our exploratory research question, aiding the identification of a broad set of causes for users' misperceptions and solutions to solve the matter (including the importance of the usage context itself). Thus, as this work did not aim to provide design and development guidelines for a specific product in a particular domain, we believe that the results aided the identification of relevant research gaps for future work (see section 8.2). However, studies aiming to address a more specific issue might need to contextualize the problem to the Delphi group.

Another adaptation that could have been employed is the utilization of different question types in the second questionnaire. As we only allowed the group to assess the statements through Likert scales, our results are limited to their opinions on the adequacy of such causes and solutions to the mental model issue. Nonetheless, other questions could have been available, such as ranking or scoring types of questions. By combining these question types, our findings could have further explored how the professionals understand the impact level of each cause for users' misperceptions and how they prioritize solutions. However, such a questionnaire design could have turned the second round into a long task, coming along with the risk of a high number of participants dropping out mid-study.

8.2. Future work

This research suggests some open gaps that are yet to be addressed. Firstly, our studies results did not deeply examine the influence of users' backgrounds on their mental models. It is still unknown exactly which characteristics are influential to users' perceptions and what variables are particularly affected. Although some studies in the SLR have indicated some aspects such as technical knowledge, age, and previous voice interaction experience, the results were conflicting and indicated the need for further assessment. Similarly, few studies in the SLR addressed the effects of device type and placement on users' perceptions of VAs' expertise and expectations for features. We also observed that the direction of such a relationship is still unclear (i.e., is it the device type/ placement that affects users' perceptions or vice versa?), being a research gap.

Our findings also indicate that several usage domains could be investigated to identify task requirements and determine adequate levels of transparency and anthropomorphism for VAs. Studies could examine varied usage contexts for voice interaction, such as healthcare interfaces, smart speakers for home automation, and in-vehicle VAs.

As for development improvements, we consider that future research could evaluate the applicability of currently existing usability best practices and tools for designing and evaluating VAs. Likewise, developers should attempt to understand which skills and knowledge could be necessary to design VAs and assess whether the market fulfills such requirements. Finally, it could be beneficial to create new development methodologies that support the varied professionals involved in VAs' projects to align their understandings of relevant problems and adequate solutions.

9

References

ABDOLRAHMANI, A.; STORER, K. M.; ROY, A. R. M.; KUBER, R.; BRANHAM, S. M. Blind Leading the Sighted. **ACM Transactions on Accessible Computing**, vol. 12, no. 4, p. 1–35, 20 Jan. 2020. DOI 10.1145/3368426.

AMAZON. What is a Voice User Interface? 2020. Available at: https://developer.amazon.com/pt-BR/alexa/alexa-skills-kit/vui. Accessed on: 10 Dec. 2020.

AMMARI, T.; KAYE, J.; TSAI, J. Y.; BENTLEY, F. Music, Search, and IoT: How people (really) use voice assistants. **ACM Transactions on Computer-Human Interaction**, vol. 26, no. 3, 2019. DOI 10.1145/3311956.

BALASURIYA, S. S.; SITBON, L.; BAYOR, A. A.; HOOGSTRATE, M.; BRERETON, M. Use of voice activated interfaces by people with intellectual disability. 4 Dec. 2018., cited By 6. **Proceedings of the 30th Australian Conference on Computer-Human Interaction**. New York, NY, USA: ACM, 4 Dec. 2018. p. 102–112. DOI 10.1145/3292147.3292161.

BARNUM, C. **Usability Testing Essentials. Ready, Set...** Burlington, MA: Elsevier, 2011. DOI 10.1016/B978-0-12-375092-1.00023-4.

BENETEAU, E.; GUAN, Y.; RICHARDS, O. K.; ZHANG, M. R.; KIENTZ, J. A.; YIP, J.; HINIKER, A. Assumptions Checked: How Families Learn About and Use the Echo Dot. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, New York, NY, USA, vol. 4, no. 1, Mar. 2020. DOI 10.1145/3380993.

BENETEAU, E.; RICHARDS, O. K.; ZHANG, M.; KIENTZ, J. A.; YIP, J.; HINIKER, A. Communication Breakdowns Between Families and Alexa. 2019. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2019. p. 1–13. DOI 10.1145/3290605.3300473.

BHOWMIK, A. Senses, Perception, and Natural Human-Interfaces for Interactive Displays. *In*: BHOWMIK, A. (ed.). **Interactive displays: Natural Human-Interface Technologies**. West Sussex, UK: Wiley, 2015. p. 1–26.

BONFERT, M.; SPLIETHÖVER, M.; ARZAROLI, R.; LANGE, M.; HANCI, M.; PORZEL, R. If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. 2018. **Proceedings of the 20th ACM International Conference on Multimodal Interaction**. New York, NY, USA: Association for Computing Machinery, 2018. p. 95–102. DOI 10.1145/3242969.3242995.

BURBACH, L.; HALBACH, P.; PLETTENBERG, N.; NAKAYAMA, J.; ZIEFLE, M.; CALERO VALDEZ, A. “Hey, Siri”, “Ok, Google”, “Alexa”. Acceptance-Relevant Factors of Virtual Voice-Assistants. 2019-July., Jul. 2019.

2019 IEEE International Professional Communication Conference (ProComm): IEEE, Jul. 2019. vol. 2019-July, p. 101–111. DOI 10.1109/ProComm.2019.00025.

CAMBRE, J.; REIG, S.; KRAVITZ, Q.; KULKARNI, C. "All Rise for the AI Director": Eliciting Possible Futures of Voice Technology through Story Completion. 2020. **Proceedings of the 2020 ACM Designing Interactive Systems Conference**. New York, NY, USA: Association for Computing Machinery, 2020. p. 2051–2064. DOI 10.1145/3357236.3395479.

CHEN, M.-L.; WANG, H.-C. How Personal Experience and Technical Knowledge Affect Using Conversational Agents. 5 Mar. 2018. **Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion**. New York, NY, USA: ACM, 5 Mar. 2018. p. 1–2. DOI 10.1145/3180308.3180362. Available at: https://doi.org/10.1145/3180308.3180362.

CHÉRIF, E.; LEMOINE, J.-F. Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. **Recherche et Applications en Marketing (English Edition)**, vol. 34, no. 1, p. 28–47, 13 Mar. 2019. DOI 10.1177/2051570719829432.

CHIN, H.; MOLEFI, L. W.; YI, M. Y. Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse. 21 Apr. 2020. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 21 Apr. 2020. p. 1–13. DOI 10.1145/3313831.3376461.

CHO, M.; LEE, S.; LEE, K.-P. Once a Kind Friend is Now a Thing. 18 Jun. 2019. **Proceedings of the 2019 on Designing Interactive Systems Conference**. New York, NY, USA: ACM, 18 Jun. 2019. p. 1557–1569. DOI 10.1145/3322276.3322332.

CLARK, L.; DOYLE, P.; GARAIALDE, D.; GILMARTIN, E.; SCHLÖGL, S.; EDLUND, J.; AYLETT, M.; CABRAL, J.; MUNTEANU, C.; EDWARDS, J.; R COWAN, B. The State of Speech in HCI: Trends, Themes and Challenges. **Interacting with Computers**, vol. 31, no. 4, p. 349–371, 2019. https://doi.org/10.1093/iwc/iwz016.

CLARK, L.; PANTIDI, N.; COONEY, O.; DOYLE, P.; GARAIALDE, D.; EDWARDS, J.; SPILLANE, B.; GILMARTIN, E.; MURAD, C.; MUNTEANU, C.; WADE, V.; COWAN, B. R. What Makes a Good Conversation? 2 May 2019. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2 May 2019. p. 1–12. DOI 10.1145/3290605.3300705.

COWAN, B. R.; BRANIGAN, H. P.; OBREGÓN, M.; BUGIS, E.; BEALE, R. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human−computer dialogue. **International Journal of Human-Computer Studies**, vol. 83, p. 27–42, 2015. DOI https://doi.org/10.1016/j.ijhcs.2015.05.008. Available at: http://www.sciencedirect.com/science/article/pii/S1071581915001020.

COWAN, B. R.; DOYLE, P.; EDWARDS, J.; GARAIALDE, D.; HAYES-BRADY, A.; BRANIGAN, H. P.; CABRAL, J.; CLARK, L. What's in an accent? 2019., i2. **Proceedings of the 1st International Conference on Conversational**

User Interfaces - CUI '19. New York, New York, USA: ACM Press, 2019. p. 1–8. DOI 10.1145/3342775.3342786.

COWAN, B. R.; PANTIDI, N.; COYLE, D.; MORRISSEY, K.; CLARKE, P.; AL-SHEHRI, S.; EARLEY, D.; BANDEIRA, N. "What can i help you with?" 4 Sep. 2017. **Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services**. New York, NY, USA: ACM, 4 Sep. 2017. p. 1–12. DOI 10.1145/3098279.3098539.

DA COSTA BRITO, Lara; QUARESMA, Maria Manuela Rupp; "O design centrado no usuário nas metodologias ágeis", p. 125-139 . In: **Anais do 17º Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano-Tecnologia e o 17 º Congresso Internacional de Ergonomia e Usabilidade de Interfaces e Interação Humano-Computador**. São Paulo: Blucher, 2019.

DE BARCELOS SILVA, A.; GOMES, M. M.; DA COSTA, C. A.; DA ROSA RIGHI, R.; BARBOSA, J. L. V.; PESSIN, G.; DE DONCKER, G.; FEDERIZZI, G. Intelligent personal assistants: A systematic literature review. **Expert Systems with Applications**, vol. 147, p. 113193, 2020. DOI 10.1016/j.eswa.2020.113193.

DELBECQ, A. L.; VAN DE VEN, A. H.; GUSTAFSON, D. H. **Group techniques for program planning: A guide to nominal group and delphi processes**. Glenview, IL: Scott, Foresman and Company, 1975.

DOYLE, P. R.; EDWARDS, J.; DUMBLETON, O.; CLARK, L.; COWAN, B. R. Mapping perceptions of humanness in intelligent personal assistant interaction. 2019. **Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2019**, Taipei Taiwan: ACM, 2019. https://doi.org/10.1145/3338286.3340116.

DUBIEL, M.; HALVEY, M.; GALLEGOS, P. O.; KING, S. Persuasive Synthetic Speech. 22 Jul. 2020., cited By 0. **Proceedings of the 2nd Conference on Conversational User Interfaces**. New York, NY, USA: ACM, 22 Jul. 2020. p. 1–9. DOI 10.1145/3405755.3406120.

FERREIRA, J. J.; MONTEIRO, M. S. What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. In: Marcus A., Rosenzweig E. (eds) Design, User Experience, and Usability. Design for Contemporary Interactive Environments. HCII 2020. Lecture Notes in Computer Science, vol 12201. Springer, Cham. https://doi.org/10.1007/978-3-030-49760-6_4

FESTERLING, J.; SIRAJ, I. Alexa, What Are You? Exploring Primary School Children's Ontological Perceptions of Digital Voice Assistants in Open Interactions. **Human Development**, vol. 64, no. 1, p. 26–43, 2020. DOI 10.1159/000508499.

FISH, L. S.; BUSBY, D. M. The Delphi Method. *In*: SPRENKLE, D. H.; PIERCY, F. P. (eds.). **Research methods in family therapy**. 2nd ed: Guilford Publications, 2005. p. 238–253.

FOURNEY, A.; DUMAIS, S. T. Automatic Identification and Contextual Reformulation of Implicit System-Related Queries. 2016. **Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval** [...]. New York, NY, USA: Association for Computing Machinery, 2016. p. 761–764. DOI 10.1145/2911451.2914701.

GARG, R.; SENGUPTA, S. He Is Just Like Me. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, New York, NY, USA, vol. 4, no. 1, p. 1–24, 18 Mar. 2020. DOI 10.1145/3381002. Available at: https://doi.org/10.1145/3381002.

GIANNAROU, L.; ZERVAS, E. Using Delphi technique to build consensus in practice. **International Journal of Business Science and Applied Management**, vol. 9, no. 2, p. 65–82, 2014.

GIL, A. **Métodos e técnicas de pesquisa social**. 6th ed. São Paulo: Editora Atlas, 2008.

GOOGLE. **Be Cooperative... Like Your Users**. Available at: <https://developers.google.com/actions/design/be-cooperative>. Accessed in: 5 set. 2017.

GOOGLE. **Instilling User Confidence Through Confirmations and Acknowledgements.** Available at: <https://developers.google.com/actions/design/instilling-user-confidence>. Accessed in: 5 set. 2017b.

GRICE, P. **Studies in the Way of Words**. First pape. London, England: Harvard University Press, 1991.

GUZMAN, A. L. Voices in and of the machine: Source orientation toward mobile virtual assistants. **Computers in Human Behavior**, vol. 90, p. 343–350, Jan. 2019. DOI 10.1016/j.chb.2018.08.009.

HACKOS, J. T.; REDISH, J. **User and Task Analysis for Interface Design**. Toronto: Wiley, 1998.

HOLZINGER, A. From Machine Learning to Explainable AI. Aug. 2018. **2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)**, 2018, pp. 55-66, doi: 10.1109/DISA.2018.8490530.

HOY, M. B. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. **Medical Reference Services Quarterly**, vol. 37, no. 1, p. 81–88, 2018. https://doi.org/10.1080/02763869.2018.1404391.

HUTCHBY, I.; WOOFFITT, R. **Conversation analysis: Principles, practices and applications**. Cambrigde: Polity press, 1998.

HUXOHL, T.; POHLING, M.; CARLMEYER, B.; WREDE, B.; HERMANN, T. Interaction Guidelines for Personal Voice Assistants in Smart Homes. Oct. 2019. **2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)**, IEEE, Oct. 2019. p. 1–10. DOI 10.1109/SPED.2019.8906642.

IBGE. Educação 2018 PNAD Contínua. **Pesquisa Nacional por Amostra de Domicílios Cont**, vol. 2018, no. 2, p. 1–12, 2019. .

INVOCA. **The Rise of Voice**., 2018. Available at: https://www.witlingo.com/the-rise-of-voice-timeline/. Accessed on: 29 April 2019

JAVED, Y.; SETHI, S.; JADOUN, A. Alexa's Voice Recording Behavior. 26 Aug. 2019., cited By 0. **Proceedings of the 14th International Conference on Availability, Reliability and Security** [...]. New York, NY, USA: ACM, 26 Aug. 2019. p. 1–10. DOI 10.1145/3339252.3340330.

JUNG, H.; KIM, H.; HA, J.-W. Understanding Differences between Heavy Users and Light Users in Difficulties with Voice User Interfaces. 22 Jul. 2020. **Proceedings of the 2nd Conference on Conversational User Interfaces** [...]. New York, NY, USA: ACM, 22 Jul. 2020. p. 1–4. DOI 10.1145/3405755.3406170.

KENDALL, L.; CHAUDHURI, B.; BHALLA, A. Understanding Technology as Situated Practice: Everyday use of Voice User Interfaces Among Diverse Groups of Users in Urban India. **Information Systems Frontiers**, vol. 22, no. 3, p. 585–605, 7 Jun. 2020. DOI 10.1007/s10796-020-10015-6.

KIERAS, D. E.; BOVAIR, S. The Role of a Mental Model in Learning to Operate a Device. **Cognitive Science**, vol. 8, p. 255–273, 1984.

KIM, J.; JEONG, M.; LEE, S. C. "Why did this voice agent not understand me?" 21 Sep. 2019., e1; i1. **Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings** [...]. New York, NY, USA: ACM, 21 Sep. 2019. p. 146–150. DOI 10.1145/3349263.3351513.

KIRSCHTHALER, P.; PORCHERON, M.; FISCHER, J. E. What Can i Say?: Effects of Discoverability in VUIs on Task Performance and User Experience. 2020., **CUI '20: 2nd Conference on Conversational User Interfaces,** Bilbao, Spain: Association for Computing Machinery, 2016, https://doi.org/10.1145/3405755.3406119

KISELEVA, J.; WILLIAMS, K.; JIANG, J.; HASSAN AWADALLAH, A.; CROOK, A. C.; ZITOUNI, I.; ANASTASAKOS, T. Understanding User Satisfaction with Intelligent Assistants. 2016a. **Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval** [...]. New York, NY, USA: Association for Computing Machinery, 2016. p. 121–130. DOI 10.1145/2854946.2854961.

KISELEVA, J.; WILLIAMS, K.; JIANG, J.; HASSAN AWADALLAH, A.; CROOK, A. C.; ZITOUNI, I.; ANASTASAKOS, T. Understanding User Satisfaction with Intelligent Assistants. 13 Mar. 2016b. **Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval** [...]. New York, NY, USA: ACM, 13 Mar. 2016. p. 121–130. DOI 10.1145/2854946.2854961.

KITCHENHAM, B. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Duham, UK: University of Durham, 2007.

KROEMER, K. H. E.; GRANDJEAN, E. **Fitting the task to the human - Atextbook of occupational ergonomics**. 5th ed. Philadelphia, PA: Taylor & Francis, 1997. vol. 4.

KUZMINYKH, A.; SUN, J.; GOVINDARAJU, N.; AVERY, J.; LANK, E. Genie in the Bottle. 21 Apr. 2020. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2020. p. 1–13. DOI 10.1145/3313831.3376665.

LAU, J.; ZIMMERMAN, B.; SCHAUB, F. Alexa, Are You Listening? **Proceedings of the ACM on Human-Computer Interaction**, New York, NY, USA, vol. 2, no. CSCW, p. 1–31, Nov. 2018. DOI 10.1145/3274371.

LEE, S.; CHO, M.; LEE, S. What if Conversational Agents Became Invisible?Comparing Usersa Mental Models According to Physical Entity of AI Speaker. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, vol. 4, no. 3, 2020. DOI 10.1145/3411840.

LETHO, M.; NAH, F.; YI, J. DECISION-MAKING MODELS, DECISION SUPPORT, AND PROBLEM SOLVING. *In*: SALVENDY, G. (ed.). **Handbook of Human Factors and Ergonomics**. 4th ed. New Jersey: John Wiley & Sons, 2012. p. 192–242.

LI, Z.; RAU, P.-L. P. Effects of Self-Disclosure on Attributions in Human–IoT Conversational Agent Interaction. **Interacting with Computers**, vol. 31, no. 1, p. 13–26, 1 Jan. 2019. DOI 10.1093/iwc/iwz002.

LI, Z.; RAU, P.-L. P.; HUANG, D. Self-Disclosure to an IoT Conversational Agent: Effects of Space and User Context on Users' Willingness to Self-Disclose Personal Information. **Applied Sciences**, vol. 9, no. 9, p. 1887, 8 May 2019. DOI 10.3390/app9091887.

LINSTONE, H.; TURROF, M. **The Delphi Method: Techniques and Applications**. Newark, NJ: Addison-Wesley, 1975.

LIU, S. Forecast of eCommerce transactions value via voice assistants worldwide in 2021 and 2023., 2021. Available at: https://www.statista.com/statistics/1256695/ecommerce-voice-assistant-transactions/#:~:text=The%20total%20worldwide%20transaction%20value,billion%20U.S.%20dollars%20in%202023. Accessed on: 2 January 2022

LOPATOVSKA, I. Personality Dimensions of Intelligent Personal Assistants. 2020., I2. **Proceedings of the 2020 Conference on Human Information Interaction and Retrieval** [...]. New York, NY, USA: Association for Computing Machinery, 2020. p. 333–337. DOI 10.1145/3343413.3377993.

LOPATOVSKA, I.; GRIFFIN, A. L.; GALLAGHER, K.; BALLINGALL, C.; ROCK, C.; VELAZQUEZ, M. User recommendations for intelligent personal assistants. **Journal of Librarianship and Information Science**, vol. 52, no. 2, p. 577–591, 8 Jun. 2020. DOI 10.1177/0961000619841107.

LOPATOVSKA, I.; OROPEZA, H. User interactions with "Alexa" in public academic space. **Proceedings of the Association for Information Science and Technology**, vol. 55, no. 1, p. 309–318, Jan. 2018. DOI 10.1002/pra2.2018.14505501034.

LOPATOVSKA, I.; RINK, K.; KNIGHT, I.; RAINES, K.; COSENZA, K.; WILLIAMS, H.; SORSCHE, P.; HIRSCH, D.; LI, Q.; MARTINEZ, A. Talk to me: Exploring user interactions with the Amazon Alexa. **Journal of Librarianship and Information Science**, vol. 51, no. 4, p. 984–997, 7 Dec. 2019. DOI 10.1177/0961000618759414.

LOPATOVSKA, I.; WILLIAMS, H. Personification of the Amazon Alexa. 2018. **Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18** [...]. New York, New York, USA: ACM Press, 2018. p. 265–268. DOI 10.1145/3176349.3176868.

LOVATO, S. B.; PIPER, A. M.; WARTELLA, E. A. Hey Google, Do Unicorns Exist? 12 Jun. 2019. **Proceedings of the 18th ACM International Conference**

on Interaction Design and Children [...]. New York, NY, USA: ACM, 12 Jun. 2019. p. 301–313. DOI 10.1145/3311927.3323150.

LUGER, E.; SELLEN, A. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. 2016. **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: Association for Computing Machinery, 2016. p. 5286–5297. DOI 10.1145/2858036.2858288.

MASSARO, W. D. Language and information processing. *In*: MASSARO, W. D. (ed.). **Understanding Language - An information-processing analysis of speech perception, reading, and psycholinguistics**. New York, NY, USA: Academic Press, 1975. p. 3–28. https://doi.org/10.1016/b978-0-12-478350-8.50007-6.

MAUÉS, M. P. **Marcela Pedroso Maués Um olhar sobre os assistentes virtuais personificados e a voz como interface**. 2019. Pontifical Catholic University of Rio de Janeiro, 2019.

MCLEAN, G.; OSEI-FRIMPONG, K. Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. **Computers in Human Behavior**, vol. 99, p. 28–37, 2019. DOI https://doi.org/10.1016/j.chb.2019.05.009.

MEEKER, M. **Internet Trends 2016**, 2016. Available at: https://www.kleinerperkins.com/perspectives/2016-internet-trends-report. Accessed on: 29 April 2019.

MILLER, T. Explanation in artificial intelligence: Insights from the social sciences. **Artificial Intelligence**, vol. 267, p. 1–38, Feb. 2019. DOI 10.1016/j.artint.2018.07.007.

MOAR, J. **The digital assistants of tomorrow**. Hampshire, UK:, 2019. Available at: https://www.juniperresearch.com/document-library/white-papers/the-digital-assistants-of-tomorrow. Accessed on: 29 April 2019

MOAR, J.; ESCHERICH, M. Hey Siri, how will you make money? 2020. Available at: https://www.juniperresearch.com/whitepapers/hey-siri-how-will-you-make-money Accessed on: 2 January 2022

MOORE, R. J.; ARAR, R. **Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework**. New York, NY: Association for Computing Machinery, 2019. DOI 10.1145/3304087.

MORAES, A. Algumas estratégias para a implementação da pesquisa em design considerando sua importância para a consolidação do ensino de design. **Estudos em Design**, vol. Maio, Número Especial, 1997.

MORIUCHI, E. Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. **Psychology & Marketing**, vol. 36, no. 5, p. 489–501, 15 May 2019. DOI 10.1002/mar.21192.

MOTTA, I.; QUARESMA, M. OPPORTUNITIES AND ISSUES IN THE ADOPTION OF VOICE ASSISTANTS BY BRAZILIAN SMARTPHONE USERS. **Ergodesign & HCI**, vol. 7, no. Especial, p. 138, 31 Dec. 2019. DOI 10.22570/ergodesignhci.v7iEspecial.1312.

_____. Understanding Task Differences to Leverage the Usability and Adoption of Voice Assistants (VAs). In: Soares M.M., Rosenzweig E., Marcus A. (eds) Design, **User Experience, and Usability: Design for Contemporary Technological Environments. HCII 2021. Lecture Notes in Computer Science**, vol 12781. Springer, Cham. https://doi.org/10.1007/978-3-030-78227-6_35

_____. Users' Error Recovery Strategies in the Interaction with Voice Assistants (VAs). In: Black N.L., Neumann W.P., Noy I. (eds) **Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021). IEA 2021. Lecture Notes in Networks and Systems**, vol 223. Springer, Cham. https://doi.org/10.1007/978-3-030-74614-8_82

MYERS, C.; FURQAN, A.; NEBOLSKY, J.; CARO, K.; ZHU, J. Patterns for how users overcome obstacles in Voice User Interfaces. 2018-April., 2018a. **Conference on Human Factors in Computing Systems - Proceedings** [...]. [*S. l.*: *s. n.*], 2018. vol. 2018-April, p. 1–7. https://doi.org/10.1145/3173574.3173580.

MYERS, C.; FURQAN, A.; NEBOLSKY, J.; CARO, K.; ZHU, J. Patterns for How Users Overcome Obstacles in Voice User Interfaces. 19 Apr. 2018b., E. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 19 Apr. 2018. p. 1–7. DOI 10.1145/3173574.3173580.

MYERS, C. M. Adaptive Suggestions to Increase Learnability for Voice User Interfaces. 2019., E. **Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion** [...]. New York, NY, USA: Association for Computing Machinery, 2019. p. 159–160. DOI 10.1145/3308557.3308727.

MYERS, C. M.; FURQAN, A.; ZHU, J. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. 2 May 2019., E. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 2 May 2019. p. 1–9. DOI 10.1145/3290605.3300277.

MYERS, C. M.; GRETHLEIN, D.; FURQAN, A.; ONTAÑÓN, S.; ZHU, J. Modeling Behavior Patterns with an Unfamiliar Voice User Interface. 7 Jun. 2019., cited By 0. **Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization** [...]. New York, NY, USA: ACM, 7 Jun. 2019. p. 196–200. DOI 10.1145/3320435.3320475.

NASS, C.; BRAVE, S. **Wired for Speech – How Voice Activates and Advances the Human–Computer Relationship**. Cambridge, MA: MIT Press, 2005.

NASS, C.; MOON, Y. Machines and Mindlessness: Social Responses to Computers. **Journal of Social Issues**, vol. 56, no. 1, p. 81–103, 2000.

NASS, C.; STEUER, J.; TAUBER, E. R. Computers Are Social Actors. 1994. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: Association for Computing Machinery, 1994. p. 72–78. DOI 10.1145/191666.191703.

NORMAN, D. **The design of everyday things**. Revised an. New York, NY, USA: Basic books, 2013.

OH, Y. H.; CHUNG, K.; JU, D. Y. Differences in Interactions with a Conversational Agent. **International Journal of Environmental Research and Public Health**, vol. 17, no. 9, p. 3189, 4 May 2020. DOI 10.3390/ijerph17093189.

PARK, S.; LIM, Y. Investigating User Expectations on the Roles of Family-shared AI Speakers. 21 Apr. 2020., cited By 0. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2020. p. 1–13. DOI 10.1145/3313831.3376450.

PEARL, C. **Designing Voice User Interfaces: Principles of Conversational Experiences**. S: O'Reilly, 2016., E-book

PETERSEN, E. **Is it too soon...?**, 2020. Available at: https://www.tiktok.com/@lizemopetey/video/6870511001536646405. Accessed on: 2 January 2022.

PITARDI, V.; MARRIOTT, H. R. Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. **Psychology & Marketing**, vol. 38, no. 4, p. 626–642, 20 Apr. 2021. DOI 10.1002/mar.21457.

PORCHERON, M.; FISCHER, J. E.; REEVES, S.; SHARPLES, S. Voice Interfaces in Everyday Life. 2018-April., 21 Apr. 2018., i2. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2018. vol. 2018-April, p. 1–12. DOI 10.1145/3173574.3174214.

PORCHERON, M.; FISCHER, J. E.; SHARPLES, S. "Do Animals Have Accents?" 25 Feb. 2017., E. **Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing** [...]. New York, NY, USA: ACM, 25 Feb. 2017. p. 207–219. DOI 10.1145/2998181.2998298.

PRADHAN, A.; LAZAR, A.; FINDLATER, L. Use of Intelligent Voice Assistants by Older Adults with Low Technology Use. **ACM Trans. Comput.-Hum. Interact.**, New York, NY, USA, vol. 27, no. 4, Sep. 2020. DOI 10.1145/3373759.

PRADHAN, A.; MEHTA, K.; FINDLATER, L. "Accessibility Came by Accident." 21 Apr. 2018., E. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2018. p. 1–13. DOI 10.1145/3173574.3174033.

PRIDMORE, J.; ZIMMER, M.; VITAK, J.; MOLS, A.; TROTTIER, D.; KUMAR, P. C.; LIAO, Y. Intelligent Personal Assistants and the Intercultural Negotiations of Dataveillance in Platformed Households. **Surveillance & Society**, vol. 17, no. 1/2, p. 125–131, 31 Mar. 2019. DOI 10.24908/ss.v17i1/2.12936.

PURINGTON, A.; TAFT, J. G.; SANNON, S.; BAZAROVA, N. N.; TAYLOR, S. H. "Alexa is my new BFF": Social roles, user satisfaction, and personification of the Amazon Echo. **Conference on Human Factors in Computing Systems - Proceedings**, vol. Part F1276, p. 2853–2859, 2017. https://doi.org/10.1145/3027063.3053246.

ROBART, A. Looking Ahead to the Voice Era. 2017. Available at: https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/Looking-Ahead-to-the-Voice-Era. Accessed on: 29 Apr. 2019.

RONG, X.; FOURNEY, A.; BREWER, R. N.; MORRIS, M. R.; BENNETT, P. N. Managing Uncertainty in Time Expressions for Virtual Assistants. 2017. **Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: Association for Computing Machinery, 2017. p. 568–579. DOI 10.1145/3025453.3025674.

RUIZ, C.; QUARESMA, M. UX Aspects of AI Principles: The Recommender System of VoD Platforms. In M. M. Soares, E. Rosenzweig, & A. Marcus (Eds.), Design, User Experience, and Usability: Design for Contemporary Technological Environments. HCII 2021. Lecture Notes in Computer Science, vol 12781 (pp. 535–552). Springer, Cham. https://doi.org/10.1007/978-3-030-78227-6_38

SACKS, H.; SCHEGLOFF, E. A.; JEFFERSON, G. A Simplest Systematics for the Organization of Turn-Taking for Conversation. **Language**, vol. 50, no. 4, p. 696, Dec. 1974. DOI 10.2307/412243.

SCHEGLOFF, E. A.; JEFFERSON, G.; SACKS, H. The Preference for Self-Correction in the Organization of Repair in Conversation. **Language**, vol. 53, no. 2, p. 361, 1977. https://doi.org/10.2307/413107.

SCHEGLOFF, E. A.; SACKS, H. Opening up closings. **Semiotica**, vol. 8, no. 4, p. 289–327, 1973.

SOUZA, B.; QUARESMA, M. O perfil do UX designer: um panorama da visão do mercado de trabalho. Mar. 2019. **In: 13º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design, 2018, Joinville. Anais do 13º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design.** Joinville: Univille, 2018. **v. 1.**. São Paulo: Editora Blucher, Mar. 2019. p. 6121–6121. DOI 10.5151/ped2018-7.1_AIC_03.

STATISTA. Total number of Amazon Alexa skills in selected countries as of January 2020. 2020. Available at: https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/. Accessed on: 2 Nov. 2021.

TRAJKOVA, M.; MARTIN-HAMMOND, A. “Alexa is a Toy”: Exploring Older Adults’ Reasons for Using, Limiting, and Abandoning Echo. 21 Apr. 2020. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2020. p. 1–13. DOI 10.1145/3313831.3376760.

VAILSHERY, L. S. Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions). 2021. Available at: https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/. Accessed on 2 January 2022.

VAILSHERY, L. S. Voice technology - Statistics & Facts. 2022. Available at: https://www.statista.com/topics/6760/voice-technology/#dossierKeyfigures. Accessed on 12 January 2022.

VTYURINA, A.; FOURNEY, A. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. 2018-April., 21 Apr. 2018., cited By 19. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 21 Apr. 2018. vol. 2018-April, p. 1–7. DOI 10.1145/3173574.3173782.

WEBER, P.; LUDWIG, T. (Non-)Interacting with conversational agents. 6 Sep. 2020. **Proceedings of the Conference on Mensch und Computer** [...]. New York, NY, USA: ACM, 6 Sep. 2020. p. 321–331. DOI 10.1145/3404983.3405513.

WEST, M.; KRAUT, R.; HAN EI, C. **I'd blush if I could**., 2019. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1. Accessed on 29 April 2019.

WHITE-SMITH, H.; CUNHA, S.; KORAY, E.; KEATING, P. **Technology Tracker - Q1 2019**, 2019. Available at: https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-03/techtracker_report_q12019_final_1.pdf. Accessed on 29 April 2019.

WICKENS, C. D.; LEE, J. D.; LIU, Y.; GORDON-BECKER, S. **Introduction to human factors engineering**. 2nd ed. London, England: Pearson, 2014.

_____. Cognition. In **An Introduction to Human Factors Engineering**. second. London: Pearson, 2014. p. 100–134.

WILSON, J. R.; RUTHERFORD, A. Mental models: Theory and application in human factors. **Human Factors**, vol. 31, no. 6, p. 617–634, 1989. https://doi.org/10.1177/001872088903100601.

WU, Y.; EDWARDS, J.; COONEY, O.; BLEAKLEY, A.; DOYLE, P. R.; CLARK, L.; ROUGH, D.; COWAN, B. R. Mental Workload and Language Production in Non-Native Speaker IPA Interaction. 22 Jul. 2020. **Proceedings of the 2nd Conference on Conversational User Interfaces** [...]. New York, NY, USA: ACM, 22 Jul. 2020. p. 1–8. DOI 10.1145/3405755.3406118.

XU, Y.; WARSCHAUER, M. Exploring Young Children's Engagement in Joint Reading with a Conversational Agent. 2020a. **Proceedings of the Interaction Design and Children Conference** [...]. New York, NY, USA: Association for Computing Machinery, 2020. p. 216–228. DOI 10.1145/3392063.3394417.

_____. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. 2020b. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: Association for Computing Machinery, 2020. p. 1–13. DOI 10.1145/3313831.3376416.

YANG, X.; AURISICCHIO, M.; BAXTER, W. Understanding Affective Experiences with Conversational Agents. 2 May 2019. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems** [...]. New York, NY, USA: ACM, 2 May 2019. p. 1–12. DOI 10.1145/3290605.3300772.

YAROSH, S.; THOMPSON, S.; WATSON, K.; CHASE, A.; SENTHILKUMAR, A.; YUAN, Y.; BRUSH, A. J. B. Children asking questions: Speech interface reformulations and personification preferences. **IDC 2018 - Proceedings of the 2018 ACM Conference on Interaction Design and Children**, p. 300–312, 2018. https://doi.org/10.1145/3202185.3202207.

ZWAKMAN, D. S.; PAL, D.; ARPNIKANONDT, C. Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa. **SN Computer Science**, vol. 2, no. 1, p. 28, 11 Feb. 2021. DOI 10.1007/s42979-020-00424-4.

Appendix 1 – Primary studies accepted for the SLR

Paper	Group(s)	Research technique	Interface	Participant type
(ABDOLRAHMANI *et al.*, 2020)	1	Interview	VA	People with visual impairment
(AMMARI *et al.*, 2019)	1, 2, 3	Interview; Log data	VA	General
(BALASURIYA *et al.*, 2018)	1, 2, 3	Interview; Observation	VA	People with Intellectual Disability
(BENETEAU *et al.*, 2019)	3	Interview;Long-term experiment	VA	Families
(BENETEAU *et al.*, 2020)	1, 4	Interview; Long-term experiment	VA	Families
(BONFERT *et al.*, 2018)	2, 3	Wizard-of-Oz	Sim	General
(CAMBRE *et al.*, 2020)	1, 2	Survey; Storywriting task	VA	General
(CHEN; WANG, 2018)	5	Experiment	VA	Heavy/Tech & Light/Non-tech users
(CHÉRIF; LEMOINE, 2019)	2	Survey	Sim	General
(CHIN; MOLEFI; YI, 2020)	2, 3	Experiment	VA	General
(CHO; LEE; LEE, 2019)	1, 2, 3	Interview; Long-term experiment; Diary	VA	General
(CLARK *et al.*, 2019)	1, 2	Interview	VA	General
(COWAN *et al.*, 2015)	2, 3	Experiment	Sim	General
(COWAN *et al.*, 2017)	1, 2, 3	Focus group	VA	Infrequent users
(COWAN *et al.*, 2019)	3	Experiment	Sim	General
(DOYLE *et al.*, 2019)	1, 2, 3	Experiment; Interview	VA	General
(DUBIEL *et al.*, 2020)	2	Survey; Experiment	Sim	General
(FESTERLING; SIRAJ, 2020)	1, 2	Focus group; Game	VA	Children
(FOURNEY; DUMAIS, 2016)	3	Log data	VA	General

(GARG; SENGUPTA, 2020)	1, 3	Interview; Log data	VA	Families
(GUZMAN, 2019)	2	Interview	VA	General
(HUXOHL *et al.*, 2019)	1, 2, 4	Survey	VA	General
(JAVED; SETHI; JADOUN, 2019)	2, 5	Survey	VA	General
(JUNG; KIM; HA, 2020)	5	Experiment	Sim	Heavy & Light users
(KENDALL; CHAUDHURI; BHALLA, 2020)	1	Interview	VA	Different socioeconomic backgrounds
(KIRSCHTHALER; PORCHERON; FISCHER, 2020)	3	Wizard-of-Oz	Sim	General
(KISELEVA *et al.*, 2016)	2	Experiment	VA	General
(KUZMINYKH *et al.*, 2020)	2	Interview;Visualization exercise	VA	General
(LAU; ZIMMERMAN; SCHAUB, 2018)	2, 5	Diary study; Interview	VA	Users and non-users
(LEE; CHO; LEE, 2020)	1, 2	Experiment; Interview; Drawing task	Sim	General
(LI; RAU, 2019)	2	Wizard-of-Oz	Sim	General
(LI; RAU; HUANG, 2019)	1	Wizard-of-Oz	Sim	Individual or Pairs of users
(LOPATOVSKA; OROPEZA, 2018)	1	Field experiment; Survey; Interview	VA	Students
(LOPATOVSKA, I.; WILLIAMS, 2018)	5	Diary study	VA	General
(LOPATOVSKA *et al.*, 2019)	1, 2, 5	Survey; Diary study	VA	General
(LOPATOVSKA, 2020)	2, 4	Diary study; Field experiment	VA	General
(LOPATOVSKA *et al.*, 2020)	1, 2	Focus group	VA	General
(LOVATO; PIPER; WARTELLA, 2019)	1, 2, 3	Interview; Long-term study	VA	Children
(LUGER; SELLEN, 2016)	1, 2, 3, 4, 5	Interview	VA	General
(MYERS, *et al.*, 2018)	3	Experiment	Sim	General
(MYERS; FURQAN; ZHU, 2019)	4, 5	Experiment	Sim	General

(MYERS *et al.*, 2019)	5	Experiment	Sim	General
(OH; CHUNG; JU, 2020)	1, 2, 5	Interview; Long-term study	VA	Young adults and Elders
(PARK; LIM, 2020)	1	Participatory study; Interviews	VA	Families
(PORCHERON; FISCHER; SHARPLES, 2017)	3	Interview; Observation	VA	Groups
(PORCHERON *et al.*, 2018)	1, 2, 3	Long-term study	VA	Families
(PRADHAN; MEHTA; FINDLATER, 2018)	1, 3, 4	Consumer review analysis; Interview	VA	People with disabilities
(PRIDMORE *et al.*, 2019)	1, 2	Focus groups	VA	Americans and Dutch
(RONG *et al.*, 2017)	3	Experiment; Interfaces	Sim	General
(TRAJKOVA; MARTIN-HAMMOND, 2020)	1, 4	Focus groups	VA	Elders
(VTYURINA; FOURNEY, 2018)	3	Wizard-of-Oz	Sim	General
(WEBER; LUDWIG, 2020)	1, 2, 4	Interview	VA	General
(WU *et al.*, 2020)	3	Experiment	VA	Native/ non-native English speakers
(XU; WARSCHAUER, 2020b)	1, 2, 5	Experiment; Interview; Drawing task	VA	Children
(XU; WARSCHAUER, 2020a)	3	Experiment	Sim	Children
(YANG; AURISICCHIO; BAXTER, 2019)	1	Survey	VA	General
(YAROSH *et al.*, 2018)	2, 3, 4	Field experiment; Interview	Sim	Children and Parents

Appendix 2 – Exploratory interviews' free and Inform consent term (English)

FREE AND INFORMED CONSENT TERM

Research's title: Leveraging users' mental models of Voice Assistants (VAs) through system explicitness on VA responses
Leading researcher: Isabela Canellas da Motta
Advisor: Prof. Manuela Quaresma, D.Sc.
Design Graduate Program
Pontifical Catholic University of Rio de Janeiro/ Department of Arts and Design

We would like to invite you to participate as a volunteer in an **interview**.

Justification

Our motive in conducting this research is to generate new knowledge regarding important design parameters for the development of voice assistants. Due to its focus on user-centered design, the result of this work may go beyond the academic limits, becoming an effective contribution for interaction designers through recommendations for the design of voice assistants.

Objective

In this research, we intend to discuss the issue of users' mental models of voice assistants. We plan to understand the main design features that elicit misunderstandings for users about the system and raise possible solutions to address this problem. We understand that the knowledge and experience of voice assistant experts from different fields - researchers and developers - is fundamental to understand these concepts and their impact on design decisions. Once the research is completed, the researcher also intends to publish the research in academic journals and academic conference proceedings.

Procedure

If you agree to participate, we will do the following activities with you. First, you will need to fill out a form with your basic information. Then, you will be invited by email to a virtual meeting with the researcher in charge. This conversation will be led by the researcher, who will ask questions about the participants' perceptions of problems in users' mental model, their causes, and possible solutions. Participants will also be encouraged to share their experiences and learnings from conducting projects and/or research with voice assistants and interfaces. The whole procedure will be conducted remotely by the responsible researcher. The procedure should take about 90 minutes.

Risks

This research has some risks: possible discomfort or embarrassment in sharing your opinions. However, to decrease the chance of these risks happening, we assure you that the procedure aims solely to understand the opinions of researchers/voice interface developers, so you will not be tested or judged. In addition, all procedures will be performed remotely, and therefore there is no risk of COVID-19 contamination.

Benefits

You will not benefit directly by participating in this study. However, your participation is vital to understanding the perceptions of voice assistant developers/ researchers.

Costs and compensation

In participating in this study, you will neither have any cost nor will you receive any financial advantage.

Data collection, confidentiality and secrecy

All answers given by the participants will be recorded for future analysis. The researcher will not disclose your name and the data will be confidential, restricted only to the researchers. The research's results will be available to you when it is completed. Your name or material indicating your participation will not be released without your permission. You will not be identified in any resulting publication.

Authorization for the use of image and statements

The body of collected data (images and audio) will not be disclosed. Do you allow the usage of your statements for academic purposes - papers, articles, classes, websites, presentations in conferences?

( ) Yes, I authorize the disclosure of my statements.

( ) No, I do not authorize the disclosure of my statements.

Participants' rights

You will have all the information you want about this research, and you are free to participate or refuse to participate. Even if you want to participate now, you can back out or stop participating at any time. Your participation is voluntary, and your refusal to participate will not result in any penalty. This consent form is printed in two original copies, one of which will be kept on file by the researcher, and the other will be provided to you. This document will be sent to you by email, and, to express your agreement, just answer the email agreeing to the terms of this document.

The data collected in the research will be kept on file with the responsible researcher for five (5) years. After this period, the researcher will evaluate the documents for their final destination, according to the legislation in force. The researchers will treat your identity with professional standards of confidentiality, using the information only for academic and scientific purposes. This term respects the Resolution 510/16 CS and was evaluated by PUC-Rio's board of ethics. If you have any questions about this research, you can also contact us by phone (55 021) 3527-1005 or by email isabela.canellas@gmail.com or mquaresma@puc-rio.br.

I declare that I agree to participate in the research and that I have been given the opportunity to read and clarify my doubts.

Your name (in full):

_____

Rio de Janeiro _____, 20__.

Signature:

_____

Responsible researcher's signature:

_____

Isabela Motta

Advisor's signature:

_____

Manuela Quaresma

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

Department of Arts and Design
PUC-Rio's Design Graduate Program as partial requirement for obtaining the Master in Design degree.

PUC-Rio's Board of Ethics in Research
Rua Marquês de São Vicente, 225 – Edifício Kennedy, 2° floor, Gávea, Rio de Janeiro, RJ. CEP: 22453-900. Telephone: (21) 3527-1618.
The Board has the attribution of analysis from the ethical point of view the research projects of the University's professors, researchers, and students, when requested.

Appendix 3 – Exploratory interviews' free and Inform consent term (Portuguese)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título da Pesquisa: Melhorando o modelo mental de usuários de assistentes de voz por meio da transparência em respostas do sistema
Pesquisador responsável: Isabela Canellas da Motta
Professora Orientadora: Prof. Dra. Manuela Quaresma
Programa de Pós-Graduação em Design
Pontifícia Universidade Católica do Rio de Janeiro/ Departamento de Artes e Design

Gostaríamos de convidar você a participar como voluntário (a) de uma **entrevista**.

Justificativa

O motivo que nos leva a realizar esta pesquisa é gerar novos conhecimentos no que diz respeito a importantes parâmetros de design para o desenvolvimento de assistentes de voz. O resultado deste trabalho, pelo seu foco no design centrado no usuário, poderá ultrapassar os limites acadêmicos, tornando-se uma efetiva contribuição para designers de interação, por meio de recomendações para o design de assistentes de voz.

Objetivo

Nesta pesquisa, pretendemos discutir a questão dos modelos mentais que usuários constroem sobre assistentes de voz. Planejamos entender quais são as principais características de design que geram desentendimentos para usuários a cerca do sistema, além de levantar possíveis soluções para abordar esse problema. Entendemos que o conhecimento e a experiência de especialistas em assistentes de voz de diversas áreas - pesquisadores e desenvolvedores - é fundamental para entender esses conceitos e como decisões projetuais são tomadas e impactadas por eles. Terminada a investigação, a pesquisadora pretende também publicar a pesquisa em revistas acadêmicas e em anais de congressos acadêmicos.

Procedimentos

Caso você concorde em participar, vamos fazer as seguintes atividades com você. Primeiro, você precisará preencher um formulário com suas informações básicas. Então, você será convidado, por email, para uma reunião virtual com a pesquisadora responsável. Essa conversa será conduzida pela pesquisadora, que irá fazer perguntas a cerca da percepção dos participantes sobre problemas no modelo mental de usuários, suas causas, e possíveis soluções. Os participantes também serão encorajados a compartilhar suas experiências e aprendizados com a condução de projetos e/ou pesquisas com assistentes e interfaces de voz. Todo o procedimento será conduzido remotamente pela pesquisadora responsável. O procedimento deverá ter em torno de 60-75 minutos.

Riscos

Esta pesquisa tem alguns riscos: possíveis desconfortos ou constrangimentos em compartilhar suas opiniões. Mas, para diminuir a chance desses riscos acontecerem, asseguramos que procedimento do qual você participará visa somente entender as opiniões de pesquisadores/ desenvolvedores de interfaces de voz, assim, você não será testado(a) ou julgado(a). Todos os procedimentos serão realizados remotamente, não havendo riscos de contaminação da COVID-19.

Benefícios

Você não irá se beneficiar de forma direta ao participar deste estudo. No entanto, sua participação é vital para a compreensão das percepções de desenvolvedores de assistentes de voz.

Custos e compensação

Para participar deste estudo, você não vai ter nenhum custo e nem receberá qualquer compensação financeira.

Informações coletadas, confidencialidade e sigilo

Todas as respostas dadas pelos participantes serão registradas para análise futura. O pesquisador não vai divulgar seu nome e os dados serão confidenciais, restritos apenas às pesquisadoras. Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. Você não será identificado (a) em nenhuma publicação que possa resultar.

Autorização para uso de imagem e declarações

O material que constitui o corpo de dados coletados (imagens e áudio) não será divulgado. Você autoriza o uso de suas declarações para finalidades acadêmicas - artigos acadêmicos, aulas, papers, sites, apresentações em simpósios ou congressos científicos relacionados ao tema?

( ) Autorizo a divulgação das minhas declarações.

( ) Não autorizo a divulgação das minhas declarações.

Direitos dos participantes

Você terá todas as informações que quiser sobre esta pesquisa e estará livre para participar ou recusar-se a participar. Mesmo que você queira participar agora, você pode voltar atrás ou parar de participar a qualquer momento. A sua participação é voluntária e o fato de não querer participar não vai trazer qualquer penalidade ou mudança na forma em que você é atendido (a). Este termo de consentimento encontra-se em duas vias originais, sendo que uma será arquivada pelo pesquisador responsável e a outra será fornecida a você. Esse documento será enviado a você por email, e, para expressar seu aceite, basta responder o email concordando com os termos do documento.

Os dados coletados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos. Decorrido este tempo, o pesquisador avaliará os documentos para a sua destinação final, de acordo com a legislação vigente. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, utilizando as informações somente para os fins acadêmicos e científicos. Este termo respeita a Resolução 510/16 CS e foi avaliado pela Câmara de Ética em Pesquisa da PUC-Rio. Se você tiver alguma dúvida sobre esta pesquisa, você também pode entrar em contato com a pesquisadora responsável pelo telefone (21) 2266-2178 ou pelo email isabela.canellas@gmail.com ou com a professora orientadora (tel: (21)3527-1005 e email: mquaresma@puc-rio.br.).

Declaro que concordo em participar da pesquisa e que me foi dada à oportunidade de ler e esclarecer as minhas dúvidas.

Seu nome (por extenso): _____

Assinatura: _____

Rio de Janeiro ____de _____ de 20__.

Assinatura da pesquisadora responsável:

_____

Isabela Motta

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

Departamento de Artes e Design
Programa de Pós-graduação em design da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Design

Câmara de Ética em Pesquisa da PUC-Rio
Rua Marquês de São Vicente, 225 – Edifício Kennedy, 2º andar, Gávea, Rio de Janeiro, RJ. CEP: 22453-900. Telefone: (21) 3527-1618.
A Câmara tem por atribuição analisar do ponto de vista ético os projetos de pesquisa dos professores, pesquisadores e discentes da Universidade, quando solicitada.

Appendix 4 – Delphi's free and Inform consent term (English)

FREE AND INFORMED CONSENT TERM

Research's title: Leveraging users' mental models of Voice Assistants (VAs) through system explicitness on VA responses
Leading researcher: Isabela Canellas da Motta
Advisor: Prof. Manuela Quaresma, D.Sc.
Design Graduate Program
Pontifical Catholic University of Rio de Janeiro/ Department of Arts and Design

We would like to invite you to participate as a volunteer in a **Delphi questionnaire**.

Justification

Our motive in conducting this research is to generate new knowledge regarding important design parameters for the development of voice assistants. Due to its focus on user-centered design, the result of this work may go beyond the academic limits, becoming an effective contribution for interaction designers through recommendations for the design of voice assistants.

Objective

In this research, we intend to gather opinions from experts in voice interfaces from several areas - researchers and developers - about the main challenges in developing responses for voice assistants. Thus, we plan to gather experts' opinions about the main issues in such responses and possible solutions to these issues. Through the Delphi questionnaire, we seek to develop a list of recommendations that reflect experts' consensus on good design practices for voice assistants. Upon completing the research, the researcher also intends to publish the findings in academic journals and conference proceedings.

Procedure

If you agree to participate, we will do the following activities with you. First, you will fill out a form with your basic information and possible nominations of other participants for this study. You will then be part of a group of experts who will participate in the study anonymously. In the first stage, you will receive a questionnaire with two questions about voice assistants by email. In the second stage, you will receive another questionnaire, in which we will ask you to evaluate design recommendations for voice assistants. Finally, in a third phase, participants will receive the questionnaire with the results of the group evaluation and will have the opportunity to revise or keep their previous evaluation, all anonymously. Each questionnaire will take two weeks to complete and will be accompanied by detailed instructions on how to complete it.

Risks

This research has some risks: possible discomfort or embarrassment in sharing your opinions in the Delphi questionnaire. However, to mitigate these risks, your answers will be completely anonymous, and no other participant will be able to recognize your identity in the generated content, neither during nor after the study. Furthermore, the procedure you participate in is only intended to understand the opinions of voice interface researchers/developers, so you will not be tested or judged.

Benefits

You will not benefit directly by participating in this study. However, you will have first-hand access to the study's results, in the form of design recommendations that reflect experts' consensus

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

on how to design voice assistants. We hope that such knowledge may be valuable for your future projects or research.

Costs and compensation

In participating in this study, you will neither have any cost nor will you receive any financial advantage.

Data collection, confidentiality and secrecy

All answers given by the participants will be recorded for future analysis. The researcher will not disclose your name and the data will be confidential, restricted only to the researchers. The research's results will be available to you when it is completed. Your name or material indicating your participation will not be released without your permission. You will not be identified in any resulting publication.

Authorization for the use of image and statements

The body of collected data will not include audio or video, just your statements in the questionnaire. Do you allow the usage of your statements for academic purposes - papers, articles, classes, websites, presentations in conferences?

( ) Yes, I authorize the disclosure of my statements.

( ) No, I do not authorize the disclosure of my statements.

Participants' rights

You will have all the information you want about this research, and you are free to participate or refuse to participate. Even if you want to participate now, you can back out or stop participating at any time. Your participation is voluntary, and your refusal to participate will not result in any penalty. This consent form is printed in two original copies, one of which will be kept on file by the researcher, and the other will be provided to you. This document will be sent to you by email, and, to express your agreement, just answer the email agreeing to the terms of this document.

The data collected in the research will be kept on file with the responsible researcher for five (5) years. After this period, the researcher will evaluate the documents for their final destination, according to the legislation in force. The researchers will treat your identity with professional standards of confidentiality, using the information only for academic and scientific purposes. This term respects the Resolution 510/16 CS and was evaluated by PUC-Rio's board of ethics. If you have any questions about this research, you can also contact us by phone (55 021) 3527-1005 or by email isabela.canellas@gmail.com or mquaresma@puc-rio.br.

I declare that I agree to participate in the research and that I have been given the opportunity to read and clarify my doubts.

Your name (in full):

_____

Responsible researcher's signature:

_____

Isabela Motta

Rio de Janeiro _____, 20__.

Signature:

_____

Department of Arts and Design
PUC-Rio's Design Graduate Program as partial requirement for obtaining the Master in Design degree.

PUC-Rio's Board of Ethics in Research
Rua Marquês de São Vicente, 225 – Edifício Kennedy, 2° floor, Gávea, Rio de Janeiro, RJ. CEP: 22453-900. Telephone: (21) 3527-1618.
The Board has the attribution of analysis from the ethical point of view the research projects of the University's professors, researchers, and students, when requested.

Appendix 5 – Delphi's free and Inform consent term (Portuguese)

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título da Pesquisa: Melhorando o modelo mental de usuários de assistentes de voz por meio da transparência em respostas do sistema
Pesquisador responsável: Isabela Canellas da Motta
Professora Orientadora: Prof. Dra. Manuela Quaresma
Programa de Pós-Graduação em Design
Pontifícia Universidade Católica do Rio de Janeiro/ Departamento de Artes e Design

Gostaríamos de convidar você a participar como voluntário (a) de um **questionário Delphi**.

Justificativa

O motivo que nos leva a realizar esta pesquisa é gerar novos conhecimentos no que diz respeito a importantes parâmetros de design para o desenvolvimento de assistentes de voz. O resultado deste trabalho, pelo seu foco no design centrado no usuário, poderá ultrapassar os limites acadêmicos, tornando-se uma efetiva contribuição para designers de interação, por meio de recomendações para o design de assistentes de voz.

Objetivo

Nesta pesquisa, pretendemos reunir opiniões de especialistas em interfaces de voz de diversas áreas - pesquisadores e desenvolvedores - sobre os principais desafios para o desenvolvimento de respostas de assistentes de voz. Dessa forma, planejamos levantar a opinião de *experts* sobre os principais problemas em tais respostas e possíveis soluções para esses problemas. Buscamos, por meio do questionário Delphi, desenvolver uma lista de recomendações que reflita o consenso de especialistas sobre boas práticas de design para assistentes de voz. Terminada a investigação, a pesquisadora pretende também publicar a pesquisa em revistas acadêmicas e em anais de congressos acadêmicos.

Procedimentos

Caso você concorde em participar, vamos fazer as seguintes atividades com você. Primeiro, você precisará preencher um formulário com suas informações básicas e possíveis indicações de outros participantes para esta pesquisa. Assim, você fará parte de um grupo de especialistas que irá participar do questionário de forma anônima. Na primeira etapa, você receberá, por email, um questionário com duas perguntas sobre assistentes de voz. Em uma segunda fase, você receberá outro questionário, em que pediremos para você avaliar recomendações de design para assistentes de voz. Finalmente, em uma terceira fase, os participantes irão receber novamente o questionário com os resultados da avaliação do grupo, e terão a oportunidade de revisar ou manter sua avaliação anterior, sempre de forma anônima. Cada questionário terá prazo de duas semanas para ser respondido e será acompanhado de instruções detalhadas sobre como preenchê-lo.

Riscos

Esta pesquisa tem alguns riscos: possíveis desconfortos ou constrangimentos em compartilhar suas opiniões no questionário Delphi. Mas, para diminuir a chance desses riscos acontecerem, suas respostas serão totalmente anônimas, e nenhum dos outros participantes será capaz de identificar sua identidade no conteúdo gerado, tanto durante, quanto após a realização do estudo. Além disso, o procedimento do qual você participará visa somente entender as opiniões de pesquisadores/ desenvolvedores de interfaces de voz, assim, você não será testado(a) ou julgado(a).

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

Benefícios

Você não irá se beneficiar de forma direta ao participar deste estudo. No entanto, você terá acesso aos resultados do estudo em primeira mão, na forma de recomendações de design que refletem um consenso de especialistas sobre como projetar assistentes de voz. Consideramos que tal conhecimento pode ser valioso para seus futuros projetos ou pesquisas.

Custos e compensação

Para participar deste estudo, você não vai ter nenhum custo e nem receberá qualquer vantagem financeira.

Informações coletadas, confidencialidade e sigilo

Todas as respostas dadas pelos participantes serão registradas para análise futura. O pesquisador não vai divulgar seu nome e os dados serão confidenciais, restritos apenas às pesquisadoras. Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. Você não será identificado (a) em nenhuma publicação que possa resultar.

Autorização para uso de imagem e declarações

O material que constitui o corpo de dados coletados não inclui imagens ou áudio, apenas suas declarações escritas no questionário. Você autoriza o uso de suas declarações para finalidades acadêmicas - artigos acadêmicos, aulas, papers, sites, apresentações em simpósios ou congressos científicos relacionados ao tema?

( ) Autorizo a divulgação das minhas declarações.
( ) Não autorizo a divulgação das minhas declarações.

Direitos dos participantes

Você terá todas as informações que quiser sobre esta pesquisa e estará livre para participar ou recusar-se a participar. Mesmo que você queira participar agora, você pode voltar atrás ou parar de participar a qualquer momento. A sua participação é voluntária e o fato de não querer participar não vai trazer qualquer penalidade ou mudança na forma em que você é atendido (a). Este termo de consentimento encontra-se em duas vias originais, sendo que uma será arquivada pelo pesquisador responsável e a outra será fornecida a você. Esse documento será enviado a você por email, e, para expressar seu aceite, basta responder o email concordando com os termos do documento.

Os dados coletados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos. Decorrido este tempo, o pesquisador avaliará os documentos para a sua destinação final, de acordo com a legislação vigente. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, utilizando as informações somente para os fins acadêmicos e científicos. Este termo respeita a Resolução 510/16 CS e foi avaliado pela Câmara de Ética em Pesquisa da PUC-Rio. Se você tiver alguma dúvida sobre esta pesquisa, você também pode entrar em contato com a pesquisadora responsável pelo telefone (21) 2266-2178 ou pelo email isabela.canellas@gmail.com ou com a professora orientadora (tel: (21)3527-1005 e email: mquaresma@puc-rio.br.).

Artes & Design PUC-Rio

Pontifícia Universidade Católica do Rio de Janeiro

Declaro que concordo em participar da pesquisa e que me foi dada à oportunidade de ler e esclarecer as minhas dúvidas.

Seu nome (por extenso): Assinatura:

_____ _____

Rio de Janeiro ____de _____ de 20__.

Assinatura da pesquisadora responsável:

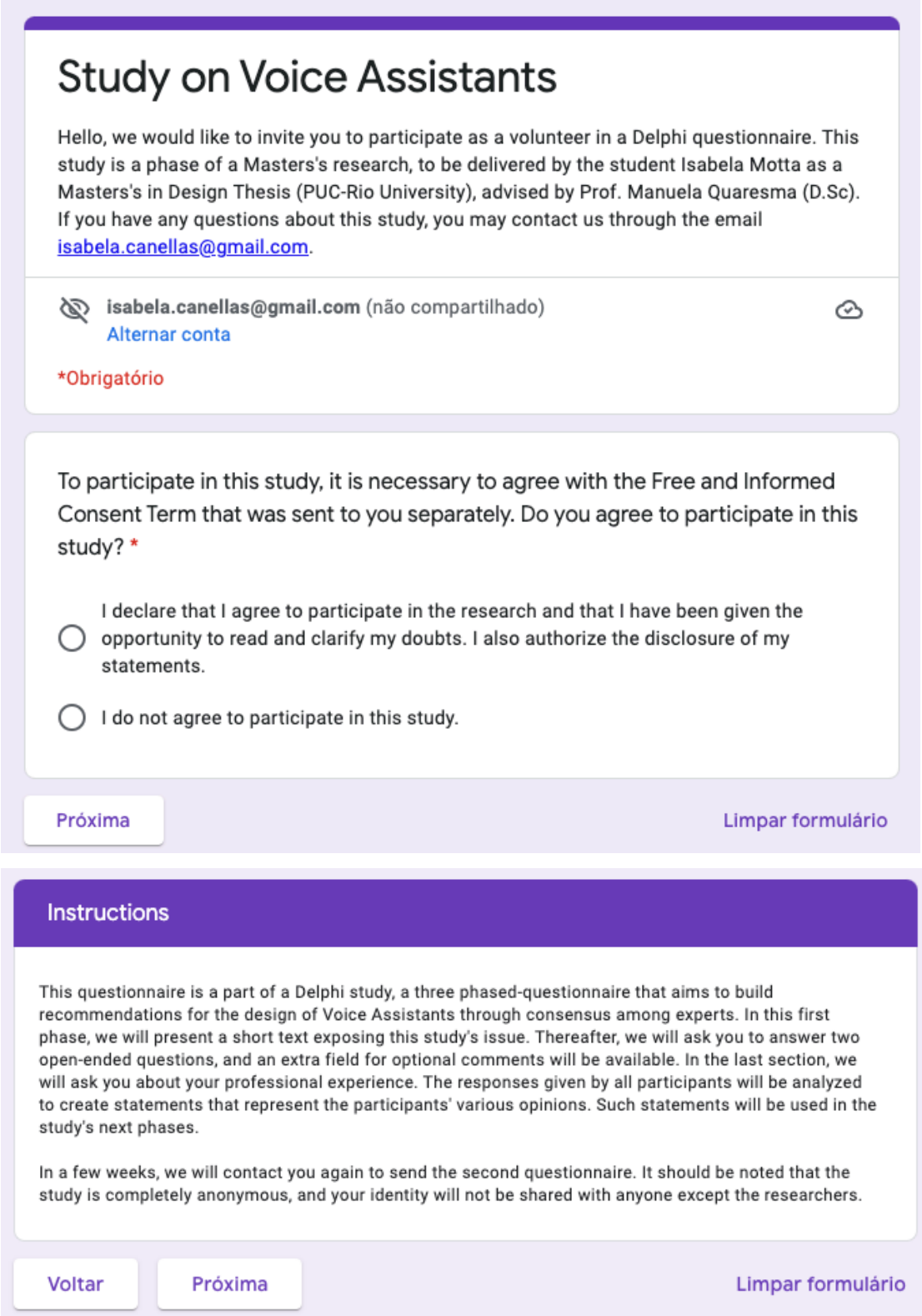
_____

Isabela Motta

Departamento de Artes e Design
Programa de Pós-graduação em design da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Design

Câmara de Ética em Pesquisa da PUC-Rio
Rua Marquês de São Vicente, 225 – Edifício Kennedy, 2º andar, Gávea, Rio de Janeiro, RJ. CEP: 22453-900. Telefone: (21) 3527-1618.
A Câmara tem por atribuição analisar do ponto de vista ético os projetos de pesquisa dos professores, pesquisadores e discentes da Universidade, quando solicitada.

Appendix 6 – Delphi's first questionnaire (English)



Study on Voice Assistants

Hello, we would like to invite you to participate as a volunteer in a Delphi questionnaire. This study is a phase of a Masters's research, to be delivered by the student Isabela Motta as a Masters's in Design Thesis (PUC-Rio University), advised by Prof. Manuela Quaresma (D.Sc). If you have any questions about this study, you may contact us through the email isabela.canellas@gmail.com.

isabela.canellas@gmail.com (não compartilhado)
Alternar conta

*Obrigatório

To participate in this study, it is necessary to agree with the Free and Informed Consent Term that was sent to you separately. Do you agree to participate in this study? *

- I declare that I agree to participate in the research and that I have been given the opportunity to read and clarify my doubts. I also authorize the disclosure of my statements.
- I do not agree to participate in this study.

Próxima Limpar formulário

Instructions

This questionnaire is a part of a Delphi study, a three phased-questionnaire that aims to build recommendations for the design of Voice Assistants through consensus among experts. In this first phase, we will present a short text exposing this study's issue. Thereafter, we will ask you to answer two open-ended questions, and an extra field for optional comments will be available. In the last section, we will ask you about your professional experience. The responses given by all participants will be analyzed to create statements that represent the participants' various opinions. Such statements will be used in the study's next phases.

In a few weeks, we will contact you again to send the second questionnaire. It should be noted that the study is completely anonymous, and your identity will not be shared with anyone except the researchers.

Voltar Próxima Limpar formulário

Users' mental models of Voice Assistants

Please, carefully read the text below:

Users' mental models of a system are paramount for the usage and interaction with such products. In this study, we consider mental models as "a type of conceptual model created by a user to represent how a product or system works, including a series of expectations about its components, functioning, and appropriate usage." Such perceptions are important since they represent the user's understanding of a product, impacting how they perform tasks and dictating performance levels. In users' interactions with Voice Assistants (Siri, Alexa, etc), mental models are also vital.

Nevertheless, users' mental models might not be aligned with Voice Assistants' actual capabilities. Evidence shows that users are not aware of relevant information for data privacy, do not correctly understand the Assistants' general functioning and actions, and have trouble understanding error sources and recovering from failures. Likewise, users have unrealistic expectations for these systems' intelligence and technical, social, and conversational capabilities. As a consequence, users face hardships throughout interactions and get frustrated, which leads to the underutilization or complete abandonment of the Voice Assistant.

Considering the issue above, please answer the questions below. For each question, please try to provide complete answers. Feel free to add examples and real situations from your work experience with conversational agents to your answers.

In your opinion, what are the causes that lead users to form mental models that are unaligned with Voice Assistants' real capabilities? Please, state at least three causes and, for each one, explain why it is relevant. *

Sua resposta

In your opinion, what solutions could solve the issue of users' incorrect mental models of Voice Assistants? Please, present at least three solutions and, for each one, explain why it is appropriate. *

Sua resposta

You may leave extra comments about the research topic or feedback on this study, if you wish.

Sua resposta

Voltar Próxima Limpar formulário

Participants' profile data

What is your name? *

Sua resposta

What is your email address? This is important so that we may contact you for the study's next phases. *

Sua resposta

In which country do you currently work? *

Sua resposta

Do you currently work or have worked with research or development of voice interfaces (IVRs, GPSs etc) or Voice Assistants (Siri, Alexa, etc)? Please, cite your job position. *

Sua resposta

In which type of institution do you work/ have you worked with voice interfaces/ assistants? *

- ◯ Enterprise
- ◯ University
- ◯ Outro: _____

How long has it been since you last worked with voice interfaces/ assistants? *

- ◯ I am currently involved with voice interfaces-related projects.
- ◯ Less than a year ago.
- ◯ Three years ago.
- ◯ Five years ago.
- ◯ Over five years ago.
- ◯ I never worked with such projects.

For how long have you worked with voice interfaces/ assistants? *

- ◯ For less than a year.
- ◯ Between 1 and 2 years.
- ◯ Between 2 and 3 years.
- ◯ Between 3 and 5 years.
- ◯ For more than 5 years.
- ◯ I have never worked with such interfaces.

What is your highest COMPLETE level of education? *

- ○ High School
- ○ Graduation
- ○ Masters
- ○ PhD
- ○ Postdoctorate

What is your graduate field? (e.g., Computer Science, Engineering, Design etc). *

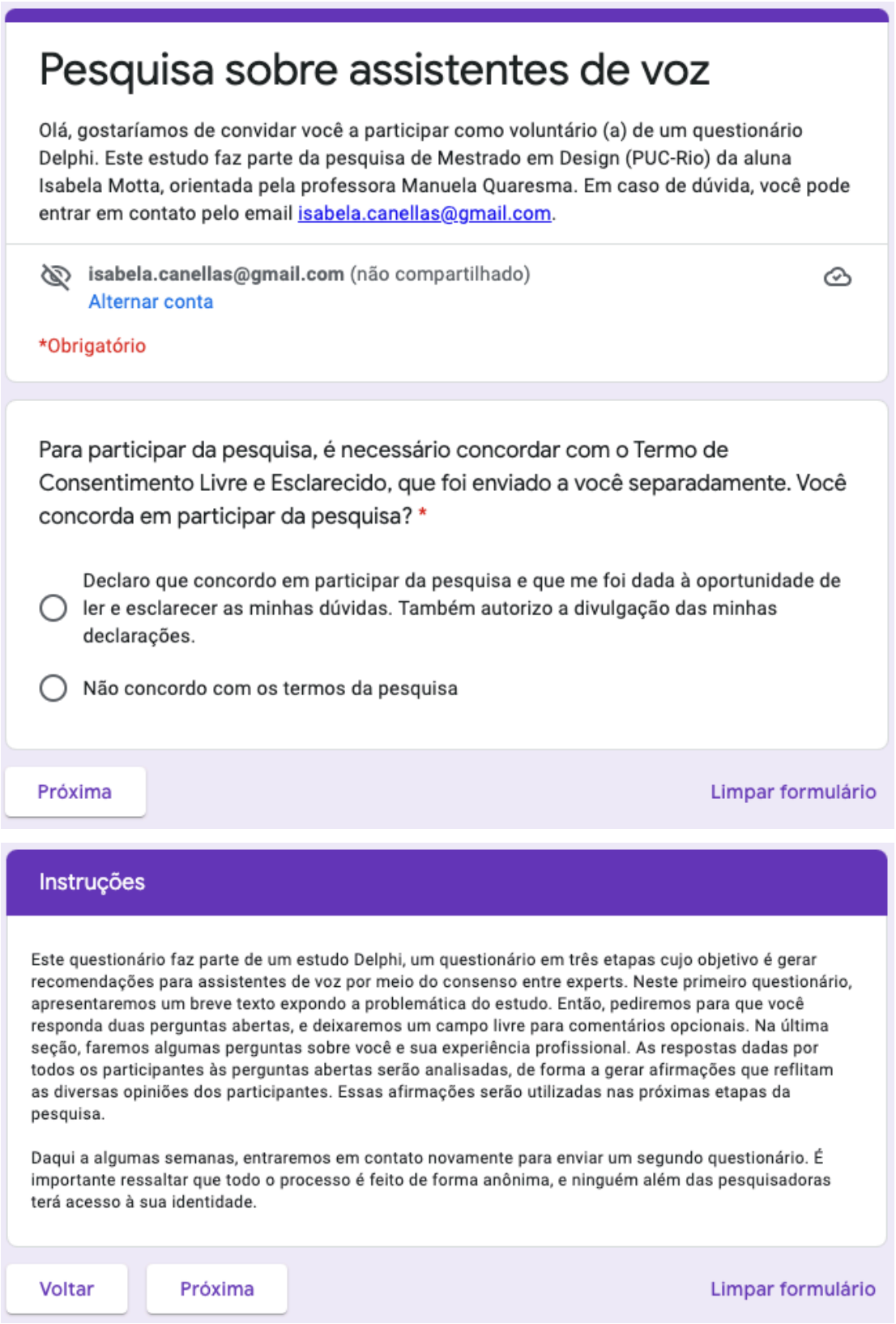
Sua resposta

It is very important for this study to reach a large number of participants. Thus, if you are aware of any other professional involved in voice interfaces' research or development, please, write their name below. You may also share our email address so that they may request this form's link (isabela.canellas@gmail.com).

Sua resposta

Voltar Enviar Limpar formulário

Appendix 7 – Delphi's first questionnaire (Portuguese)



Pesquisa sobre assistentes de voz

Olá, gostaríamos de convidar você a participar como voluntário (a) de um questionário Delphi. Este estudo faz parte da pesquisa de Mestrado em Design (PUC-Rio) da aluna Isabela Motta, orientada pela professora Manuela Quaresma. Em caso de dúvida, você pode entrar em contato pelo email isabela.canellas@gmail.com.

isabela.canellas@gmail.com (não compartilhado)
Alternar conta

*Obrigatório

Para participar da pesquisa, é necessário concordar com o Termo de Consentimento Livre e Esclarecido, que foi enviado a você separadamente. Você concorda em participar da pesquisa? *

- Declaro que concordo em participar da pesquisa e que me foi dada à oportunidade de ler e esclarecer as minhas dúvidas. Também autorizo a divulgação das minhas declarações.
- Não concordo com os termos da pesquisa

Próxima Limpar formulário

Instruções

Este questionário faz parte de um estudo Delphi, um questionário em três etapas cujo objetivo é gerar recomendações para assistentes de voz por meio do consenso entre experts. Neste primeiro questionário, apresentaremos um breve texto expondo a problemática do estudo. Então, pediremos para que você responda duas perguntas abertas, e deixaremos um campo livre para comentários opcionais. Na última seção, faremos algumas perguntas sobre você e sua experiência profissional. As respostas dadas por todos os participantes às perguntas abertas serão analisadas, de forma a gerar afirmações que reflitam as diversas opiniões dos participantes. Essas afirmações serão utilizadas nas próximas etapas da pesquisa.

Daqui a algumas semanas, entraremos em contato novamente para enviar um segundo questionário. É importante ressaltar que todo o processo é feito de forma anônima, e ninguém além das pesquisadoras terá acesso à sua identidade.

Voltar Próxima Limpar formulário

Modelos mentais de usuários de assistentes de voz

Por favor, leia com atenção:

Os modelos mentais que os usuários formam sobre um sistema são fundamentais para o uso e a interação com esses produtos. Neste estudo, consideramos modelos mentais como "um tipo de modelo conceitual criado pelo usuário para representar como um produto ou sistema funciona, incluindo uma série de expectativas sobre os componentes do sistema, seu funcionamento e uso apropriado". Tais percepções são importantes porque refletem o entendimento do usuário sobre o produto, influenciando como as pessoas realizam tarefas e ditando níveis de desempenho. Na interação de usuários com assistentes de voz (Siri, Alexa, etc), os modelos mentais também são vitais.

No entanto, é possível que os modelos mentais dos usuários não estejam alinhados com as reais capacidades dos assistentes de voz. Evidências mostram que os usuários desconhecem informações relevantes para a privacidade de seus dados, não entendem corretamente o funcionamento geral e as ações do sistema e têm dificuldades em entender fontes de erros e se recuperar de tais falhas. Além disso, os usuários têm expectativas equivocadas para a inteligência e as capacidades técnicas, sociais e conversacionais desses assistentes. Como consequência, os usuários enfrentam dificuldades na interação e ficam frustrados, o que leva à subutilização ou total abandono do assistente de voz.

Considerando a questão exposta acima, por favor responda às duas perguntas abaixo. Para cada uma delas, tente oferecer respostas mais completas o possível. Sinta-se livre para complementar suas respostas com exemplos e situações reais da sua experiência de trabalho com interfaces conversacionais, quando pertinente.

Na sua opinião, quais são as causas que levam os usuários de assistentes de voz a formarem modelos mentais desalinhados com as reais capacidades desses sistemas? Por favor, apresente pelo menos três causas e, para cada causa apresentada, explique porquê as julga relevantes. *

Sua resposta

Na sua opinião, quais soluções poderiam resolver a questão dos modelos mentais incorretos sobre os assistentes de voz? Por favor, apresente pelo menos três soluções, e, para cada uma delas, explique porquê as julga adequadas. *

Sua resposta

Você pode deixar comentários extras sobre o tema ou fornecer algum feedback sobre esta pesquisa no espaço abaixo se desejar.

Sua resposta

Voltar Próxima Limpar formulário

Dados do participante

Qual é o seu nome? *

Sua resposta

Qual é o seu email? É importante deixar uma forma de contato para que possamos dar continuidade a sua participação na pesquisa posteriormente. *

Sua resposta

Você trabalha ou já trabalhou com pesquisa ou desenvolvimento de interfaces de voz (sistemas de resposta interativa/ telefônicos, GPS, etc) ou assistentes de voz (Siri, Alexa, Google Assistant)? Por favor, cite a função desempenhada. *

Sua resposta

Em que tipo de instituição você trabalha/ trabalhou com essas interfaces/ assistentes de voz? *

- ◯ Empresa
- ◯ Universidade
- ◯ Outro: _____

Há quanto tempo você esteve envolvido em tais projetos? *

- ◯ Estou envolvido em tais projetos atualmente
- ◯ Há menos de 1 ano
- ◯ Há 3 anos
- ◯ Há 5 anos
- ◯ Há mais de 5 anos
- ◯ Nunca trabalhei com pesquisa ou desenvolvimento de interação por voz

Por quanto tempo você esteve envolvido em tais projetos? *

- ◯ Por menos de 1 ano
- ◯ Entre 1 e 2 anos
- ◯ Entre 2 e 3 anos
- ◯ Entre 3 a 5 anos
- ◯ Mais de 5 anos
- ◯ Nunca trabalhei com pesquisa ou desenvolvimento de interação por voz

Qual é seu maior nível de escolaridade COMPLETO? *

- ◯ Ensino Médio
- ◯ Graduação
- ◯ Mestrado
- ◯ Doutorado
- ◯ Pós-Doutorado

Qual é a sua área de formação? *

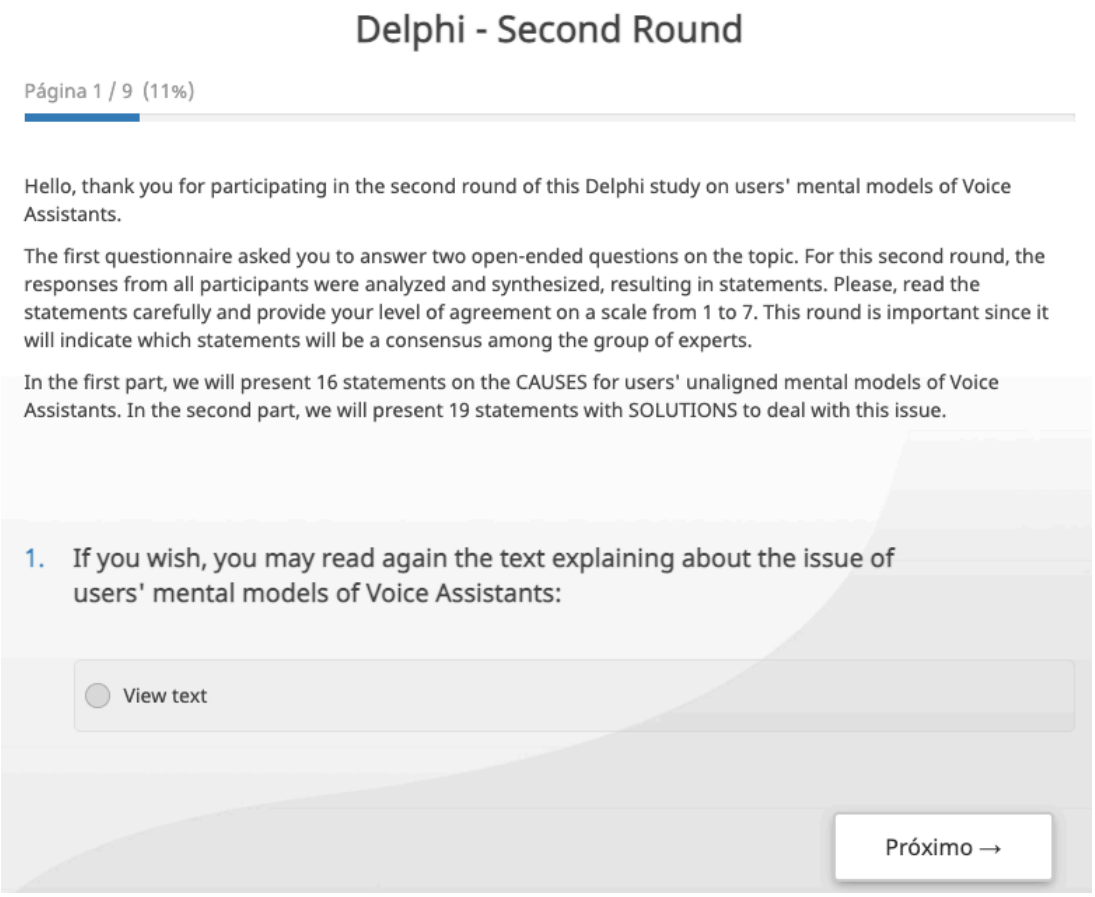
Sua resposta

É de enorme importância para este estudo que haja um grande número de participantes. Assim, se você conhecer outro desenvolvedor(a)/ pesquisador(a) de interfaces de voz ou assistentes de voz que possa ter interesse em participar, por favor, escreva seu nome abaixo. Se desejar, você também pode compartilhar o e-mail isabela.canellas@gmail.com com o interessado, e enviaremos este formulário a ele(a).

Sua resposta

Voltar Enviar Limpar formulário

Appendix 8 – Delphi's second questionnaire (English)



2. Please answer to which level you agree with the statements below on a 1 to 7 scale *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree						Totally agree
	1	2	3	4	5	6	7
Bad previous experiences with other voice interfaces create negative expectations for the Voice Assistants on users.							
Characteristics that induce anthropomorphism (e.g., voice, name, gender, metaphors, humor) cause users to expect Voice Assistants would be as capable as a human.							
Users do not know the Voice Assistants' technical limitations and require the Assistant to perform tasks and recognize commands beyond its capabilities.							
Privacy concerns hinder users from interacting with the Assistants for long enough to construct correct mental models.							
Users construct their mental models of conversations through human interactions, but Voice Assistants have lower conversational skills, letting down expectations and causing frustration.							

3. You may add comments about the statements above if you wish.

← Anterior

Próximo →

4. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree						Totally agree
	1	2	3	4	5	6	7
Voice Assistants do not tell users when they update, nor explain the updates in their skills.							
Voice Assistants do not present initial instructions to users, leaving them without knowing what to expect from the product.							
Voice Assistants do not explain their limitations for certain actions, such as recognizing the conversational context.							
Voice Assistants are too complex and demand too much of the users' cognition.							
Voice Assistants do not explain to users about their skills, how they should be utilized, how they process commands, how they make decisions, or how to recover from failures.							

5. You may add comments about the statements above if you wish.

← Anterior

Próximo →

6. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree						Totally agree
	1	2	3	4	5	6	7
Developers conduct little research about user experience, which is paramount to creating conversational flows.							
Brands present little institutional content about the Assistants and Artificial Intelligence, leading users to buy the product without being aware of its capabilities.							
Marketing raises users' expectations by exaggerating the Voice Assistant's social skills and intelligence, showing use cases that are too simple and fluid.							
In Science Fiction, systems powered with Artificial Intelligence are pictured as futuristic, intelligent, sensitive, talkative, and capable, creating unaligned perceptions about current Assistants.							
Voice Assistants from different brands have different skills, leading to the belief that an Assistant might have the same skills as the others.							
Users do not look for information about the Voice Assistant before buying it, especially for tasks that are deemed unnecessary.							

7. You may add comments about the statements above if you wish.

← Anterior

Próximo →

PART 2 (1/4)

Below, we present statements that summarize the groups' opinions on the SOLUTIONS to align users' mental models with the Voice Assistants' actual capabilities.

Please state your level of agreement with each statement below on a 1 to 7 scale.

You may also leave optional comments on the statements.

8. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree						Totally agree
	1	2	3	4	5	6	7
Voice Assistants should make it clear for users that they are talking to a machine and not a human.							
Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).							
Voice Assistants should provide examples and explanations about their skills' scope and action execution, decision making, and learning processes.							
Voice Assistants should present usage tips throughout interactions, including mechanisms to clarify the conversation context.							
The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.							

9. You may add comments about the statements above if you wish.

← Anterior

Próximo →

10. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree 1	2	3	4	5	6	Totally agree 7
The Voice Assistant should explain to users about the privacy of their data to help them decide which tasks to perform.							
Manufacturers and professionals from the Artificial Intelligence field should offer information about such technology to the population in an accessible manner (e.g., institutional material).							
Voice Assistants should present initial instructions, tutorials, and information about new supported actions and new ways to formulate commands.							
Marketing on Voice Assistants should stick to these systems' actual capabilities, presenting common and realistic use cases.							
Developers should improve speech recognition technology (e.g., synonyms, intents, words in other languages, accents, localization, and different voice types and users).							

11. You may add comments about the statements above if you wish.

← Anterior

Próximo →

12. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree 1	2	3	4	5	6	Totally agree 7
Developers should create error recovery mechanisms (ex: inform what was misunderstood and how to reformulate the command).							
There should be a platform that shows failed past interactions to help users understand the reasons for errors and the system's scope.							
Developers of conversational flows should receive training on the Assistans' technical limitations so they can produce appropriate flows.							
Developers should understand users (e.g., profiles, goals, contexts, behavior, semantics, mental models) to create solutions that address their needs and context.							
Developers should research mechanisms in human conversations and adapt them to interactions with Voice Assistants.							

13. You may add comments about the statements above if you wish.

← Anterior

Próximo →

14. Please answer to which level you agree with the statements below on a 1 to 7 scale. *

1 = Totally disagree; 4 = Neither agree, nor disagree; 7 = Totally agree.

	Totally disagree						Totally agree
	1	2	3	4	5	6	7
Developers should always apply best practices of usability and voice interaction when designing Voice Assistants.							
Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).							
Users should receive training on how to use Voice Assistants, including supported language patterns.							
No solution should be applied to Voice Assistants since users will naturally learn how to interact.							

15. You may add comments about the statements above if you wish.

← Anterior

Próximo →

Thank you for answering this questionnaire!

The results of this study's second round will be available soon.

16. Please provide your name and email below: *

17. If you wish to add any comments on this study, please use the field below:

← Anterior

Final

Appendix 9 – Delphi's second questionnaire (Portuguese)

Delphi - Segunda Rodada

Olá, obrigada por participar da segunda etapa deste estudo Delphi sobre modelos mentais de usuários de Assistentes de Voz.

O primeiro questionário pediu para que você respondesse duas perguntas abertas sobre o assunto. Para esta segunda etapa, as respostas de todos os participantes foram analisadas e sintetizadas em formato de afirmações. Por favor, leia as afirmações com atenção e assinale seu nível de concordância, em escala de 1 a 7. Essa etapa é importante porque irá determinar quais das afirmações serão um consenso entre o grupo de experts.

Na primeira parte, serão apresentadas 16 afirmações sobre as CAUSAS para o desalinhamento dos modelos mentais de usuários de Assistentes de Voz. Na segunda parte, serão apresentadas 19 afirmações contendo SOLUÇÕES para lidar com essa questão.

1. Se desejar, você pode ver novamente o texto explicativo sobre a questões dos modelos mentais:

Ver texto

Próximo →

PARTE 1 (1/3)

Abaixo, serão apresentadas afirmações que resumem as opiniões do grupo sobre as CAUSAS que levam os usuários de Assistentes de Voz a formarem modelos mentais desalinhados com as reais capacidades do sistema.

Por favor, marque seu nível de concordância com cada uma das frases abaixo, em uma escala de 1 a 7.

Você também pode adicionar comentários opcionais a cada uma das frases.

2. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
Experiências prévias ruins com outras interfaces de voz geram expectativas negativas nos usuários quanto os Assistentes de Voz							
Características que induzem o antropomorfismo (ex: voz, nome, gênero, metáforas, humor) levam os usuários a esperarem que os Assistentes sejam tão capazes quanto um humano.							
Os usuários desconhecem os limites técnicos dos Assistentes de Voz e da Inteligência Artificial, e exigem que o Assistente desempenhe tarefas e reconheça comandos além de sua capacidade.							
Preocupações com privacidade impedem os usuários de interagir o suficiente com os Assistentes para formar modelos mentais corretos.							
Os usuários constróem seu modelo mental sobre conversações por meio de interações humanas, mas os Assistentes têm menor capacidade conversacional, gerando quebra de expectativa e frustração.							

3. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

4. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
Os Assistentes de Voz não avisam aos usuários quando atualizam, e não explicam sobre as atualizações em suas capacidades.							
Os Assistentes de Voz não apresentam instruções iniciais para os usuários, deixando-os sem saber o que esperar do produto.							
Os Assistentes de Voz não explicam sobre limitações para determinadas ações, como o reconhecimento do contexto conversacional.							
Os Assistentes de Voz são muito complexos e exigem demais da cognição dos usuários.							
Os Assistentes de Voz não explicam aos usuários sobre suas capacidades, como devem ser utilizados, como processam os comandos, como tomam decisões e como se recuperar de erros.							

5. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

6. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
Os desenvolvedores realizam poucas pesquisas sobre a experiência dos usuários, fator crucial para a criação de fluxos conversacionais.							
As marcas apresentam pouco conteúdo institucional sobre os Assistentes e a Inteligência Artificial, levando os usuários a comprarem o produto sem conhecer suas capacidades.							
O marketing aumenta as expectativas dos usuários ao exagerar a capacidade social e a inteligência dos Assistentes de Voz e oferecer exemplos de uso muito simples e fluidos.							
Na ficção científica, sistemas de Inteligência Artificial são retratados como futurísticos, inteligentes, sensíveis, comunicativos e capazes, gerando percepções desalinhadas sobre os Assistentes atuais.							
Assistentes de Voz de diferentes marcas têm habilidades diferentes, levando à crença de que um Assistente possa ter as mesmas habilidades dos demais.							
Os usuários não se informam sobre os Assistentes de Voz antes de usá-los, principalmente, para tarefas que julgam desnecessárias.							

7. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

PARTE 2 (1/4)

Abaixo, serão apresentadas afirmações que resumem as opiniões do grupo sobre as SOLUÇÕES para alinhar os modelos mentais dos usuários de Assistentes de Voz às reais capacidades do sistema.

Por favor, marque seu nível de concordância com cada uma das frases abaixo, em uma escala de 1 a 7.

Você também pode adicionar comentários opcionais a cada uma das frases.

8. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
O Assistente de Voz deve deixar claro que os usuários estão falando com uma máquina, e não com um humano.							
Os desenvolvedores devem evitar características que humanizam os Assistentes de Voz (ex: nome, gênero, voz natural, metáforas).							
O Assistente de Voz deve fornecer exemplos e explicações sobre seu escopo de capacidades e seus processos de execução de ações, tomada de decisão e aprendizagem.							
O Assistente de Voz devem apresentar dicas de uso durante as interações, incluindo mecanismos para clarificar o contexto da conversa							
O Assistente de Voz deve deixar claro a importância da colaboração entre usuário e sistema, permitindo que o usuário ensine conteúdos para o assistente.							

9. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

10. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente 1	2	3	4	5	6	Concordo plenamente 7
O Assistente deve explicar aos usuários sobre a privacidade de seus dados para ajudá-lo a decidir quais tarefas executar.							
Empresas e profissionais da área de Inteligência Artificial devem fornecer à população informações sobre essa tecnologia de forma acessível (ex: material institucional).							
Os Assistentes de Voz devem apresentar instruções iniciais, tutoriais e informações sobre novas ações suportadas e novas formas de formular comandos.							
O marketing dos Assistentes de Voz deve ser fiel às suas reais capacidades, apresentando cenários de uso comuns e realistas.							
Desenvolvedores devem melhorar a tecnologia de reconhecimento da fala (ex: sinônimos, intenções, palavras em outros idiomas, sotaques, regionalizações e diferentes tipos de vozes e usuários).							

11. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

12. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
Os desenvolvedores devem criar mecanismos de recuperação de erros (ex: informar o que foi incompreendido e como reformular o comando).							
Deve existir uma plataforma que mostre falhas em interações passadas para ajudar os usuários a entender o motivo de erros e o escopo do sistema.							
Desenvolvedores de fluxos conversacionais devem receber treinamento sobre as limitações técnicas dos Assistentes para produzir fluxos apropriados.							
Os desenvolvedores devem entender os usuários (ex: perfils, objetivos, contextos, comportamentos, semântica, modelos mentais) e adequar as soluções às suas necessidades e contextos.							
Os desenvolvedores devem pesquisar mecanismos de conversação humanos e adaptá-los às interações com Assistentes de Voz.							

13. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

14. Marque seu nível de concordância com as afirmações abaixo em uma escala de 1 a 7. *

1 = Discordo plenamente; 4 = Não concordo, nem discordo; 7 = Concordo plenamente;

	Discordo plenamente						Concordo plenamente
	1	2	3	4	5	6	7
Os desenvolvedores sempre devem aplicar boas práticas de usabilidade e de interação por voz ao projetar os Assistentes.							
Os usuários devem se informar melhor sobre os Assistentes antes de começarem a usá-los (ex: ler conteúdos oficiais e extra-oficiais do produto).							
Os usuários devem receber treinamento sobre como utilizar os Assistentes de Voz, incluindo os padrões de linguagem suportados.							
Nenhuma solução deve ser aplicada aos Assistentes de Voz porque os usuários vão aprender a interagir naturalmente.							

15. Se desejar, adicione um comentário sobre as frases anteriores.

← Anterior

Próximo →

Obrigada por responder este questionário!

Os resultados da segunda etapa estarão disponíveis em breve.

16. Por favor, insira seu nome e email abaixo. *

17. Se desejar adicionar algum comentário sobre esta pesquisa, utilize o campo abaixo:

← Anterior

Final

Appendix 10 – Delphi's third questionnaire (English)

Delphi - Third Round

Página 1 / 5 (20%)

Hello, thank you for participating in the third round of this Delphi study on users' mental models of Voice Assistants.

The first questionnaire asked you to answer two open-ended questions on the topic. Such responses were analyzed and synthesized, resulting in statements. In the second round, the group was asked to rate these statements according to the participants' level of agreement on a scale from 1 to 7. The group rated a total of 35 statements, 16 on the CAUSES for users' unaligned mental models of Voice Assistants, and 19 with SOLUTIONS to deal with this issue. The goal was to understand which statements would reach a CONSENSUS among the group.

In this third round, we will present the results from the second round (descriptive statistics). You may review the results and leave comments to argue in favor/ against the group's opinion. Such comments are OPTIONAL.

1. If you wish, you may read again the text explaining about the issue of users' mental models of Voice Assistants:

View text

Próximo →

How to review the results?

For each statement, we will present the following descriptive statistics

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
8	Statement	6,00	6	1	75%	0%	25%

- #: the number of the statement by order of presentation in round 2;

- Mean: the average value of the group's score of agreement;

- Median: the number which divides the sample by 50%. The median shows the group's tendency to agree or disagree with the statement. A median of 1, 2, or 3 indicates disagreement. A median of 4 indicates uncertainty/ neutrality; A median of 5, 6, or 7 indicates agreement.

- Inter Quartile Range (IQR): represents the level of dispersion on the groups' opinions. The smaller the IQR, the smaller the dispersion and the stronger the consensus.

- Percentage of agreement: represents the percentage of participants who agreed with the statement (that is, the sum of participants who rated a statement as 5, 6, or 7, divided by the total).

- Percentage of uncertainty/ neutrality: represents the percentage of participants who were neutral towards the statement (that is, participants who rated a statement as 4, divided by the total).

- Percentage of disagreement: represents the percentage of participants who disagreed with the statement (that is, the sum of participants who rated a statement as 1, 2, or 3, divided by the total).

What is considered a consensus?

In this study, we consider strong consensus statements that had 1) an IQR of 1 or lower, and 2) a percentage of agreement/ disagreement of at least 75%. These are marked in green on the following pages.

However, we also considered that some statements reached some level of consensus if they satisfied one of the two criteria above. These are marked in yellow on the following pages.

← Anterior

Próximo →

PART 1

Below, we present the groups' assessment of the statements that summarize the CAUSES that lead users of Voice Assistants to form mental models that are unaligned with these systems' actual capabilities.

Please review the results and provide comments, if you wish.

2. Do you need to view the instructions again?

View

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
1	Bad previous experiences with other voice interfaces create negative expectations for the Voice Assistants on users.	6,22	6	1	100%	0%	0%
2	Characteristics that induce anthropomorphism (e.g., voice, name, gender, metaphors, humor) cause users to expect Voice Assistants would be as capable as a human.	5,61	6	1	89%	6%	6%
5	Users construct their mental models of conversations through human interactions, but Voice Assistants have lower conversational skills, letting down expectations and causing frustration.	6,11	6	1	89%	11%	0%
14	In Science Fiction, systems powered with Artificial Intelligence are pictured as futuristic, intelligent, sensitive, talkative, and capable, creating unaligned perceptions about current Assistants.	6,22	6,5	1	89%	11%	0%
9	Voice Assistants are too complex and demand too much of the users' cognition.	2,67	2	1	11%	11%	78%

3. You may add comments about the statements that reached STRONG CONSENSUS (above) if you wish.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
3	Users do not know the Voice Assistants' technical limitations and require the Assistant to perform tasks and recognize commands beyond its capabilities.	6,17	6,5	1,75	94%	6%	0%
13	Marketing raises users' expectations by exaggerating the Voice Assistant's social skills and intelligence, showing use cases that are too simple and fluid.	5,78	6	1,75	78%	17%	6%
10	Voice Assistants do not explain to users about their skills, how they should be utilized, how they process commands, how they make decisions, or how to recover from failures.	5,72	6	2	89%	6%	6%
8	Voice Assistants do not explain their limitations for certain actions, such as recognizing the conversational context.	5,61	6	2	83%	6%	11%

4. You may add comments about the statements that reached SOME CONSENSUS (above) if you wish.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
12	Brands present little institutional content about the Assistants and Artificial Intelligence, leading users to buy the product without being aware of its capabilities.	4,56	5	1,75	61%	11%	28%
16	Users do not look for information about the Voice Assistant before buying it, especially for tasks that are deemed unnecessary.	5,06	5,5	2	56%	33%	11%
6	Voice Assistants do not tell users when they update, nor explain the updates in their skills.	4,94	6	2,5	72%	0%	28%
15	Voice Assistants from different brands have different skills, leading to the belief that an Assistant might have the same skills as the others.	5,17	5	2,75	67%	22%	11%
4	Privacy concerns hinder users from interacting with the Assistants for long enough to construct correct mental models.	3,83	3	2,75	33%	11%	56%
11	Developers conduct little research about user experience, which is paramount to creating conversational flows.	4,22	4,5	3	50%	11%	39%
7	Voice Assistants do not present initial instructions to users, leaving them without knowing what to expect from the product.	5,00	5,5	3,75	56%	17%	28%

5. You may add comments about the statements that DID NOT REACH CONSENSUS (above) if you wish.

← Anterior

Próximo →

PART 2

Below, we present the groups' assessment of the statements that summarize the SOLUTIONS to align users' mental models with the Voice Assistants' actual capabilities.

Please review the results and provide comments, if you wish.

6. Do you need to view the instructions again?

View

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
25	Marketing on Voice Assistants should stick to these systems' actual capabilities, presenting common and realistic use cases.	6,28	6	1	100%	0%	0%
27	Developers should create error recovery mechanisms (ex: inform what was misunderstood and how to reformulate the command).	6,44	7	1	94%	6%	0%
30	Developers should understand users (e.g., profiles, goals, contexts, behavior, semantics, mental models) to create solutions that address their needs and context.	6,33	7	1	94%	6%	0%
31	Developers should research mechanisms in human conversations and adapt them to interactions with Voice Assistants.	6,11	6	1	94%	6%	0%
26	Developers should improve speech recognition technology (e.g., synonyms, intents, words in other languages, accents, localization, and different voice types and users).	6,11	6,5	1	89%	0%	11%
32	Developers should always apply best practices of usability and voice interaction when designing Voice Assistants.	6,28	7	1	89%	11%	0%
24	Voice Assistants should present initial instructions, tutorials, and information about new supported actions and new ways to formulate commands.	6,00	6,5	1	83%	6%	11%
29	Developers of conversational flows should receive training on the Assistans' technical limitations so they can produce appropriate flows.	6,11	7	1	83%	11%	6%

7. You may add comments about the statements that reached STRONG CONSENSUS (above) if you wish.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
34	Users should receive training on how to use Voice Assistants, including supported language patterns.	4,00	4	1	33%	44%	22%
20	Voice Assistants should present usage tips throughout interactions, including mechanisms to clarify the conversation context.	6,00	6,5	1,75	89%	6%	6%
35	No solution should be applied to Voice Assistants since users will naturally learn how to interact.	2,72	2,5	1,75	17%	6%	78%
19	Voice Assistants should provide examples and explanations about their skills' scope and action execution, decision making, and learning processes.	6,06	6,5	2	94%	6%	0%
17	Voice Assistants should make it clear for users that they are talking to a machine and not a human.	5,56	6,5	2	78%	11%	11%
22	The Voice Assistant should explain to users about the privacy of their data to help them decide which tasks to perform.	5,67	6	2	78%	17%	6%
23	Manufacturers and professionals from the Artificial Intelligence field should offer information about such technology to the population in an accessible manner (e.g., institutional material).	5,72	6	2	78%	17%	6%
28	There should be a platform that shows failed past interactions to help users understand the reasons for errors and the system's scope.	5,50	5,5	2	78%	17%	6%

8. You may add comments about the statements that reached SOME CONSENSUS (above) if you wish.

#	Statement	Mean	Median	IQR	Agreement	Uncertainty	Disagreement
18	Developers should avoid characteristics that humanize the Voice Assistant (e.g., name, gender, natural voice, metaphors).	3,83	4	2	39%	22%	39%
21	The Voice Assistant should clarify the importance of the collaboration between user and system, allowing users to teach content to the Assistant.	5,50	6	2,75	72%	22%	6%
33	Users should inform themselves better about the Assistants before utilizing them (e.g., read official and unofficial content about the product).	4,61	4	4	44%	22%	33%

9. You may add comments about the statements that DID NOT REACH CONSENSUS (above) if you wish.

← Anterior

Próximo →

Thank you for answering this questionnaire!

The final results of this study will be available soon.

10. Please provide your name:

11. If you wish to add any comments on this study, please use the field below:

← Anterior

Final

Appendix 11 – Delphi's third questionnaire (Portuguese)

Delphi - Third Round

Página 1 / 5 (20%)

Olá, obrigada por participar da terceira etapa deste estudo Delphi sobre modelos mentais de usuários de Assistentes de Voz.

O primeiro questionário pediu para que você respondesse duas perguntas abertas sobre o assunto. Essas respostas foram analisadas e sintetizadas em formato de afirmações. Na segunda etapa, o grupo atribuiu uma nota para essas frases de acordo com seu nível de concordância, em escala de 1 a 7. O grupo avaliou um total de 35 frases, sendo 16 sobre as CAUSAS para o desalinhamento dos modelos mentais de usuários de Assistentes de Voz e 19 contendo SOLUÇÕES para lidar com essa questão. O objetivo foi entender quais frases seriam um CONSENSO entre o grupo.

Nesta terceira etapa, nós apresentaremos os resultados da segunda etapa (estatísticas descritivas). Você pode revisar os resultados e deixar comentários adicionais para argumentar a favor ou contra a opinião do grupo. Esses comentários são OPCIONAIS.

1. Se desejar, você pode ver novamente o texto explicativo sobre a questões dos modelos mentais:

○ Ver texto

Próximo →

Como revisar os resultados?

Para cada frase, iremos apresentar as seguintes estatísticas descritivas

A	B	C	D	E	F	G	H
#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
8	Frase	6,00	6	1	75%	0%	25%

- #: representa o número da frase, por ordem de apresentação na fase 2;

- Média: O valor médio do nível de concordância do grupo;

- Mediana: o número que divide a amostra em 50%. A mediana mostra a tendência do grupo a concordar ou discordar da frase. Uma mediana de 1, 2 ou 3 indica discordância. Uma mediana de 4 indica neutralidade. Uma mediana de 5, 6 ou 7 indica concordância.

- Amplitude Inter Quartílica (IQR): Representa o nível de dispersão nas opiniões do grupo. Quanto menor o IQR, menor é a dispersão e mais forte é o consenso.

- Porcentagem de concordância: representa a porcentagem de participantes que concordaram com a frase (isto é, a soma dos participantes que classificaram a frase como 5, 6 ou 7, dividido pelo número total de participantes).

- Porcentagem de incerteza/ neutralidade: representa a porcentagem de participantes que foram neutros em relação a frase (isto é, a soma dos participantes que classificaram a frase como 4, dividido pelo número total de participantes).

- Porcentagem de discordância: representa a porcentagem de participantes que foram neutros em discordaram da frase (isto é, a soma dos participantes que classificaram a frase como 1, 2 ou 3, dividido pelo número total de participantes).

O que está sendo considerado um consenso?

Neste estudo, nós consideramos um consenso forte as frases que 1) tiveram IQR de 1 ou menor, e 2) um percentual de concordância/ discordância de 75% ou maior. Tais frases estão marcadas em verde nas páginas seguintes.

Porém, nós também consideramos que algumas frases alcançaram algum nível de consenso se elas satisfizeram um dos dois critérios citados acima. Tais frases estão marcadas em amarelo nas páginas seguintes.

← Anterior

Próximo →

PARTE 1

Abaixo, serão apresentadas as avaliações do grupo quanto as afirmações que resumem as as CAUSAS que levam os usuários de Assistentes de Voz a formarem modelos mentais desalinhados com as reais capacidades do sistema.

Por favor, revise os resultados e adicione comentários, se desejar.

2. Você deseja ver as instruções novamente?

- ○ Ver

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
1	Experiências prévias ruins com outras interfaces de voz geram expectativas negativas nos usuários quanto os Assistentes de Voz	6,22	6	1	100%	0%	0%
2	Características que induzem o antropomorfismo (ex: voz, nome, gênero, metáforas, humor) levam os usuários a esperarem que os Assistentes sejam tão capazes quanto um humano.	5,61	6	1	89%	6%	6%
5	Os usuários constróem seu modelo mental sobre conversações por meio de interações humanas, mas os Assistentes têm menor capacidade conversacional, gerando quebra de expectativa e frustração.	6,11	6	1	89%	11%	0%
14	Na ficção científica, sistemas de Inteligência Artificial são retratados como futurísticos, inteligentes, sensíveis, comunicativos e capazes, gerando percepções desalinhadas sobre os Assistentes atuais.	6,22	6,5	1	89%	11%	0%
9	Os Assistentes de Voz são muito complexos e exigem demais da cognição dos usuários.	2,67	2	1	11%	11%	78%

3. Você pode adicionar comentários opcionais sobre as frases que atingiram um CONSENSO FORTE (acima) se desejar.

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
3	Os usuários desconhecem os limites técnicos dos Assistentes de Voz e da Inteligência Artificial, e exigem que o Assistente desempenhe tarefas e reconheça comandos além de sua capacidade.	6,17	6,5	1,75	94%	6%	0%
13	O marketing aumenta as expectativas dos usuários ao exagerar a capacidade social e a inteligência dos Assistentes de Voz e oferecer exemplos de uso muito simples e fluidos.	5,78	6	1,75	78%	17%	6%
10	Os Assistentes de Voz não explicam aos usuários sobre suas capacidades, como devem ser utilizados, como processam os comandos, como tomam decisões e como se recuperar de erros.	5,72	6	2	89%	6%	6%
8	Os Assistentes de Voz não explicam sobre limitações para determinadas ações, como o reconhecimento do contexto conversacional.	5,61	6	2	83%	6%	11%

4. Você pode adicionar comentários opcionais sobre as frases que atingiram ALGUM CONSENSO (acima) se desejar.

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
12	As marcas apresentam pouco conteúdo institucional sobre os Assistentes e a Inteligência Artificial, levando os usuários a comprarem o produto sem conhecer suas capacidades.	4,56	5	1,75	61%	11%	28%
16	Os usuários não se informam sobre os Assistentes de Voz antes de usá-los, principalmente, para tarefas que julgam desnecessárias.	5,06	5,5	2	56%	33%	11%
6	Os Assistentes de Voz não avisam aos usuários quando atualizam, e não explicam sobre as atualizações em suas capacidades.	4,94	6	2,5	72%	0%	28%
15	Assistentes de Voz de diferentes marcas têm habilidades diferentes, levando à crença de que um Assistente possa ter as mesmas habilidades dos demais.	5,17	5	2,75	67%	22%	11%
4	Preocupações com privacidade impedem os usuários de interagir o suficiente com os Assistentes para formar modelos mentais corretos.	3,83	3	2,75	33%	11%	56%
11	Os desenvolvedores realizam poucas pesquisas sobre a experiência dos usuários, fator crucial para a criação de fluxos conversacionais.	4,22	4,5	3	50%	11%	39%
7	Os Assistentes de Voz não apresentam instruções iniciais para os usuários, deixando-os sem saber o que esperar do produto.	5,00	5,5	3,75	56%	17%	28%

5. Você pode adicionar comentários opcionais sobre as frases que NÃO ATINGIRAM CONSENSO (acima) se desejar.

← Anterior

Próximo →

PARTE 2

Abaixo, serão apresentadas as avaliações do grupo sobre as afirmações que resumem as SOLUÇÕES para alinhar os modelos mentais dos usuários de Assistentes de Voz às reais capacidades do sistema.

Por favor, revise os resultados e adicione comentários adicionais, se desejar.

6. Você deseja ver as instruções novamente?

○ Ver

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
25	O marketing dos Assistentes de Voz deve ser fiel às suas reais capacidades, apresentando cenários de uso comuns e realistas.	6,28	6	1	100%	0%	0%
27	Os desenvolvedores devem criar mecanismos de recuperação de erros (ex: informar o que foi incompreendido e como reformular o comando).	6,44	7	1	94%	6%	0%
30	Os desenvolvedores devem entender os usuários (ex: perfils, objetivos, contextos, comportamentos, semântica, modelos mentais) e adequar as soluções às suas necessidades e contextos.	6,33	7	1	94%	6%	0%
31	Os desenvolvedores devem pesquisar mecanismos de conversação humanos e adaptá-los às interações com Assistentes de Voz.	6,11	6	1	94%	6%	0%
26	Desenvolvedores devem melhorar a tecnologia de reconhecimento da fala (ex: sinônimos, intenções, palavras em outros idiomas, sotaques, regionalizações e diferentes tipos de vozes e usuários).	6,11	6,5	1	89%	0%	11%
32	Os desenvolvedores sempre devem aplicar boas práticas de usabilidade e de interação por voz ao projetar os Assistentes.	6,28	7	1	89%	11%	0%
24	Os Assistentes de Voz devem apresentar instruções iniciais, tutoriais e informações sobre novas ações suportadas e novas formas de formular comandos.	6,00	6,5	1	83%	6%	11%
29	Desenvolvedores de fluxos conversacionais devem receber treinamento sobre as limitações técnicas dos Assistentes para produzir fluxos apropriados.	6,11	7	1	83%	11%	6%

7. Você pode adicionar comentários opcionais sobre as frases que atingiram um CONSENSO FORTE (acima) se desejar.

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
34	Os usuários devem receber treinamento sobre como utilizar os Assistentes de Voz, incluindo os padrões de linguagem suportados.	4,00	4	1	33%	44%	22%
20	O Assistente de Voz devem apresentar dicas de uso durante as interações, incluindo mecanismos para clarificar o contexto da conversa	6,00	6,5	1,75	89%	6%	6%
35	Nenhuma solução deve ser aplicada aos Assistentes de Voz porque os usuários vão aprender a interagir naturalmente.	2,72	2,5	1,75	17%	6%	78%
19	O Assistente de Voz deve fornecer exemplos e explicações sobre seu escopo de capacidades e seus processos de execução de ações, tomada de decisão e aprendizagem.	6,06	6,5	2	94%	6%	0%
17	O Assistente de Voz deve deixar claro que os usuários estão falando com uma máquina, e não com um humano.	5,56	6,5	2	78%	11%	11%
22	O Assistente deve explicar aos usuários sobre a privacidade de seus dados para ajudá-lo a decidir quais tarefas executar.	5,67	6	2	78%	17%	6%
23	Empresas e profissionais da área de Inteligência Artificial devem fornecer à população informações sobre essa tecnologia de forma acessível (ex: material institucional).	5,72	6	2	78%	17%	6%
28	Deve existir uma plataforma que mostre falhas em interações passadas para ajudar os usuários a entender o motivo de erros e o escopo do sistema.	5,50	5,5	2	78%	17%	6%

8. Você pode adicionar comentários opcionais sobre as frases que atingiram ALGUM CONSENSO (acima) se desejar.

#	Frase	Média	Mediana	IQR	Concordância	Incerteza	Discordância
18	Os desenvolvedores devem evitar características que humanizam os Assistentes de Voz (ex: nome, gênero, voz natural, metáforas).	3,83	4	2	39%	22%	39%
21	O Assistente de Voz deve deixar claro a importância da colaboração entre usuário e sistema, permitindo que o usuário ensine conteúdos para o assistente.	5,50	6	2,75	72%	22%	6%
33	Os usuários devem se informar melhor sobre os Assistentes antes de começarem a usá-los (ex: ler conteúdos oficiais e extra-oficiais do produto).	4,61	4	4	44%	22%	33%

9. Você pode adicionar comentários opcionais sobre as frases que NÃO ATINGIRAM CONSENSO (acima) se desejar.

← Anterior

Próximo →

Obrigada por responder este questionário!

Os resultados finais do estudo estarão disponíveis em breve.

10. Por favor, informe seu nome:

11. Se desejar adicionar algum comentário sobre esta pesquisa, utilize o campo abaixo:

← Anterior

Final

Appendix 12 – Approval of PUC-Rio's board of ethics

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

CÂMARA DE ÉTICA EM PESQUISA DA PUC-Rio
Parecer da Comissão da Câmara de Ética em Pesquisa da PUC-Rio 66/2021 – Protocolo 70/2021

A Câmara de Ética em Pesquisa da PUC-Rio foi constituída como uma Câmara específica do Conselho de Ensino e Pesquisa conforme decisão deste órgão colegiado com atribuição de avaliar projetos de pesquisa do ponto de vista de suas implicações éticas.

Identificação:
Título: "Melhorando o modelo mental de usuários de assistentes de voz (AVs) por meio da transparência em respostas do sistema" (Departamento de Artes & Design da PUC-Rio)
Autora: Isabella Canellas da Motta (Mestranda do Departamento de Artes & Design da PUC-Rio)
Orientadora: Maria Manuela Quaresma (Professora do Departamento de Artes & Design da PUC-Rio)

Apresentação: Pesquisa descritiva que visa oferecer recomendações para o design de respostas de AVs que alinham os modelos mentais de usuários com as reais capacidades do sistema. Adotará um método de abordagem quali quanti junto a pesquisadores e desenvolvedores que trabalhem no desenvolvimento de AVs atuais, e experts na área. Prevê a aplicação de entrevista semi estruturada usando a plataforma de vídeo chamada Zoom ou a Google Meet. Também, via online utilizará o Método Delphi – Questionário usado para comunicação de um grupo de especialistas no trabalho de um problema complexo, preservando o anonimato. Tem apoio teórico na revisão de literatura sobre: Modelos mentais na interação humano-computador (IHC) e Modelos mentais de usuários sobre AVs.

Aspectos éticos: O projeto e o Termo de Consentimento Livre e Esclarecido apresentados estão de acordo com os princípios e valores do Marco Referencial, Estatuto e Regimento da Universidade no que se refere às responsabilidades de seu corpo docente e discente. O Termo expõe com clareza os objetivos da pesquisa e os procedimentos a serem seguidos. Garante o sigilo e a confidencialidade dos dados coletados. Informa sobre a possibilidade de interrupção na pesquisa sem aplicação de qualquer penalidade ou constrangimento.

Parecer: Aprovado

Prof. José Ricardo Bergmann
Presidente do Conselho de Ensino e Pesquisa da PUC-Rio

Ilda Lopes Rodrigues da Silva

Profª Ilda Lopes Rodrigues da Silva
Coordenadora da Comissão da Câmara de Ética em Pesquisa da PUC-Rio

Rio de Janeiro, 31 de agosto de 2021

Vice-Reitoria para Assuntos Acadêmicos
Câmara de Ética em Pesquisa da PUC-Rio – CEPq/PUC-Rio
Rua Marquês de São Vicente, 225 - Gávea – 22453-900
Rio de Janeiro – RJ – Tel. (021) 3527-1612 / 3527-1618
e-mail: vrac@puc-rio.br