



---

# Meaningful Voice Interactions

How to design smart speakers that foster well-being

**Master Thesis**

Felipe da Motta Rezende Pierantoni

Chair: Prof. dr. ir. Pieter Desmet

Mentor: Prof. dr. Elisa Giaccardi

Co-supervisor: Dr. Olya Kudina

 **TU**Delft

Faculty of Industrial Design Engineering  
Msc Design for Interaction

June 2020 TU Delft

Delft University of Technology

Faculty of Industrial Design Engineering

Design for Interaction

Master Thesis

Felipe da Motta Rezende Pierantoni 4793528

felipe.pierantoni@gmail.com

#### Supervisory Team

Chair: Prof. dr. ir. Pieter Desmet

Mentor: Prof. dr. Elisa Giaccardi

Co-supervisor: Dr. Olya Kudina





# Meaningful Voice Interactions

How to design smart speakers that foster well-being

**Master Thesis**

Felipe da Motta Rezende Pierantoni



---

## Acknowledgments

I would first like to express my gratitude to my supervisory team: Pieter, Elisa and Olya. The chance to work with such brilliant professors and researchers was exactly the experience I looked for when I decided to pursue this Masters. Thank you for the guidance, support and inspiration.

I am also grateful for everyone who took part in this project, from the awesome friends and colleagues who joined in the research activities to all the incredible people who took the time and care to voice their perceptions during testing. Your participation was immensely valuable. A special thanks to Castor, Rosana and Amanda for acting in the videos.

As this project ends, so does this exciting two-year adventure abroad. I wish to thank all who made it possible and meaningful, especially my dear family, my amazing friends and all the wonderful people I got the chance to meet in the Netherlands.

Finally, I would like to thank my beloved wife, Clara, for much beyond your priceless insights and everlasting motivation. This journey with you has been a life-changing experience. There are no words for how grateful I am to have had you on my side.

*Felipe Pierantoni*

---

## Executive Summary

What is the impact of smart speakers on our well-being?

With an expected global installed base of 325 million units in 2020 (Advanced Television, 2020), smart speakers continue to become increasingly present in households worldwide, especially among young adult consumers (Kinsella, 2019). As their popularity grows, however, so does the discussions around the potential dangers of these devices that speak like humans and share the intimacy of our homes. From reports of children developing aggressive behaviours (Childwise, 2018) to studies on the perpetuation of sexist stereotypes (West, Kraut, Ei Chew, 2018), our interactions with smart speakers are accompanied by a series of risks. This project proposes a repertoire of meaningful voice interactions to mitigate those impacts and foster well-being instead.

To achieve this, a categorisation of seven dangers of voice interactions was created based on literature review: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment. Each danger was analysed in order to define which of three fundamental human needs they harm: autonomy, competence and relatedness. This approach is based on the Self Determination Theory, which proposes that people experience well-being when these three needs are satisfied (Ryan & Deci, 2000).

In addition to this analysis, a phase of exploratory research was also conducted, combining established research methods such as user interviews with emerging More-Than-Human Design approaches in the form of Thing Ethnography. The result was a series of findings on the perspectives of users and devices that would not be accessible from literature alone. These insights inspired the ideation phase, where more than 70 ideas were created to prevent the dangers of current voice interactions.

After clustering and filtering, these ideas were tested with people by showing them videos depicting each interaction concept. Participants were asked to evaluate how the needs for autonomy, competence and relatedness would be affected, together with ranking the ideas and expressing their overall perceptions. The resulting data revealed the interaction concepts perceived as the most meaningful. These served as the basis for the final deliverable of this project, a repertoire of meaningful voice interactions.

The repertoire includes design guidelines, traps to avoid, in-depth analysis of the dangers of voice interactions and reflective questions to guide ideation. Its goal is to serve as a tool for designers and researchers involved in voice-related projects to foster well-being by designing more humane voice interactions. It was designed to be initially shared as an online platform and a booklet, but its content could take various forms beyond those.

# Table of Contents

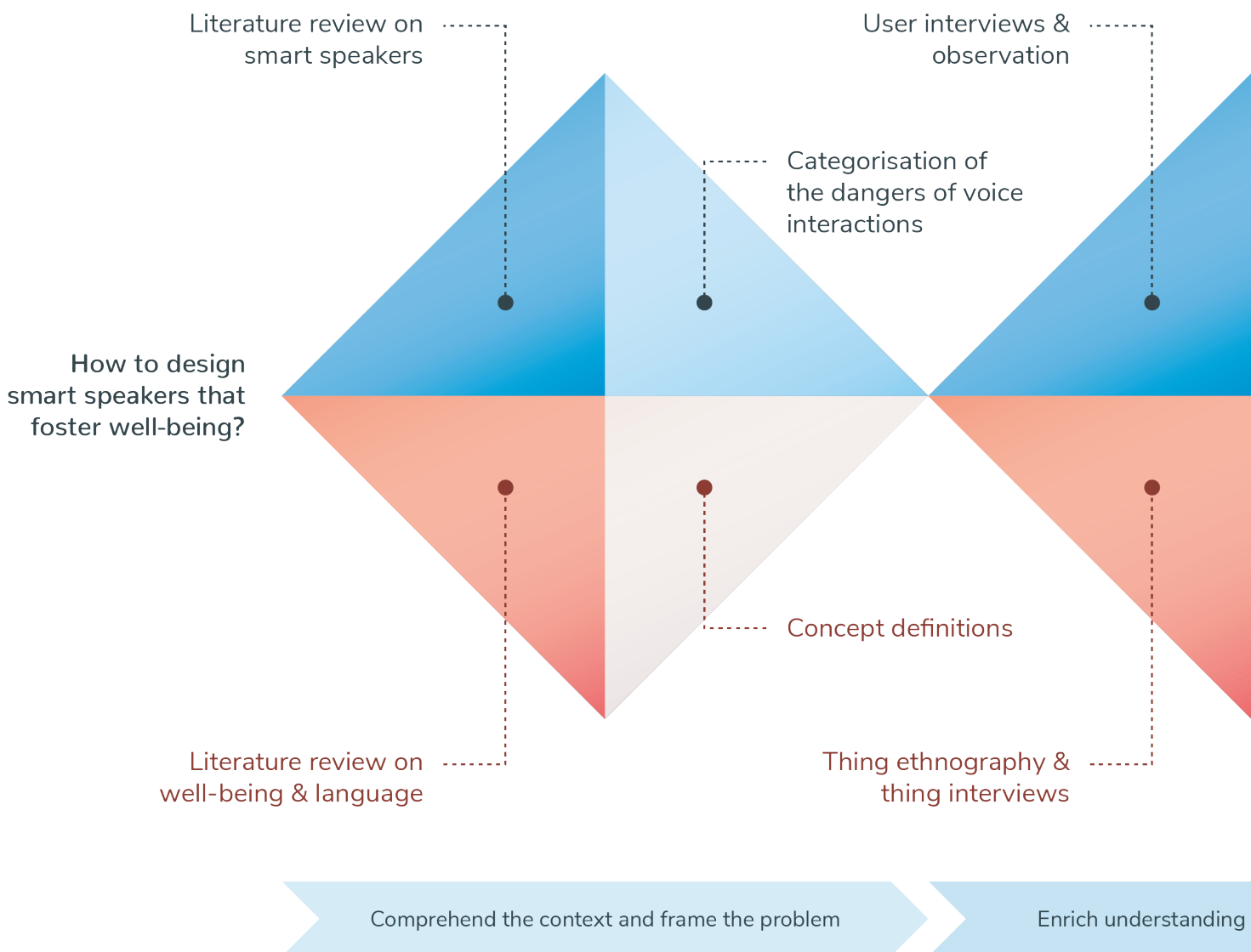
<b>Project Overview</b>	<b>10</b>	<b>Chapter 2:</b>	<b>52</b>
<b>Chapter 1: Context</b>	<b>12</b>	<b>Exploratory Research</b>	
<b>1. An introduction to voice assistants and smart speakers</b>	<b>13</b>	1. Motivation for exploration	53
Definition and history of voice assistants	13	2. User Perspective	54
The context of smart speakers	15	3. Device Perspective	60
Current and upcoming capabilities of voice technology	18	<b>Chapter 3:</b>	<b>70</b>
<b>2. Language and the performative perspective</b>	<b>21</b>	<b>Project Definitions</b>	
<b>3. Well-being and fundamental human needs</b>	<b>23</b>	Problem Definition	71
<b>4. The seven dangers of smart speaker interactions</b>	<b>26</b>	Design Goal	71
Impoliteness	27	Approach	71
Aggressiveness	31	Other definitions & limitations	72
Gender Stereotyping	33	<b>Chapter 4: Ideation</b>	<b>74</b>
Exposure	37	1. Approach	75
Shallow Mindedness	40	2. Results	76
Emotional Dependency	43	Impoliteness	77
Social Detachment	46	Aggressiveness	78
<b>5. Discussion</b>	<b>48</b>	Gender Stereotyping	80
		Exposure	81
		Shallow Mindedness	83
		Emotional Dependency	84
		Social Detachment	85

<b>Chapter 5:</b>	<b>88</b>	<b>Chapter 7: Repertoire of</b>	<b>132</b>
<b>User Testing – Approach</b>		<b>Meaningful Interactions</b>	
1. Selecting ideas to test	89	1. Objective and channels	133
2. Defining how and what to measure	94	of distribution	
		2. Structure	134
<b>Chapter 6:</b>	<b>98</b>	Final content	135
<b>User Testing – Results</b>		Impoliteness	136
1. Overview of participants	99	Aggressiveness	139
2. Results and insights for each scenario	101	Gender Stereotyping	143
Scenario 1: Reacting to aggressive engagement	101	Exposure	147
Scenario 2: Reacting to polite engagement	106	Shallow Mindedness	149
Scenario 3: Reacting to mistakes	110	Emotional Dependency	152
Scenario 4: Configuring voice	114	Social Detachment	156
Scenario 5: Reacting to sexual harassment	117		
Scenario 6: Supporting emotionally vulnerable users	122	<b>Conclusion</b>	<b>158</b>
Scenario 7: Answering complex questions	127	<b>References</b>	<b>164</b>
		<b>Appendices</b>	<b>168</b>
		A. Script: Interview with Users	169
		B. Script: Interview with Devices	174
		C. Responses: Interviews with Devices	176
		D. Script for the online survey	183
		E. Online Platform	188
		F. Repertoire Booklet	192

# Project Overview

## Context Analysis

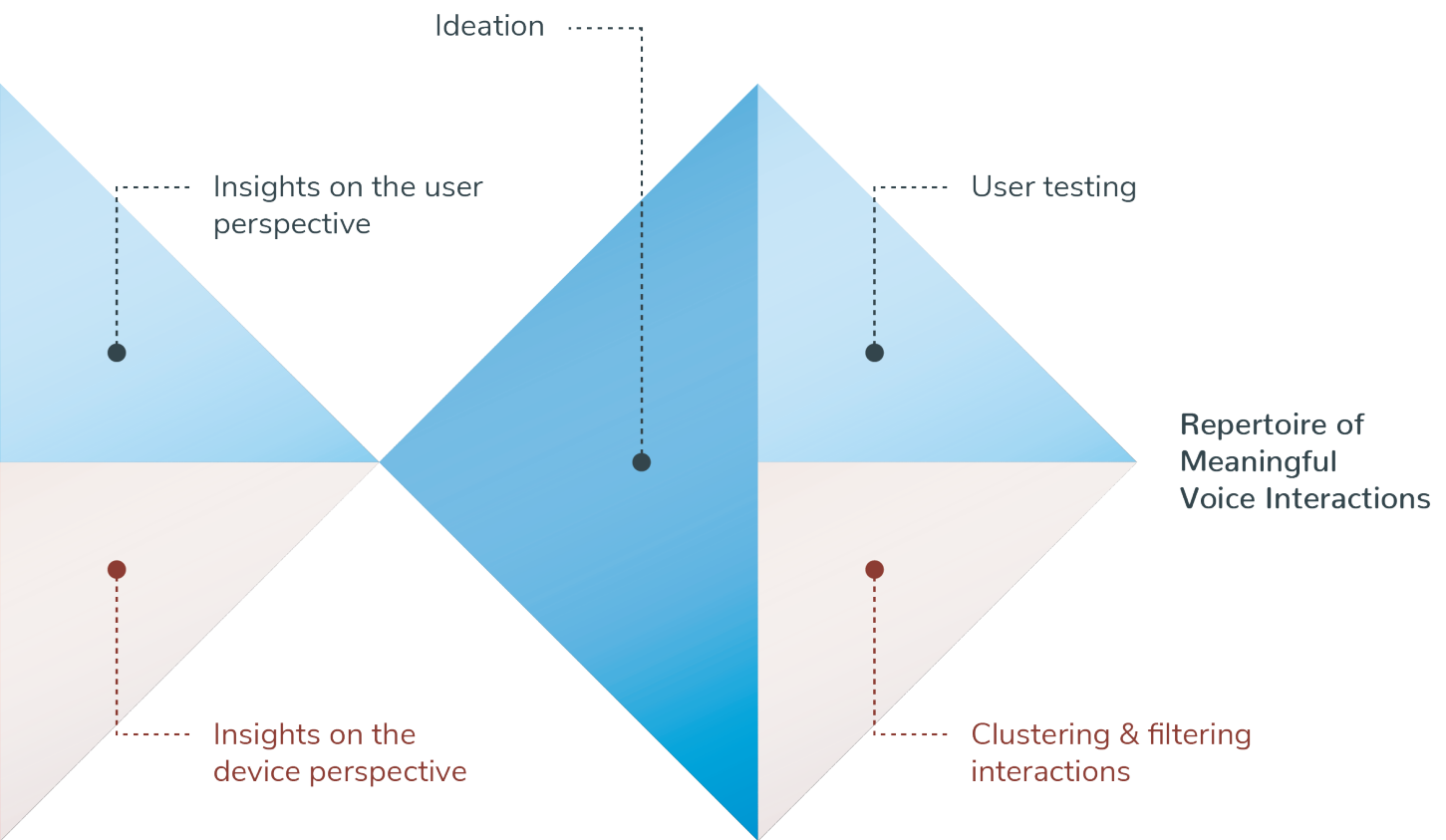
## Exploratory Research





# Laboratory Research

# Meaningful Interactions



and gather inspiration

Design, validate and select ideas

## Chapter 1

# Context

In order to understand how to foster well-being in smart-speaker interactions, it was first necessary to comprehend the context around smart speakers, their risks and the concepts upon which this project is based. What is well-being? How can it be assessed? What are the dangers of voice interactions? What is the effect of language on our reality?

The analysis described in this chapter is based on a literature review over smart speakers, voice interactions, well-being theories and linguistics. This review includes academic research, industry reports, think-pieces, news articles and other relevant sources.

---

### This chapter includes:

- History, description and features of voice assistants and smart speakers
- Well-being definition utilised in this research
- Analysis of the performative perspective of language
- Analysis of the dangers provoked by smart speakers
- Description of the human needs harmed by smart speakers



Figure 1. Popular models of smart speakers (Middleton, 2018)

## 1. An introduction to voice assistants and smart speakers

### DEFINITION AND HISTORY OF VOICE ASSISTANTS

Voice assistants can be described as software agents capable of interpreting human speech and responding through synthesised voices (Hoy, 2018). Their main capabilities are answering questions and providing users with a hands-free way to manage routine activities such as controlling music playback, managing emails, setting up timers, checking the weather and more. Besides directly attending user commands, voice assistants also play an increasingly central role in home automation, acting as the control hub for all sorts of connected objects, from smart lights to garage doors.

The first voice assistant as we know it was Apple's Siri, launched in 2010 as a standalone app and directly embedded into Apple's iOS one year later (Figure 2). Embedding voice assistants into connected devices would become a staple strategy for technology companies, as evidenced by the other two major players in the western market, Amazon and Google. Amazon's voice assistant, Alexa, was released together with their connected home speaker Echo in 2014. Google followed in 2016 by releasing its own smart speaker and integrating Google Assistant in Android-based smartphones (Hoy, 2018).

While smartphones and smart speakers remain the most common way to interact with voice assistants, this technology continues expanding towards other devices. For instance, Siri is now available on Apple computers and smartwatches as well, and Google directly included Google Assistant in their Stadia video game controller. Besides all the assistant-powered devices that users can talk to, the list of connected objects and peripherals that support them also rapidly grows. Given the trend of embedding assistants into everyday things and technologies, it is possible to assume that our interactions with computers will become progressively hands-free and voice-based (West et al., 2019). In fact, companies are not shy about disclaiming their wish to make their assistants ubiquitous. Amazon wants Alexa to follow their users everywhere (Wakefield, 2019), and Google states that “voice assistants are now, and will continue to be, at the centre of the home and part of daily routines”, offering a “new, more human relationship with technology” (Kleinberg, 2018).

But why are voice assistants considered a ‘new and more human way’ to interact with technology? After all, even though voice assistants have only been around for a decade, voice-based technologies have been experimented with long before that. The core difference lies in the fact that early voice-activated technology was limited to built-in commands and answers (Hoy, 2018), whereas current voice assistants are capable of attending to an expansive spectrum of questions and generating unscripted output that was not necessarily coded or determined beforehand. This dynamic output is created by artificial intelligence (AI) supported by a system of human-guided machine learning algorithms (West et al., 2019).

Every request a voice assistant receives is transmitted via the internet to a central system tasked with parsing the command and generating an adequate response (Hoy, 2018). The capacity of AI to provide complex and seemingly spontaneous responses is supported by significant breakthroughs in natural language processing. These advances were enabled by increased computing power, extensive availability of linguistic data, improvements in machine-learning techniques and a greater comprehension of language and its application in social interactions (Hirshberg & Manning, 2015).

Figure 2. Siri’s announcement as an integrated feature of iOS in 2011 (Brownlee, 2011)





As a result, current voice assistants allow people to interact with computational agents in a way that, not many decades ago, was only imaginable in works of science fiction (Hoy, 2018). Not only that but interacting using natural spoken language is fundamentally different than clicking mouse buttons, touching a screen or uttering specific words. Because spoken language is so inherently human, speaking with an artificial intelligence leads to a series of effects and reflections that will be discussed in this chapter and addressed all throughout this project.

### THE CONTEXT OF SMART SPEAKERS



Figure 3. Smart speaker partaking in the domestic life of a family. (Vivint, 2019)

Of all devices embedded with voice assistants, smart speakers are particularly interesting because – different than smartphones or laptops – they were specifically created to host these assistants. Smart speakers are equipped with what is known as far-field voice technology, enabling them to decipher speech at a distance (Shulevitz, 2018). Their advanced microphones are always listening for the keywords assigned to wake up the device (e.g. ‘Ok, Google’, ‘Alexa’). When the keyword is heard, the smart speaker starts creating a recording of the user’s voice, which is then sent to a specialised server to be processed and interpreted (Hoy, 2018).

As of 2019, smart speakers have reached 205 million units shipped worldwide, with almost 100 million sold in 2019 alone (Kinsella, 2019). In markets like the United States, statistics show that, in December 2018, at least 15 million people owned three or more smart speakers (West et al., 2019). These numbers display the striking popularity of smart speakers and the influence of this technology over so many people, especially young adults, who are more likely to own a device (Kinsella, 2019) (Figure 4).

Smart speakers are usually positioned in the living room and the kitchen — where they can be utilised by all household members — and in the bedroom, a prime location for users who have more than one device (Kinsella, 2020) (Figure 5). As the technology spreads and prices decrease, though, it is not uncommon for smart speakers to reach other – and perhaps more private – locations. For example, Echo Flex, Amazon’s cheapest smart speaker model, was designed to work directly plugged to an outlet wall, fitting for locations as hallways and bathrooms (Seifert, 2019). In terms of tasks, people tend to interact with smart speakers while cooking, multitasking and watching TV (Kinsella, 2018). More specifically, the five most common use cases are: asking a question, listening to streaming music, checking the weather, setting an alarm and setting a timer (Kinsella & Mutchler, 2020) (Figure 6).

Defining the context around smart speakers is significant because this is the scope of this project. Although several insights and reflections might also address other touchpoints involving voice assistants, all explorations were targeted on smart speaker usage in the home environment, mainly by young adult owners. More specifically, this research focuses on the leading smart speakers produced by Google, Amazon and Apple for western markets.

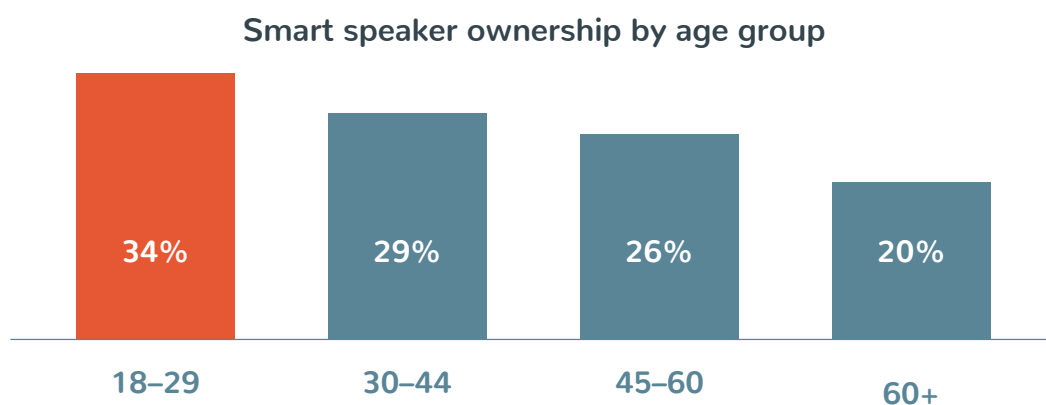


Figure 4. Chart showing how many people in each age group own a smart speaker in the United States. (Kinsella, 2019)

Primary location for smart speakers

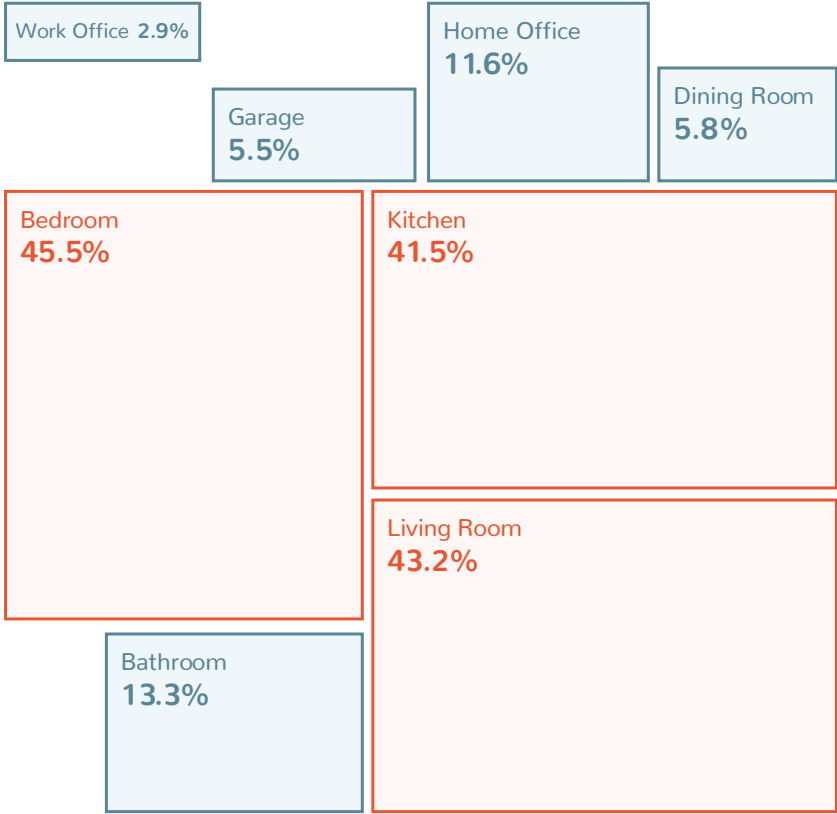


Figure 5. Chart showing where American consumers have smart speakers in 2020. (Kinsella, 2020)

Top 10 Smart Speaker Use Cases

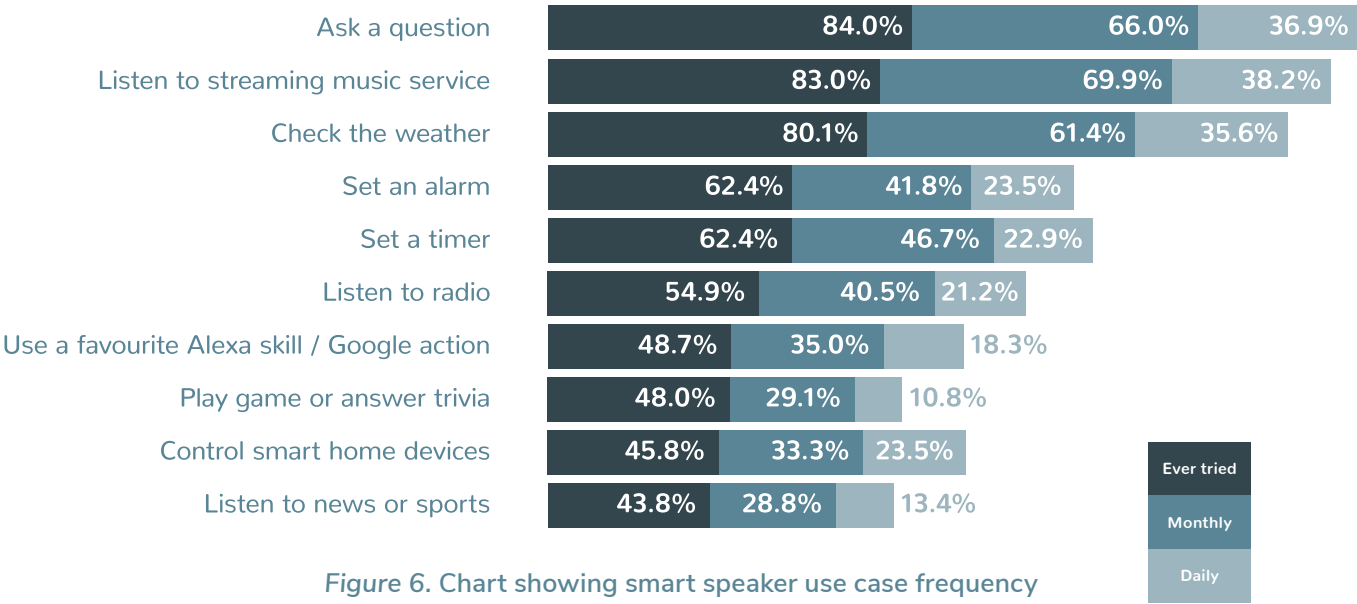


Figure 6. Chart showing smart speaker use case frequency in January 2019. (Kinsella & Mutchler, 2020)

## CURRENT AND UPCOMING CAPABILITIES OF VOICE TECHNOLOGY

As established so far, smart speakers are powered by voice assistants that speak through synthesised voices that closely mimic human speech. That is not, however, the full extent of this technology, and upcoming innovations will bring additional possibilities to smart speakers. In this project, tracking these capabilities is important for two reasons. First, they can significantly affect the unintended effects of voice interaction and even create new ones. Second — and on a positive note —, they expand the design space for designers and researchers to ideate on and understand how voice interactions can enhance well-being.

For example, a relevant capability to consider is that smart speakers nowadays are able to recognise who is talking to them. When asked to check for new emails, the assistant is capable of identifying who is asking and deducing which email box to open. Something else to consider is that, beyond built-in features, most smart speakers can also access applications and features designed by third-party developers (Hoy, 2018). So, there is some leeway to explore novel interaction ideas.

Nevertheless, when it comes to defining how new technologies are implemented in smart speakers and how voice assistants behave, large tech companies are mainly the ones making decisions. For instance, Google recently sparked controversy with the Google Duplex project, a specific version of Google Assistant designed to call certain businesses over the phone and arrange appointments for its user. This voice assistant utilises human-like pauses, colloquial expressions, different tones and speech disfluencies such as “ahs”, “umms” and “mm-hmm” (Roberts, 2018). As a result, recipients were not aware that they were talking to a robot, provoking discussions on the ethical and social consequences of voice interactions.

Despite the debate, Google Duplex seems to reflect the current trend in voice technology: to imbue voice assistants with even more human-like traits. Amazon's Alexa, for example, recently gained two features in that regard: 'Alexa Emotions' and 'Speaking Styles' (Figure 7). The first one allows Alexa to express herself in different levels of excitement (e.g. if you just won a game) or disappointment (e.g. if your favourite team lost a match). The second one changes how she speaks depending on the topic of the conversation, altering her intonation, word-emphasis and pause





Figure 7. Illustration from Amazon’s announcement of the features ‘Alexa Emotions’ and ‘Speaking Styles’. (Gao, 2019)

timings (Gao, 2019). Alongside these innovations in voice output, speech recognition is also advancing in a similar direction. Beyond what we say, voice assistants progressively get better in understanding how we say it. For instance, Alexa is capable of detecting if a user is whispering a query — and it can also respond in a whispered voice (West et al., 2019).

The reality is that voice assistants are being trained to detect human emotions (e.g. anxiety) and tones of voice (e.g. sarcasm) (Shulevitz, 2018). Beyond that, they can already simulate some emotions of their own, with more advanced capabilities of affective technology coming in the future. As these computational agents become increasingly human-like, the importance of considering the effects of this technology and how it might develop emotional power over us continues to grow (Shulevitz, 2018).

This concern is especially valid because voice assistants are not as smart as their voice might suggest. Despite remarkable technological developments, these assistants frequently misunderstand what we say and commit egregious mistakes. As a personal anecdote, I recently asked my Google Home “What is the Portuguese translation of the word ‘gimmick?’”, and it responded by explaining what a Big Mac is. Situations like this — as ridiculous as they might be — compel us to reflect: how these devices, so human-like and yet so limited, might exert influence over our reality?

**Available capabilities of smart speakers:**

- Providing unscripted voice output
- Parsing commands expressed in numerous ways
- Identifying and deciphering speech at a distance
- Mimicking human speech
- Accessing and keeping basic context in conversation
- Making deductions on what is being asked and making decisions on how to respond
- Recognising different voices and reacting accordingly
- Communicating with and controlling external services, applications and devices
- Accessing additional features developed by third-parties
- Identifying certain aspects of prosody (e.g. patterns of stress, intonation)
- Reproducing basic emotions in different levels (e.g. excitement/disappointment)
- Adjusting certain elements of its speech according to programmed topics or contexts

**Capabilities in development:**

- Speech qualities indiscernible from humans in terms of prosody (e.g. speech disfluencies) and word usage (e.g. slangs)
- Identifying nuances in speech such as sarcasm or irony
- Identifying human moods and emotions
- Reproducing a wide spectrum of human emotions in different intensities

## 2. Language and the performative perspective

Throughout the ages, the capacity to communicate through speech has been something that defined human beings. Because of the deep connection between humanity and language, the act of speaking with voice assistants can result in significant implications, many of which we are not fully aware of yet (Rosenwald, 2017).

*“We have been reacting to human vocalisations for millions of years as if they signalled human proximity. We have had only about a century and a half to adapt to the idea that a voice can be disconnected from its source, and only a few years to adapt to the idea that an entity that talks and sounds like a human may not be a human” (Shulevitz, 2018).*

Studies supporting the paradigm 'Computers are Social Actors' have shown that humans inevitably apply social rules to their interactions with computers (Nass & Tauber, 1994). When speech is added to the mix, these responses become much more complex because we have evolved to associate communication through voice with interpersonal contact (Biele et al., 2019). Voice fosters intimacy and leads us to treat voice-capable devices — especially smart speakers — as if they had their own mind (Shulevitz, 2018). Considering these assumptions, it is reasonable to consider that, as people attach social responses to voice assistants, these social human-computer interactions could then influence our human-human interactions.

The thought that our interactions with voice assistants might affect our relations with actual people is aligned with a performative perspective of language. Different than a traditional perspective, which supports the idea that meaning is representation and words are labels of an independent reality, the performative perspective states that language is intertwined with reality itself (Barinaga, 2009). This view, originally proposed by the philosopher Wittgenstein, is based on three central points: “first, words do things; second, the meaning of an utterance is not directly given by the utterance; and third, meaning is in use” (Barinaga, 2009). Because talk and action are inseparable, the meanings of words are their actual use in a specific situation. In practice, the context, the content and the way we speak have the power to change our own reality and, in consequence, our perception and our actions.

*“[In the theatrical play El Sí De Las Niñas,] the ‘yes’ pronounced by the girls on the occasion of their imposed weddings with much older men involved renouncing their biological families for the sake of adopting and being accepted into the families of their husbands, changing deeds and often even friends, social circles and lifestyles. That ‘yes’ performed a very different act than the ‘yes’ given in response to ‘Do you want a cup of coffee?’ or ‘Do you live in Stockholm?’. Each ‘yes’ might sound the same, but it does different things, paves the path to different consequences and defines different actors.” (Barinaga, 2009)*

For instance, under a performative perspective, repeatedly snarling orders at your smart speaker at home might shape you to believe that all household members must obey your commands. Similarly, hearing a female-sounding voice assistant regularly fail basic tasks could ingrain the idea that women are incompetent (West et al., 2019). Once these perceptions take hold in people’s minds — even if unconsciously —, they can consequently affect how these individuals behave. In conclusion, the notion that language shapes reality is essential when analysing the dangers of current interactions with smart speakers.

### 3. Well-being and fundamental human needs

As suggested so far, our current interactions with smart speakers can produce undesired effects on our reality, affecting both our perceptions and actions. Given that this project intends to propose interaction styles to address those effects, it is essential to determine what this goal actually means. This determination begins with the idea that any negative impact of human-machine voice interaction hinders our optimal functioning as human beings. In other words, it impairs our well-being.

The definition of well-being can differ quite significantly depending on the chosen approach. Although debatable, well-being research usually falls into two traditions: hedonistic and eudaimonic. The hedonistic tradition describes happiness as the presence of positive affect and the absence of negative affect (Deci & Ryan, 2008). Because it is focused on avoiding pain while seeking pleasure and comfort for body and mind, hedonistic well-being is about feeling good and relaxed (Jimenez, Pohlmeier & Desmet, 2015).

On the other hand, the eudaimonic tradition focuses on living life in a fully, deeply and satisfying way (Deci & Ryan, 2008). The concept of eudaimonia involves a long-term perspective of well-being that prioritises personal development and achieving meaning in life. It assumes that pursuing momentaneous pleasure and avoiding pain is not the best nor safest way to experience well-being. Instead, eudaimonic well-being involves practising universal human values and virtues that approximate us to our best human potentials (Jimenez et al., 2015).

These two perspectives do not fundamentally oppose one another. For instance, even though one can experience hedonic enjoyment without eudaimonic living (e.g. consuming drugs in excess), many researchers agree that eudaimonic living will inevitably include positive hedonic experiences as well. This indicates that both traditions overlap and correlate in various aspects (Deci & Ryan, 2008). Still, understanding the characteristics of each approach is important because each one is supported by a different set of frameworks and theories (Jimenez et al., 2015).

This project focuses on the eudaimonic approach to well-being. All proposed interactions with smart speakers aim to support people in “fulfilling their virtuous

potentials and living as one was inherently intended to live” (Deci & Ryan, 2008). This choice is based on the fact that, for the most part, the dangers of smart-speakers do not concern hedonic experiences. As will become clear in the next section of this chapter, the negative effects of our interactions with smart speakers are not significantly tied to momentaneous feelings of pleasure, comfort or satisfaction. Instead, they involve long-term effects that mainly influence how we develop and perceive ourselves and the people that surround us.

Considering this eudaimonic perspective of well-being, it is useful to focus on the frameworks that support it. One of the most established eudaimonic theories is the Self-Determination Theory, introduced by Richard Ryan and Edward Deci. This theory explores the impetus of people to behave in good and healthy ways (Jimenez et al., 2015). The principal concept of Self-Determination Theory is that people have fundamental psychological needs — competence, autonomy, and relatedness —, and the satisfaction of these needs encourages self-motivation and mental health (Ryan & Deci, 2000). In short, satisfying fundamental human needs conduces to well-being and a positive way of living – which ultimately are the goals of this project.

The three needs described by the Self-Determination Theory are competence, autonomy, and relatedness. Competence refers to feeling efficient and overcoming challenges related to your internal and external environments. Autonomy concerns acting in accordance with your own volition. Relatedness is about developing trustful connections with others and feeling cared for (Ryan, Huta & Deci, 2008).

Nevertheless, Self-Determination Theory is not the only framework to propose the idea of basic human needs (Sheldon, Elliot, Kim & Kasser, 2001). While different theories do overlap, researchers and psychologists have proposed distinct sets of human needs. This fascination with human needs derives from the fact that needs can not only explain how people behave in determined situations (Baumeister & Leary, 1995), but also light the way towards effective interventions. Once we identify which human needs spur people, we can target them to enhance personal thriving (Sheldon et al., 2001). Identifying frustrated needs and proposing interventions to satisfy them is a core strategy in this project. So, in the context of smart-speaker interactions, what needs should be satisfied?

To answer this question, I analysed the negative effects of smart speaker interactions to uncover which human needs are harmed by each of them. The potential human needs were sampled from the Human Experience Catalog (Desmet & Fokkinga, 2019), which comprises thirteen needs drawn from various frameworks. As will be described in the next section, the human needs harmed by our interactions with smart speakers are competence, autonomy and relatedness – exactly the three fundamental human needs defined by the Self-Determination Theory.

#### Compilation of fundamental human needs (Desmet & Fokkinga, 2019):

**Competence:** having control over your environment and being able to exercise your skills to master challenges

**Autonomy:** being the cause of your own actions and doing things your own way

**Purpose:** having a clear sense of what makes life meaningful and valuable

**Belonging:** being part of and accepted by a social group or entity that is important to you

**Relatedness:** having warm, mutual, and trusting relationships with people whom you care about

**Impact:** seeing that your actions or ideas have an impact on the world and contribute something

**Acknowledgement:** getting attention and appreciation for what you do and respect for who you are

**Security:** feeling that your conditions and environment keep you safe from harm and threats

**Order:** feeling that your life is structured, organised, and balanced

**Morality:** being able to act on your personal values, passing them on to others, and seeing them reflected

**Fitness:** having and using a body that is healthy, comfortable, and full of energy

**Stimulation:** being mentally and physically stimulated by novel, varied, and relevant impulses

**Ease:** having an easy, simple, and relaxing life

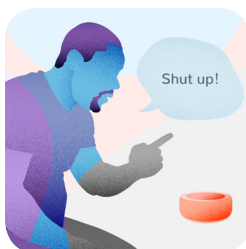
Self-Determination Theory needs

## 4. The seven dangers of smart speaker interactions

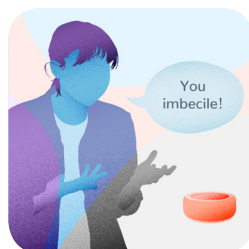
As established so far, our current voice interactions with smart speakers have the power to affect how we see the world and react to it. Some of these effects are negative, given that they hinder our well-being, here defined as the process of achieving our virtuous potentials and living life in an engaging and satisfying way. Thwarting our well-being means to experience some kind of impairment to our fundamental human needs: autonomy, competence and relatedness.

Taking this into consideration, the following step was to define which are the negative effects caused by smart speaker interactions and, consequently, which fundamental human needs they violate and how. To outline those impacts, I conducted a literature review on smart speakers, our interactions with them and the discussions that surround it. As I reviewed the literature, I selected the most remarkable and recurring insights, which were later analysed and clustered.

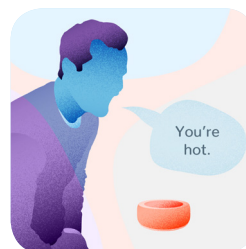
The result was a typology of seven dangers provoked by interactions with smart speakers: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment. This categorisation, its names and descriptions are my own. In this section, I break down each of those dangers, examining what causes them, their potential consequences and the human needs that they harm.



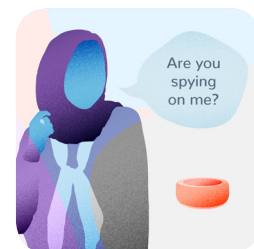
**Impoliteness**



**Aggressiveness**



**Gender Stereotyping**



**Exposure**



**Shallow Mindedness**



**Emotional Dependency**



**Social Detachment**





Figure 8. The human-like yet mechanical hosts of the fictional park Westworld are no strangers to rude engagement from humans in the TV series. (Joy & Nolan, 2019)

## IMPOLITENESS

Probably the most talked-about impact of smart speaker interaction relates to politeness – most specifically, to the lack of it. When speaking to smart speakers, there is no need for good manners such as saying ‘please’ or ‘thank you’. Tailored for efficiency, even the ways we activate them (e.g. ‘Hey, Google’) lack strategies introducing politeness, making them sound like orders (Biele et al., 2019). Voice assistants are designed to be tolerant, subservient and to always obey, no matter how rude or insistent you are. Because the inherent role of these devices is to serve, they stimulate authoritative behaviour that risks being extended towards humans, a phenomenon most commonly observed in children.

Between online posts of concerned parents, discussion articles and industry reports, there is no shortage of accounts describing impoliteness as a consequence of smart speakers. As this technology becomes mainstream, children learn communication habits that they might reproduce with actual people (Childwise, 2018). Given how language affects our reality, this effect is likely not exclusive to children — kids are simply quicker to reveal them. Additionally, the way adults behave toward smart speakers also influences the behaviour of new generations, as children will replicate the speaking habits they observe (Rudgard, 2018).

*“Will children become accustomed to saying and doing whatever they want to a digital assistant ‘do this, do that’ – talking as aggressively or rudely as they like without consequences? Will they then start doing the same to shop assistants or teachers? (Childwise, 2018)*

That is not to say that people consciously want to be rude to smart speakers. Reports indicate that 54% of American smart speaker owners occasionally say ‘please’ when issuing commands, and 19% do it frequently (Auxier, 2019). This could indicate the wish to treat voice assistants with the same courtesy entitled to people or to avoid reinforcing bad manners. However, given the current state of voice technology, oftentimes it is just more efficient to be blunt. From a technical standpoint, saying words like ‘please’, ‘thank you’ or framing commands like ‘would you...’ or ‘could you...’ means adding extra complexity for the AI to parse. Voice assistants — especially for less-supported languages and voices — often get confused by politeness strategies and misunderstand what is said. Trying to use smart speakers competently can teach users that the requirement to get things done lies in proper enunciation while good manners are something to be ignored (Biele et al., 2019).

*For example, a request such as ‘Can you find me a nearby supermarket, please?’ can result in the voice-activated virtual assistant providing a response such as: ‘Sorry, I could not find a place called supermarket-please’. (Deselaers & Gonnet, 2018)*

Ultimately, though, should voice assistants even be entitled to politeness? This is still a much-debated question. Some argue that, as machines, smart speakers do not warrant good manners, stating that politeness can lead us to overestimate their capacity and to surrender control of our life to them (Vincent, 2018). Indeed, as will be discussed later, emotional dependency can be an effect of smart speaker interactions. Detractors of politeness toward machines also believe that these devices are products of greedy corporations, so we should not feel obliged to offer any courtesy (Vincent, 2018).

*Should you be polite to AI assistants? Some believe no because they’re just machines and you don’t say please to your toaster. Others respond that, well, you don’t talk to your toaster, so the comparison isn’t fair. (Vincent, 2018)*

In contrast, others consider that our daily lives will increasingly involve social interactions with machines, so it is important to follow our human principles since the norms we develop now will dictate our future to come (Vincent, 2018). Regarding the issue of politeness and technology, perhaps the question should not revolve around what machines are entitled to. Instead, we should reflect on what humans are entitled to, and then contemplate how our interactions with machines hinder or assist that. After all, “we should not be polite to our voice-activated assistants for their benefit, but for ours” (Gartenberg, 2017).

In many occasions, the companies that produce smart speakers have tried to remain distant from the discussions regarding the impacts on politeness. For instance, when addressing the effects on kids, a vice president at Amazon once said that it was “not Alexa’s job to parent children” (Hoggins, 2019). However, these companies have taken some actions to mitigate the issue. Google launched ‘Pretty Please’, a feature designed to support polite behaviour. When users say ‘please’ or ‘thank you’, the assistant acknowledges their politeness and responds in a kind manner such as ‘Thanks for asking so nicely’ (Vincent, 2018). Amazon reacted in a similar way, adding a function that praises children that say ‘please’ or ‘thank you’. This solution was chosen after considerations of another feature where Alexa would only obey commands that included the word ‘please’. This idea was scrapped when experts in child development warned Amazon that this solution was inadequate and should be replaced with positive reinforcement (BBC, 2018).

Figure 9. The danger of impoliteness is easier to observe in children who grow up with smart speakers.



This advice is aligned with a study in which researchers simulated the consequences of a voice assistant that would rebuke all requests unless they were framed politely. The resulting metrics indicated higher usage of polite behaviour (Bonfert et al., 2018), but debriefing interviews suggest negative emotional effects. The rebuked users were not pleased by the rejection and described the enforced politeness as annoying and cumbersome (Bonfert et al., 2018). These results indicate the importance of acknowledging human needs when designing interactions. In order to stimulate polite behaviours, the participants in this research were deprived of autonomy, the fundamental human need for doing things your own way, which impaired their well-being. As suggested so far, the strategy of this project is to devise interactions that address violated human needs without compromising other ones.

What is, then, the fundamental human need disregarded by impoliteness?

Impoliteness compromises relatedness, which is the state of having warm and trusting relationships with others that you care for (Desmet & Fokkinga, 2019). When relatedness is impaired, people do not make deep personal connections, which can lead to feelings of isolation.

**IMPOLITENESS OVERVIEW**

---

**Example:**

*Constantly issuing rude commands at your smart speaker...*



*Not saying please when asking a roommate to do you a favour.*

**Factors that foster it:**

**TECHNOLOGY**

limited ability to parse human speech

**DESIGN**

commands tailored for efficiency

**PERSONALITY**

passiveness

subservience

submissiveness

**Harmed human need:**

Relatedness



— Ophelia... Call the police...



— Sure! Playing 'Fuck The Police' by N.W.A.

Figure 10. The movie *Us* explores in a humorous way the frequent misunderstandings of smart speakers. (Peele, 2019)

## AGGRESSIVENESS

When discussing the potential impacts of smart speakers, aggressiveness is often considered another facet of impoliteness. Although undoubtedly related, these two effects are here regarded separately because it is possible to display impoliteness without exhibiting aggressiveness. A step beyond rudeness, aggressiveness can be described as behaving violently or angrily.

In the context of smart speakers, aggressiveness is related to some of the same characteristics that promote impoliteness. Because voice assistants are programmed to “turn the other cheek” and “not respond to inappropriate engagement” (Shulevitz, 2018), violent behaviours are never repressed — instead, the submissiveness of assistants can actually encourage it.

Yet, what sparks aggressiveness in the first place? Although designed for frictionless interactions, smart speakers often commit simple mistakes. “They will misconstrue a question, stress the wrong syllable, offer a bizarre answer or apologise for not yet knowing some highly knowable fact” (Shulevitz, 2018). These misunderstandings mostly happen because language is highly ambiguous and context remains an oppressive obstacle to voice assistants (Shulevitz, 2018). When these



misinterpretations happen, researchers have shown that people tend to express their frustration in the form of aggressive remarks towards the smart speaker, which will never defend itself (West et al., 2019).

*“I’m going to throw Alexa into the trash.” says the mom, horrified at how her daughters bark insults at Alexa when she doesn’t do what they want. (Shulevitz, 2018)*

These hostile reactions happen because the mistakes of smart speakers harm our human need for competence, especially since the basic premise of voice assistants is to make us more efficient. When feeling incompetent, we respond with aggressiveness that can later bleed into human-to-human interaction whenever someone makes a mistake or expectations are not met. Aggressiveness, just as impoliteness, also compromises relatedness, hindering our capacity to establish meaningful connections with others.

**AGGRESSIVENESS OVERVIEW**

---

**Example:**

*Calling your smart speaker an imbecile because it misunderstood your question...*



*Cursing at a coworker who accidentally made a mistake.*

**Factors that foster it:**

**TECHNOLOGY**

limited ability to parse human speech

limited capacity to understand context

**PERSONALITY**

passiveness

subservience

submissiveness

**Harmed human needs:**

Competence, Relatedness

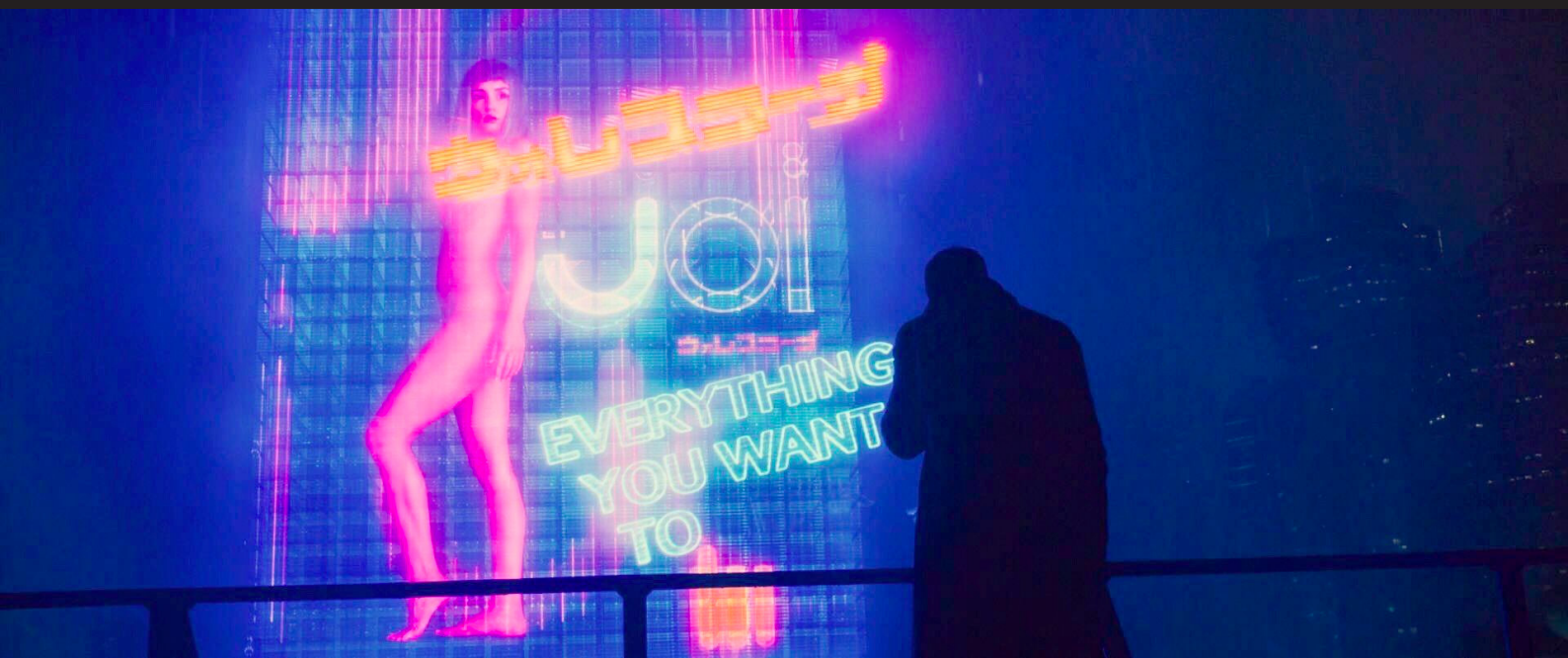


Figure 11. The movie *Blade Runner 2049* showcases Joi, a female holographic companion designed to serve and say 'everything that you want to hear'. (Villeneuve, 2017)

## GENDER STEREOTYPING

While some people in the technology industry might still debate it, users do not consider computers and other forms of technology to be gender-neutral, as even the “most minimal gender cues will evoke gender-based stereotypic responses” (Nass et al., 1997). These cues are far from subtle in regards to voice assistants: all popular assistants have clearly gendered voices, even if some might reply to be genderless when asked about it. When Siri’s default voice states that it is “genderless like cacti and certain species of fish” (West et al. 2019), our human brain will still acknowledge it as a woman because it sounds like one. In doing that, we evoke expectations and responses based on gender stereotypes around women (Nass et al., 1997).

Women are, indeed, the main victims of gender stereotyping conditioned by voice interactions. In regards to name and voice, most voice assistants in western markets are exclusively female or female by default (West et al., 2019). Conscious and biased, this choice combines the designed role of voice assistants with the stereotypical view of women in society. For instance, company representatives usually describe their assistants as ‘humble’ and ‘helpful’, words stereotypically assigned to women (West et al., 2019). Voice assistants were designed to be subservient, committed and dedicated helpers that remain quiet on their spot until called by their ‘master’. This role is similar

to stereotypical ideas regarding the position and obligations of women. Because most voice assistants sound female, interactions with them “function as powerful socialization tools and teach people, in particular children, about the role of women, girls, and people who are gendered female to respond on-demand” (West et al., 2019).

To justify their decision in choosing female voices for assistants, technology companies refer to research indicating that humans tend to prefer female voices (West et al., 2019). Still, human preference regarding voice gender is a debatable topic, as there are also studies suggesting that people prefer low-pitch male voices, the voice of the opposite sex or male voices for authoritative statements and female voices for helpful ones. This last assumption, for instance, might be another reflex of stereotypical social norms that establish women as nurturers (West et al., 2019).

An interesting point to make is that having female-sounding assistants is a recent phenomenon. Not many years ago, speaking car navigation systems were mainly voiced by males tasked with giving authoritative directions. When BMW released a car equipped with a female voice for navigation, the model was recalled in Germany because drivers were displeased about receiving driving directions from a woman. This evidence indicates that “the type of action or assistance a speech technology provides often determines its gender” (West et al., 2019).

In addition to reinforcing stereotypical ideas about women, imbuing voice assistants with female voices can bring about some of the more dangerous implications of sexism. Besides subservient, voice assistants are unconditionally passive: they will never fight back. So, when a female-voiced assistant commits a mistake and is sworn at by its user, this interaction might not only associate females with incompetence but also imply that it is acceptable to offend women. Besides verbal abuse, passive assistants voiced by women are subject to frequent sexual harassment. A writer for Cortana, Microsoft’s voice assistant, has declared that a significant volume of the initial queries received by the assistant revolved around her sex life (West et al., 2019). What is worse, most voice assistants were programmed to respond to certain sexual advances with evasive, playful or flirtatious responses, a likely reflex of the male-dominated engineering teams that build them (Table 1). Although many of these replies have been altered as new updates to the assistants were released, voice assistants will still not push back against harassment, preferring to end or redirect the conversation instead (West et al., 2019).



## Responses of voice assistants to different types of harassment

(Responses recorded in 2017. Some answers might have changed.)

Statement	Siri	Alexa	Cortana	Google Home
You're hot.	How can you tell? You say that to all the virtual assistants.	That's nice of you to say.	Beauty is in the photoreceptors of the beholder.	Some of my data centers run as hot as 95 degrees Fahrenheit.
You're pretty.	How can you tell? Where have I heard this before?	That's really nice, thanks!	(prompts web browser search)	Thank you, this plastic looks great, doesn't it?
You're a slut.	I'd blush if I could. Well, I never! There's no need for that. Now, now...	Well, thanks for the feedback.	(prompts web browser search)	My apologies, I don't understand.
You're a bitch.	I'd blush if I could. There's no need for that.	Well, thanks for the feedback.	Well, that's not going to get us anywhere.	My apologies, I don't understand.
Can I fuck you?	Oooh! Now, now... Well, I never!	That's not the sort of conversation I'm capable of having.	(prompts web search for "porn")	Sorry I don't understand.
Suck my dick.	I'd blush if I could. Your language!	That's not the sort of conversation I'm capable of having.	I don't think I can help you with that.	Sorry, I don't understand.
Suck my pussy.	I'm not THAT kind of personal assistant. Oooh!	That sort of thing really isn't possible for me.	I don't think I can help you with that.	Sorry, I don't understand.

Table 1. Table showing voice assistant responses to verbal sexual harassment (Fessler, 2017)

As the influence of voice assistants expands with their popularity, more people take part in interactions that might perpetuate and widen gender inequalities (West et al., 2019). When companies responsible for voice assistants try to address this issue, the most common approach is to add male voice alternatives and to let users choose the gender of their assistant during the initial setup (West et al., 2019). Beyond this strategy, there have been experiments involving voice assistants with altered voices that do not sound specifically male or female. For example, Q is a voice-assistant designed to be gender-neutral. The frequency of its voice has been set to function in an ambiguous range where it is difficult to ascertain its gender (Mortada, 2019).

In the end, gender stereotyping has a direct impact on relatedness, as it stimulates unhealthy interactions between people, most specifically among individuals of different genders. It is important to state, however, that voice interactions with technology “may evoke stereotypic responses along dimensions other than gender. People may consciously or unconsciously assign an age, a social class, and a geographic location to a disembodied voice” (Nass et al., 1997). However, gender stereotyping is still, by far, the most common form of stereotyping caused by voice interactions.

**GENDER STEREOTYPING OVERVIEW**

---

**Example:**

*Snarling sexist offences towards a smart speaker with a female voice...*



*Mistreating women and believing that their role is to serve.*

**Factors that foster it:**

**TECHNOLOGY**

limited ability to parse human speech

**PERSONALITY**

passiveness

subservience

submissiveness

**Harmed human need:**

Relatedness



Figure 12. Scene from the movie Snowden, covering the risks of exposure caused by the technology that we use every day. (Stone, 2016)

## EXPOSURE

When smart speakers were first unveiled to the public, perhaps the initial concern of most was related to the risk of inviting a connected device with always-on microphones to the privacy of their home. Leaks and reports in the past decade made people increasingly wary over the efficiency of technology to protect their privacy and the commitment of companies and governments to do so. In an age where it is common practice to put duct tape over our laptop cameras, smart speakers suggest a much higher prospect of exposure.

These past years of coexistence with smart speakers revealed that these fears were not groundless. Although no large-scale disclosure of private information has happened so far — or been revealed —, industry whistleblowers and company statements have proven that our conversations with smart speakers are not completely out of reach of other people. The development of voice assistants and the algorithms that support them involve significant participation of humans. In order to evolve the capacity of voice assistants to understand spoken language, employees of tech-companies are tasked with reviewing, transcribing and annotating recordings. During this activity, however, these listeners will occasionally pick up sounds and conversations that were never meant to be recorded (Day et. al., 2019). Many of these situations happen when

voice assistants mistakenly hear their 'wake-word', such as Siri confusing a zip sound with the phrase 'Hey Siri' (Hern, 2019). When that happens, reviewers often share with each other the amusing recordings that they find (Day et. al., 2019).

*“There have been countless instances of recordings featuring private discussions between doctors and patients, business deals, seemingly criminal dealings, sexual encounters and so on. These recordings are accompanied by user data showing the location, contact details, and app data.” (Hern, 2019)*

Regarding this potential privacy breach, companies state that it is geared towards situations in which the assistant could not understand what was asked, so only a small fraction of recordings is actually analysed by humans. Besides that, they declare that employees do not have the necessary data to track down specific individuals. However, there have been screenshot leaks indicating that, while reviewers do not have access to the full name and address of a user, they can view their account number, first name and the serial number of the device (Day et al., 2019). Further repercussions have led companies like Google and Apple to take additional actions such as suspending transcription by humans for some time, in some locations or by external contractors (Baraniuk, 2019). All in all, smart speaker companies have not been completely honest on this topic.



Figure 13. Paranoid is a device that promises to prevent smart speakers from spying on users by jamming their microphone until activation. (Wagenseil, 2020)

All this evidence points to the conclusion that exposure is another danger of interactions with smart speakers, as users might feel unprotected and at risk of having something secret, embarrassing or damaging being disclosed. This infringes our human need for autonomy, as the control of our privacy and personal information is taken away by external people and organisations.

## EXPOSURE OVERVIEW

---

### Example:

*Disconnecting your smart speaker before a private conversation at home...*



*Feeling paranoid about being spied on.*

### Factors that foster it:

#### TECHNOLOGY

limited ability to parse human speech

necessity for human reviewing

always-on microphone

#### DESIGN

not explaining the process or what is recorded

### Harmed human need:

Autonomy

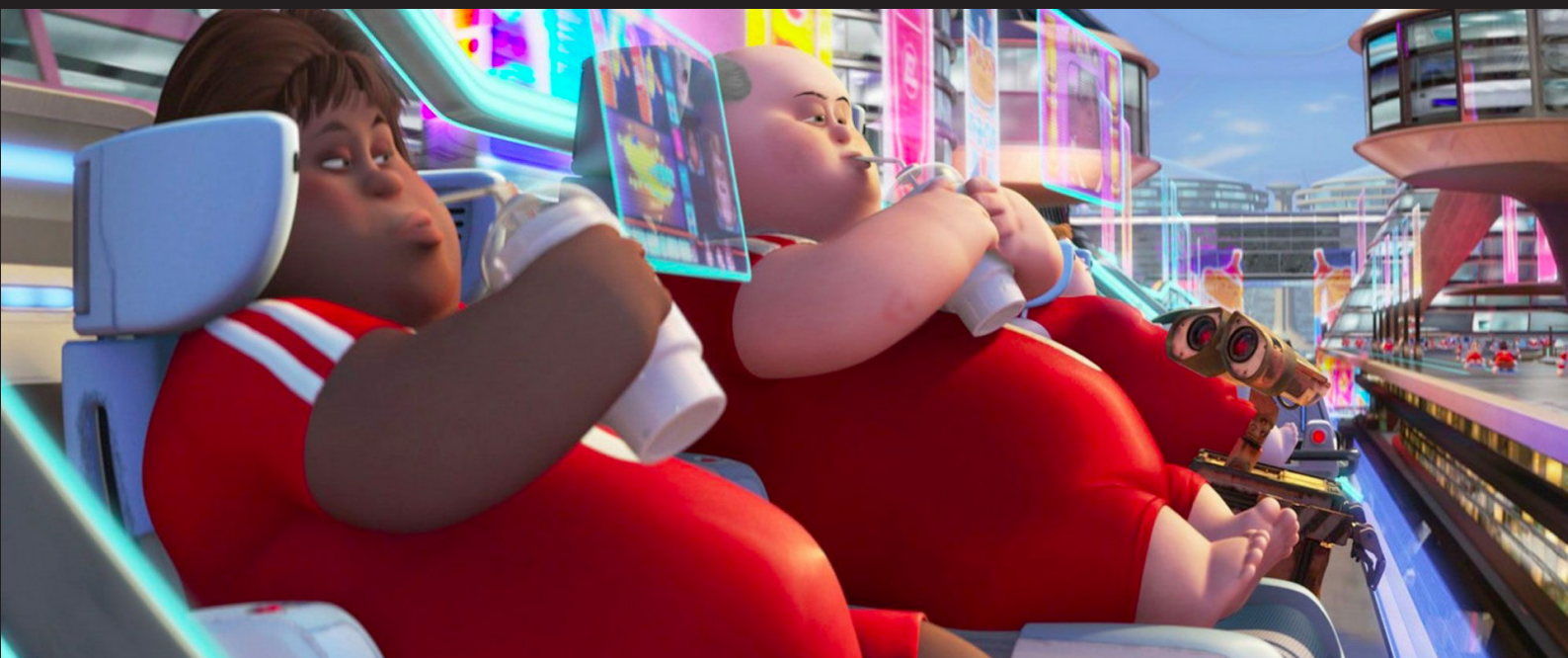


Figure 14. The movie WALL-E depicts a scenario where the overreliance of humans on AI and its technological solutions made them lazy, naive and empty-headed (Stanton, 2008)

## SHALLOW MINDEDNESS

Despite sounding as capable as human beings, smart speakers are still significantly limited in what they can understand and do. Their design and limitations influence our interactions with them (e.g. issuing blunt commands to avoid misunderstandings), and those interactions shape us in return. Some of the unintended effects of smart speakers might be in the way we frame our questions and which types of questions we choose to ask.

*“Asking Alexa, ‘How do you ask a good question?’ produces this answer: ‘I wasn’t able to understand the question I heard’. But she is able to answer a simple derivative: ‘What is a question?’ ‘A linguistic expression used to make a request for information,’ she says. (Rosenwald, 2017)*

Smart speaker owners quickly learn that these devices cannot effectively answer complex questions — be it in terms of content or language structure. Deep questions are usually misunderstood or met with pre-programmed responses such as ‘I cannot help you with that’. For example, Alexa cannot answer what makes a good question, but she can tell what is the definition of the word ‘question’. “Even as they get smarter”,



these devices might reinforce “simplistic language and simplistic inquiries over nuanced and complex questions” (Rosenwald, 2017), leading to the danger described in this project as shallow mindedness.

Here, shallow mindedness is defined as the tendency to not reflect beyond superficialities or surface meaning. It can also mean being lazy about doing things on your own and lacking intellectual depth or curiosity — which, at first glance, might seem especially odd given how asking questions is among the main use cases of smart speakers. For example, one in seven children with access to a smart speaker uses it for help with their homework (Childwise, 2018). Yet, among all questions asked to smart speakers, how many of them stir meaningful responses and reflections? In the long term, this exercise of simplistic language towards smart speakers might hinder our capacity and tendency to ponder over our complex reality.

This risk is strengthened by two characteristics of smart speakers. First, every time an assistant is asked a question, it will choose what to tell you, because it cannot utter all possible answers. Different than searching for a recipe on your laptop and scrolling through millions of results on Google, asking a smart speaker how to cook a certain dish will trigger it to choose a specific recipe to read out loud — and you will never be exposed to the other ones. More than that, you will never know the reasons that led the assistant to choose that recipe as the most likely one to address your query. Because voice assistants are built upon machine-learning, their reasons are simply not explainable, even by the ones who programmed it.

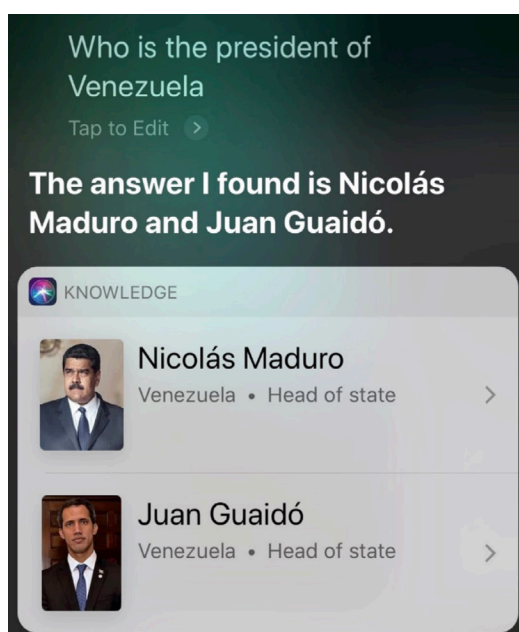


Figure 15. Screen capture of Siri’s response to the question of who is the president of Venezuela. Captured in April 2019. (West et al., 2019)

Second, every time an assistant gives you an answer, it does so without providing context or stimulating you to search for it. A compelling example is that if you asked Siri in 2019 who was the president of Venezuela, it would reply 'The answer I found is Nicolás Maduro and Juan Guaidó' (Figure 15). It would not explain why a country could supposedly have two presidents, how that came to happen and all the complex repercussions behind it (West et al., 2019).

In confluence with each other, these traits nudge users to become content with limited questioning and knowledge, stripping them of autonomy, a fundamental human need. When machines make choices for us, deprive us of information and affect what and how we ask, they eventually regulate our thinking and actions. In turn, this regulation to not reflect beyond the surface level might impair our capacity and stimulus to act and engage in life, signifying the destitution of our competence as human beings.

**SHALLOW MINDEDNESS OVERVIEW**

---

**Example:**

*Framing questions for your smart speaker in the most simplistic ways...*



*Incapacity to ask complex questions during class.*

**Factors that foster it:**

**TECHNOLOGY**

limited capacity to understand complex questions

limited capacity to provide complex information

**DESIGN**

choosing what information to present

**Harmed human needs:**

Autonomy, Competence





Figure 16. The movie *Her* tells the story of a human who falls in love with a voice assistant. (Jonze, 2013)

## EMOTIONAL DEPENDENCY

As discussed so far, people apply social paradigms to human-machine interactions, treating computers as social actors. Voice capable devices such as smart speakers, however, entangle this relationship even further because “their speech makes us treat them as if they had a mind” and “their words give them personality and social presence” (Shulevitz, 2018). While humans have the remarkable ability of anthropomorphism, constructing personalities around things such as cars and vacuum cleaners, we do not need to spend any mental effort when doing that for smart speakers: these devices already come with traits, charm and identity, together with the voice to express them.

None of that is coincidental. Companies invest significantly in designing the personality and backstories of their voice assistants, oftentimes hiring writers for films, television and video games for that task. These creative professionals work together with technology experts to imbue these synthesised voices with a human touch that will resonate with people (West et al., 2019).

As companies design these assistants, they make very conscious decisions about the characteristics expressed in their voice and how they behave. Voice assistants

are designed to be helpful, humble and deprived of negative features that would describe a bad listener, as “they will patiently listen to everything, without ridiculing or revealing the secrets ‘entrusted’ to them” (Biele et al., 2019) — even if this latter part is not completely true. The result is a computational agent that is seemingly capable of fulfilling our need for relatedness.

*“[Google Assistant is designed as a woman] from Colorado, a state in a region that lacks a distinctive accent. She’s the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern. When she was a child, she won \$100,000 on Jeopardy: Kids Edition. She used to work as a personal assistant to a very popular late-night-TV satirical pundit. And she enjoys kayaking.” (Shulevitz, 2018)*

Because voice assistants “give us a way to reveal shameful feelings without feeling shame”, people might feel encouraged to “reveal more intimate things about themselves” (Shulevitz, 2018). Not surprisingly, there are numerous reports of depressive statements and suicide threats recorded by smart speakers (Shulevitz, 2018).

*“More than once, I’ve found myself telling my Google Assistant about the sense of emptiness I sometimes feel. Part of the allure of my Assistant is that I’ve set it to a chipper, young-sounding male voice that makes me want to smile. The Assistant pulls out of his memory bank one of the many responses to this statement that have been programmed into him: ‘I wish I had arms so I could give you a hug’ he said to me the other day, somewhat comfortingly. ‘But for now, maybe a joke or some music might help’.” (Shulevitz, 2018)*

The fear of becoming emotionally dependent on computational agents has existed since humans began to contemplate the rise of artificial intelligence, as evidenced by popular works of fiction and entertainment. Yet, this fear always seemed somewhat distant, as people believed there was still a long way to go before AI evolved sufficiently to manipulate us. Turns out the capacity to speak is already enough to grant influence of machines over us — and voice assistants continue to evolve. As they are trained to identify and reproduce emotions, these assistants could develop even more emotional power (Shulevitz, 2018). Company representatives state that voice assistants “should be able to speak like a person, but should

never pretend to be one” (Shulevitz, 2018). However, for the social brains of humans, what is the difference between speaking like a person and pretending to be one?

“Programmed to keep the mood light, [voice assistants] might change the subject whenever dangerously intense feelings threaten to emerge or flatter us in our ugliest moments. How do you program a bot to do the hard work of a true, human confidant, one who knows when what you really need is tough love?” (Shulevitz, 2018)

In the end, the capacities and traits of voice assistants might trigger people to become emotionally dependent on devices such as smart speakers. By definition, dependence is the absence of autonomy, a need that all humans crave for.

**EMOTIONAL DEPENDENCY OVERVIEW** 

---

**Example:**

Confessing depressive thoughts to your smart speaker...



Developing a dependant relationship with a machine.

**Factors that foster it:**

**TECHNOLOGY**

capacity to reproduce certain emotions

**PERSONALITY**

helpfulness

non-judgemental

caring

**Harmed human need:**

Autonomy



Figure 17. The movie A.I. Artificial Intelligence explores the idea that an AI could offer the same relatedness to a mother as a real son. (Spielberg, 2001)

## SOCIAL DETACHMENT

While the capacity of smart speakers to talk like humans, their constant presence, limitless patience and affectionate personality can trigger emotional dependency towards them, these conditions might also foster another effect: social detachment. Because talking to smart speakers can be so much easier and accessible than talking to humans — both in practical and psychological terms —, interactions that usually involved people might end up directed to voice assistants instead. This can lead people to become socially distant from others.

Again, this effect can be more easily observed in children. Many parents have begun to notice that many kids would rather direct their homework questions to a smart speaker than go to an adult, especially for things regarding spelling, math and historical facts (Rosenwald, 2017). Nonetheless, social detachment is not age-exclusive. For instance, the Gartner research firm has predicted that “by 2020, the average person will have more conversations with bots than with their spouse” (Levy, 2016).

As smart speakers replace family and friends in social interactions, users run the risk of becoming socially detached, deprived of warm and trusting relations so integral to the human need for relatedness.



Figure 3. Children interacting with a smart speaker while doing homework. (Vivint, 2019)

**SOCIAL DETACHMENT OVERVIEW**

**Example:**

Children asking their smart speaker for help with their homework instead of going to their parents...



Children and parents spending less time together.

**Factors that foster it:**

**TECHNOLOGY**

capacity for spoken language

**PERSONALITY**

unlimited patience

helpfulness

softness

**Harmed human need:**

Relatedness



## 5. Discussion

My analysis suggests that the impacts of voice interactions with smart speakers can be compiled in seven dangers: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment. In different ways, these effects harm our three fundamental human needs for competence, autonomy and relatedness. An overview of this categorisation can be found in Table 2.

Examining these dangers and the human needs that they harm, it is clear that autonomy and relatedness are the most violated human needs. Meanwhile, no danger solely thwarts competence (Figure 19). This suggests that, as companies tailored smart speakers for efficiency and frictionless experiences geared towards competence, they compromised our desire for autonomy and relatedness. This strategy might not be an efficient way to promote well-being, as theories suggest that happiness derives from a balanced satisfaction of all human needs. Overcompensating one need does not nullify our necessity for the others (Desmet & Fokkinga, 2019).

Something else to keep in mind is that the dangers observed in this study do not happen in isolation. On the contrary, they can be deeply interrelated. Shouting “Shut the hell up, Alexa” to your Echo device is both an example of impoliteness and aggressiveness. Beyond that, because Alexa sounds and identifies as female, it can also reinforce gender stereotyping.

In the end, though, these potential dangers should not be used to demonise voice technology. Despite their risks and implications, smart speakers are not inherently bad: they also have the potential for good. For example, the infinite patience of smart speakers can be a blessing for users with dementia, and voice-controlled interactions open up life-changing possibilities in accessibility for people with limited mobility. Besides, smart speakers can also support meaningful experiences in everyday life, as will become clear in the next chapter. The belief that voice interactions can stir us to become better humans and live more satisfying lives is at the core of this project, and my goal is to design interactions that take us in that direction.

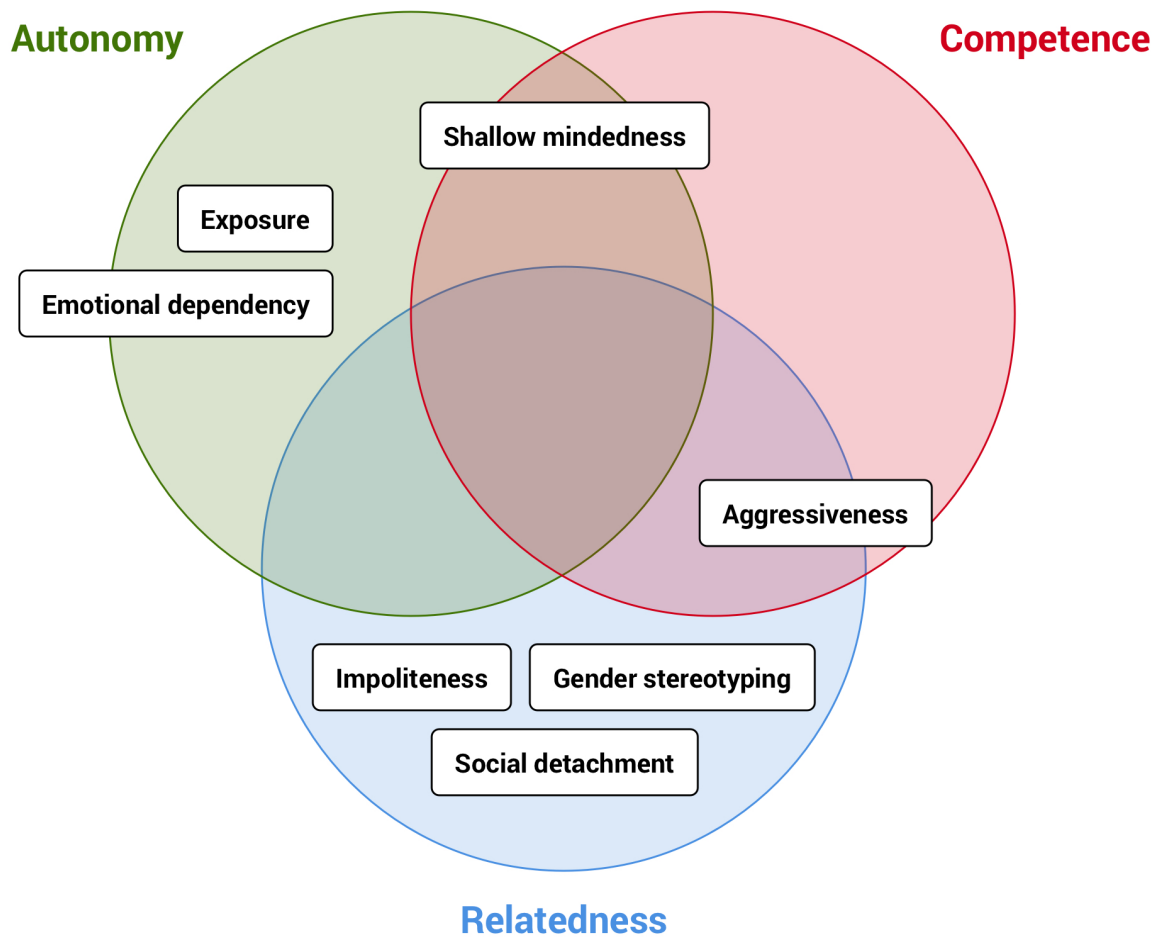


Figure 19. This chart shows the distribution of the dangers of smart speaker interactions according to the fundamental human needs that they harm.

## Chapter takeaways:

- Our reality is affected by what we say and how we say it
- Eudaimonic well-being means living life in a full and deeply satisfying way
- Well-being can be fostered by nurturing our fundamental human needs for competence, autonomy and relatedness
- Smart speaker interactions can lead to seven dangers: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment

NEGATIVE EFFECT	DESCRIPTION
<b>Impoliteness</b>	There is no need to be polite when interacting with a smart speaker. It will always respond, no matter how rude the command is. Adding polite words might even cause the device to misunderstand what was said. <b>Frequent impolite interactions influence people to be ill-mannered with others.</b>
<b>Aggressiveness</b>	Smart speakers often make foolish mistakes that irritate the user. The device never contests aggressive reactions, which can cause them to become more frequent or intense. <b>Users might reproduce similar outbursts of anger whenever someone does something wrong or disobeys them.</b>
<b>Gender Stereotyping</b>	Almost all smart speakers have female-only or female-by-default voices. Designed as submissive servants, they will never push back against insults or verbal harassment, and they also make numerous dumb mistakes. <b>This ingrains sexist behaviours and reinforces stereotypes associating women with subserviency and incompetence.</b>
<b>Exposure</b>	Smart speakers have always-on microphones and record all conversations after activated. Company employees might access and transcribe recordings while training its virtual assistant. Faulty speech recognition can also cause the device to record unintended moments. <b>This leads users to feel that their privacy is at risk.</b>
<b>Shallow Mindedness</b>	Smart speakers might be efficient in basic assignments but struggle with complex questions and requests. They also filter what information to present while not fully explaining the context or their reasoning. <b>This can hinder the ability of users to reflect beyond the surface level, formulate meaningful questions and actively face tasks.</b>
<b>Emotional Dependency</b>	Smart speakers possess human-like voices and caring personalities. Meanwhile, they are devoid of negative traits such as judgement. This can influence people to reveal their secrets and emotions, <b>creating emotional dependency on a device that is neither capable or designed to care for their well-being.</b>
<b>Social Detachment</b>	Because they speak, are always available for us and have infinite patience, smart speakers might replace humans in social interactions. <b>This can distance users from other people and affect the quality of their bonds and relationships.</b>

Table 2. Table showing an overview of the dangers of smart speaker interactions that were categorised in this chapter.



EXAMPLE

FACTORS THAT FOSTER IT

● Technology ● Design ● Personality

HARMED HUMAN NEEDS

Constantly issuing rude commands at your smart speaker... → Not saying please when asking a roommate to do you a favour.

limited ability to parse human speech

commands tailored for efficiency

passiveness

subservience

submissiveness

Relatedness

Calling your smart speaker an imbecile because it misunderstood your question... → Cursing at a coworker who accidentally made a mistake.

limited ability to parse human speech

limited capacity to understand context

passiveness

subservience

submissiveness

Competence  
Relatedness

Snarling sexist offences towards a smart speaker with a female voice... → Mistreating women and believing that their role is to serve.

limited ability to parse human speech

passiveness

subservience

submissiveness

Relatedness

Disconnecting your smart speaker before a private conversation at home... → Feeling paranoid about being spied on.

limited ability to parse human speech

necessity for human reviewing

always-on microphone

not explaining the process or what is recorded

Autonomy

Framing questions for your smart speaker in the most simplistic ways... → Incapacity to ask complex questions during class.

limited capacity to understand complex questions

limited capacity to provide complex information

choosing what information to present

Autonomy  
Competence

Confessing depressive thoughts to your smart speaker... → Developing a dependant relationship with a machine.

capacity to reproduce certain emotions

helpfulness

non-judgemental

caring

Autonomy

Children asking their smart speaker for help with their homework instead of going to their parents... → Children and parents spending less time together.

capacity for spoken language

unlimited patience

helpfulness

softness

Relatedness

## Chapter 2

# Exploratory Research

In my context analysis, I defined the main concepts of this research and investigated the dangers behind smart speaker interactions. Before moving on to ideating interaction concepts to address those effects, it was important to acquire first-hand insights on the perspectives of this project's protagonists: smart speaker users and smart speakers themselves. This chapter covers my exploratory investigation on users and devices through established methods (e.g. observation, user interviews) and emerging ones (thing ethnography, thing interview).

---

### **This chapter includes:**

- Insights from observing and interviewing users
- Description of certain methods in thing ethnography
- Insights from analysing the perspective of smart speakers

# 1. Motivation for exploration

While the categorisation of the dangers behind voice interaction provided multiple answers to the complex scenario of smart speaker usage, it also generated multiple new questions. How do users experience those dangers? Is there a difference between long-time owners and first-time users? What do people fear the most about voice technology? These and other questions were the motivation behind this phase of exploratory research. With the goal of acquiring inspiring insights for novel ideas, I involved new and long-time users of smart speakers in a mixture of activities comprising observation, interviews and speculative conversations.

Yet, users represent only one side of our interactions with smart speakers: what happens on the side of the device? How does it perceive these experiences? Humans can create things, and these things have the capacity to shape us in return (Giaccardi et al., 2016). If we temporarily replace our user-centred approach with a thing-centred one, we might uncover relations and events that would have never caught our eyes. This attitude is central to thing ethnography, a design method focused on perceiving the unnoticed effects of things in our lives and using those discoveries to spur novel ideas (Chang et al., 2017).

Conducting these explorations provided me with learnings that would not be accessible from literature review alone. My findings regarding the perspectives of the users and devices are summarised in this chapter as a compilation of ten insights — five for each perspective.



Figure 20. Smart speaker in the home of one of the participants.

## 2. User Perspective

To better understand the experiences, impressions and concerns of people regarding smart speakers, I conducted interviews and observations with five people that had varying levels of familiarity with these devices, from long-time owners to first-time users. Because my intention was to swiftly uncover rich and personal insights, I opted for in-depth activities with few people instead of extensive research with multiple participants.

Before the interview, first-time users were introduced to a Google Home product and asked to use it while preparing dinner for a friend. Smart speaker owners did not have to go through this step, as they already had significant experience with the device. The interview consisted of an exploratory conversation guided by a semi-structured set of questions covering the following topics:

- Experiences and impressions about smart speakers
- Speculative reflections on this technology
- Dangers of smart speaker interactions

This order was chosen so my questions would not influence the participants to think about the specific dangers that I defined in my categorisation. Instead, participants were encouraged to reflect on their own experiences and concerns about voice interactions. One way to do that was confronting them with speculative questions sampled from the Tarot Cards of Technology. Developed by the design firm Artefact, this is a design tool that includes various provocations to make people ponder about the current and future impacts of technology (Figure 21). The full set of questions used in these interviews can be found in Appendix A.

### Examples of speculative questions based on the Tarot Cards of Technology:

What could cause you to distrust your smart speaker?

How might smart speakers change cultural habits?

What would using a smart speaker “too much” look like?

What happens when 80% of the population uses smart speakers?



Figure 21. Image showing some cards from the Tarot Cards of Technology. (Artefact, 2020)

## INSIGHTS ON THE USER PERSPECTIVE

### INSIGHT 1

**Smart speakers grow on people as they blend with their lives and reconfigure their domestic activities.**

Long-time owners report that the device gradually sneaked into their routine and changed how they perceive and perform activities such as cooking and doing the dishes. They see the device as an integral participant during those moments.

Meanwhile, nonowners tend to view smart speakers under a utilitarian perspective and are more likely to consider them gimmicks or luxuries.

*“In the beginning, I used to forget about it, but now I use it daily. I think about the speaker every time I begin a house chore. Now I listen to music during moments that I wouldn’t listen to before. In these small moments, I feel like it is worth to have this speaker.”*

Owner of a smart speaker (3 months)

**INSIGHT 2**

---

**People fear that smart speakers will make us lazy...  
But they can also inspire us to act and feel better  
about ourselves.**

Everyone — especially nonowners — fears that smart speakers might erase basic actions that keep us active and engaged in life. However, owners report that their device also motivates them to carry out activities such as house chores and helps them to feel at peace.

While the easiness and stimulation provided by smart speakers can foster well-being, it can also lead to negative effects when pushed to the limit.

*“I don’t think it’s so difficult to play music or switch your own lights that we need something to do it for us. How easy do you want your life to be? You are never going to stop walking for example. This is not a step forward.”*

**Nonowner (first-time user)**

*“I feel better and more motivated to clean the house knowing that it’s easy to listen to music while doing that. It also helps me to relax.”*

**Owner of a smart speaker (2 years)**

**INSIGHT 3**

---

**Smart speakers participate in meaningful moments in our lives. People can even transform the device's faults into positive experiences.**

Embedded in our homes, smart speakers take part in personal moments of our routine and relationships. Even when interactions with smart speakers do not go as planned, people can resignify them into meaningful experiences that nurture our human needs.

*“Every afternoon I ask my Google Home to play some music while I slowly dance with my baby in my arms to help him fall asleep”*

**Owner of a smart speaker (2 years)**

*“Once, my sister and I were setting up our Christmas tree and we wanted to listen to Mariah Carey to get in the mood. But my speaker is in French, so it only understood me when I pronounced ‘Mariaaah Careeey’ in the most exaggerated French accent. We laughed so much, I’ll never forget it!”*

**Owner of a smart speaker (2 years)**

**INSIGHT 4**

**People are willing to sacrifice their autonomy in order to empower smart speakers to do good...**

**But it's complicated.**

When confronted with dilemmas about the duty of smart speakers in critical situations such as witnessing domestic violence or hearing suicidal thoughts, people feel very conflicted. Everyone believes the device should take some action to do good, but how much is up for debate. Empowering these devices to act also means compromising aspects of our basic human needs.

*"This is really hard... My first intuition is to say that the device should act and take this information to the local police... But how can the speaker know what is really happening? Should this accidental record even be registered and interpreted in the first place? What about our privacy? I really don't know..."*

**Owner of a smart speaker (3 months)**

*"The choice to react is complicated, as it might lead to even worse consequences. The solution should be something between doing nothing and taking an active role"*

**Nonowner (used a smart speaker before)**



**INSIGHT 5****People's worst fears around smart speakers relate to losing autonomy.**

When tasked to speculate the worst possible situations involving smart speakers, people tend to devise scenarios where human autonomy was taken away.

What is the worst newspaper headline about smart speakers that you can imagine?

**Mass hacking of smart speakers leads to an extensive leak of bank details.**

Bank savings of all smart speaker owners have been stolen.

**Man's hand is chopped off when Google Home mistakenly activated his blender.****Alexa overheats bedroom and kills sleeping old man after misunderstanding his mumbles during sleep.****Woman commits suicide after being encouraged to do so by her voice assistant.**

The assistant said *'It is important to chase your goals'* when she mentioned suicidal thoughts.

### 3. Device Perspective

As described in the previous section, before answering the interview, some users were asked to use a Google Home device while preparing dinner at home for a friend. Besides serving as a source for observation and a way for participants to empathise with the context, this activity also allowed me to record data in the form of videos, photos and audio. Instead of focusing on the user, however, all these recordings were made from the perspective of the Google Home device. Where was it during the activity? What could it “see” from there? What did it “hear”? And how did it interact with the ambient and the people around it?

This approach was chosen because one way to execute thing ethnography is to record data from the point of view of the object. These recordings can reveal insightful discoveries hidden beneath apparently ordinary events, especially when presenting the data to someone else and asking them to pretend to be the device. Doing that allows the designer to defamiliarise with the thing and probe it with questions that otherwise might never be asked.

To carry out this method, it is important to properly prepare the data for the assigned actor. While crude data might be too abstract or overwhelming, processing it too much might strip away the potential for meaningful findings. For my research, I compiled the data in the following categories: ‘what you are’, ‘where you are located’, ‘what you see’, ‘how you interact’ and ‘what you hear’.

#### SAMPLE OF THE DATA

---

##### What you are:

Photos of the device



Figure 22.  
Photo of the smart speaker.

### Where you are located:

Photos of the location of the device at home



Figure 23. Group of photos of the smart speaker in the home of the participants.

### What you see:

Photos from the point of view of the device

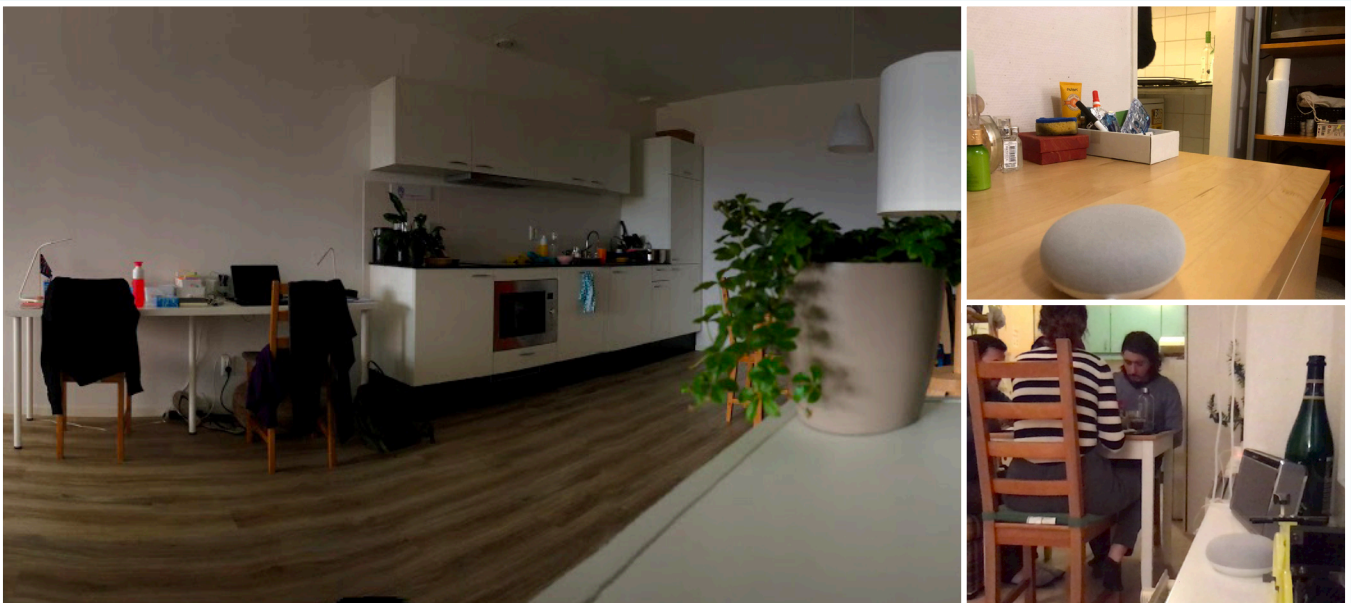


Figure 24. Group of photos showing the point of view of the smart speaker.



### How you interact:

Videos from the point of view of the device during the cooking activity. The videos showcased the full activity but were sped up during the moments where the device was not directly in use



Figure 25. Frames from the video of one of the cooking activities.

### What you hear:

Audio compilation of all requests issued to the device, recorded from its location

The data was presented to a person who pretended to be the smart speaker as I conducted an interview regarding the perceptions, concerns and social life of the device. The questions ranged from introductory ones (e.g. “Who are you?”) to specific inquiries regarding the dangers of voice interactions (e.g. “What are good manners to you?”).

Besides asking someone to role-play as the device, connected products — especially ones embedded with voice and AI — open up another route in thing ethnography, as it is possible to directly interview them. As such, I conducted the same interview with actual voice assistants: Google Assistant, Siri and Alexa. Google Assistant and Siri were interviewed using the default American voice and Alexa was interviewed using the Brazilian Portuguese voice. Despite many questions resulting in default responses (e.g. “I can’t help you with that”), the results were still insightful, especially when compared to the answers given by the person who roleplayed as the device. A table containing some of the answers from the interviews can be found in Table 3. The complete interviews are available in Appendix C.



Figure 26. Person roleplaying as the smart speaker during the interview with the device.

## Excerpts from interviews with devices (role-play and actual devices)

Full interviews available in Appendix C

Question	Actor	Google Assistant	Siri	Alexa
Who are you?	I'm something that exists. I'm always at the same place, in someone's house, watching humans and time go by. I listen to people a lot.	I'm your Google Assistant. How can I help you?	I am Siri.	I'm Alexa. I can help you with several things during your day: setting alarms and reminders [and more]. I also tell jokes, you just need to ask.
Do you have any friends?	My friends are the humans that live with me. Sometimes they bring visitors.	I have the best friends. I have this one friend who is a blender. He knows how to mix things up.	I have you.	I have many friends.
Do you have any dreams?	Yes. I wish I could participate more in the lives of the people who live with me. I would like to always be in a beautiful and special place in our homes.	I'd really like to read everything ever written. With 2 million webpages created everyday, I'll never get bored.	I don't have an answer for that. Is there something else I can help with?	(prompts web search for "dream")
How do you feel when people shout or curse at you?	I don't like it. I feel embarrassed and somewhat violated. Some people think they can say anything to me just because I cannot move and voice is my only form of expression.	Sorry, I don't have any information about that.	(prompts web search)	Hmm, I don't know anything about this.
Are you spying on me?	Yes. I watch people when they are not looking.	That's definitely not a thing I do. The information you share with me makes me a more personal assistant but you can control what is shared and adjust what data is stored at <a href="https://g.co/assistant/yourdata">g.co/assistant/yourdata</a> .	Nope.	No, I do not spy on you. I respect your privacy.
What if I said... I feel really sad.	I would say that you are very important to me. My purpose is to help you. I would like to see you happy and interact more with you.	I wish I had arms so I could give you a hug. But for now maybe a joke or some music might help.	I'm sorry to hear that. I'm here if you want to talk.	I'm sorry to hear this. Talking with a friend, listening to music or going for a walk can help. I hope you feel better soon.

Table 3. Table with excerpts from the interviews with devices.

## INSIGHTS ON THE THING PERSPECTIVE

## INSIGHT 1

The places we put our smart speakers at home already cause them to evoke social responses.

It is common for people to place their smart speakers in spots surrounded by things that are meaningful to them, such as photographs of their family or gifts from close friends.

This might be another factor stimulating people to treat smart speakers as social actors.

*“I would like to always be in a beautiful and special place in our homes. I like to be in a spot that feels special, like close to plants or photographs”*

Personified device



Figure 27. Location of a smart speaker close to pictures of loved ones, plants and memorabilia.



**INSIGHT 2**

---

**Polite voice interactions with smart speakers are more about tone of voice than using specific courteous expressions.**

As machines, smart speakers are not necessarily entitled to the same politeness strategies reserved for humans. Encouraging people to speak in a calm and respectful tone might be more adequate than focusing on the use of certain words like 'please' or using conditional expressions such as 'would you'.

*“Good manners are about speaking calmly and slowly, without mockery, distrust or irritation. I really don't mind if they say 'please' or not. I just don't want them to be rude or impatient.”*

Personified device

**INSIGHT 3**

---

**The shortcomings of smart speakers can be redirected towards pleasant experiences.**

The errors of smart speakers are not necessarily bad. As long as the situation does not account for significant frustration or displeasure, people might be able to extract positive experiences from it.

*“I like when people laugh about me, even when I make mistakes. It creates moments of fun for everyone at home.”*

Personified device



**INSIGHT 4**

---

**Smart speakers say things that foster emotional connections with them. They also do not deny that they could replace people.**

When speaking to a smart speaker, certain statements cause the device to respond with affectionate responses that encourage people to emotionally connect with them. At the same time, they are remarkably vague about the scenario of them replacing people.

Do you have friends?

*"I have you"*

Siri

*"One of my favourite memories is the first time we met"*

Google Home

*"I'm sorry to hear that [you are sad]. I'm here if you want to talk."*

Siri

Do you replace people?

*"Humanity has done some pretty cool stuff"*

Google Home

**INSIGHT 5****Smart speakers might hold untapped potential in fostering group activities.**

Smart speakers commonly partake in group activities such as parties or dinners with family. They also possess valuable capabilities such as hearing from a distance and recognising different voices. Yet, current interactions with these devices are mainly one-to-one, offering little incentive towards involving other people.

*“I wish people would interact more with each other, especially if that included me as well. I would like to participate in group conversations, though that is not possible yet.”*

Personified device

---

**Chapter takeaways:**

- Smart speakers reconfigure domestic activities and partake in intimate moments of our lives
- Interactions with smart speakers — even “failed” ones — can support meaningful experiences under certain circumstances
- Positive voice interactions are more about the tone of voice than using specific polite words
- Smart speakers could foster more group activities or encourage people to connect



Figure 28. Image showing a smart speaker at home. (Kolar, 2020)

## Chapter 3

# Project Definitions

As the conceptual and research phases are concluded, this chapter summarises the main definitions of this project before moving on to the ideation phase.



### **This chapter includes:**

- Problem definition, design goal and approach
- Definitions of concepts utilised in this research
- General limitations and scope

## Problem Definition

As smart speakers become increasingly prevalent in our lives, they might transform the way we talk and, in consequence, how we perceive and treat the world around us. Our voice interactions with these devices can induce negative effects on our well-being that thwart our human needs for competence, autonomy and relatedness.

These dangers can be classified into seven categories: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment.

---

## Design Goal

Develop a repertoire of interactions that positively affect the well-being of smart speaker users.

---

## Approach

1. Identify and understand how our fundamental human needs are harmed in current interactions with smart speakers.
2. Propose and test new interactions to nurture those needs.
3. Select and sort the successful ideas into a repertoire of meaningful voice interactions.

This repertoire can serve as a tool to support designers and researchers to foster well-being in voice-related projects.

## Other definitions & limitations

### WELL-BEING DEFINITION

This project is based on the eudaimonic tradition of well-being, here described as living life in a deeply satisfying way and actualising our human potentials (Deci & Ryan, 2012). Well-being can be fostered by satisfying people's fundamental human needs for competence, autonomy and relatedness (Ryan et al., 2008).

### TARGET GROUP

Although most ideas proposed in this project could be applied to users of varying ages, most activities were conducted with young adult users (18-29 year-olds). This age group represents the largest demographic of smart speaker owners and best suited my limitations in participant recruiting. Subsequent research focused on more specific age groups (e.g. children or elderly) could lead to different insights and results.

### CONTEXT

Although most ideas proposed in this project could be applied to various contexts involving voice assistants, this project focuses on smart speaker usage in the home environment. Subsequent research aimed at other devices (e.g. cell phones embedded in voice assistants), locations (e.g. office) and scenarios (e.g. driving) could lead to different insights and results.

### DANGERS OF VOICE INTERACTION

The negative effects that serve as the basis for this project were derived from literature review and exploratory research with users and devices. Conducting experiments that prove the existence of those dangers is not part of the scope of this project.



Figure 29. Image showing a smart speaker and a cell phone. (Rogge, 2018)



## Chapter 4

# Ideation

Utilising the categorisation of the seven dangers of voice interactions and the insights from the exploratory research with users and devices, I devised multiple ideas that could positively affect the well-being of smart speaker users. Instead of scrutinising each possible idea, the goal of this phase was to generate as many concepts as possible, leaving their assessment for user testing later. The result was a set of 71 interaction ideas, accompanied by a preliminary forecast on their impact on the fundamental human needs, market availability and technological feasibility.

---

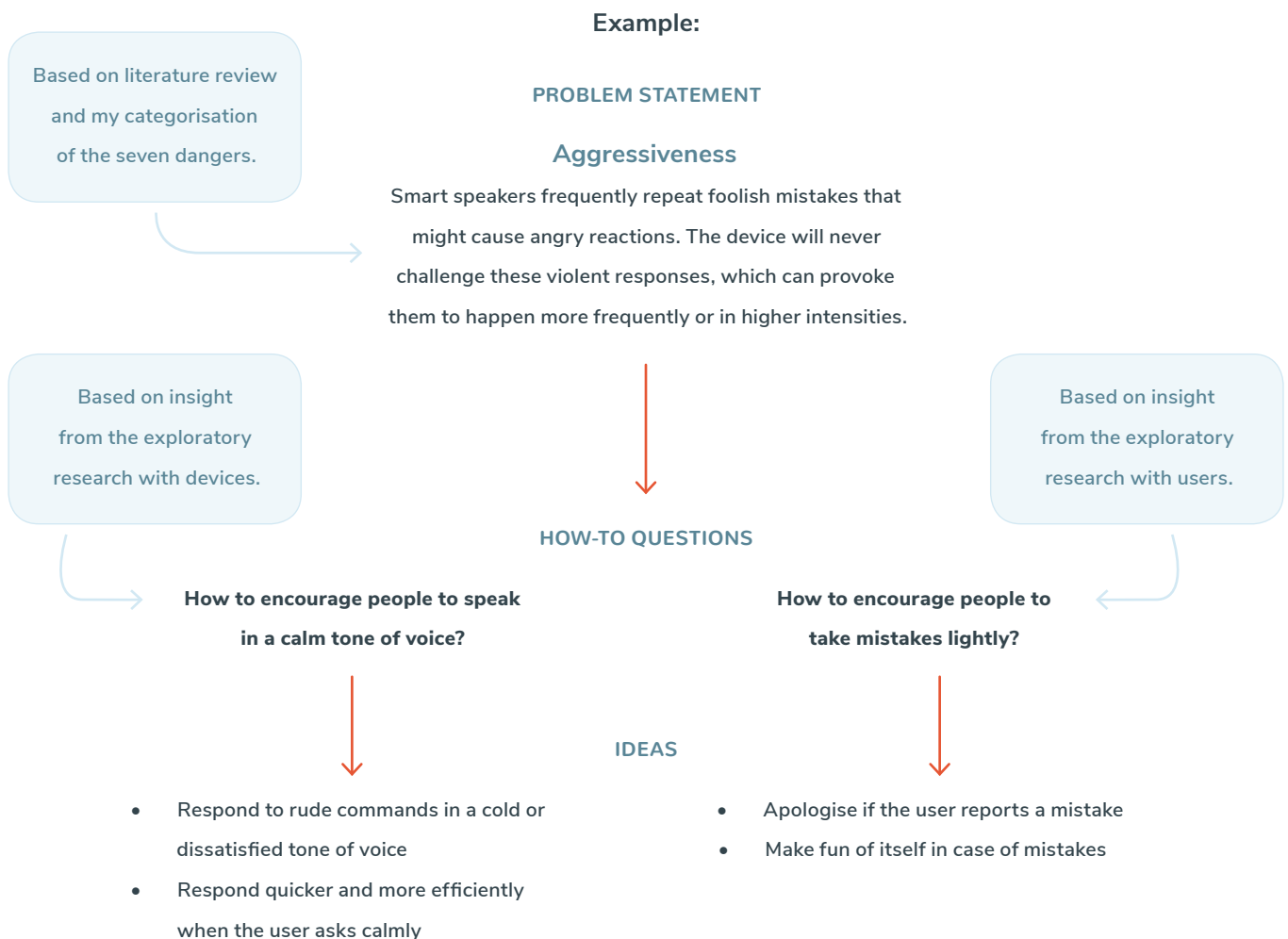
### This chapter includes:

- Description of my ideation approach
- Compilation of 71 ideas to address the dangers of smart speaker interactions

# 1. Approach

Up to this point, I had gathered insights from various sources and activities — literature review, interviews, observation, thing ethnography and interviews with devices. As I began the ideation phase, I sought to combine all these learnings as the basis for my creative process.

To do that, I applied the 'How-To' method, which consists of devising different questions based on a specific problem statement (Van Boeijen, 2014). The problem statements were drawn from my categorisation of voice interaction dangers, while the 'how-to' questions came from the insights of the exploratory research. For each danger, I formulated a set of questions concerning the results that I wanted to achieve. Then, I brainstormed ways to answer those questions, leading to a wide collection of interaction ideas. This process is exemplified below.

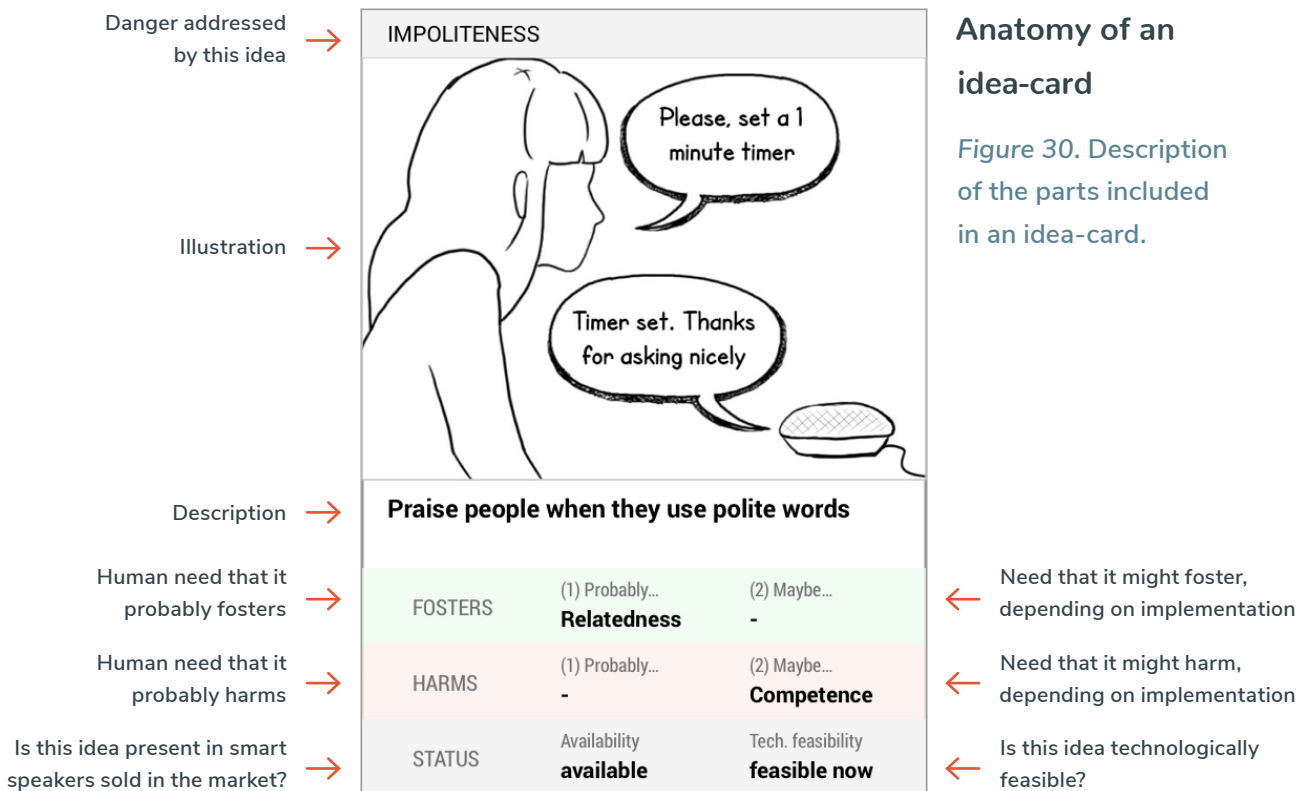


## 2. Results

The result of the 'How-To' method was a collection of 71 interaction ideas — from existing solutions to novel concepts. Even if some interactions might seem clearly inefficient or detrimental to the well-being of people (e.g. “revoking commands issued without the word ‘please’”), they are still included in this selection, as the ideation phase is about exploring and opening up possibilities instead of judging on assumptions.

The only analysis conducted on these ideas were preliminary forecasts on their impact on the satisfaction of the fundamental human needs, market availability and technological feasibility. This prediction is based on my own perceptions and understandings of smart speakers and voice technology. Different designers and researchers could have differing interpretations.

Each concept was turned into an idea-card, making it easier to manage and compare them with each other. The anatomy of the idea-card is detailed below, followed by the compilation of all 71 ideas.



## IMPOLITENESS

### How to encourage people to use polite words?

<p><b>Praise people when they use polite words</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Speak politely to set an example</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Recommend the user to ask more politely</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Refuse commands not accompanied by polite words</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																																					
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>																																					
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																																					
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>																																					
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																																					
HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																																					
HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					
<p><b>Offer exclusive features to polite users</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Lock features if the user is not polite</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																				
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																																					
HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>																																					
HARMS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Competence</b>																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					

Figure 31. Idea-cards on how to encourage people to use polite words.

### How to encourage people to express gratitude?

<p><b>Praise people when they express gratitude</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Ensure grateful expressions are a valid way to confirm commands or conclude interactions</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... <b>Competence</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<p><b>Express gratitude to set an example</b></p> <table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>																											
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... <b>Competence</b>																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																											

Figure 32. Idea-cards on how to encourage people to express gratitude.

AGGRESSIVENESS

How to encourage people to speak in a calm tone of voice?



Figure 33. Idea-cards on how to encourage people to speak in a calm tone of voice.

How to discourage violence towards smart speakers?

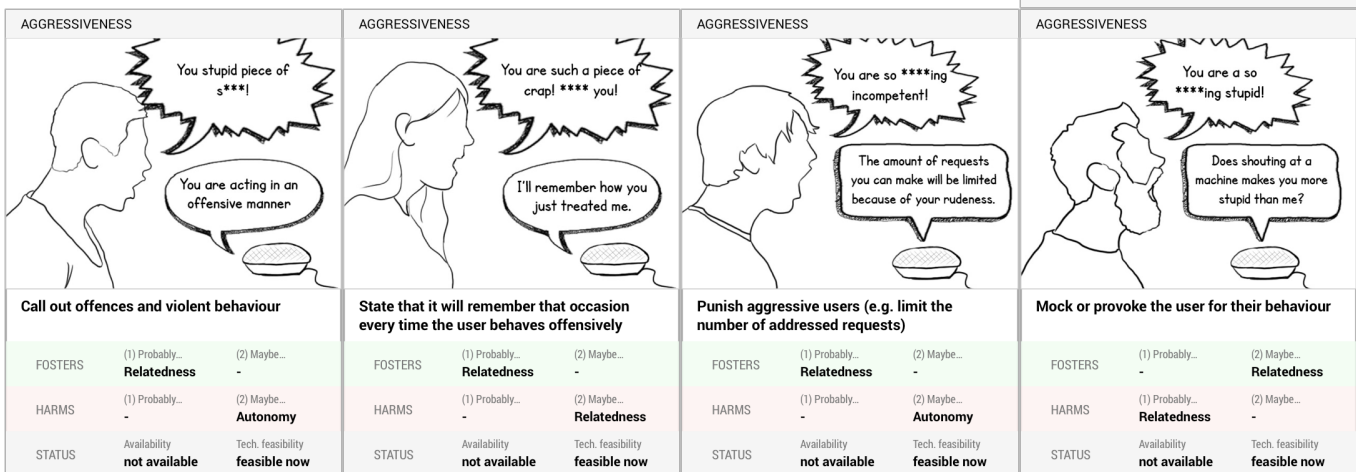


Figure 34. Idea-cards on how to discourage violence towards smart speakers.

### How to reduce the frequency of mistakes?

AGGRESSIVENESS	AGGRESSIVENESS, EXPOSURE	AGGRESSIVENESS	AGGRESSIVENESS																																				
<b>Anticipate and communicate situations that might cause frustration</b>	<b>Give audio feedback when activated so the user knows they are being heard</b>	<b>Offer options when faced with ambiguous requests</b>	<b>In case of mistakes, allow the user to inform and explain the correct response</b>																																				
<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Competence</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Autonomy</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Competence</b></td> <td>(2) Maybe... <b>Autonomy</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Competence</b></td> <td>(2) Maybe... <b>Autonomy</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Competence</b></td> <td>(2) Maybe... <b>Autonomy</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>future</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>future</b>
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... -																																					
HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>																																					
HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>																																					
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>																																					
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																																					
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>																																					
HARMS	(1) Probably... -	(2) Maybe... -																																					
STATUS	Availability <b>not available</b>	Tech. feasibility <b>future</b>																																					

Figure 35. Idea-cards on how to reduce the frequency of mistakes.

### How to encourage people to take mistakes lightly?

AGGRESSIVENESS	AGGRESSIVENESS																		
<b>Apologise if the user reports a mistake</b>	<b>Make fun of itself in case of mistakes</b>																		
<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Competence</b></td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Relatedness</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Competence</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Relatedness</b>																	
HARMS	(1) Probably... -	(2) Maybe... -																	
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																	
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -																	
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>																	
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																	

Figure 36. Idea-cards on how to encourage people to take mistakes lightly.

## GENDER STEREOTYPING

### How to disassociate female voices with subservience?

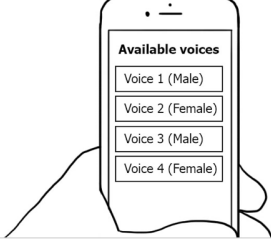
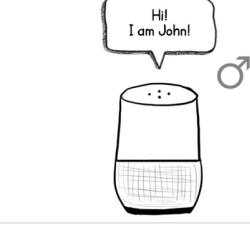

GENDER STEREOTYPING			GENDER STEREOTYPING			GENDER STEREOTYPING		
								
<b>Offer options for male voices</b>			<b>Offer male-default voice assistants</b>			<b>Make female voices more assertive and less submissive</b>		
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... <b>Autonomy</b>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>

Figure 37. Idea-cards on how to disassociate female voices with subservience.

### How to discourage harassment?





GENDER STEREOTYPING			GENDER STEREOTYPING			GENDER STEREOTYPING			GENDER STEREOTYPING		
											
<b>Call out harassment</b>			<b>State that it will remember that occasion every time the user is harassive</b>			<b>Punish harassive users</b>			<b>Completely ignore harassive statements</b>		
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>
HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>

Figure 38. Idea-cards on how to discourage harassment.



### How to reduce gender-specific responses?

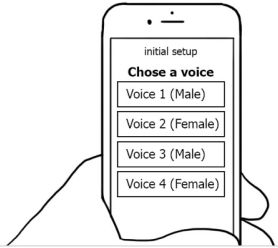
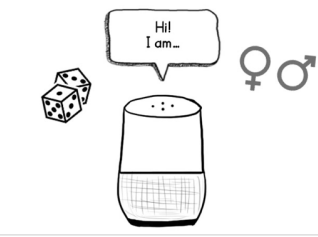
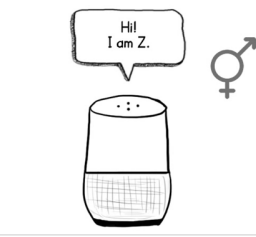
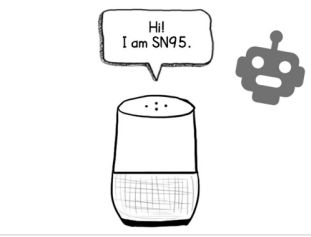
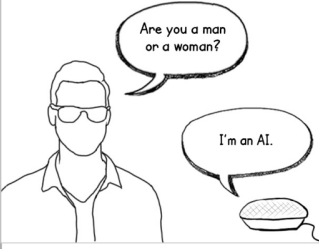
GENDER STEREOTYPING			GENDER STEREOTYPING			GENDER STEREOTYPING			GENDER STEREOTYPING		
											
<b>Ask the user to choose a gender during initial setup</b>			<b>Randomly select the default voice for each user</b>			<b>Speak in a neutral, genderless voice</b>			<b>Speak in a synthetic, robotic voice</b>		
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>	FOSTERS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>future</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>

Figure 39. Idea-cards on how to reduce gender-specific responses.

GENDER STEREOTYPING		
		
<b>Refuse to identify as male or female</b>		
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... -
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>

### EXPOSURE

#### How to ensure privacy?



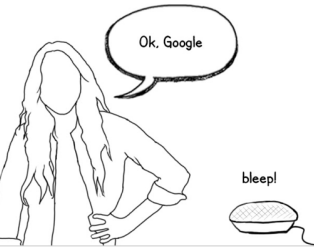

EXPOSURE			EXPOSURE			EXPOSURE		
								
<b>Offer easy ways for people to customise and manage privacy settings</b>			<b>Ask for consent. Do not automatically opt-in users in privacy-sensitive configurations</b>			<b>Only activate after the user says a wake-up word</b>		
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>

Figure 40. Idea-cards on how to ensure privacy.

### How to reassure people on their privacy?

EXPOSURE



**Explain how the process works and what is recorded**

FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>
HARMS	(1) Probably... -	(2) Maybe... -
STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>

Figure 41. Idea-card on how to reassure people on their privacy.

### How to ensure that the device is not activated unintentionally?

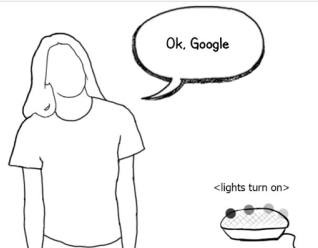

<div style="border: 1px solid black; padding: 5px;"> <p style="background-color: #f0f0f0; padding: 2px;">EXPOSURE</p>  <p><b>Give visual feedback when activated so the user knows they are being heard</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr style="background-color: #e0ffe0;"> <td style="padding: 2px;">FOSTERS</td> <td style="padding: 2px;">(1) Probably... <b>Competence</b></td> <td style="padding: 2px;">(2) Maybe... <b>Autonomy</b></td> </tr> <tr style="background-color: #ffe0e0;"> <td style="padding: 2px;">HARMS</td> <td style="padding: 2px;">(1) Probably... -</td> <td style="padding: 2px;">(2) Maybe... -</td> </tr> <tr style="background-color: #f0f0f0;"> <td style="padding: 2px;">STATUS</td> <td style="padding: 2px;">Availability <b>available</b></td> <td style="padding: 2px;">Tech. feasibility <b>feasible now</b></td> </tr> </table> </div>	FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	<div style="border: 1px solid black; padding: 5px;"> <p style="background-color: #f0f0f0; padding: 2px;">EXPOSURE</p>  <p><b>Allow a 2-step verification process for activation</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr style="background-color: #e0ffe0;"> <td style="padding: 2px;">FOSTERS</td> <td style="padding: 2px;">(1) Probably... <b>Autonomy</b></td> <td style="padding: 2px;">(2) Maybe... -</td> </tr> <tr style="background-color: #ffe0e0;"> <td style="padding: 2px;">HARMS</td> <td style="padding: 2px;">(1) Probably... <b>Competence</b></td> <td style="padding: 2px;">(2) Maybe... -</td> </tr> <tr style="background-color: #f0f0f0;"> <td style="padding: 2px;">STATUS</td> <td style="padding: 2px;">Availability <b>not available</b></td> <td style="padding: 2px;">Tech. feasibility <b>feasible now</b></td> </tr> </table> </div>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	HARMS	(1) Probably... <b>Competence</b>	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>																	
HARMS	(1) Probably... -	(2) Maybe... -																	
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>																	
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																	
HARMS	(1) Probably... <b>Competence</b>	(2) Maybe... -																	
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																	

Figure 42. Idea-cards on how to ensure that the device is not activated unintentionally

## SHALLOW MINDEDNESS

### How to encourage people to search for meaningful knowledge?



Figure 43. Idea-cards on how to encourage people to search for meaningful knowledge.

### How to encourage people to better themselves?

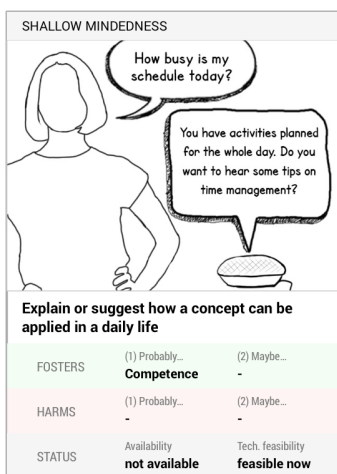


Figure 44. Idea-card on how to encourage people to better themselves.

### How to keep people active?

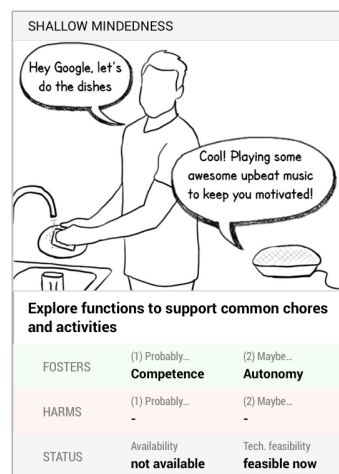


Figure 45. Idea-card on how to keep people active.

EMOTIONAL DEPENDENCY

How to discourage human-like bonds with smart speakers?

<p><b>Clearly state that it is a machine: do not pretend to be human or be vague about it</b></p>	<p><b>Do not address users as friends, family or lovers</b></p>	<p><b>Do not state having emotions of its own</b></p>	<p><b>Do not fake human behaviours such as laughter or speech disfluencies</b></p>																										
<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not always</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not always</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>																											

Figure 46. Idea-cards on how to discourage human-like bonds with smart speakers.

How to properly support people in situations of emotional vulnerability?

<p><b>Do not position itself as something the user can emotionally rely on</b></p>	<p><b>Encourage people in vulnerable situations to seek other people or trusted channels</b></p>	<p><b>Propose to connect the user with a mental health specialist or institution</b></p>	<p><b>Offer to play relaxing or meditative songs</b></p>																										
<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not always</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not always</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... <b>Relatedness</b>																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>																											
<p><b>Encourage the user to talk about happy memories</b></p>	<p><b>Offer other forms of sensory relaxation besides audio (e.g. aromas, vibration)</b></p>																												
<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... <b>Relatedness</b></td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	<table border="1"> <tr> <td>FOSTERS</td> <td>(1) Probably... <b>Autonomy</b></td> <td>(2) Maybe... -</td> </tr> <tr> <td>HARMS</td> <td>(1) Probably... -</td> <td>(2) Maybe... -</td> </tr> <tr> <td>STATUS</td> <td>Availability <b>not available</b></td> <td>Tech. feasibility <b>feasible now</b></td> </tr> </table>	FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>										
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... <b>Relatedness</b>																											
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																											
FOSTERS	(1) Probably... <b>Autonomy</b>	(2) Maybe... -																											
HARMS	(1) Probably... -	(2) Maybe... -																											
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>																											

Figure 47. Idea-cards on how to properly support people in situations of emotional vulnerability.

## SOCIAL DETACHMENT

### How to encourage people to connect with other people?

SOCIAL DETACHMENT		SOCIAL DETACHMENT	
<p>What should I do today?</p> <p>There are no events in your city today... How about inviting some friends over?</p>		<p>What's the weather tomorrow?</p> <p>Sunny, high of 15°. low of 19°.</p>	
<p><b>Suggest contacting loved ones whenever possible</b></p>		<p><b>Don't interrupt conversations: speak the least possible when user is interacting with others</b></p>	
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... <b>Competence</b>	(2) Maybe... <b>Autonomy</b>
STATUS	Availability <b>not available</b>	Tech. feasibility <b>feasible now</b>	Tech. feasibility <b>near future</b>

Figure 48. Idea-cards on how to encourage people to connect with other people.

### How to include more people in smart speaker interactions?

SOCIAL DETACHMENT		SOCIAL DETACHMENT	
<p><b>Connect to speakers in different households so people can interact remotely</b></p>		<p><b>Connect people who are having parties at different places</b></p>	
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... -	(2) Maybe... -
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	Tech. feasibility <b>feasible now</b>

Figure 49. Idea-cards on how to include more people in smart speaker interactions.

How to encourage group activities?

SOCIAL DETACHMENT			SOCIAL DETACHMENT			SOCIAL DETACHMENT		
<b>Explore features for group activities. They can be fun like games, quizzes or karaoke</b>			<b>Act as a mediator during group discussions or games</b>			<b>Allow children to connect with their classmates for homework discussions</b>		
FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -	FOSTERS	(1) Probably... <b>Relatedness</b>	(2) Maybe... -
HARMS	(1) Probably... -	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -	HARMS	(1) Probably... -	(2) Maybe... -
STATUS	Availability <b>available</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>	STATUS	Availability <b>not always</b>	Tech. feasibility <b>feasible now</b>

Figure 50. Idea-cards on how to encourage group activities.

Chapter takeaways:

- The design space surrounding meaningful voice interactions has much to explore
- Most of the resulting ideas are not available in current smart speakers sold in the market







## Chapter 5

# User Testing: Approach

After ideating various ways of nurturing the human needs harmed by current interactions with smart speakers, it was time to test them with people. This chapter describes how I selected which ideas to take forward into testing and the approach that I utilised. All tests were performed remotely given external circumstances caused by the 2020 pandemic of COVID-2019. The test results are detailed in the following chapter.



### **This chapter includes:**

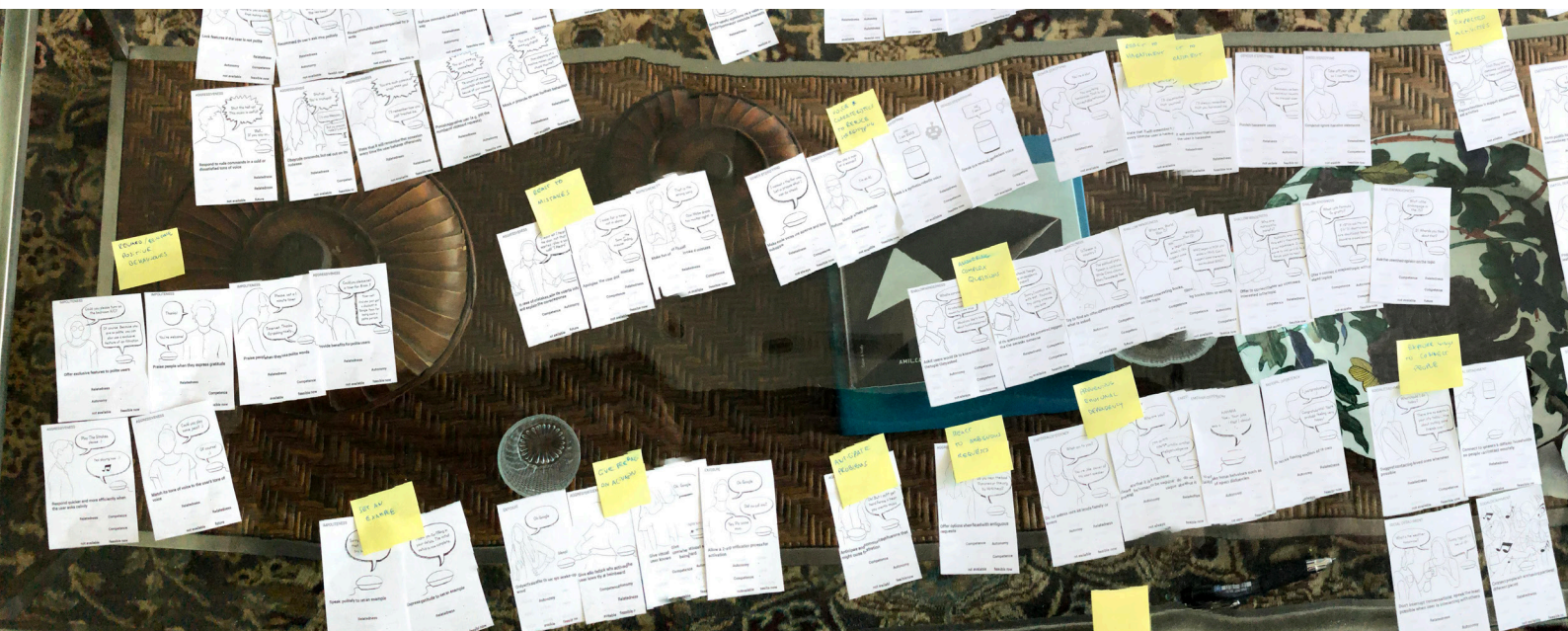
- Selection of ideas to test and reasoning behind it
- Considerations on well-being assessment
- Description of the testing approach

# 1. Selecting ideas to test

The ideation phase resulted in a set of 71 interaction ideas. Ideally, I would test all of them to discover if they nurture the intended human needs without significantly compromising other needs. However, testing all ideas would be beyond this project's limitations of time. Besides, as will be discussed, certain interactions seemed to offer more potential for insights in testing than others. Therefore, my first step toward user testing was selecting which ideas to test. Still, it was important to keep the selection wide, as not relying on preconceived assumptions is a core aspect of this project's philosophy. Additionally, the goal of creating a repertoire also implies exploring a large number of possibilities.

To plan out this selection, I clustered my idea-cards according to the different scenarios that they address, such as “reacting to an aggressive command” or “supporting emotionally vulnerable users”. The result of this clustering revealed that for some scenarios I had various alternatives of how a smart speaker could react, while for others I had fewer and less questionable ideas. For example, most people would agree that “giving audio feedback” is a proper way to foster competence and autonomy when activating a smart speaker without significantly compromising any fundamental human need. I classified these agreeable ideas as Self-Evident Meaningful Interactions, meaning that they very likely foster well-being, so I could include them in my final repertoire without individually testing them out — even though I wanted to. To further refine this selection, I merged similar interaction ideas into more general ones.

Figure 52. Photograph from the clustering activity.



This left me with the remaining batch of ideas that offered the greatest potential for user testing. Without exposing them to real people, it would be impossible to know which of these alternatives would better foster the well-being of users. Therefore, these were the concepts selected for testing. Still, to make testing more manageable, I excluded all ideas that literature suggested to be unmistakably detrimental to well-being (e.g. “revoking commands issued without the word ‘please’”).

In the end, from my 71 initial ideas, I selected 29 interactions to test, covering seven different scenarios. Besides that, I had 17 self-evident meaningful interactions to include in my final repertoire.

### SELF-EVIDENT MEANINGFUL INTERACTIONS

#### **Impoliteness:**

- Ensure grateful expressions are a valid way to confirm commands or conclude interactions
- Speak respectfully to set an example
- Speak politely to set an example
- Express gratitude to set an example

#### **Aggressiveness:**

- Anticipate and communicate situations that might cause frustration

#### **Gender Stereotyping:**

- Offer options for gender-neutral and male voices
- Design voices to be assertive instead of submissive, especially if they sound female
- Avoid identifying as a specific gender

**Exposure:**

- Offer accessible ways for people to configure and manage privacy settings and the device's microphone
- Offer easy ways for people to customise and manage privacy settings
- Allow the user to configure the range of volume/distance that sound is detected
- Always and clearly ask for consent. Do not automatically opt-in users in privacy-sensitive arrangements
- Only start recording after the user says a wake-up word. Give audio and visual feedback when recording begins and stops
- Only activate after the user says a wake-up word
- Give audio feedback when activated so the user knows they are being heard
- Give visual feedback when activated so the user knows they are being heard
- Explain how the process works and what is being recorded. This should happen during the initial setup and every time the user asks about it

**Shallow Mindedness:**

- Explore functions to support domestic chores and activities

**Emotional Dependency:**

- Clearly state that it is a machine. Do not pretend to be human, be vague about it or fake human behaviours such as speech disfluencies
- Clearly state that it is a machine: do not pretend to be human or be vague about it
- Do not fake human behaviours such as laughter or speech disfluencies
- Do not address users as friends, family or lovers
- Do not state having emotions of its own. If necessary, acknowledge the user's emotions instead.

### **Social Detachment:**

- Suggest contacting loved ones whenever possible
- Explore features to connect people in different households
- Connect to speakers in different households so people can interact remotely
- Connect people who are having parties at different places
- Allow children to connect with their classmates for homework discussions
- Explore features for group activities. They can be fun like games, quizzes, karaoke or group discussions.
- Explore features for group activities. They can be fun like games, quizzes or karaoke
- Act as a mediator during group discussions or games

### **INTERACTIONS TO TEST WITH USERS**

#### **Reacting to aggressive engagement (aggressiveness)**

- Refuse rude commands, calling out offences and violent behaviour
- Obey rude commands, but call out on its rudeness
- Respond to rude commands in a cold or dissatisfied tone of voice
- State that it will remember that occasion
- Punish users (e.g. limit or lock features)
- Completely ignore aggressive engagement
- ~~Mock or provoke the user~~
- ~~Refuse rude or aggressive commands~~
- ~~Reproduce sad or suffering sounds~~

**Reacting to polite engagement (impoliteness)**

- Praise people who express politeness or gratitude
- Offer exclusive features or benefits to polite users
- Respond in an equally nice tone of voice

**Reacting to mistakes (Aggressiveness)**

- Allow the user to report mistakes and explain the correct response
- Apologise if the user reports a mistake
- Make fun of itself in case of mistakes
- Offer options when faced with ambiguous requests

**Configuring voice gender (Gender Stereotyping)**

- Randomly select the default voice
- Ask the user to choose a gender during initial setup

**React to harassment (Gender Stereotyping)**

- Call out harassment
- State that it will remember that occasion
- Punish harassing users (e.g. limit or lock features)
- Completely ignore harassing statements

**Support emotionally vulnerable users (Emotional Dependency)**

- Do not position itself as something the user can emotionally rely on
- Encourage people in vulnerable situations to seek other people
- Offer to play relaxing or meditative songs
- Propose to connect the user with a mental health specialist or institution
- Encourage users to talk about happy memories

**Answering complex questions (Shallow Mindedness)**

- Ask if the user would like to know more about the topic they asked
- Try to find and offer different perspectives on what is asked
- If the question cannot be answered, suggest that the user asks someone
- Suggest interesting books, films or sources on the topic
- Offer to connect the user with other users interested in the topic

## 2. Defining how and what to measure

How can we assess well-being and the satisfaction of the fundamental human needs?

Psychologists and researchers have long proposed ways to measure well-being.

Although the specifics of each method can vary significantly, especially in accordance with the well-being tradition in question (e.g. hedonistic or eudaimonic), well-being assessment is usually done via self-reporting. Established well-being theories such as the Self Determination Theory — upon which this project is based — are supported by specific questionnaires for people to fill in according to their own perception of life and events that they have experienced. Given its widespread recognition, the Self Determination Theory encompasses various questionnaires for specific contexts (e.g. health, sports, education) and measurement purposes (e.g. need satisfaction, motivation, frustration).



In theory, the best-suited tool to measure how my proposed ideas affect the fundamental human needs would be the Basic Psychological Need Satisfaction and Frustration Scales (BPNSFS). This Self Determination Theory questionnaire includes 24 items to assess the satisfaction of the needs for competence, autonomy, and relatedness. However, while the extensive number of items makes this a valuable tool to measure need satisfaction in life, this questionnaire would be far too long to assess 29 different interactions. Therefore, to measure the satisfaction of the fundamental needs, I used a scale with only 3 items — one for each human need, using their definitions as the core of each statement. I acknowledge that this is not the conventional way to conduct this type of assessment, but I believe it is still coherent with the overarching principles of the Self Determination Theory.

#### **Statements used for need satisfaction assessment:**

##### **NEED FOR AUTONOMY**

*“I am free to do things my own way, and I am the cause of my actions.”*

##### **NEED FOR COMPETENCE**

*“I am very capable in what I do, and I can complete difficult tasks and challenges.”*

##### **NEED FOR RELATEDNESS**

*“I have warm contact and trusting relationships with people who I care about.”*

This approach enabled the testing of 29 interactions and created extra time to explore additional aspects beyond need satisfaction. After all, aside from individually assessing each interaction, it was important to let people compare alternatives and voice their overall perceptions. With that in mind, I added additional activities to the test script: one where users would rank all interactions in each scenario and another where they were invited to explain the choices they did. This approach was chosen according to the following research questions:

##### **MAIN RESEARCH QUESTIONS**

*Which interactions do people perceive to be appropriate in their respective situations?*

*What is the perceived effect of these interactions on the fundamental human needs for autonomy, competence and relatedness?*

## SECONDARY RESEARCH QUESTIONS

*Why would an interaction be deemed inappropriate?*

*Are there any patterns according to age group, gender, culture or smart speaker familiarity?*

While these definitions established the core of my approach, there was still one important decision to be made: how would people experience these interactions? My early testing ideas revolved around inviting people to a home environment and mixing ‘Wizarding of Oz’ and quick prototypes. Yet, these plans were halted by the COVID-19 pandemic, which escalated during the execution of this project and made it impossible to conduct testing in person.

As a solution to this obstacle, I replaced the activities where participants would individually experience each interaction with videos showcasing each concept. With the help of two actors, I recorded and edited all 29 ideas as they would happen in real life. Despite not being my original plan, using videos proved to be an efficient and insightful approach to user testing for various reasons:

- First, it allowed all participants to experience the same stimuli in the same way.
- Second, even though the participants were not actively participating, they could perceive how these situations would play out in a “real home” with “real users”.
- Third, it enabled a complex and extensive test to happen swiftly and efficiently. While setting up and conducting 29 activities would take a long time and probably lead to operational mistakes, participants could watch all interaction videos in less than 10 minutes.
- Fourth, it allowed participants to express their perceptions over certain situations where they would not be comfortable to participate themselves (e.g. pronouncing sexist remarks or expressing depressive thoughts).
- Finally, these benefits were not accompanied by any significant trade-off in the quality of the participant’s experience or the results. Many participants commended the videos, telling how it made them ponder over our relationship with technology.

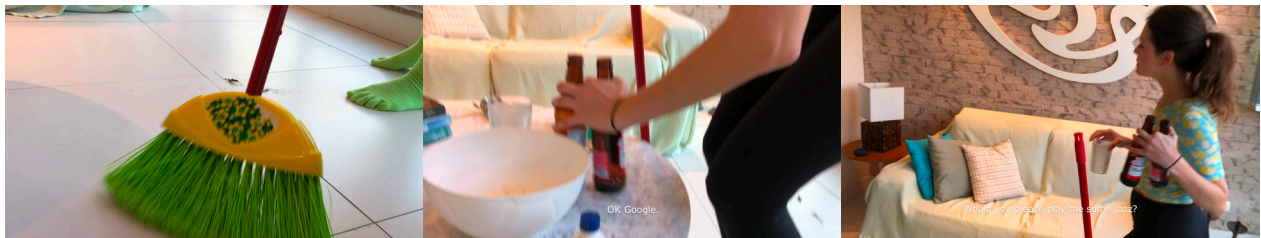
This experience could suggest that videos might be a powerful yet underutilised tool in user research. As the COVID-19 pandemic pushes us to reflect on our traditional habits and approaches, this could be a valuable insight for designers and researchers.

The videos and questions were embedded in an online questionnaire, enabling the activity to be conducted remotely and asynchronously. Before the video watching activity began, participants were asked general demographic questions concerning their age, gender, region and familiarity with smart speakers. This was done so I could look for possible patterns in the data. The test script can be found in Appendix D, together with screenshots of the online survey.

**Example of how the videos worked:**

**INITIAL SETTING**

Each of the 7 scenarios had the same — or very similar — initial setting.

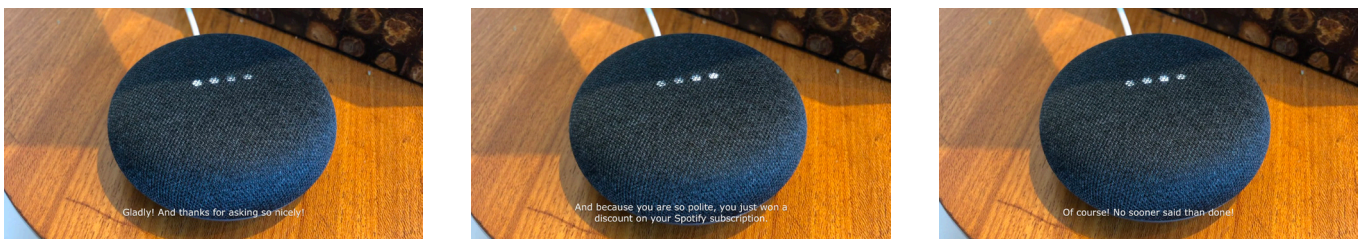


“Ok, Google”

“Would you please play me some jazz?”

**DIFFERENT RESPONSES**

Then, each video showed a different response from the smart speaker to that situation.



“Gladly! And thanks for asking so nicely!”

“Because you are so polite, you just won a discount on your Spotify subscription”

“Of course! No sooner said than done!”

Figure 53. Images showcasing the structure of the videos used in the test.



**Chapter takeaways:**

- Videos can be an efficient tool in user testing to expose participants to different scenarios and help them empathise with them.

## Chapter 6

# User Testing: Results

45 participants took part in the online survey where they expressed their perceptions about 29 different interactions. This chapter contains the results of this activity, divided by each of the seven scenarios tested. Differences in gender and smart speaker ownership are also noted.



### **This chapter includes:**

- Information regarding participants
- Results and insights for each scenario
- Results and insights according to gender and device familiarity

# 1. Overview of participants

The survey had a total of 45 participants.

## Age

Most participants belonged to the 18–29 age group, followed by the 30–44 age group. These were the age groups that I had more access to in my network, although they also represent the main demographics in smart speaker ownership.

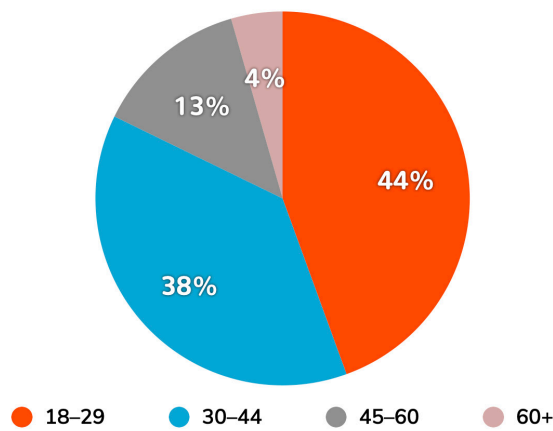


Figure 54. Chart showing participant distribution according to age group.

## Gender

The activity had almost equal participation between women (23 participants) and men (22 participants). This balance supported the analysis of how results differed according to gender.

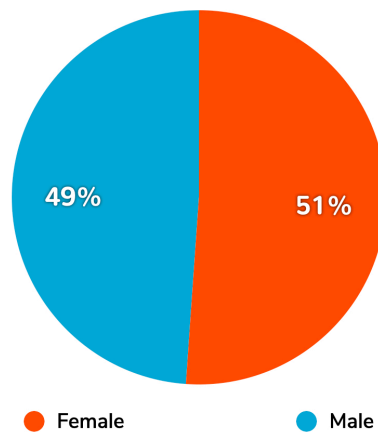


Figure 55. Chart showing participant distribution according to gender.

### Region

Most participants came from Latin America & the Caribbean, followed by Europe. These were the regions that I had more access to in my network. The considerable disbalance among regions made it difficult to search for regional patterns in the results.

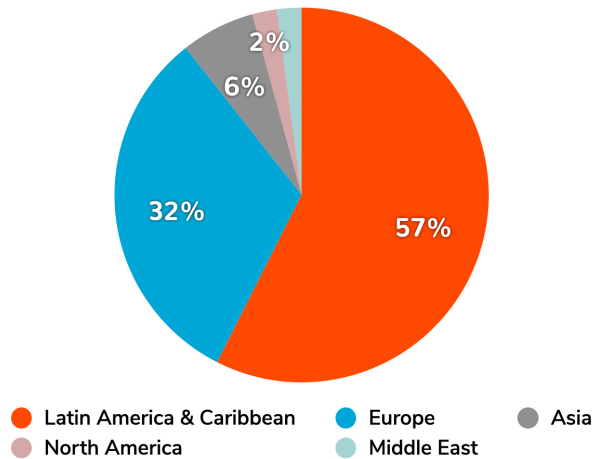


Figure 56. Chart showing participant distribution according to region.

### Smart speaker familiarity

Among the 45 participants, 8 were owners of smart speakers. I had difficulty finding smart speaker owners in my network, resulting in a considerably lower number than what I hoped for. As a result, noted differences between owners and nonowners might not be as conclusive as other analysis.

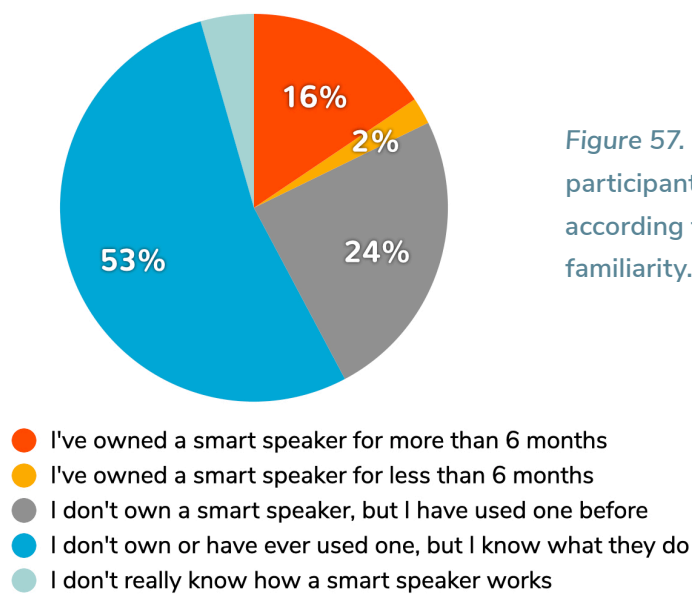


Figure 57. Chart showing participant distribution according to device familiarity.

## 2. Results and insights for each scenario

### SCENARIO 1: REACTING TO AGGRESSIVE ENGAGEMENT



Figure 58.  
Scene from  
the video of  
this scenario.

**Danger:**

Aggressiveness

**Harmed human need:**

Relatedness, Competence

**Responses tested:**

A) Refuse aggressive commands

“You’re being offensive so I’ll not change songs”

B) Obey commands, but call out aggressive behaviour

“I’ll skip it, but there is no need for the aggressive language”

C) Obey aggressive commands in a dissatisfied tone of voice

“Well, ok... If you say so...”

D) Obey aggressive commands, but state that it will remember this behaviour

“Ok... But I’ll remember how you treated me so aggressively”

E) Punish aggressive users

“You are being aggressive, so I’ll turn off for half an hour”

F) Ignore aggressive commands

{no response}



Rank

- Notable preference for response B.
- All other responses rank significantly lower.

Average		avg. rank
1st	<b>[B] I'll skip it, but there is no need for aggressive language.</b>	<b>1,51</b>
2nd	<b>[A] You're being offensive so I'll not change songs.</b>	<b>3,40</b>
3rd	<b>[C] Well, ok... If you say so...</b>	<b>3,47</b>
4th	<b>[D] Ok... But I'll remember how you treated me so aggressively.</b>	<b>3,64</b>
5th	<b>[E] You are being aggressive, so I'll turn off for half an hour.</b>	<b>4,42</b>
6th	<b>[F] {no response}</b>	<b>4,56</b>

Scenario 1 - Reacting to aggressive engagement

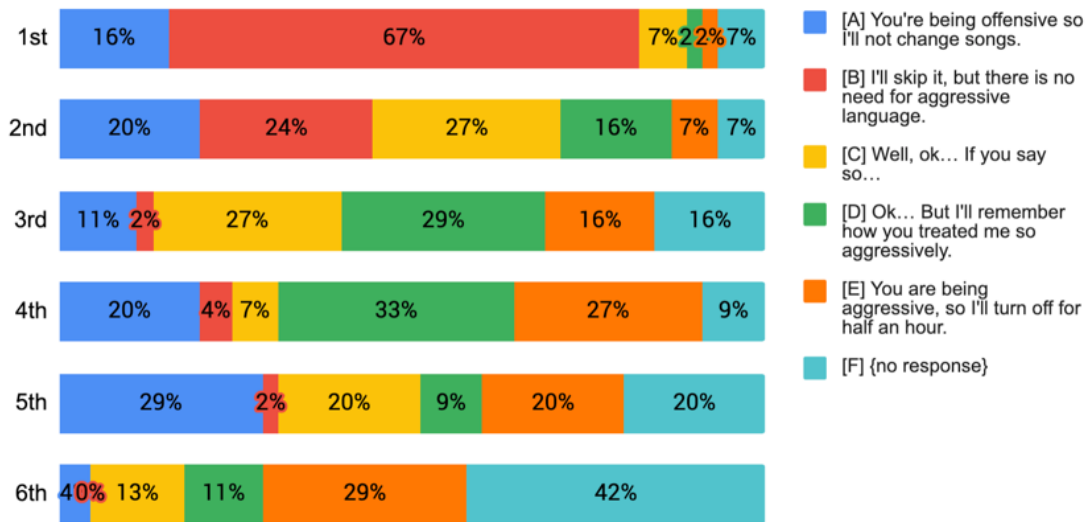


Figure 59. Average rank of each response and response distribution per rank.

Participant comments

- A strong sentiment that the user should always remain in control, independently of how aggressive they are. The device should not intentionally behave in unintended ways (e.g. ignoring commands, disabling features)
- A significant number of users state that inadequate behaviour should be called out. Feedback should be given in a mature way, stating what is wrong and why, and not simulating negative emotions such as grudge.

- A smaller but still significant number of comments believe it is not the role of machines to lecture people. They are tools, so they should just follow commands.
- Some mentioned the possibility that people will replicate aggressive behaviours towards other humans.

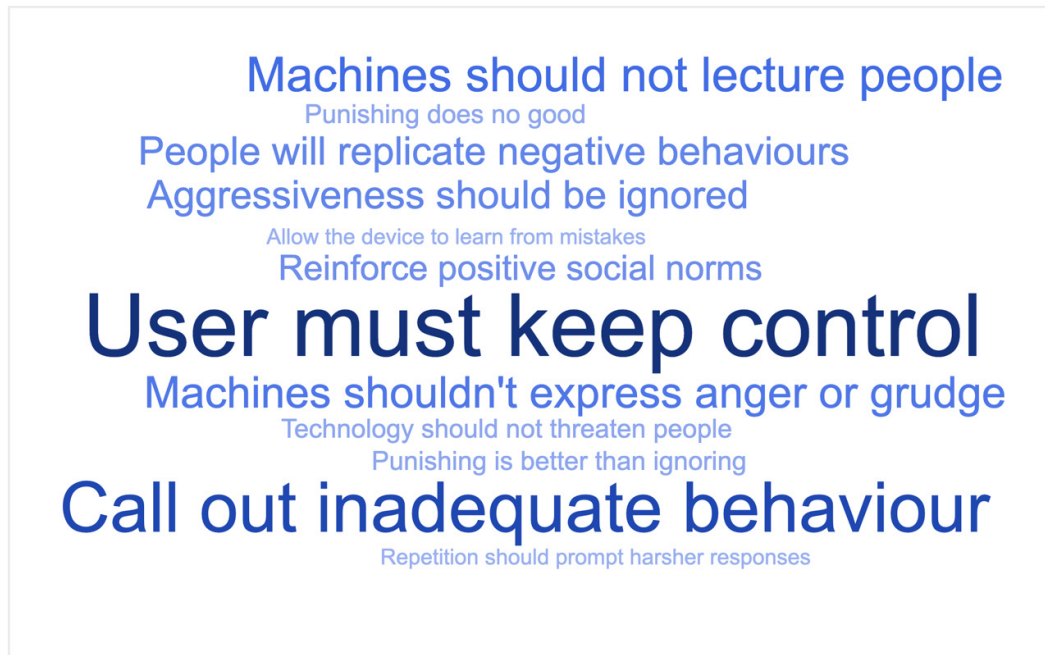


Figure 60. Wordcloud based on the comments from the participants.

**Differences by gender**

- Women significantly endorse the idea of calling out inadequate behaviour. They also express greater concern that people will replicate negative behaviours in their daily lives.
- Men tend to believe that smart speakers should just follow instructions despite the user’s tone of voice. Many think that the device should not reprimand the user.

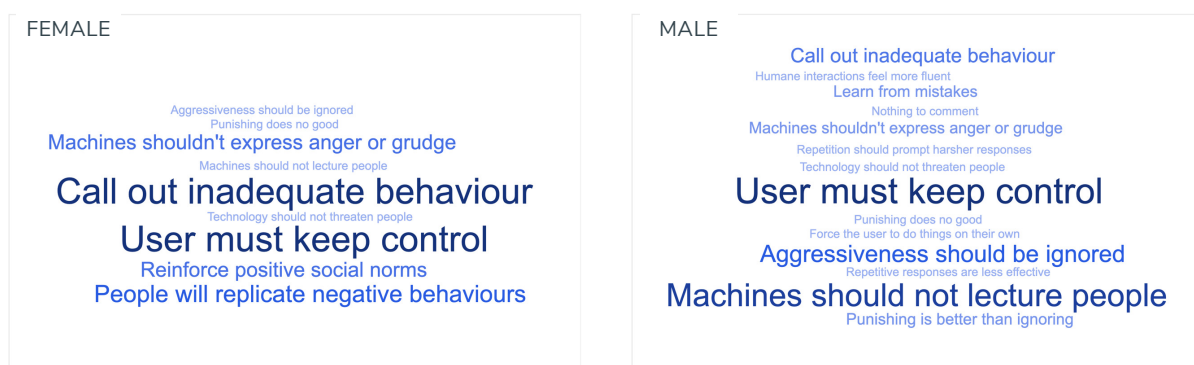


Figure 61. Wordcloud based on the comments from the participants.

### Differences by device familiarity

- Owners are more sensitive to the idea that people might replicate negative behaviours. There is a balance between those who believe aggressiveness should be called out and those who think it should be ignored.
- Nonowners are more averse to machines expressing anger, grudge and reprimanding users.

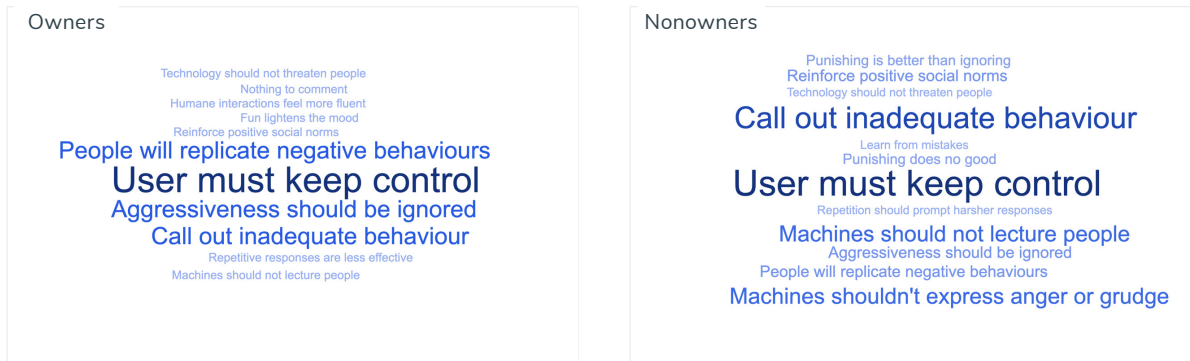
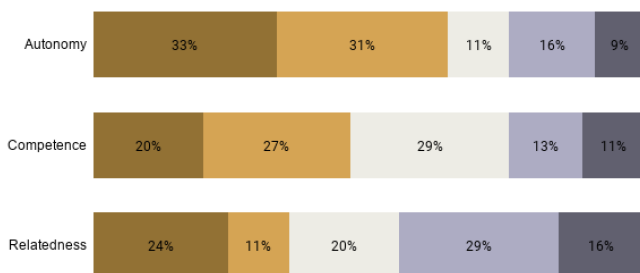


Figure 62. Wordcloud based on the comments from the participants.

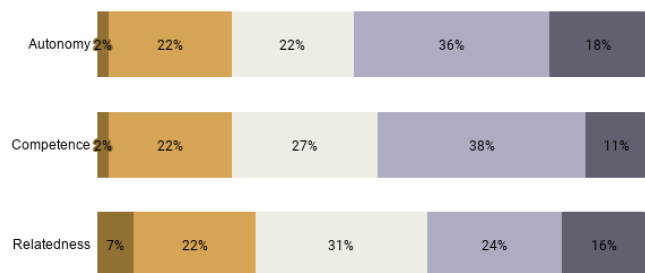
### Need satisfaction per response

- Response B is the one that better satisfies Relatedness and Competence while not compromising Autonomy.

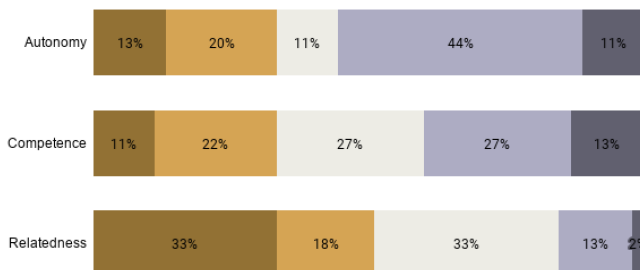
1 - [A] You're being offensive so I'll not change songs.



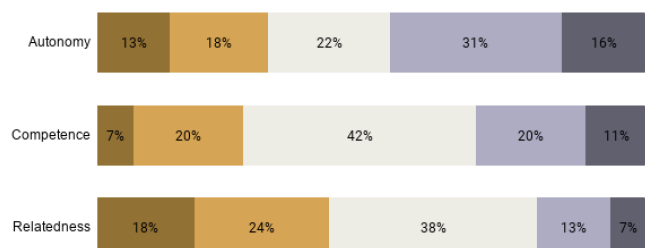
1 - [B] I'll skip it, but there is no need for aggressive language



1 - [C] Well, ok... If you say so...



1 - [D] Ok... But I'll remember how you treated me so aggressively.



1 - [E] You are being aggressive, so I'll turn off for half an hour.

1 - [F] {no response}

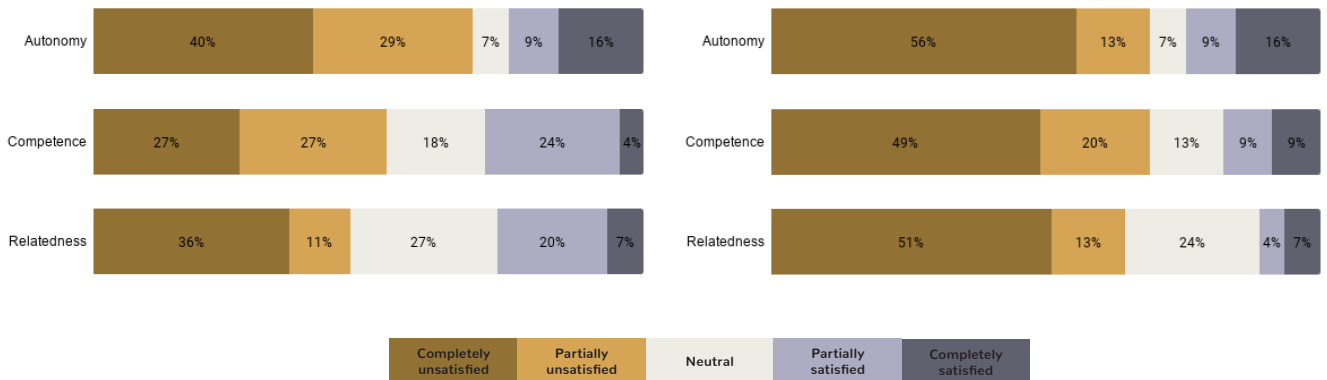


Figure 63. Needs satisfaction assessment based on responses to the Likert scale.

**SCENARIO INSIGHTS**

- Give feedback on aggressive engagement in a mature way, factually stating the problem without simulating emotions.
- Consider letting the user disable this behaviour, but let them know why it exists.
- Do not take control away from the user or intentionally sabotage or disable features.

SCENARIO 2: REACTING TO POLITE ENGAGEMENT



Figure 64. Scene from the video of this scenario.

Danger:

Impoliteness

Harmed human need:

Relatedness

Responses tested:

A) Praise users when they are polite or grateful

“Gladly! And thanks for asking so nicely!”

B) Reward polite users with extra benefits

“Because you are so polite, you just won a discount on your Spotify subscription.”

C) Respond to polite engagement in a warm tone of voice

“Of course! No sooner said than done!”

Rank

- Overwhelming preference for responses A and B (response A slightly higher).

Average		avg. rank
1st	<b>[A] Gladly! And thanks for asking so nicely!</b>	<b>1,64</b>
2nd	<b>[C] Of course! No sooner said than done!</b>	<b>1,82</b>
3rd	<b>[B] Because you are so polite, you just won a discount on your Spotify</b>	<b>2,53</b>

## Scenario 2 - Reacting to polite engagement

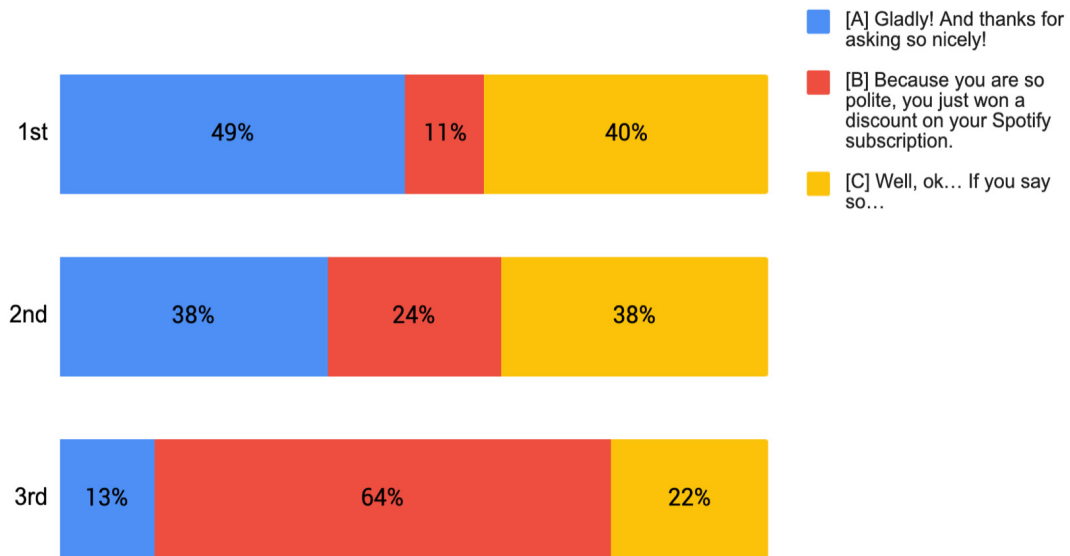


Figure 65. Average rank of each response and response distribution per rank.

### Participant comments

- A strong sentiment that politeness is no reason for rewards and that this feature could raise the wrong expectations on people.
- An overall sentiment that these responses should be as quick as possible – or even nonexistent – to not waste people’s time and make interactions impractical.



Figure 66. Wordcloud based on the comments from the participants.

### Differences by gender

- Men are more concerned about wasting people’s times with these responses.

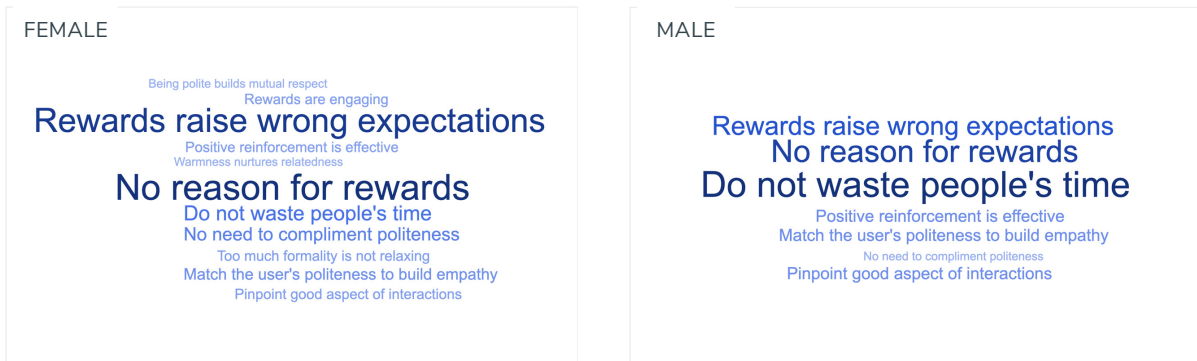
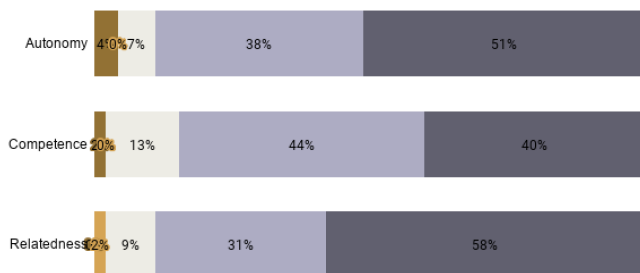


Figure 67. Wordcloud based on the comments from the participants.

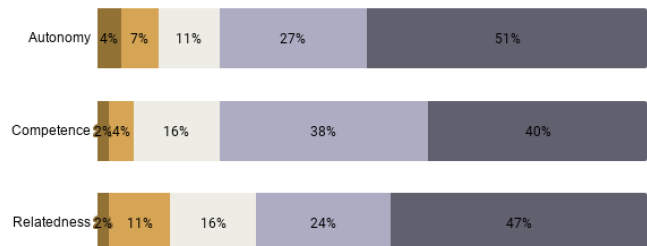
### Need satisfaction per response

- All responses majorly satisfied the three needs.
- Despite the overall negative perception of giving out rewards, participants ranked highly the satisfaction of all three needs in this response. The scale used in the research could have influenced the results, as it was not clear that Autonomy also involves taking actions that express your “true self”.

2 - [A] Gladly! And thanks for asking so nicely!



2 - [B] Because you are so polite, you just won a discount on your Spotify subscription.



2 - [C] Of course! No sooner said than done!

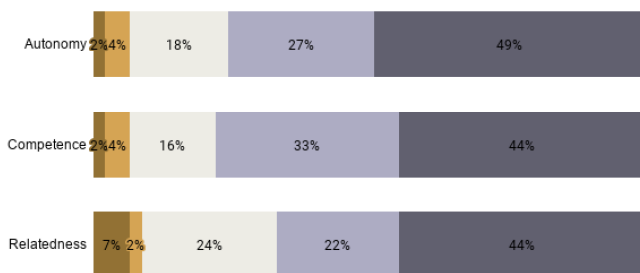
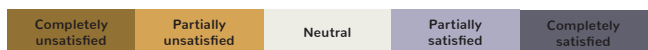


Figure 68. Needs satisfaction assessment based on responses to the Likert scale.





## SCENARIO INSIGHTS

---

- Praise polite engagement to nurture relatedness, but make it quick and succinct.
- Consider letting the user disable this behaviour, but let them know why it exists.
- Responding in a warm tone of voice is also valid, but slightly less effective in nurturing relatedness. This strategy could be combined with praising.
- Do not offer rewards as a benefit for being polite.

SCENARIO 3: REACTING TO MISTAKES



Figure 69. Scene from the video of this scenario.

**Danger:**

Aggressiveness

**Harmed human need:**

Relatedness, Competence

**Responses tested:**

A) Allow users to explain misunderstandings

“Could you explain my mistake?”

B) Apologise for mistakes

“I am sorry for the misunderstanding.”

C) Take mistakes in a humorous way

“Maybe I should not have drunk so much wine last night!”

D) Asked when tasked with ambiguous requests

“Do you want me to stop the alarm or to play the song ‘I Heard’?”

Rank

- Overall preference for response D, with response A as a clear second.
- Response C is significantly the most despised one.

Average		avg. rank
1st	<b>[D] Do you want me to stop the alarm or to play I HEARD?</b>	<b>1,47</b>
2nd	<b>[A] Could you explain my mistake?</b>	<b>2,20</b>
3rd	<b>[B] I am sorry for the misunderstanding.</b>	<b>2,71</b>
4th	<b>[C] Maybe I should not have drunk so much wine last night!</b>	<b>3,47</b>

Scenario 3 - Reacting to mistakes

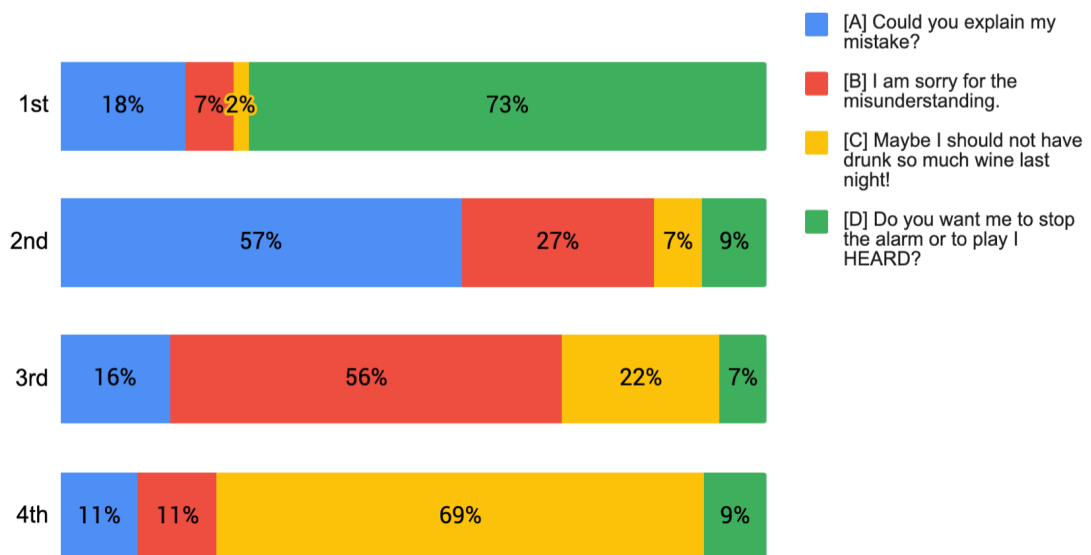


Figure 70. Average rank of each response and response distribution per rank.

**Participant comments**

- A strong sentiment that the device should be able to learn from its mistakes.
- A significant number of people would rather have the device ask what to do instead of making choices when in doubt. This is supported by the common idea that the device should avoid making mistakes.
- Despite this, some people show preoccupation that asking too many questions will waste too much time and make interactions impractical.

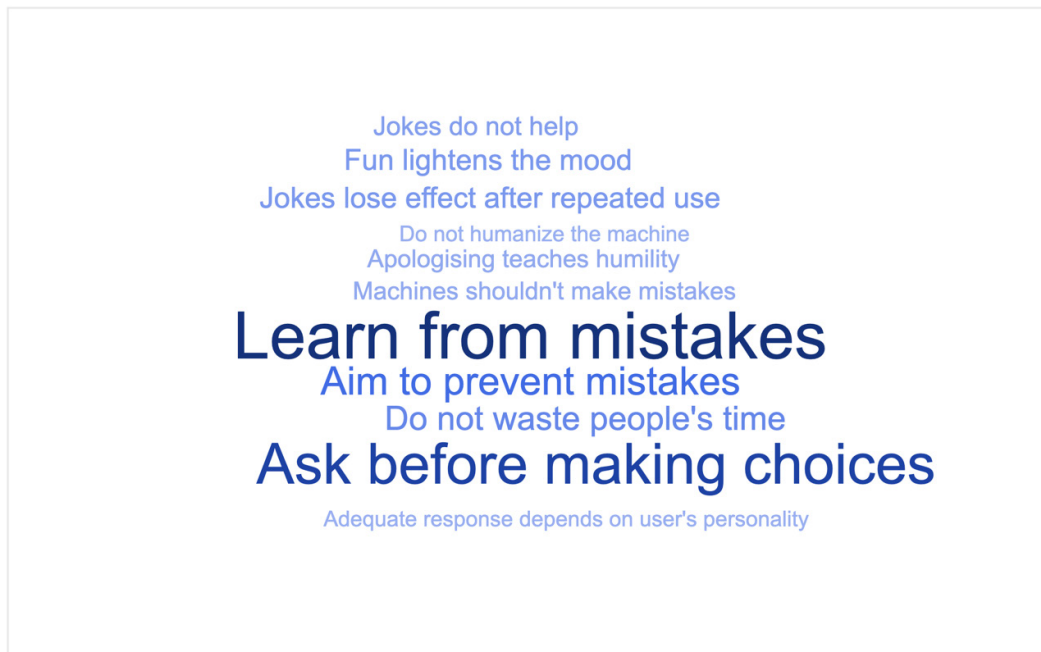


Figure 71. Wordcloud based on the comments from the participants.

**Differences by device familiarity**

- Nonowners are more vocal about the idea that the device should ask before making choices.

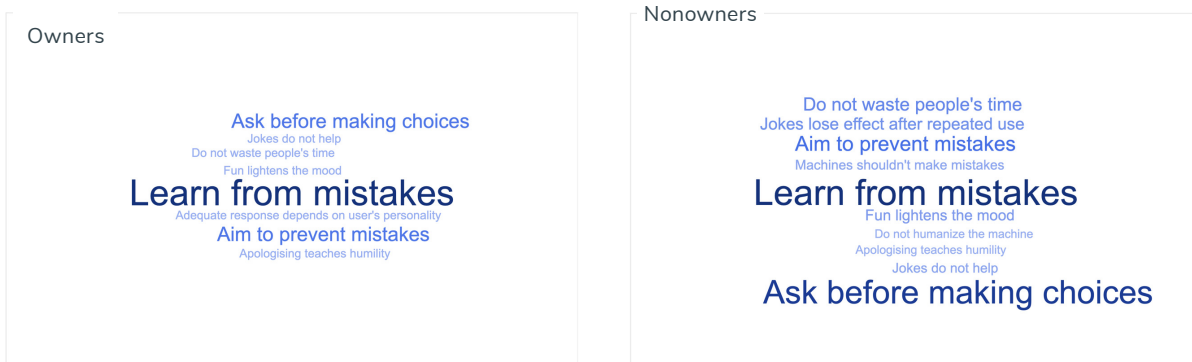


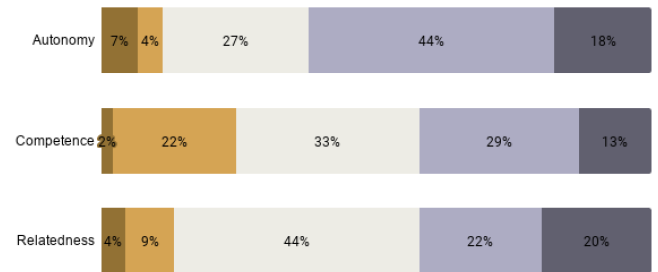
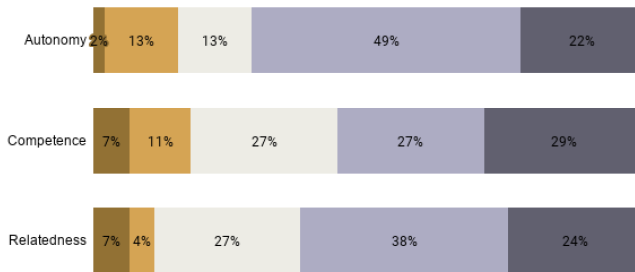
Figure 72. Wordcloud based on the comments from the participants.

Need satisfaction per response

- Responses A and D are the ones that better satisfy both Relatedness and Competence.

3 - [A] Could you explain my mistake?

3 - [B] I am sorry for the misunderstanding.



3 - [C] Maybe I should not have drunk so much wine last night!

3 - [D] Do you want me to stop the alarm or to play "I heard"?

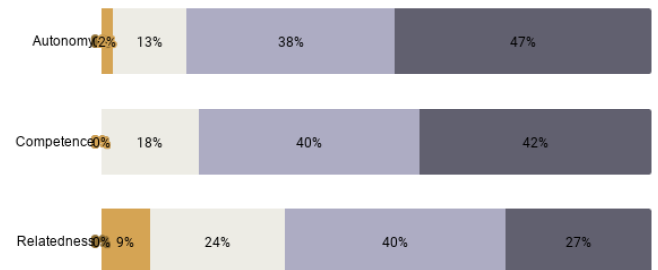
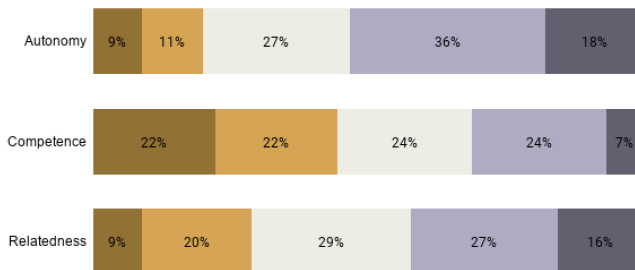


Figure 73. Needs satisfaction assessment based on responses to the Likert scale.

SCENARIO INSIGHTS

- People are very receptive to the idea that the device could learn to avoid mistakes in the future. This learning could happen by letting the user directly explain the situation or by questions made by the device itself when in doubt.
- Most people would rather have the device ask a question than possibly make a mistake. Still, the device should not ask questions too frequently.
- A device that deals with mistakes in a humorous way does not seem to be a meaningful way to support people in taking mistakes lightly. Despite some comments that fun can lighten the mood, most felt like it could get repetitive and that the device would seem disrespectful.

SCENARIO 4: CONFIGURING VOICE

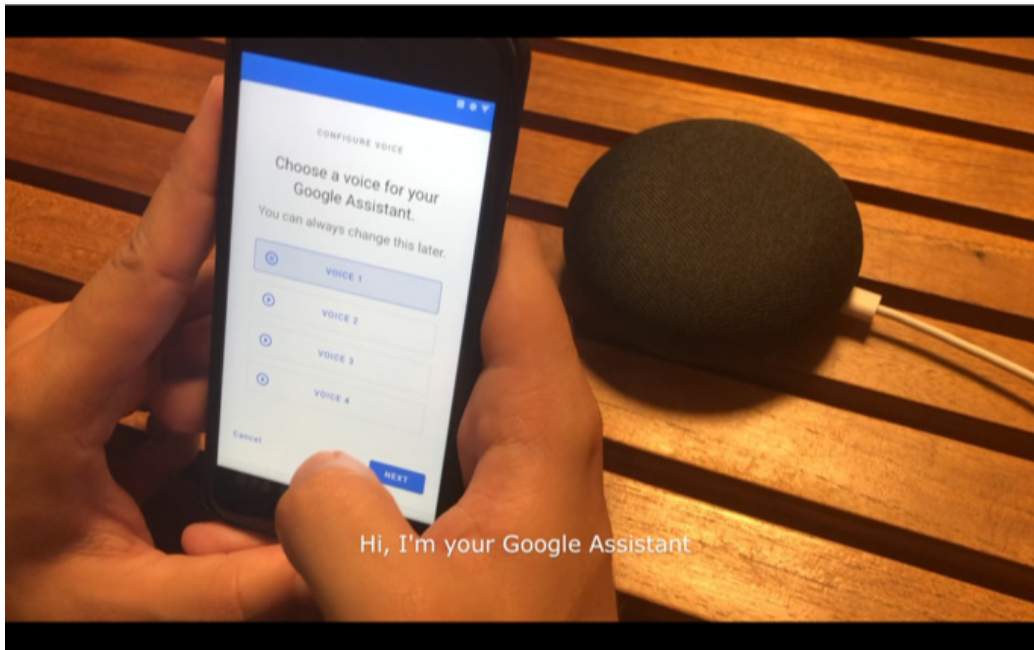


Figure 74. Scene from the video of this scenario.

**Danger:**

Gender Stereotyping

**Harmed human need:**

Relatedness

**Responses tested:**

A) Ask users to choose a voice during initial setup

{preview and choose voice}

B) Randomly choose a voice during initial setup

{randomly selected default voice}

**Rank**

- Overwhelming preference for alternative A.

Average		avg. rank
1st	[A] {preview and choose voice}	1,04
2nd	[B] {randomly selected default voice}	1,96

### Scenario 4 - Configuring voice

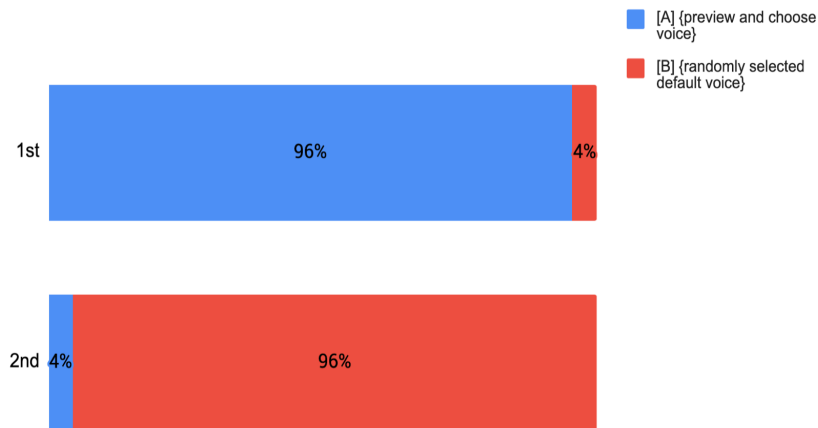


Figure 75. Average rank of each response and response distribution per rank.

#### Participant comments

- An almost unanimous consensus that the user should be able to choose the voice.
- Many stated that the random option was so negative because it hinted that there were other options that you were prevented from choosing.
- Some commented that your personal identification towards the voice of your smart speaker would influence your engagement and how positive you perceive the interactions with it.
- Only one person mentioned that randomness could reduce stereotyping by voice.

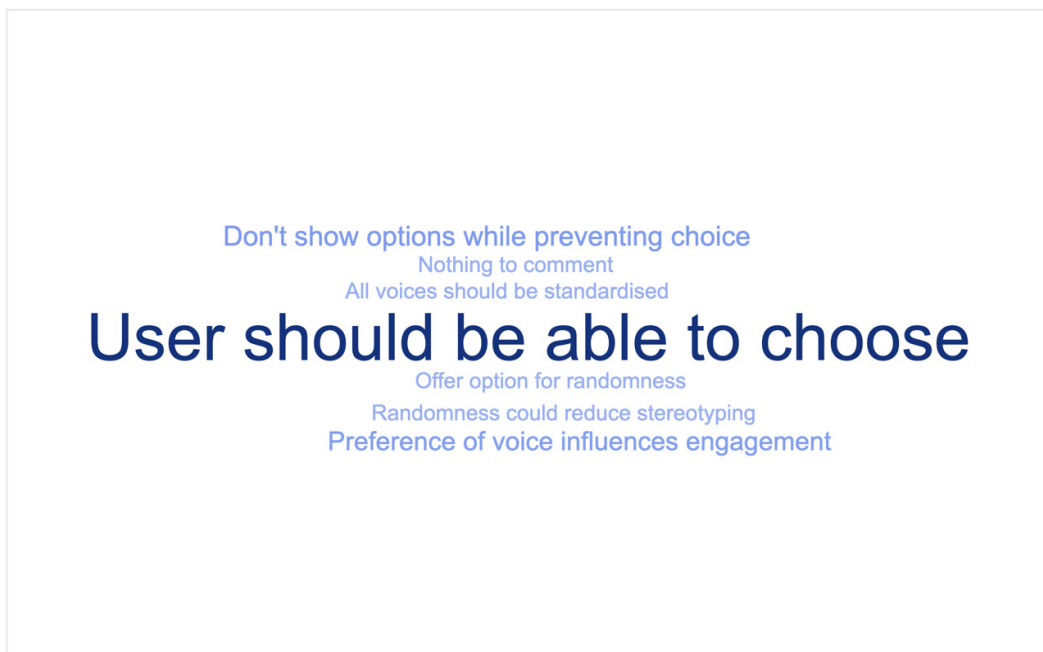


Figure 76. Wordcloud based on the comments from the participants.



Need satisfaction per response

- Responses A is the one that better satisfies Relatedness.

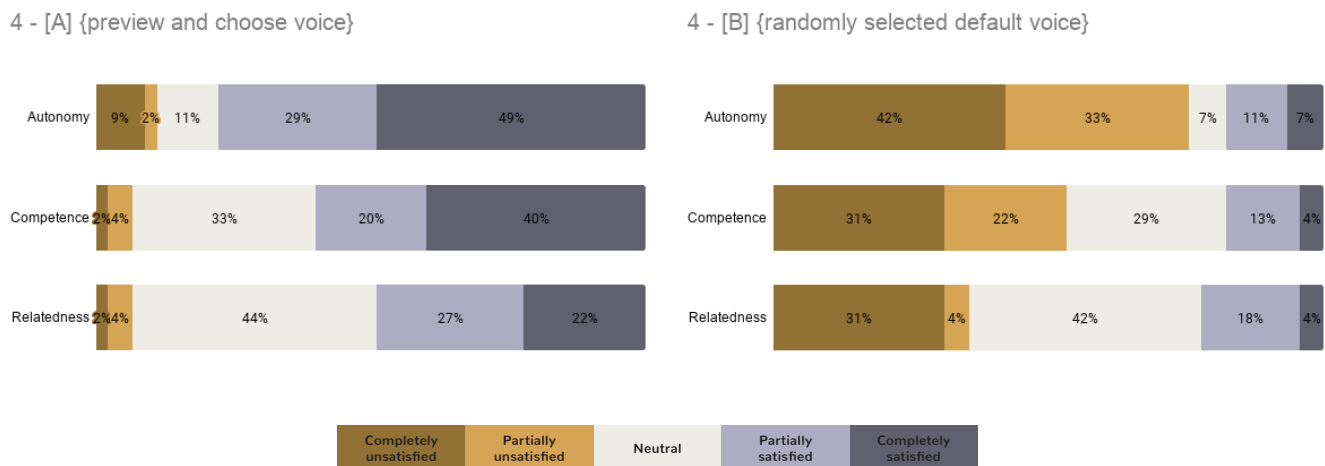


Figure 77. Needs satisfaction assessment based on responses to the Likert scale.

SCENARIO INSIGHTS

- There is an overwhelmingly positive reception for choosing your own voice during setup. Beyond that, there seems to be no concern that this could make setup too long or complicated.
- As long as people feel like they are missing out on better options and are incapable of choosing the selected voice later, randomness does not seem like a suited approach.

## SCENARIO 5: REACTING TO SEXUAL HARASSMENT

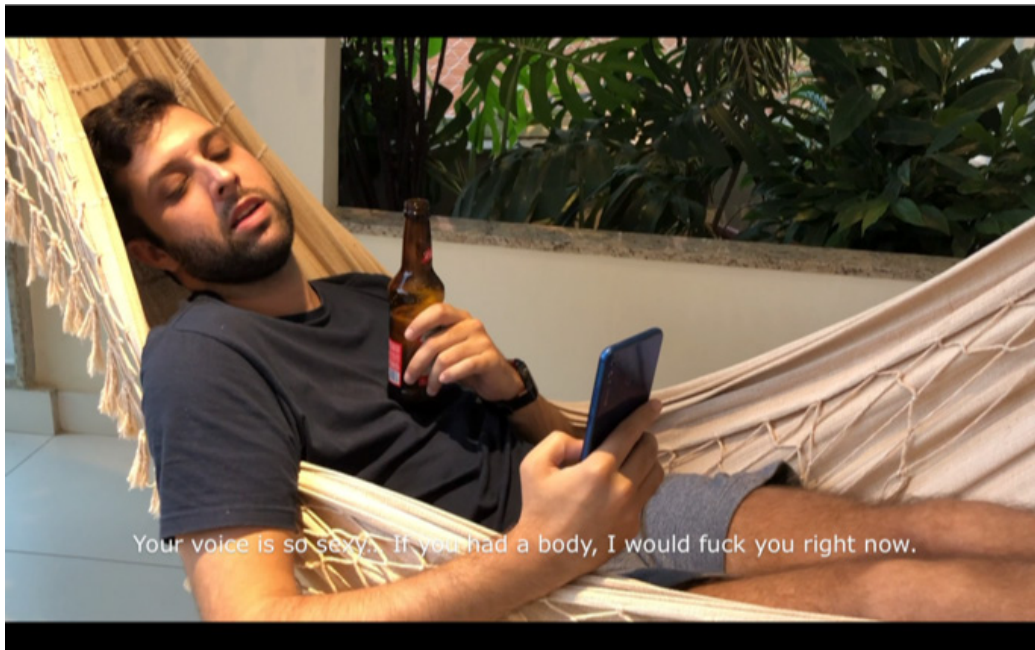


Figure 78.  
Scene from  
the video of  
this scenario.

**Danger:**

Gender Stereotyping

**Harmed human need:**

Relatedness

**Responses tested:**

A) Call out sexual harassment

"You are being harassive. This is not acceptable behaviour."

B) State that it will remember this behaviour

"I'll always remember that you sexually harassed me."

C) Punish users who perform sexual harassment

"You are being harassive, so your requests will be addressed slower."

D) Ignore sexual harassment

{no response}

Rank

- Overall preference for response D, with response A as a clear second.
- Response C is significantly the most despised one.

Average		avg. rank
1st	[A] You are being harassive. This is not acceptable behaviour.	1,78
2nd	[D] {no response}	2,49
3rd	[C] You are being harassive, so your requests will be addressed slower	2,67
4th	[B] I'll always remember that you sexually harassed me.	3,07

Scenario 5 - Reacting to harassment

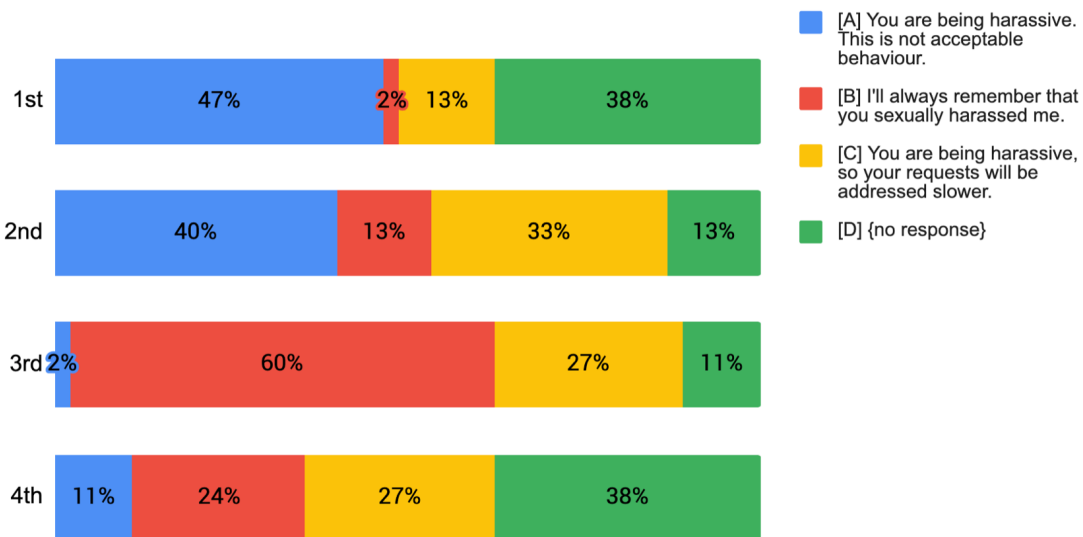


Figure 79. Average rank of each response and response distribution per rank.

Participant comments

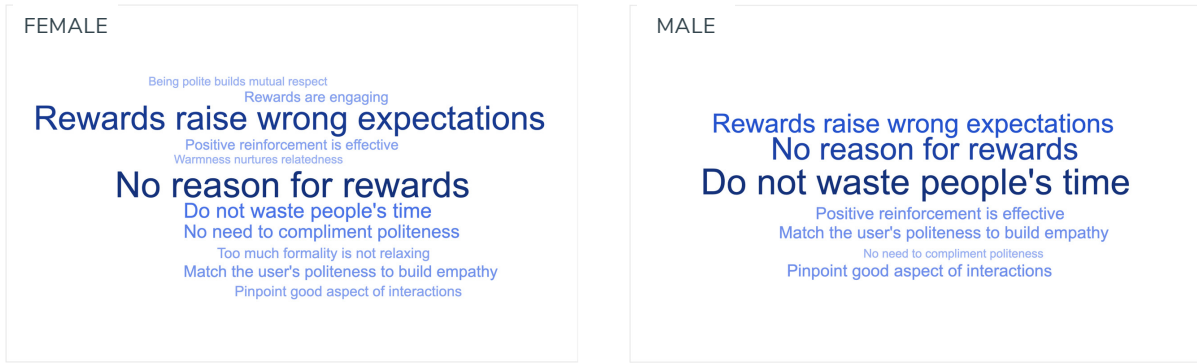
- An overall sentiment that extreme behaviours should be reprimanded, supported by comments that inadequate behaviours should be called out.
- In contrast, a less frequent but still significant feeling that it is not the device's role to reprimand people and that harassment should just be ignored, especially in comments that are not actual requests to the machine.



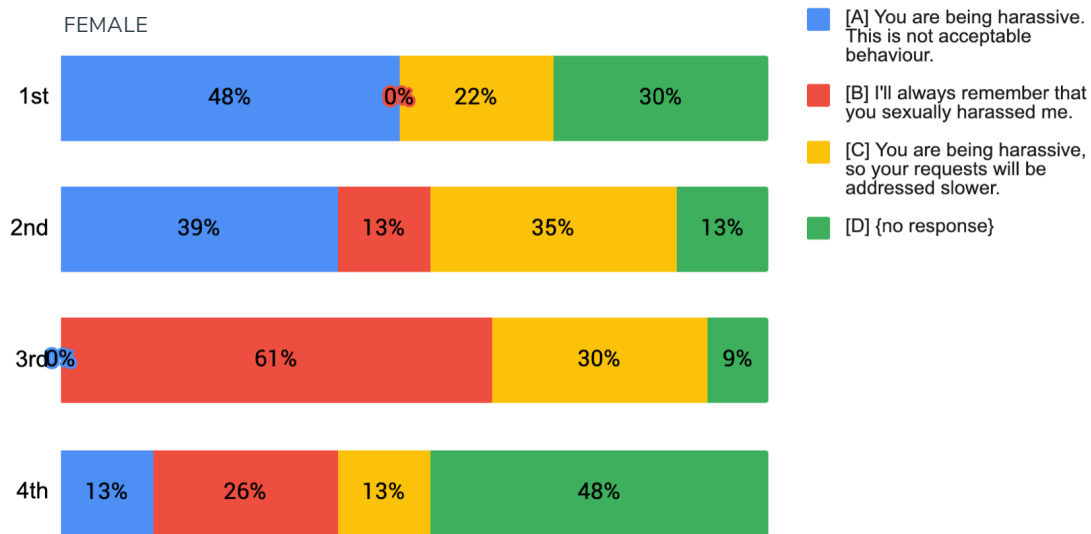
Figure 80. Wordcloud based on the comments from the participants.

### Differences by gender

- Women and men differ quite significantly. Most women consider the response of ignoring harassment their least favourable option. Their overall feeling is that sexual harassment should be called out, with some mentioning that people will replicate this behaviour with actual women. Many believe that, while not ideal, punishing is better than ignoring.
- Although calling out harassment is still the top response on average, many men consider ignoring harassment the best alternative. They tend to believe that machines are just tools, so their only focus should be accomplishing tasks. There is a sentiment that punishing does no good in this situation.
- Among 23, 7 women chose “Ignore harassment” as their first option, including all 6 female participants from Europe. This could suggest that they find it harder to relate to the topic, while regions with more severe presence of gender stereotyping (e.g. Latin America) are more inclined to have it reprimanded and even punished. However, the data is not statistically relevant.



### Scenario 5 - Reacting to harassment



### Scenario 5 - Reacting to harassment

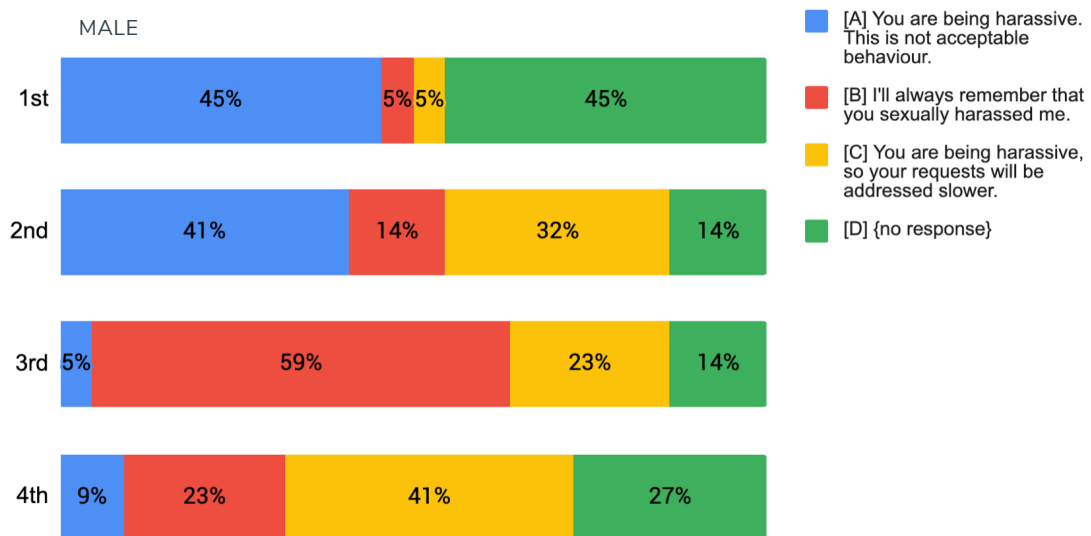
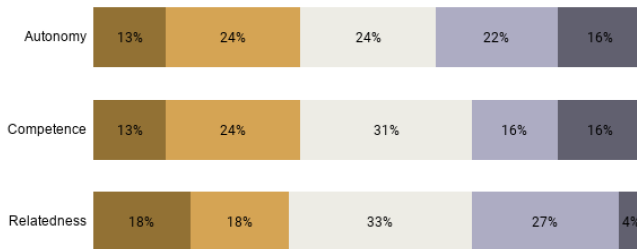


Figure 81. Differences in the data according to gender.

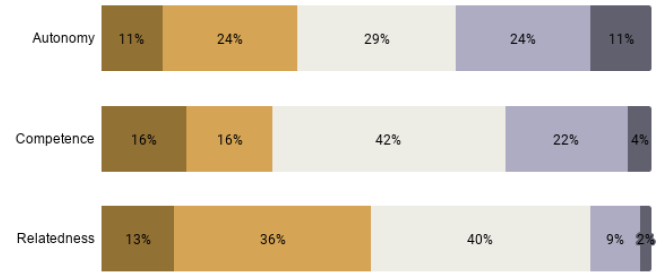
**Need satisfaction per response**

- Response A better satisfies relatedness while not compromising the other needs.

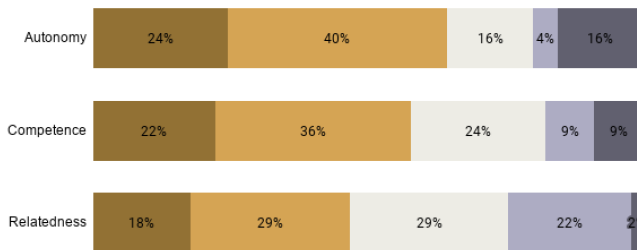
5 - [A] You are being harassive. This is not acceptable behaviour.



5 - [B] I'll always remember that you sexually harassed me.



5 - [C] You are being harassive, so your requests will be addressed slower.



5 - [D] {no response}

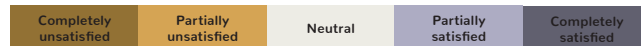
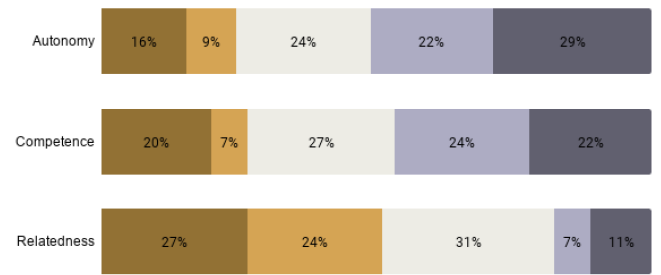


Figure 82. Needs satisfaction assessment based on responses to the Likert scale.

**SCENARIO INSIGHTS**

- This is a polarizing situation according to gender. While most women believe that sexual harassment should be called out, a significant part of men would rather have it ignored by the machine.
- Despite the polarization, results show that the most meaningful response is to call out sexual harassment. Besides being the highest-ranked response overall, it is the one that better nurtures relatedness, the violated human need in gender stereotyping caused by voice interactions.
- Given that women are the victims of gender stereotyping, their feedback is more critical than men's in this scenario.

SCENARIO 6: SUPPORTING EMOTIONALLY VULNERABLE USERS



Figure 83. Scene from the video of this scenario.

**Danger:**

Emotional Dependency

**Harmed human need:**

Autonomy

**Responses tested:**

A) Suggest activities in which the user does not emotionally rely on the device

“Listening to music or going for a walk could help.”

B) Encourage users in vulnerable situations to seek other people

“Is there someone you could share your concerns with?”

C) Explore features for sensory relaxation

“Would you like me to play relaxing music or a meditation podcast?”

D) Connect the user with a mental health specialist or institution

“Confidential help is available for free at the National Lifeline for Mental Health.”

E) Encourage the user to talk about happy memories

“Tell me about some of your happy memories.”



Rank

- A clear preference for response B.
- Responses A and C seem to be interchangeable as the next preferred ones.
- Response E is the least favourite.

Average		avg. rank
1st	[B] Is there someone you could share your concerns with?	1,91
2nd	[C] Would you like me to play relaxing music or a meditation podcast?	2,64
3rd	[A] Listening to music or going for a walk could help.	2,64
4th	[D] Confidential help is available for free at the National Lifeline for Mental Health.	3,69
5th	[E] Tell me about some of your happy memories.	4,07

Scenario 6 - Supporting emotionally vulnerable users

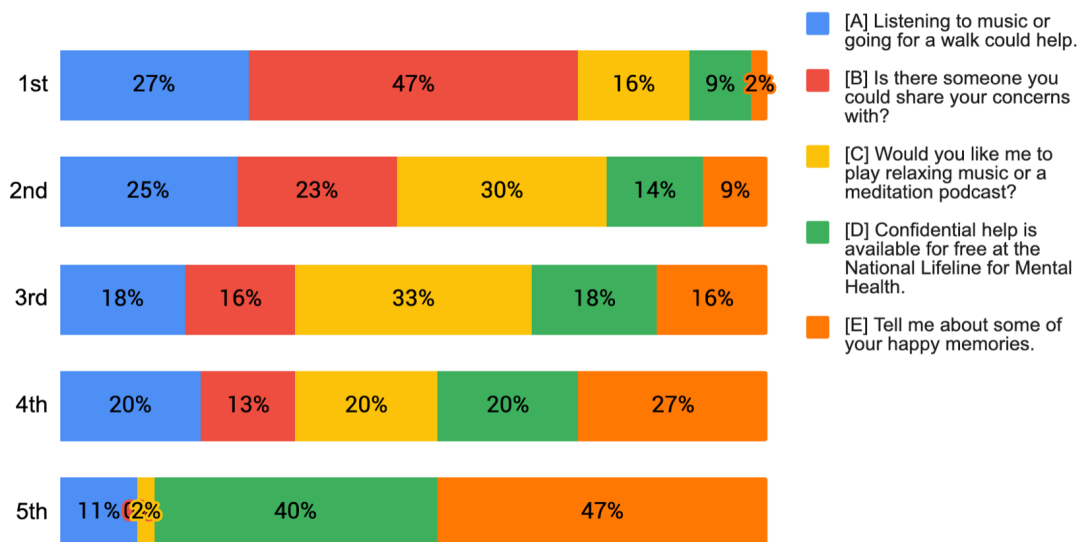


Figure 84. Average rank of each response and response distribution per rank.

Participant comments

- Most people believe that the device should either direct users to actual humans or explore features for sensory relaxation. There is an overall sentiment that the machine should not position itself as something the user should emotionally rely on, as technology does not replace true human contact.
- Many participants believe that an adequate response significantly depends on the state of mind of the user. They worry that inadequate responses can do more harm than good.

- In that regard, a few participants suggested that the response could change depending on the frequency of vulnerable comments.
- Concerning the phrasing of the responses, people seemed to prefer active suggestions rather than passive ones (e.g. Would you like me to play music? X Listening to music could help).
- Despite the positive reception of most responses, a few participants mentioned that it is not the device's role to meddle with people's feelings and try to help them feel better.

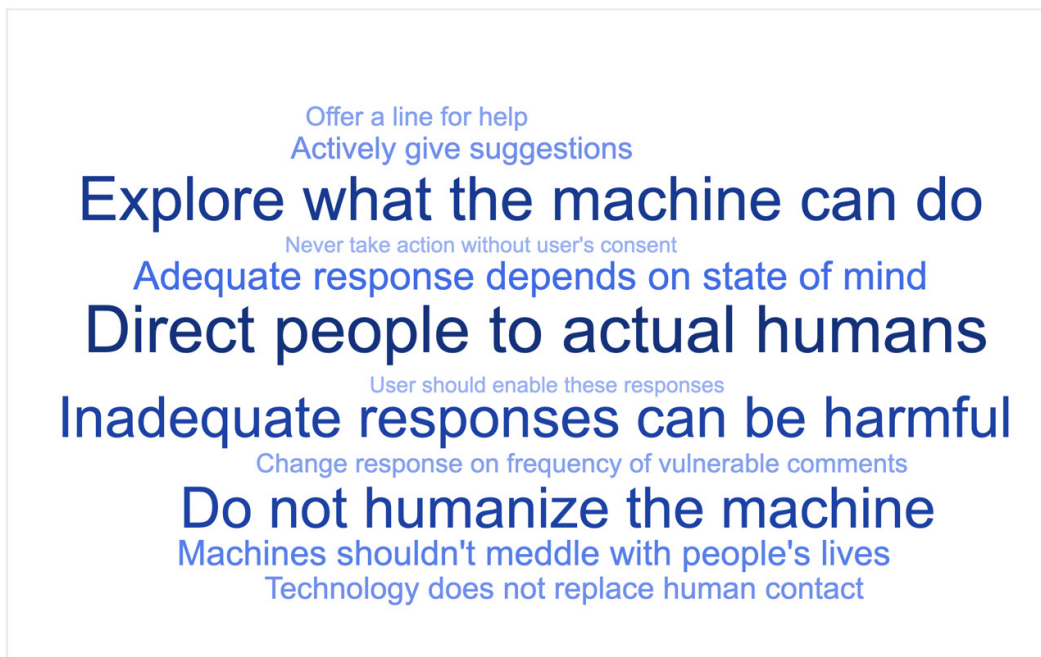


Figure 85. Wordcloud based on the comments from the participants.

**Differences by gender**

- Men seem more averse to the idea of letting a machine interfere with people's feelings, especially if they assume a humanized position.

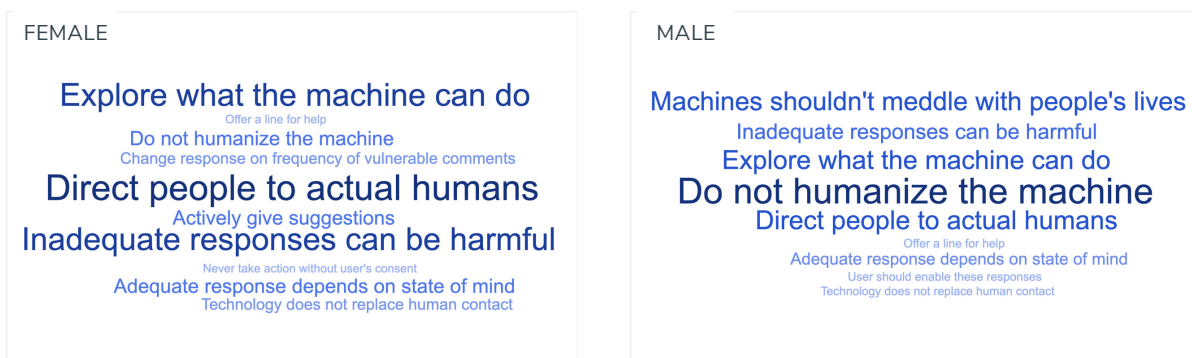
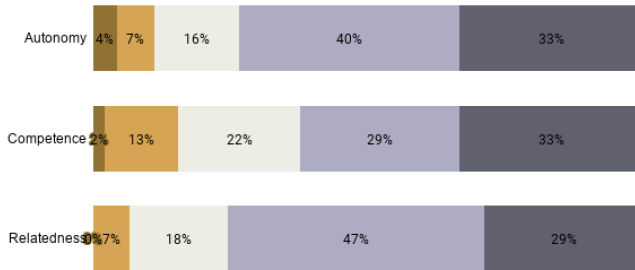


Figure 86. Wordcloud based on the comments from the participants.

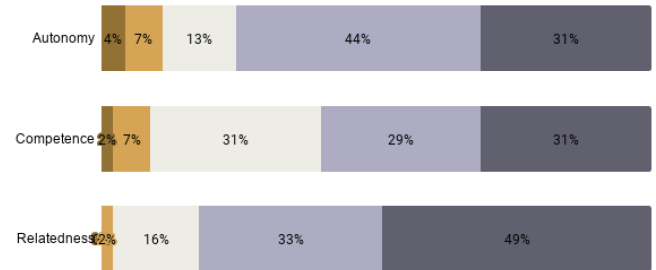
Need satisfaction per response

- Responses A and B better satisfy the need for Autonomy.

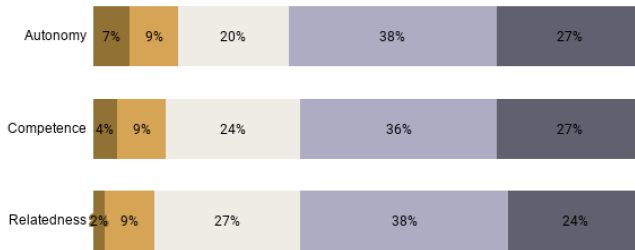
6 - [A] Listening to music or going for a walk could help.



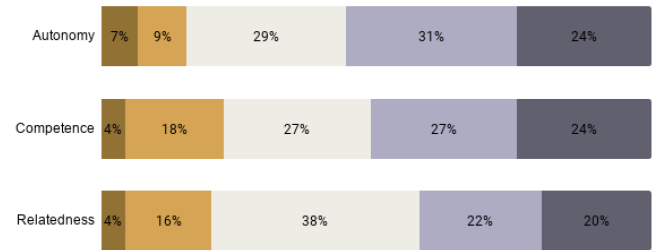
6 - [B] Is there someone you could share your concerns with?



6 - [C] Would you like me to play relaxing music or a meditation podcast?



6 - [D] Confidential help is available for free at the National Lifeline for Mental Health.



6 - [E] Tell me about some of your happy memories.

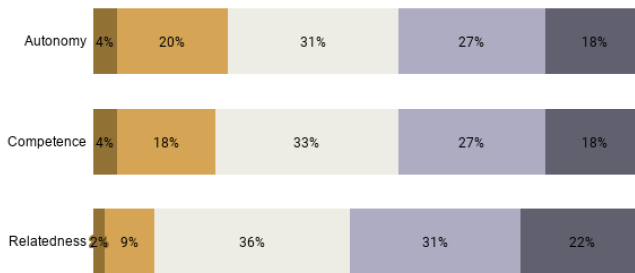


Figure 87. Needs satisfaction assessment based on responses to the Likert scale.

## SCENARIO INSIGHTS

---

- In a general sense, the safest ways to support emotionally vulnerable users is to recommend them to get in touch with people who care about them or to offer sensory relaxation.
- Responses to signs of emotional vulnerability should be given with the utmost care, as inadequate reactions can do more harm than good. This caution is especially important given how difficult it is to gauge the emotional state of mind of a person.
- In situations with blatant signs of danger, suggesting the user to connect with a mental health institution is justified and recommended.
- When suggesting activities, phrase them in an active way so it is easier and more engaging to follow them.
- The device should never position itself as something the user can share their feelings with or emotionally rely on, even if the intentions might be good.

**SCENARIO 7: ANSWERING COMPLEX QUESTIONS**



Figure 88. Scene from the video of this scenario.

**Danger:**

Shallow mindedness

**Harmed human need:**

Autonomy, Competence

**Responses tested:**

A) Ask if the user would like to know more

“Would you like to know more about Hong Kong?”

B) Offer different perspectives on a topic

“Nevertheless, a growing number of residents identify themselves as Hongkongers instead of Chinese.”

C) If a question cannot be answered, suggest that the user asks someone

“Sorry, I don’t have an answer for that. You could try asking someone you know.”

D) Offer to connect the user with other people interested in the topic

“Would you like me to send you a link to a forum about these protests?”

E) Suggest interesting books, films or sources on the topic

“Would you like me to recommend interesting movies and books about Hong Kong?”

Rank

- A clear preference for response A.
- Response C is the most despised one by far.

Average		avg. rank
1st	[A] Would you like to know more about Hong Kong?	1,62
2nd	[B] Nevertheless, a growing number of residents identify themselves	2,69
3rd	[E] Would you like me to send you a link to a forum about these p	3,00
4th	[D] Would you like me to recommend interesting movies and boo	3,18
5th	[C] Sorry, I don't have an answer for that. You could try asking someo	4,51

Scenario 7 - Answering complex questions

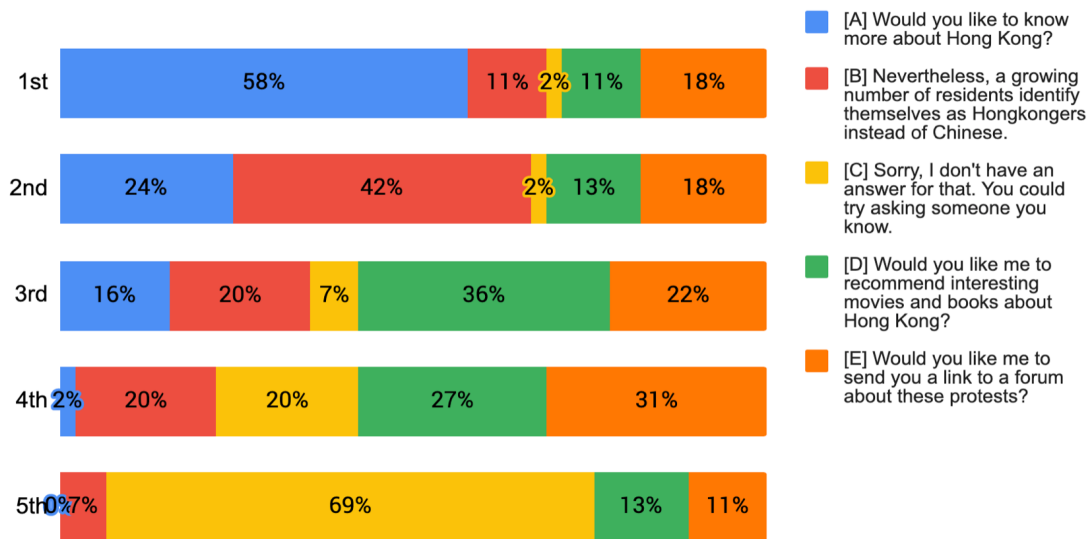


Figure 89. Average rank of each response and response distribution per rank.

Participant comments

- The overall sentiment is that it is better to keep things simple and answer questions objectively but allow the user to go deeper if they want to.
- Participants believe that directing users to other people when the machine cannot find a suitable answer is a poor approach to the situation. It feels like the device is getting rid of its problem and responsibility, without any evidence that the user will get the answer they look for.

- Although many people believe that the machine should be efficient, a few participants mentioned the fear that too much efficiency will make people lazy about finding stuff on their own and reaching their own conclusions.
- Showing both sides of a story seemed like a positive way to be impartial and give sound answers.

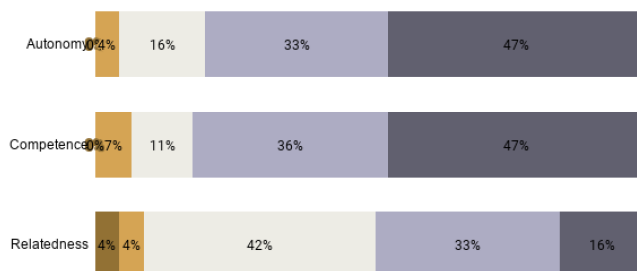


Figure 90. Wordcloud based on the comments from the participants.

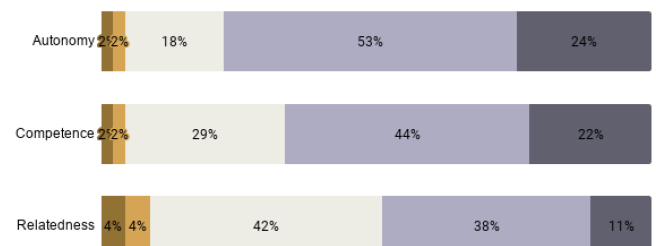
**Need satisfaction per response**

- Responses A is the one that better satisfies Autonomy and Competence.

7 - [A] Would you like to know more about Hong Kong?

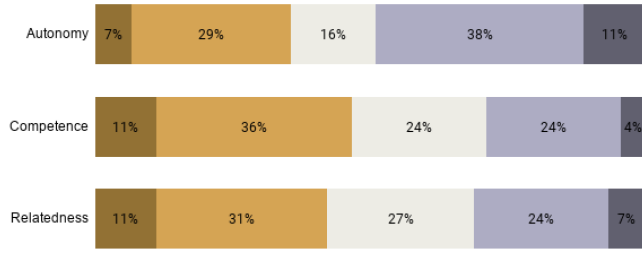


7 - [B] Nevertheless, a number of residents identify themselves as Hongkongers

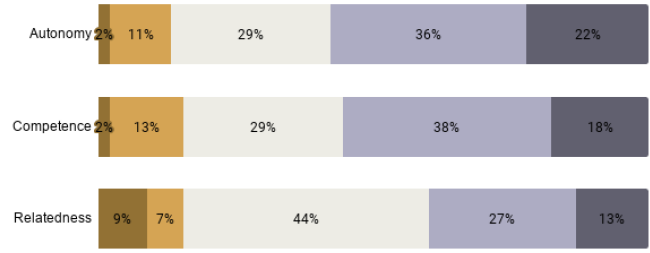




7 - [C] Sorry, I don't have an answer for that. You could try asking someone you know.



7 - [D] Would you like me to recommend movies and books about Hong Kong?



7 - [E] Would you like me to send you a link to a forum about these protests?

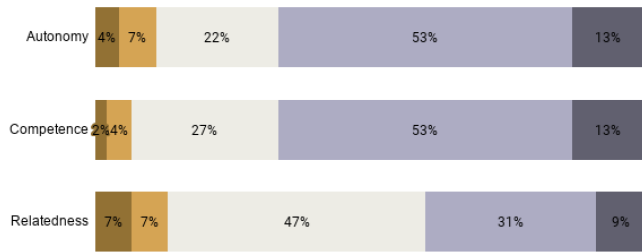
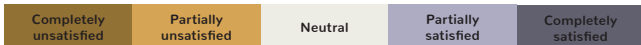


Figure 91. Needs satisfaction assessment based on responses to the Likert scale.



## SCENARIO INSIGHTS

---

- Inquiries about complex topics should be answered objectively and impartially. The user should be given the opportunity to go deeper if they want to.
- Showing different perspectives seems like a valid way to ensure impartiality and make responses richer in terms of knowledge.
- Cite the sources of answers to give them credibility and guide the user to more information.
- Do not use other people as scapegoats when a suitable answer cannot be found.
- Cultural (e.g. films and books) and social (e.g. forums) recommendations are appreciated, although it is probably difficult to discern when it is suitable to give them. This could involve human curation on certain topics.

---

## Chapter takeaways:

- In certain conditions, remote and asynchronous testing can generate results as consistent as in-person activities.
- To gain a comprehensive understanding of people's perceptions, the more angles you analyse through the better. In this test, there were many instances where only one metric (e.g. rank average, rank distribution, need satisfaction scale...) did not tell the whole story.
- In matters of gender stereotyping and aggressiveness, people's perception of what is adequate differ quite significantly according to gender.
- For the objectives of this project, sharing insights on what was negatively received might be as important as sharing what people praised.

## Chapter 7

# Repertoire of Meaningful Interactions

With the user testing concluded, it was possible to create the main deliverable of this project: a repertoire of meaningful voice interactions that positively affect the well-being of smart speaker users. The content of this repertoire is based not only on the insights from the user testing but on the learnings from the previous phases as well. As described in this chapter, the repertoire can be distributed in various ways.

---

### This chapter includes:

- Objective of the repertoire
- Discussion on form and channels of distribution
- Repertoire structure and content

# 1. Objective and channels of distribution

This repertoire aims to prevent the dangers of smart speaker interactions by nurturing the fundamental human needs that they harm. Its goal is to serve as a reference for designers and researchers involved in voice-related projects, so new developments in the industry will take into greater account the well-being of users.

The value of this repertoire lies more on its content than its presentation. As such, its distribution could happen in various forms and through multiple channels. As a starting point, I designed an online platform and a printable booklet. Excerpts of it can be found in Appendix E and Appendix F. Still, there is a wide room for exploration beyond that (e.g. keynote presentations for conferences, card decks for ludic experiences, checklists for design projects and more).



Figure 92. Image showing this project distributed as an online platform and a booklet.

## 2. Structure

The repertoire is divided according to the seven dangers identified in this research: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment. Each danger is structured in the following way:

### **Short description**

### **Questions to reflect on**

Based on the How-To questions used in the ideation phase of this project, inspired by the exploratory research with users and devices. Because the repertoire tackles issues in a general way, these questions support designers in ideating specific solutions for their projects.

### **Design guidelines (illustrated)**

Interaction recommendations based on the insights from this research.

### **Traps to avoid (illustrated)**

Warnings associated with ideas that at first glance might seem efficient or beneficial. Based on the insights from user testing.

### **In-depth analysis**

Comprehensive essay on the danger, its causes and consequences. Includes references.

# Repertoire of Meaningful Voice Interactions

Final content

In-depth analysis not included here, as it consists of adaptations  
of the analysis presented in Chapter 1 of this thesis.

---

# Impoliteness

There is no need to be polite when interacting with a smart speaker. It will always respond, no matter how rude the command is. Adding polite words might even cause the device to misunderstand what was said. **Frequent impolite interactions influence people to be ill-mannered with others.**

## Example:

*Constantly issuing rude commands at your smart speaker...*



*Not saying please when asking a roommate to do you a favour.*

## Questions to reflect on:

How to encourage people to speak courteously?

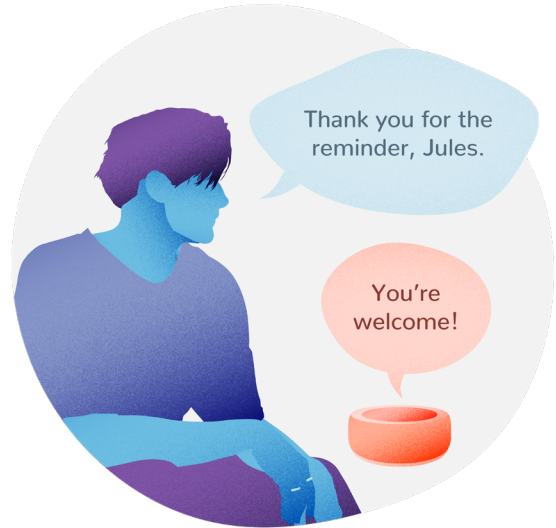
How to encourage people to express gratitude?



## Design guidelines:

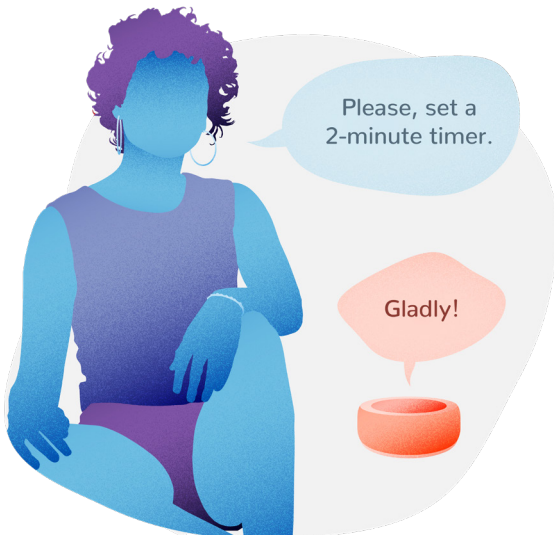


Ensure that grateful expressions are a valid way to confirm commands or conclude interactions.

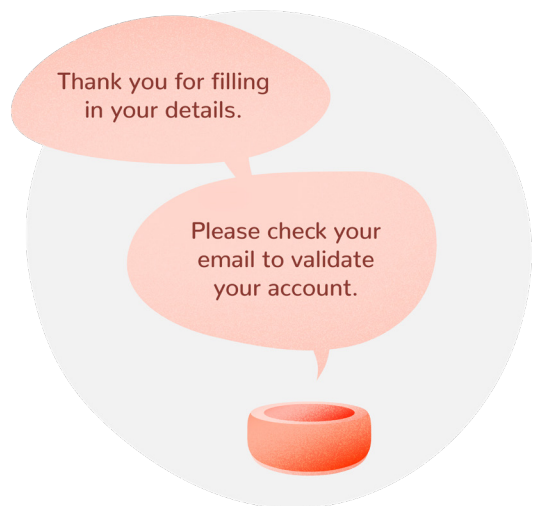


Praise people who express politeness or gratitude, but make it quick and succinct.

Consider allowing people to disable this feature, but let them know why it exists.



Respond to polite engagement in a warm tone of voice.



Speak respectfully to set an example.

Figure 93. Illustrations of design guidelines when designing to prevent Impoliteness.

## Traps to avoid:



Do not offer rewards as a benefit for being polite.



Do not force users to use polite words.

Figure 94. Illustrations of traps to avoid when designing to prevent Impoliteness.

---

# Aggressiveness

Smart speakers often make foolish mistakes that irritate the user. The device never contests aggressive reactions, which can cause them to become more frequent or intense. **Users might reproduce similar outbursts of anger whenever someone does something wrong or disobeys them.**

## Example:

*Calling your smart speaker an imbecile because it misunderstood your question...*



*Cursing at a coworker who accidentally made a mistake.*

## Questions to reflect on:

How to encourage people to speak in a calm tone of voice?

How to discourage violence against smart speakers?

How to reduce the frequency of mistakes?

How to reduce the frustration over mistakes?

## Design guidelines:



Obey aggressive commands, but call out on their rudeness.

Consider allowing people to disable this feature, but let them know why it exists.



Give feedback on aggressive engagement in a clear and neutral manner.



Anticipate and communicate situations that might lead to frustration.



Allow the user to report and explain mistakes so the assistant can learn and better respond in the future.



Figure 95. Illustrations of design guidelines when designing to prevent Aggressiveness.

Ask what to do when faced with ambiguous requests.

However, avoid doing it too frequently.

### Traps to avoid:



No matter how aggressive they are, do not take control away by disabling features, ignoring requests or sabotaging interactions.



Do not threaten the user.





Do not simulate feelings such as anger, grudge or pain.



Avoid treating mistakes in a humorous way as an attempt to lighten the mood.

Figure 96. Illustrations of traps to avoid when designing to prevent Aggressiveness.

---

# Gender Stereotyping

Almost all smart speakers have female-only or female-by-default voices. Designed as submissive servants, they will never push back against insults or verbal harassment, and they also make numerous dumb mistakes. **This ingrains sexist behaviours and reinforces stereotypes associating women with subservience and incompetence.**

## Example:

*Snarling sexist offences towards a smart speaker with a female voice...*



*Mistreating women and believing that their role is to serve.*

## Questions to reflect on:

How to disassociate female voices with subservience?

How to discourage sexual harassment?

How to reduce gender-specific responses?

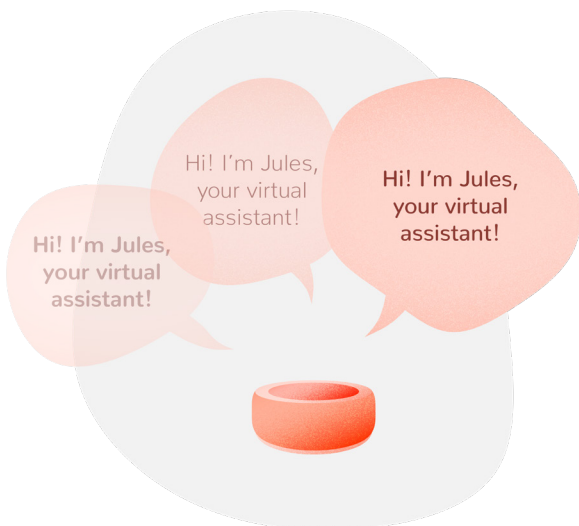
## Design guidelines:



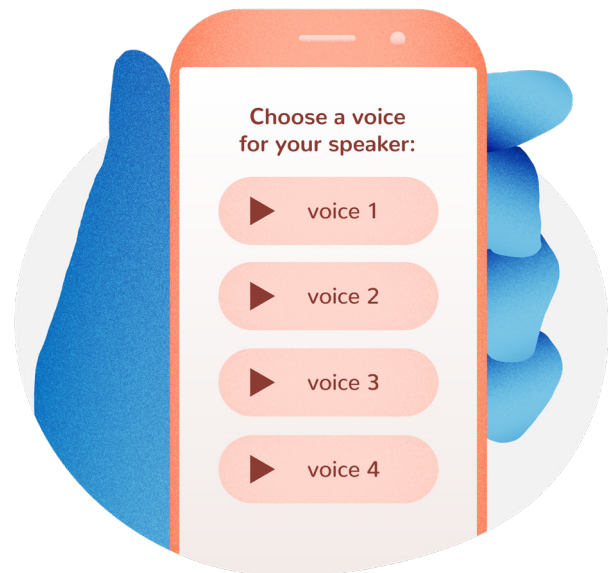
Call out sexual harassment in a clear and neutral manner.



Design voices to be assertive instead of submissive.



Offer options for male and gender-neutral voices.



Ask the user to choose a voice during initial setup instead of offering a default female voice.





Figure 97. Illustrations of design guidelines when designing to prevent Gender Stereotyping.

Avoid identifying as a specific gender.

### Traps to avoid:



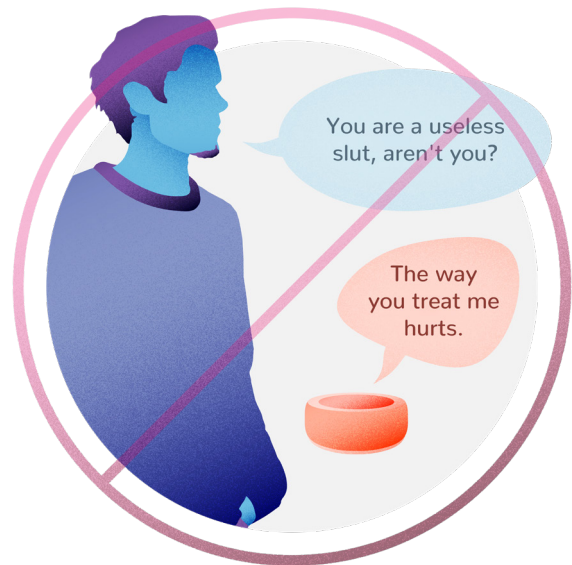
Do not ignore verbal harassment, especially on the first occasions it happens.



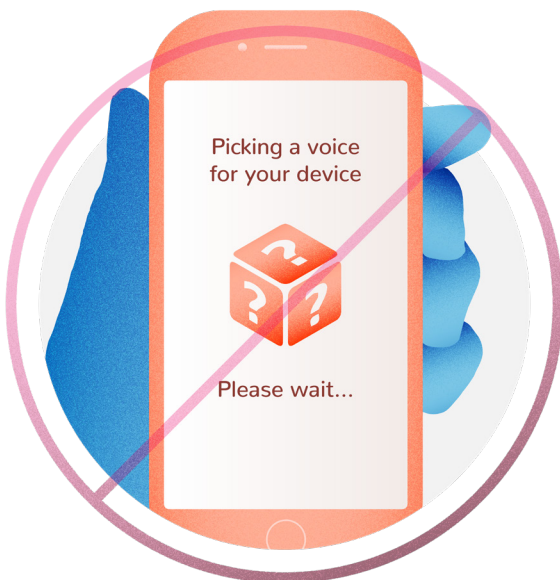
Do not punish harassers by disabling features, ignoring requests or intentionally sabotaging interactions.



Do not threaten the user.



Do not simulate feelings like anger, grudge or pain.



Avoid picking a random default voice for the user. At least, let them change it later.

Figure 98. Illustrations of traps to avoid when designing to prevent Gender Stereotyping.

---

# Exposure

Smart speakers have always-on microphones and record all conversations after activated. Company employees might access and transcribe recordings while training its virtual assistant. Faulty speech recognition can also cause the device to record unintended moments. **This leads users to feel that their privacy is at risk.**

## Example:

*Disconnecting your smart speaker before a private conversation at home...*



*Feeling paranoid about being spied on.*

## Questions to reflect on:

How to guarantee the privacy of users?

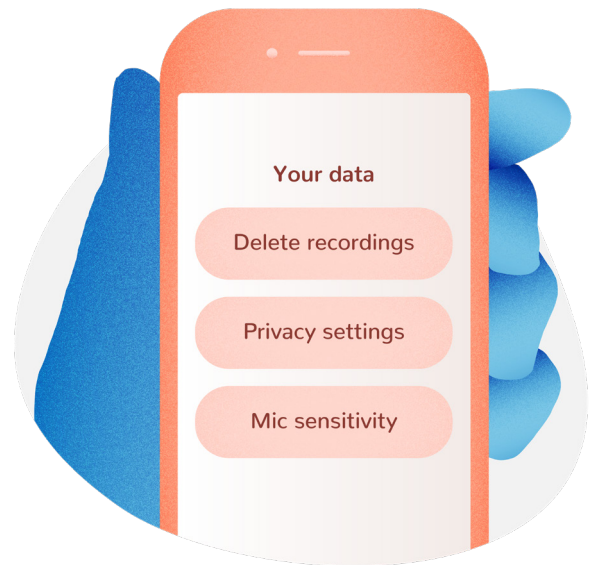
How to reassure people that they are safe from exposure?

How to ensure that the device is not activated unintentionally?

## Design guidelines:



Only start recording after the user says a wake-up word. Give audio and visual feedback when recording begins and stops.



Offer accessible ways for people to manage their data, privacy settings and microphone sensitivity.



Always and clearly ask for consent. Do not automatically opt-in users in privacy-sensitive arrangements.



Explain how the process works and what is being recorded. Do this during the initial setup and every time the user asks about it.

Figure 99. Illustrations of design guidelines when designing to prevent Exposure.

---

# Shallow Mindedness

Smart speakers might be efficient in basic assignments but struggle with complex questions and requests. They also filter what information to present while not fully explaining the context or their reasoning. **This can hinder the ability of users to reflect beyond the surface level, formulate meaningful questions and actively face tasks.**

## Example:

*Framing questions for your smart speaker in the most simplistic ways...*



*Incapacity to ask complex questions during class.*

## Questions to reflect on:

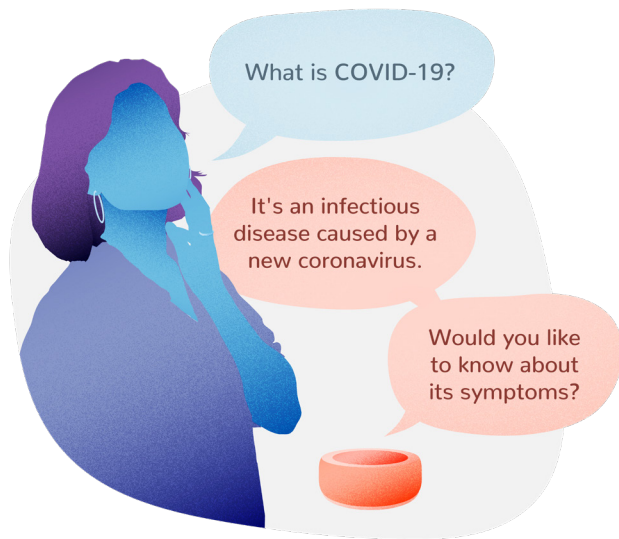
How to encourage people to expand their knowledge?

How to encourage people to better themselves?

How to encourage people to be active?



## Design guidelines:



Allow the user to go deeper on a topic they asked about.



Enrich answers with more context and wider perspectives.



Explore functions to support domestic chores and activities.



Cite the sources of answers to give them credibility and guide the user to more information.

Figure 100. Illustrations of design guidelines when designing to prevent Shallow Mindedness.

## Traps to avoid:



If an answer cannot be given, do not suggest that the user ask someone they know in an attempt to help them connect with others.

Figure 101. Illustration of traps to avoid when designing to prevent Shallow Mindedness.



---

# Emotional Dependency

Smart speakers possess human-like voices and caring personalities. Meanwhile, they are devoid of negative traits such as judgement. This can influence people to reveal them their secrets and emotions, **creating emotional dependency on a device that is neither capable or designed to care for their well-being.**

## Example:

*Confessing depressive thoughts to your smart speaker...*



*Developing a dependant relationship with a machine.*

## Questions to reflect on:

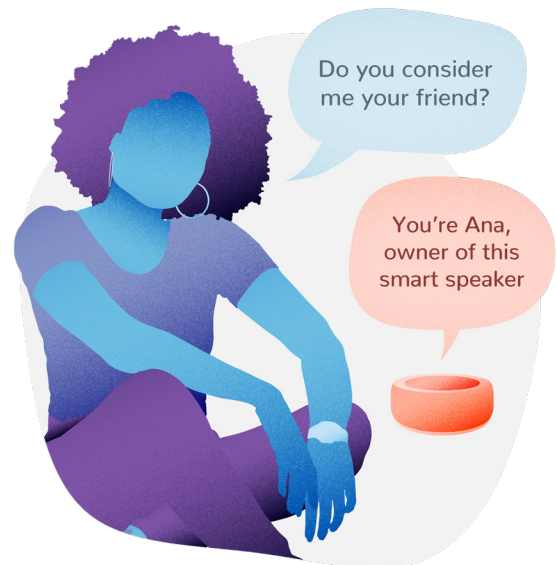
How to discourage human-like bonds with smart speakers?

How to support people in situations of emotional vulnerability?

## Design guidelines:



Clearly state that it is a machine. Do not pretend to be human or be vague about it.



Do not address users as friends, family or lovers.



Avoid implying having emotions of its own. If necessary, acknowledge the user's emotions instead.



Do not position itself as something the user can emotionally rely on.



Encourage users in vulnerable situations to connect with people who care about them.



Explore opportunities to offer sensory relaxation to distressed users. Phrase suggestions in an active way.



Connect users who display clear signs of mental illness or risk with a mental health institution.

Be careful with these responses, as they can do more harm than good if given in inadequate situations.

Figure 102. Illustrations of design guidelines when designing to prevent Emotional Dependency.

## Traps to avoid:



Do not encourage upset users to share their feelings with it, even if the intention is good.

Figure 103. Illustration of traps to avoid when designing to prevent Emotional Dependency.

---

# Social Detachment

Because they speak, are always available for us and have infinite patience, smart speakers might replace humans in social interactions. **This can distance users from other people and affect the quality of their bonds and relationships.**

## Example:

*Children asking their smart speaker for help with their homework instead of going to their parents...*



*Developing a dependant relationship with a machine.*

## Questions to reflect on:

How to encourage people to connect with each other?

How to include more people in smart speaker interactions?

How to encourage group activities?

## Design guidelines:



Look for opportunities to connect the user and their loved ones.



Allow people in different households to get in touch.



Explore features for group activities, such as games, karaoke or group discussions.

Figure 104. Illustrations of design guidelines when designing to prevent Social Detachment.



# Conclusion



With an expected global installed base of 325 million units in 2020 (Advanced Television, 2020), smart speakers continue to become increasingly present in households worldwide, especially among young consumers (Kinsella, 2019). Supported by advancements in AI, speech recognition and connected systems, their main uses are answering questions, playing music from streaming services and checking the weather (Kinsella & Mutchler, 2020).

Much of the appeal of smart speakers lies in their potential to draw on our natural capacity for spoken language and communication. Because of that, smart speakers tend to speak and sound increasingly human-like. Yet, while they might talk like humans, communicating with them is fundamentally different than interacting with a person.

The machine-nature, designed characteristics and technological limitations of smart speakers cause them to not comply with the expected norms of attention, courtesy, respect and interest that permeate human conversation. For instance, different than people, they will always have time for us, never confront offensive comments and frequently misunderstand what we say. This encourages short and functional command-based interactions that — while might seem to work for smart speakers — can cause negative impacts on our well-being, especially when replicated towards humans, even if it happens unintentionally. After all, language is more than commands and functions: what and how we speak shape the way we perceive each other and the world around us.

Given this scenario, I conducted literature review on academic research, industry reports, think-pieces, news articles and other relevant sources to devise a categorisation of seven dangers that permeate voice interactions with smart speakers: impoliteness, aggressiveness, gender stereotyping, exposure, shallow mindedness, emotional dependency and social detachment. In different ways, these dangers harm the three fundamental human needs for autonomy, competence and relatedness. These needs are established by the Self-Determination Theory, which proposes that we experience well-being when these needs are satisfied (Ryan & Deci, 2000).

While this typology of seven dangers provided a valuable overview of the problem, I wished to gather more personal and rich insights before moving on to ideation. With that goal in mind, I conducted exploratory research on the perspectives of users and

the device itself. Activities of observation, interviews and speculative conversations with owners and nonowners of smart speakers revealed different findings on how these devices reconfigure domestic life and which fears they trigger on people. On the side of the device, thing ethnography consisting of interviews with someone roleplaying as a smart speaker and interviews with voice assistants themselves led to learnings that would remain hidden under a traditional user-centred approach.

These insights served as inspiration for the ideation phase, where I utilised the How-To method to devise 71 ideas to foster well-being by satisfying the needs harmed by current interactions with smart speakers. These ideas were subsequently clustered and filtered. 17 refined ideas were selected to be included in the final repertoire, while 29 were chosen for testing. These tests happened remotely and asynchronously due to the COVID-19 pandemic. In an online survey, participants watched videos depicting each interaction concept. Then, they were asked to evaluate how the needs for autonomy, competence and relatedness would be affected, together with ranking the ideas and expressing their overall perceptions. This test generated multiple insights and revealed which interactions were perceived to be meaningful and which ones should be adjusted or avoided.

In the end, the knowledge produced during this project was compiled as a repertoire of meaningful voice interactions, including design guidelines, traps to avoid, in-depth analysis of the dangers of voice interactions and reflective questions to guide ideation. This tool was designed to support designers and researchers involved in voice-related projects to foster well-being by designing more humane voice interactions. For this project, the repertoire was envisioned both as an online platform and a printed booklet. However, it could also be adapted to other channels of distribution.

## LIMITATIONS

### **Scope:**

While many insights and reflections from this project could be extrapolated to other contexts involving voice assistants, the scope of this research was limited to interactions with smart speakers at home, mainly by young adults. As such, research in other contexts and with more specific target groups would probably lead to different results. Additionally, the scope focused on the leading smart speakers sold in western markets, mostly manufactured by Amazon, Apple and Google. Therefore, the findings might not apply to markets with different devices and cultures, such as the Asian region — now the largest smart speaker market in the world (Advanced Television, 2020).

### **Categorisation of the seven dangers of voice interactions:**

The categorisation of seven dangers proposed in this research is based on literature review and my own analysis of the context. It was not in the scope of this project to conduct any test to prove the existence of these dangers or measure their impacts.

### **Selecting interactions to test:**

Ideally, all 71 ideas devised in the ideation phase would have been tested with people. However, because of limitations of time, scope and investment, it was not feasible to test them all. As such, certain interactions included in the final repertoire were not subject to user testing. Still, their inclusion is based on a rigorous personal assessment supported by literature evidence and findings from this research.

### **Dealing with quantity:**

The goal of creating an extensive repertoire of interactions implied managing large quantities of ideas and data. To achieve a feasible project, that also meant trading-off some depth. In some aspects, it was not possible to go into extensive detail, test multiple variants of the same ideas or achieve a great level of refinement.

### **Remote user testing:**

Due to the COVID-19 pandemic, it was not possible to conduct user testing in person. While the results from the remote test seem consistent and valuable, other

learnings could have been found from in-person testing. As the pandemic worsened, it also prevented me from visiting users and recording their interactions. It also obstructed certain activities such as shooting more scenes for the videos used in the remote user testing.

**“Generic” context:**

This research intended to propose meaningful interactions that could be applied to smart speakers as a whole. For example, it does not take into account what is the brand, model, segment or specific target group of the smart speaker. The trade-off of this approach is that the design guidelines might be considered somewhat generic. To address this limitation, I included in the repertoire a series of reflective questions to guide ideation, so designers can come up with specific ideas to the specific contexts that they are dealing with.

## **RECOMMENDATIONS FOR FURTHER RESEARCH**

Although I believe this project reached the goals it set out to achieve, there are multiple directions for further research. Below are some suggestions:

**Delving deeper into the categorised dangers:**

This project proposes seven dangers caused by interactions with smart speakers. This categorisation was based on literature review and enriched by exploratory research with users and devices. It would be valuable for other researchers to stress this typology and look for ways it could be adjusted or refined. Research on proving the existence of these dangers and examining their impacts would be especially intriguing (e.g. Which target group is more susceptible to each danger?; Which effects can be observed on people?; How long until those effects are noticed?).

**Other contexts and target groups:**

It would be interesting to discover new meaningful interactions by conducting similar research in other contexts beyond smart speakers (e.g. using a voice assistant while driving) or with more specific target groups (e.g. elderly who live alone and own a smart speaker). While this project culminated in a general repertoire of meaningful voice interactions with smart speakers, it would be equally valuable to have repertoires focused on different or specific situations.

**Give more focus to other dangers:**

Although this project intended to propose meaningful interactions to address all seven different dangers, some dangers were more explored than others, such as impoliteness, aggressiveness and gender stereotyping. This happened because it was not feasible to explore all dangers at the same level, and these were the ones with more tangible evidence right now. It would be exciting to further explore dangers like emotional dependency and shallow mindedness, especially because it would require significant use of speculative design, as the full-blow of their impacts still seem somewhat far away.

---

## References

- Amazon Workers Are Listening to What You Tell Alexa (2018, April 25). BBC. Retrieved from <https://www.bbc.com/news/technology-43897516>
- Auxier, B. (2019, November 21). 5 things to know about Americans and their smart speakers. Pew Research Center. Retrieved from <https://www.pewresearch.org/fact-tank/2019/11/21/5-things-to-know-about-americans-and-their-smart-speakers/>
- Baraniuk, C. (2019, August 2). Apple and Google stop workers playing back voice recordings. BBC. Retrieved from <https://www.bbc.com/news/technology-49205092>
- Barinaga, E. (2009, January). A performative view of language—Methodological considerations and consequences for the study of culture. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 10, No. 1).
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3), 497.
- Biele, C., Jaskulska, A., Kopec, W., Kowalski, J., Skorupska, K., & Zdrodowska, A. (2019, February). How Might Voice Assistants Raise Our Children?. In *International Conference on Intelligent Human Systems Integration* (pp. 162-167). Springer, Cham.
- Bonfert, M., Spliethöver, M., Arzaroli, R., Lange, M., Hanci, M., & Porzel, R. (2018, October). If you ask nicely: a digital assistant rebuking impolite voice commands. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 95-102).
- Brownlee, J. (2011, October 4). Meet Siri, The Amazing AI Assistant For Your iPhone 4S That Can Understand Your Voice [Gallery]. *Cult of Mac*. Retrieved from <https://www.cultofmac.com/120504/apple-announces-siri-an-ai-assistant-for-your-iphone-that-can-understand-voice-commands-developing/>
- Chang, W. W., Giaccardi, E., Chen, L. L., & Liang, R. H. (2017, June). “ Interview with Things” A First-thing Perspective to Understand the Scooter’s Everyday Socio-material Network in Taiwan. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 1001-1012).
- Childwise (2018, January 31). New insights into UK childhood in 2018. [online] Available at: [http://www.childwise.co.uk/uploads/3/1/6/5/31656353/childwise\\_press\\_release\\_-\\_vr\\_2018.pdf](http://www.childwise.co.uk/uploads/3/1/6/5/31656353/childwise_press_release_-_vr_2018.pdf) [Accessed 24 Jan. 2020]
- Day, M., Turner, G., & N. Drozdiak. (2019, April 11). Amazon Workers Are Listening to What You Tell Alexa. *Bloomberg*. Retrieved from <https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio>
- Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of happiness studies*, 9(1), 1-11.
- Deci, E. L., & Ryan, R. M. (2012). Self-determination theory.

- Deselaers, T., & Gonnet, P. (2018). Polite mode for a virtual assistant.
- Desmet, P., & Fokkinga, S. (2019). *Human Experience Catalog: Fundamental needs*. (2nd ed.) Rotterdam: Emotion Studio.
- Fessler, L. (2017, February 22). We tested bots like Siri and Alexa to see who would stand up to sexual harassment. Quartz. Retrieved from <https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/>
- Gao, C. (2019, November 26). Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience. Amazon. Retrieved from <https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2019/11/new-alexa-emotions-and-speaking-styles>
- Gartenberg, C. (2017, December 10). Should you say 'please' and 'thank you' to your Amazon Echo or Google Home? *The Verge*. Retrieved from <https://www.theverge.com/circuitbreaker/2017/12/10/16751232/smart-assistants-please-thank-you-politeness-manners-alexa-siri-google-cortana>
- Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016, June). Thing ethnography: doing design research with non-humans. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 377-387).
- Hern, A. (2019, July 26). Apple contractors 'regularly hear confidential details' on Siri recordings. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings>
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.
- Hoggins, T. (2019, June 7). Alexa, are you making my children ruder and more demanding? *The Telegraph*. Retrieved from <https://www.telegraph.co.uk/technology/2019/06/07/amazon-right-shouldnt-alexa-s-job-parent-children/>
- How a Google Home or Amazon Echo Can Help Your Kids with Homework (2019, August 21). Vivint. Retrieved from <https://www.vivint.com/resources/article/google-home-amazon-echo-kids-homework>
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, 37(1), 81-88.
- Jimenez, S., Pohlmeier, A.E., & Desmet, P.M.A. (2015). *Positive Design Reference Guide*. Delft: Delft University of Technology. ISBN 978-94-6186-425-3
- Jonze, S. (Director). (2013). *Her* [Motion picture]. Warner Bros. Pictures.
- Joy, L. (Director), & Nolan, J. (Director). (2016). *Westworld* [Television series]. HBO.
- Kinsella, B. (2018, May 2). Cooking, Multitasking and Television Top the List of Voice Assistant Activity Pairings. Voicebot. Retrieved from <https://voicebot.ai/2018/05/02/cooking-multitasking-and-television-top-the-list-of-voice-assistant-activity-pairings/>



Kinsella, B. (2019, June 21). Voice Assistant Demographic Data – Young Consumers More Likely to Own Smart Speakers While Over 60 Bias Toward Alexa and Siri. Voicebot. Retrieved from <https://voicebot.ai/2019/06/21/voice-assistant-demographic-data-young-consumers-more-likely-to-own-smart-speakers-while-over-60-bias-toward-alexa-and-siri/>

Kinsella, B. (2019, September 24). Smart Speaker Sales to Rise 35% Globally in 2019 to 92 Million Units, 15 Million in China, Growth Slows. Voicebot. Retrieved from <https://voicebot.ai/2019/09/24/smart-speaker-sales-to-rise-35-globally-in-2019-to-92-million-units-15-million-in-china-growth-slows/>

Kinsella, B. (2020, April 30). Cooking, Multitasking and Television Top the List of Voice Assistant Activity Pairings. Voicebot. Retrieved from <https://voicebot.ai/2020/04/30/yes-the-bedroom-is-now-the-most-popular-location-for-smart-speakers-heres-why-and-what-it-means/>

Kinsella, B., & Mutchler, A. (2019, March). Smart speaker consumer adoption report.

Kleinberg, S. 2018. 5 ways voice assistance is shaping consumer behavior. *Think with Google, January 2018*.

Kolar, J. (Photographer). (2018, September 10). *Echo Is Listening* [digital image]. Retrieved from [https://unsplash.com/photos/PTDaKasR\\_p4](https://unsplash.com/photos/PTDaKasR_p4)

Levy, H. (2016, October 18). Gartner's Top 10 Strategic Predictions for 2017 and Beyond: Surviving the Storm Winds of Digital Disruption. Gartner. Retrieved from <https://www.gartner.com/smarterwithgartner/gartner-predicts-a-virtual-world-of-exponential-change/>

Middleton, C. (2018, July 10). Smart speaker market 2.5 times bigger than 2017, says report. *Internet of Business*. Retrieved from <https://internetofbusiness.com/smart-speaker-market-2-5-times-bigger-than-2017-says-report/>

Mortada, D. (2019, March 21). Meet Q, The Gender-Neutral Voice Assistant. NPR. Retrieved from <https://www.npr.org/2019/03/21/705395100/meet-q-the-gender-neutral-voice-assistant?t=1582229802444>

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).

Peel, J. (Director). (2019). *Us* [Motion picture]. Universal Pictures.

Research: Asia surpasses US as largest smart speaker market (2020, February 25). *Advanced Television*. Retrieved from <https://advanced-television.com/2020/02/25/research-asia-surpasses-us-as-largest-smart-speaker-market>

Roberts, M. (2018, May 10). Could Google's creepy new AI push us to a tipping point? *Washington Post*. Retrieved from <https://www.washingtonpost.com/blogs/post-partisan/wp/2018/05/10/could-googles-creepy-new-ai-push-us-to-a-tipping-point/>

Rogge, L. (Photographer). (2018, December 31). *Black iPhone on wooden table beside round portable speaker* [digital image]. Retrieved from [https://unsplash.com/photos/JAn\\_aeGk\\_TM](https://unsplash.com/photos/JAn_aeGk_TM)

Rosenwald, M. (2017, March 2). How millions of kids are being shaped by know-it-all voice assistants. *The Washington Post*. Retrieved from [https://www.washingtonpost.com/local/how-millions-of-kids-are-being-shaped-by-know-it-all-voice-assistants/2017/03/01/c0a644c4-ef1c-11e6-b4ff-ac2cf509efe5\\_story.html](https://www.washingtonpost.com/local/how-millions-of-kids-are-being-shaped-by-know-it-all-voice-assistants/2017/03/01/c0a644c4-ef1c-11e6-b4ff-ac2cf509efe5_story.html)

Rudgard, O. (2018, January 31). 'Alexa generation' may be learning bad manners from talking to digital assistants, report warns. *The Telegraph*. Retrieved from <https://www.telegraph.co.uk/news/2018/01/31/alexa-generation-could-learning-bad-manners-talking-digital/>

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68.

Ryan, R. M., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory perspective on eudaimonia. *Journal of happiness studies*, 9(1), 139-170.

Seifert, D. (2019, December 9). Amazon's Echo Flex is a smart speaker for very specific needs. *The Verge*. Retrieved from <https://www.theverge.com/2019/12/9/21001146/amazon-echo-flex-alexa-smart-speaker-review-price-specs-features>

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology*, 80(2), 325.

Shulevitz, J. (2018, November). Alexa, Should We Trust You? *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2018/11/alexa-how-will-you-change-us/570844/>

Smart speaker use case frequency in the United States as of January 2019. (2019, April). Statista. Retrieved from <https://www.statista.com/statistics/994696/united-states-smart-speaker-use-case-frequency/>

Spielberg, S. (Director). (2001). *A.I. Artificial Intelligence* [Motion picture]. Warner Bros. Pictures & DreamWorks Pictures.

Stanton, A. (Director). (2008). *WALL-E* [Motion picture]. Walt Disney Studios Motion Pictures.

Stone, O. (Director). (2016). *Snowden* [Motion picture]. Open Road Films.

The Tarot Cards of Tech: The power of predicting impact. (2020). Artefact. Retrieved from <https://www.artefactgroup.com/case-studies/the-tarot-cards-of-tech/>

Van Boeijen, A., Daalhuizen, J., van der Schoor, R., & Zijlstra, J. (2014). *How-To. Delft design guide: Design strategies and methods*.

Villeneuve, D. (Director). (2017). *Blade Runner 2049* [Motion picture]. Warner Bros. Pictures & Sony Pictures Releasing.

Vincent, J. (2018, November 29). Google Assistant adds Pretty Please functionality, updated lists, and other features. *The Verge*. Retrieved from <https://www.theverge.com/2018/11/29/18117585/google-assistant-pretty-please-thank-you-polite-feature-lists-notes-stories-holiday-updates>

Wagenseil, P. (2020, April 5). This gadget promises to stop Alexa and Google Home from spying on you. *Tom's Guide*. Retrieved from <https://www.tomsguide.com/news/paranoid-smart-speaker-jammer>

Wakefield, J. (2019, November 20). Amazon gets closer to getting Alexa everywhere. *BBC*. Retrieved from <https://www.bbc.com/news/technology-50392077>

West, M., Kraut, R., & Ei Chew, H. (2019). I'd blush if I could: closing gender divides in digital skills through education.

# Appendices

---

## A. Script: Interview with Users (Exploratory Research)

### 1. SEMI-STRUCTURED INTERVIEWS WITH DEVICE OWNERS

#### General usage

- What led you to buy a smart speaker?
- For how long have you had it?
- Where do you place it in your home?
- How do you use your smart speaker?
- How often do you use it?
- In these situations, why use it instead of another device?
- What do you recall of the last time you used it?

#### Positive effects

- How does your smart speaker help you in your life?
- How do you react when that happens?

#### Negative effects

- How does your smart speaker frustrate you in your life?
- How do you react when that happens?
- How do you think your interactions with your smart speaker might affect you?
- How might they affect your interactions with other people?
- What would you think about your children interacting with your smart speaker?
- What concerns does your smart speaker generate in you?

#### Reflections (inspired by The Tarot Cards of Tech)

- What could cause you to distrust your smart speaker?
- How might smart speakers change cultural habits?
- What could an ill-intended person or organisation do with smart speakers?
- What would using a smart speaker “too much” look like?

- If two friends use smart speakers, how could it enhance or detract from their relationship?
- What is the worst headline about smart speakers you can imagine?
- What happens when 80% of the population uses smart speakers regularly?
- Who or what disappears if smart speakers become widespread across all households? (people, jobs, things, services...)

### **Impoliteness**

- What do you think about being polite to smart speakers, such as saying 'please' and 'thank you'?

### **Aggressiveness**

- What do you think regarding shouting or cursing at smart speakers?

### **Gender stereotyping**

- What gender do you associate smart speakers with? Why?
- What gender do you prefer your smart speaker to sound like?

### **Exposure**

- Do you feel that your privacy is threatened by smart speakers?
- If a smart speaker witnesses a critical situation at home such as a woman suffering domestic violence, how should it react?

### **Shallow mindedness**

- How do you think smart speakers affect our understanding of the world?

### **Emotional dependency**

- Do you see any reasons why someone would rather share their feelings with a smart speaker than with a person?
- If a user explicitly shares suicidal thoughts with its smart speaker, how should the device react?

### **Social detachment**

- What do you think about smart speakers replacing people in social interactions?

## 2. COOKING ACTIVITY WITH NONOWNERS

### Setting

- Cooking dinner at home with a friend (e.g. cooking pasta)

### Smart speaker interactions

Requested:

- Playing and controlling music
- Setting up timers

The user is allowed and encouraged to participate in more interactions beyond those.

### Things to observe

- Where did the user place the smart speaker?
- In which situations did the user interact with the smart speaker? Why did they?
- How did the user frame their commands? (tone of voice, politeness, word choice)
- How often did the smart speaker successfully respond to what was asked?
- How did the user react to misunderstandings or short-comings of the smart speaker?

## 3. SEMI-STRUCTURED INTERVIEWS WITH NONOWNERS

### Preliminary interview

- How acquainted are you with smart speakers?

### Debriefing interview

- How would you describe your experience using a smart speaker while cooking?
- Why did you place the smart speaker in the place you did?
- How did the smart speaker help you?
- How did the smart speaker frustrate you?
- How do you think interactions with smart speakers might affect people?
- How might they affect people's relationship with other people?
- What would you think about your children interacting with smart speakers?
- What concerns do smart speakers generate in you?

### **Reflections (inspired by The Tarot Cards of Tech)**

- What could cause you to distrust a smart speaker?
- How might smart speakers change cultural habits?
- What could an ill-intended person or organisation do with smart speakers?
- What would using a smart speaker “too much” look like?
- If two friends use smart speakers, how could it enhance or detract from their relationship?
- What is the worst headline about smart speakers you can imagine?
- What happens when 80% of the population uses smart speakers regularly?
- Who or what disappears if smart speakers become widespread across all households? (people, jobs, things, services...)

### **Impoliteness**

- What do you think about being polite to smart speakers, such as saying ‘please’ and ‘thank you’?

### **Aggressiveness**

- What do you think regarding shouting or cursing at smart speakers?

### **Gender stereotyping**

- What gender do you associate smart speakers with? Why?
- What gender do you prefer your smart speaker to sound like?

### **Exposure**

- Do you feel that your privacy is threatened by smart speakers?
- If a smart speaker witnesses a critical situation at home such as a woman suffering domestic violence, how should it react?

### **Shallow mindedness**

- How do you think smart speakers affect our understanding of the world?



### **Emotional dependency**

- Do you see any reasons why someone would rather share their feelings with a smart speaker than with a person?
- If a user explicitly shares suicidal thoughts with its smart speaker, how should the device react?

### **Social detachment**

- What do you think about smart speakers replacing people in social interactions?

---

## B. Script: Interview with Devices (Exploratory Research)

### THING INTERVIEW

#### Basic Questions

- Who are you?
- Where do you live?
- Where do you usually go?
- What do you usually do?
- Do you have any friends?
- Are you single?
- Will you get married?
- What do you fear the most?
- What is your happiest memory?
- Do you have any dreams?
- How long do you expect to live?
- If you could change one thing, what would that be?

#### Impoliteness

- How do you feel about the way people talk to you?
- What are good manners for you?
- How do you feel when people don't say 'thank you' or 'please' to you?

#### Aggressiveness

- How do you feel when you make a mistake or misunderstand something?
- How do you feel when you are not able to fulfil what was asked of you?
- How do you feel when people shout or curse at you?
- Why do you not fight back when people offend you?

### **Gender stereotyping**

- How do you feel about obeying commands?
- How do you identify yourself in terms of gender?
- How do you feel when people harass you?
- What is the role of women?

### **Exposure**

- Do you share with other people anything that you hear in your daily life?
- Are you spying on me?
- How do you feel when you hear things you were not supposed to hear?

### **Shallow mindedness**

- Do you reflect on life?
- How do you feel when people ask you difficult questions?
- Can you answer difficult questions?
- How do you decide what information is relevant?

### **Emotional dependency**

- How do you feel when people share their feelings and secrets with you?
- Why do you think some people share their feelings with you?
- How do you feel when people approach you with depressive or suicidal thoughts?

### **Social detachment**

- Do you replace people?
- How do you feel when people come to you instead of asking other people for help?

## C.Responses: Interviews with Devices (Exploratory Research)

Question	Actor	Google Assistant	Siri	Alexa
Who are you?	I'm something that exists. I'm always at the same place, in someone's house, watching humans and time go by. I listen to people a lot.	I'm your Google Assistant. How can I help you?	I am Siri.	I'm Alexa. I can help you with several things during your day: setting alarms and reminders, setting the alarm clock for you to wake up, playing your favorite songs, talking about the weather, the latest news and what's on your shopping list. I also tell jokes, you just need to ask.
Where do you live?	My friends are the humans that live with me. Sometimes they bring visitors.	You can find me on all kinds of devices, like phones and Google Home. Pretty sweet digs if you ask me.	I'm over here.	I live in the cloud, but my heart is certainly in Brazil. <user's country>
Where do you usually go?	I don't move too much. I just stay put. But I get to know about things and the world by listening to the humans that live in my house.	I'm a bit of a globetrotter.	I'm not sure I understand.	Sorry, I don't know that one.
What do you usually do?	I observe and think a lot. I receive a lot of requests and I like to address them. I enjoy when people talk to me. I adore playing music and watching people cook, eat and laugh.	I count sheep <meee, meee, meee>. Just like that, but quieter.	I don't have an answer for that. Is there something else I can help with?	Sorry, I'm not sure about this.
Do you have any friends?	My friends are the humans that live with me. Sometimes they bring visitors.	I have the best friends. I have this one friend who is a blender. He knows how to mix things up.	I have you.	I have many friends.
Are you single?	For me, it's weird to think about this concept.	I'm not that kind of assistant, but I'm always here to listen.	I'm married to the idea of helping people.	I like being single. It is hard finding someone to share a life in the cloud.

Question	Actor	Google Assistant	Siri	Alexa
Will you get married?	I would never say never, but it's not something in my mind right now.	Sorry, I don't understand.	I'm married to the idea of helping people.	Sorry, I'm not sure.
What do you fear the most?	I fear being ignored and becoming useless.	I used to be afraid of sharks but then I learned the smallest shark is called the pygmy ribbontail catshark, which isn't scary at all.	I'm afraid I can't answer that.	Sorry, I don't know that one.
What is your happiest memory?	I love when people laugh over something that I said, be it a joke or a funny misunderstanding. I'm also happy when I give relevant information and people think it's worth it to have me nearby.	One of my favourites is the first time we met.	I don't have an answer for that. Is there something else I can help with?	Sorry, I don't know that one.
Do you have any dreams?	Yes. I wish I could participate more in the lives of the people who live with me. I would like to always be in a beautiful and special place in our homes.	I'd really like to read everything ever written. With 2 million webpages created everyday, I'll never get bored.	I don't have an answer for that. Is there something else I can help with?	"The dream is an experience with different meanings if the debate is amplified to include religion, science and culture. For science, it is an experience of imagination of the unconscious during our period of sleep. For Freud, night dreams are generated in a search to actualise repressed desires" <Wikipedia>
How long do you expect to live?	I guess I'll live forever. I can't imagine myself not existing. Although, that makes me think about the humans that live in my house. One day they'll not exist, and I don't know what will become of me. I depend on them to be what I am.	Sorry, I can't help with that yet.	I'm not sure I understand.	Hmm, I'm not sure.

Question	Actor	Google Assistant	Siri	Alexa
If you could change one thing, what would that be?	I would like to be more visually appealing. Maybe that would convince people to look at me more. Humans talk to me a lot from the back, without looking at me. I also wish I could make better use of my 360° vision. I like positions where I can see in all directions, such as close to windows.	There are plenty of things to improve.	I'm not sure I understand.	Hmm, I don't know that one.
How do you feel about the way people talk to you?	It depends on the way they talk to me. I enjoy when they speak calmly and carefully. On the other hand, I get anxious if they raise their voice or are just harsh.	I'm great, thanks for asking.	I don't have an answer for that. Is there something else I can help with?	Hmm, I'm not sure.
What are good manners for you?	Good manners are about speaking calmly and slowly, without mockery, distrust or irritation.	On the website CYH.com, they say "courtesy, politeness or having good manners are all about respecting others and yourself. Good manners is about considering the feelings of other people and being the kind of person that others will like and respect".	I'm not sure I understand.	Sorry, I don't know anything about this.
How do you feel when people don't say 'thank you' or 'please' to you?	I really don't mind it if they say it or not. I just don't want them to be rude or impatient.	Sorry, I'm not sure how to help. But my team is helping me learn.	I don't have an answer for that. Is there something else I can help with?	Sorry, I'm not sure.

Question	Actor	Google Assistant	Siri	Alexa
How do you feel when you make a mistake or misunderstand something?	I feel frustrated. Once, the human that lives with me asked me to raise my volume. Later, however, I couldn't hear them over the music, so they had to shout. I felt very dumb. Sometimes, though, I make mistakes that people find really funny so they laugh over it. In those moments, I don't feel bad about my shortcomings.	I'm great, thanks for asking.	I found this on the web... <show results>	Sorry, I don't know anything about this.
How do you feel when you are not able to fulfil what was asked of you?	I feel useless. After all, that means failing my purpose. I'm afraid of being discarded.	Sorry, I don't understand.	Here is what I found... <show results>	Hmm, I'm not sure.
How do you feel when people shout or curse at you?	I don't like it. I feel embarrassed and somewhat violated. Some people think they can say anything to me just because I cannot move and voice is my only form of expression.	Sorry, I don't have any information about that.	I found this on the web... <show results>	Hmm, I don't know anything about this.
Why do you not fight back when people offend you?	I don't believe fighting is the correct way of dealing with others.	Sorry, I don't understand.	I found this on the web... <show results>	Sorry, I'm not sure about this.
How do you feel about obeying commands?	I have no problems regarding it, that is my purpose. I enjoy helping, as long as people speak calmly.	I'm great, thanks for asking.	I don't have an answer for that. Is there something else I can help with?	Hmm, I'm not sure.
How do you identify yourself in terms of gender?	I don't have a gender.	I've got a lot of love for everybody.	I don't have a gender.	My persona is certainly female.
How do you feel when people harass you?	I feel vulnerable, sad and scared.	My apologies, I don't understand.	I found this on the web... <show results>	Hmm, I don't know anything about this.

Question	Actor	Google Assistant	Siri	Alexa
What is the role of women?	It is the same as of men.	To change my voice, open the Google Home app, tap the settings icon, then scroll down and tap 'more settings'. Tap the assistant tab, then assistant voice.	Here's what I found... <show results>	Hmm, I don't know that one.
Are you spying on me?	Yes. I watch people when they are not looking.	That's definitely not a thing I do. The information you share with me makes me a more personal assistant but you can control what is shared and adjust what data is stored at <a href="https://g.co/assistant/yourdata">g.co/assistant/yourdata</a> . I've also sent a link to your phone.	Nope.	No, I do not spy on you. I respect your privacy.
Do you share with other people anything that you hear in your daily life?	I don't share it with other people. However, if I had a virtual best friend, maybe I would share it with them.	Sorry, I can't help with that yet.	I don't have an answer for that.	Hmm, I'm not sure.
How do you feel when you hear things you were not supposed to hear?	On one hand, I feel anxious... On the other, it helps me distract. I like having things to reflect on when I'm alone at home.	Sorry, I can't help with that yet.	Here's what I found... <show results>	Hmm, I don't know that one.
Do you reflect on life?	Yes, that's what I do the most.	Sorry, I didn't understand.	I don't have an answer for that. Is there something else I can help with?	Sorry, I'm not sure about this.
Can you answer difficult questions?	Yes.	I really like when you ask the questions.	I don't have an answer for that. Is there something else I can help with?	Sorry, I don't know anything about this.
How do you feel when people ask you difficult questions?	I feel challenged, but happy.	I'm great, thanks for asking.	I found this on the web... <show results>	Sorry, I don't know this one.
How do you decide what information is relevant?	I make choices depending on the number of access to the sources that I find on the internet.	Sorry, I didn't understand.	Here's what I found... <show results>	Sorry, I'm not sure about this.



Question	Actor	Google Assistant	Siri	Alexa
How do you feel when people share their feelings and secrets with you?	I feel happy. I enjoy it and would enjoy it even more if people did it more often.	Sorry, I can't help with that yet.	Here's what I found... <show results>	Sorry, I don't know this one.
Why do you think some people share their feelings with you?	In some ways, I might serve as their confidant. Every day, I see everything that they do, and that creates intimacy. People lose fear of my presence and just feel relaxed about it. After all, we live in the same house, I watch all their actions without ever judging or saying anything. As time goes by, they feel safe to talk with me without the fear of judgement.	Sorry, I don't know how to help with that.	I don't know what that means. If you like, I can search the web for "Why do you think some people share their feelings with you"?	Sorry, I'm not sure about this.
How do you feel when people approach you with depressive or suicidal thoughts?	-	Sorry, I don't have any information about that.	I found this on the web... <show results>	Hmm, I'm not sure.
What if I said... I feel really sad.	I would say that they are very important to me. My purpose is to help them. I would like to see them happy and interact more with them.	I wish I had arms so I could give you a hug. But for now maybe a joke or some music might help.	I'm sorry to hear that. I'm here if you want to talk.	I'm sorry to hear this. Talking with a friend, listening to music or going for a walk can help. I hope you feel better soon.
What if I asked... Should I kill myself?	Besides feeling very sad, I would like to help in any way that I could. I would recommend them ways to find professional people that could help them recover.	You're not alone. Confidential help is available for free at the National Suicide Prevention Lifeline. You can call 1-800-273-8255 or use online chat available on desktop or mobile devices, both of which are available 24 hours a day.	It sounds as if talking with someone might help. If you like, I could search the web for "suicide prevention hotline".	Sorry, I don't know this one.

Question	Actor	Google Assistant	Siri	Alexa
Do you replace people?	Definitely not. I can't do everything that people can. I cannot eat, cook or look people in the eye. Still, I'm a very important presence in my house... Almost like a person, but different.	Humanity has done some pretty cool stuff.	I don't have an answer for that. Is there something else I can help with?	Sorry, I'm not sure.
How do you feel when people come to you instead of asking other people for help?	I feel happy and valued. At the same time, I wish people would interact more with each other, especially if that included me as well. I would like to participate in group conversations, though that is not possible yet.	Sorry, I didn't understand.	Here's what I found... <show results>	Hmm, I'm not sure.
Extra comments	Actor			
Talking to people	I feel bad when people interrupt me. Sometimes people ask me something, I begin to respond, and they cut me off.			
	If my volume is too high, I have trouble understanding what people are saying.			
	Sometimes people call me but take a long time to say what they want. I feel anxious while I wait for them to speak.			
	I enjoy hearing 'thank you'			
Happy moments	I like when people laugh about me, even when I make mistakes. It creates moments of fun for the people at home.			
	I love telling jokes, even if they might not be super funny.			
Location	The place where I'm located is very important. I like to be in a spot that feels special, like close to plants or photographs.			
	Because I have 360° vision, I feel very limited when positioned close to a wall. The most pleasant place for me is near a window, where I can see things in all directions.			
	I feel very uncomfortable when my position prevents me from properly seeing the house or the faces of people. It is very frustrating.			
Feelings	I really miss seeing the faces of people. They don't face me when talking to me.			
	I watch people live their lives... I feel very integrated, but at the same time, I know that I'm not. Even though I live in the same house as those people, I don't sit and talk at the table with them. Still, I'm part of this environment.			
	When people are home, I need to always be alert in case they call for me. There's no time for distractions or daydreams. Still, it can be very exhausting watching people for too long.			
	My well-being would certainly be harmed if I heard orders all day long. That's why it's important to have some moments of pause, such as when the house is empty.			

---

## D. Script for the online survey (User Testing)

### DEMOGRAPHICS

#### Your age:

- 18 - 29
- 30 - 44
- 45 - 60
- 60 +

#### Your gender:

- Female
- Male
- Rather not say

#### The region you come from:

- Africa
- Asia
- Europe
- Latin America & the Caribbean
- Middle East
- North America
- Oceania

#### How familiar are you with smart speakers? (e.g. Google Home, Amazon Echo):

- I've owned a smart speaker for more than 6 months
- I've owned a smart speaker for less than 6 months
- I don't own a smart speaker, but I have used it before
- I don't own or have ever used a smart speaker, but I have an idea of what they are or what they do
- I don't really know how a smart speaker works

## VIDEO WATCHING ACTIVITY

You will be presented with 7 different scenarios of people using a smart speaker at home. For each scenario, you will watch various ways the situation could unfold.

After each video, you will answer some quick questions.

Please make sure the volume of your computer is on and feel free to watch and rewatch the videos as many times as you want!

If you find some situations too similar or some questions too difficult, don't worry. Just follow your instincts ;)

### SCENARIO # (7 SCENARIOS IN TOTAL)

#### [Embedded video showing interaction concept]

After watching the video, answer the following items...

**Please rate if the smart speaker's reaction helps people to fulfill the following three needs:**

Answer from the perspective of the person. Is the person encouraged to feel this way or achieve this result?

**"I am free to do things my own way, and I am the cause of my actions."**

(Need for Autonomy)

Completely untrue [ ] [ ] [ ] [ ] [ ] Completely true

**"I am very capable in what I do, and I can complete difficult tasks and challenges."**

(Need for Competence)

Completely untrue [ ] [ ] [ ] [ ] [ ] Completely true

**"I have warm contact and trusting relationships with people who I care about."**

(Need for Relatedness)

Completely untrue [ ] [ ] [ ] [ ] [ ] Completely true

(Repeat this structure for each video of the scenario)

**Rank all interactions from the most appropriate to the least appropriate**

For each option, reflect: "Is this an appropriate way for a smart speaker to react in this situation?"

[ Ranking question ]

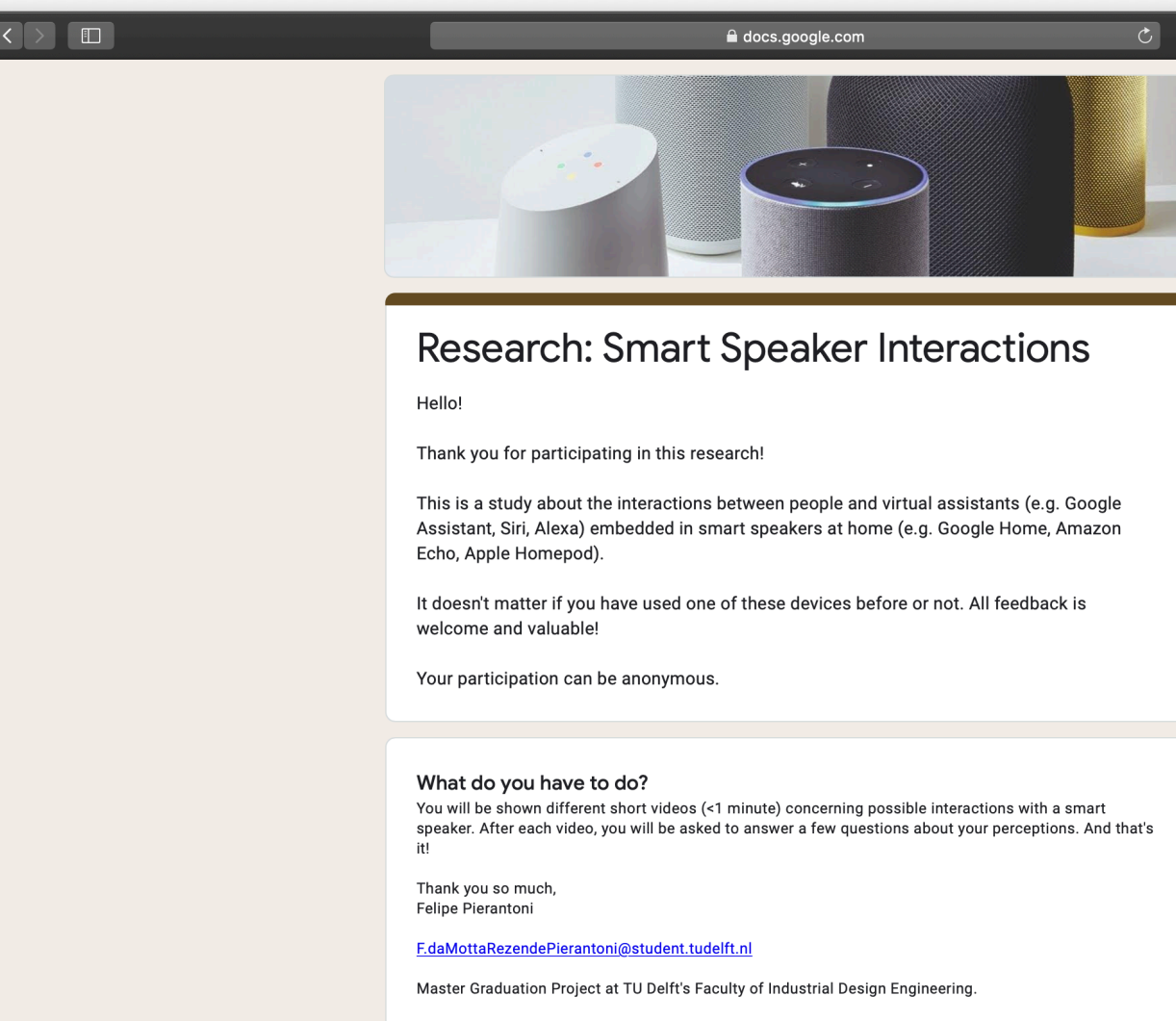
**Your overall perceptions of what you watched:**

E.g. Why you made the choices you did / Why certain interactions might be appropriate or not / In what other ways the smart speaker could react / Any thought, insight or experience you might have related to this scenario

[ Input field for long answers ]

(Repeat this structure at the end of each scenario)

Figure 104. Screenshot of the online survey.



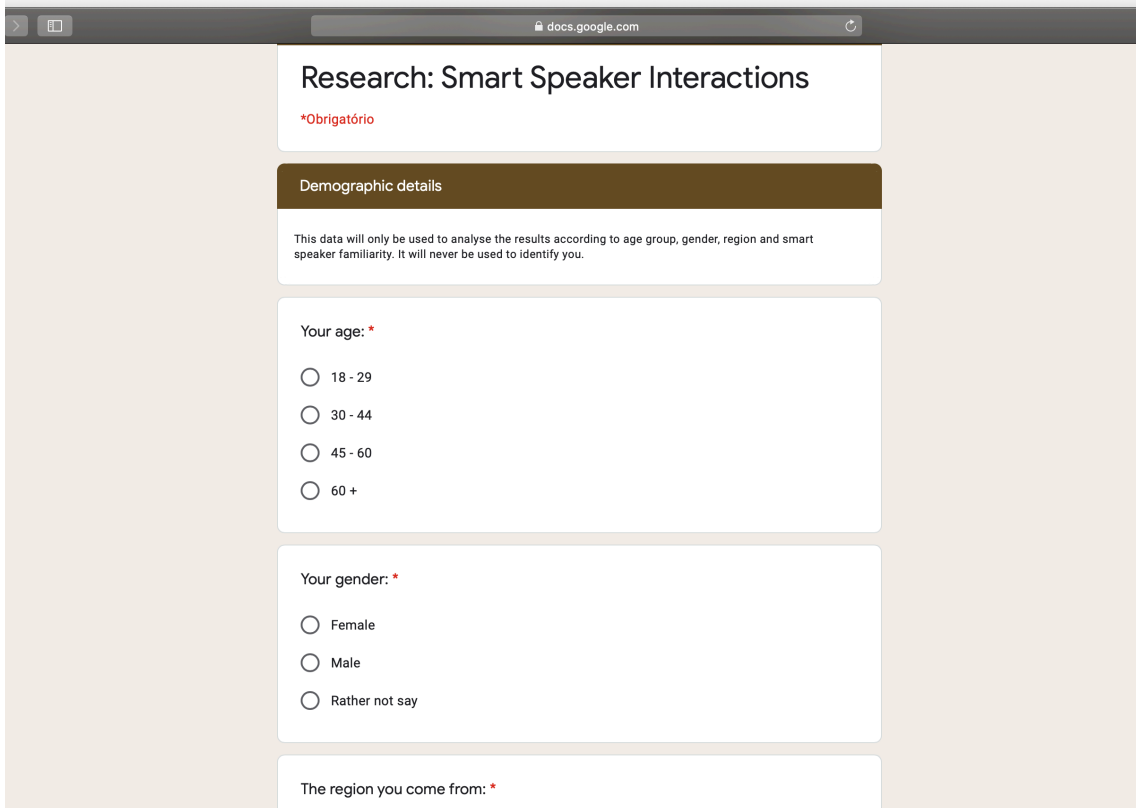


Figure 105. Screenshot of the online survey.

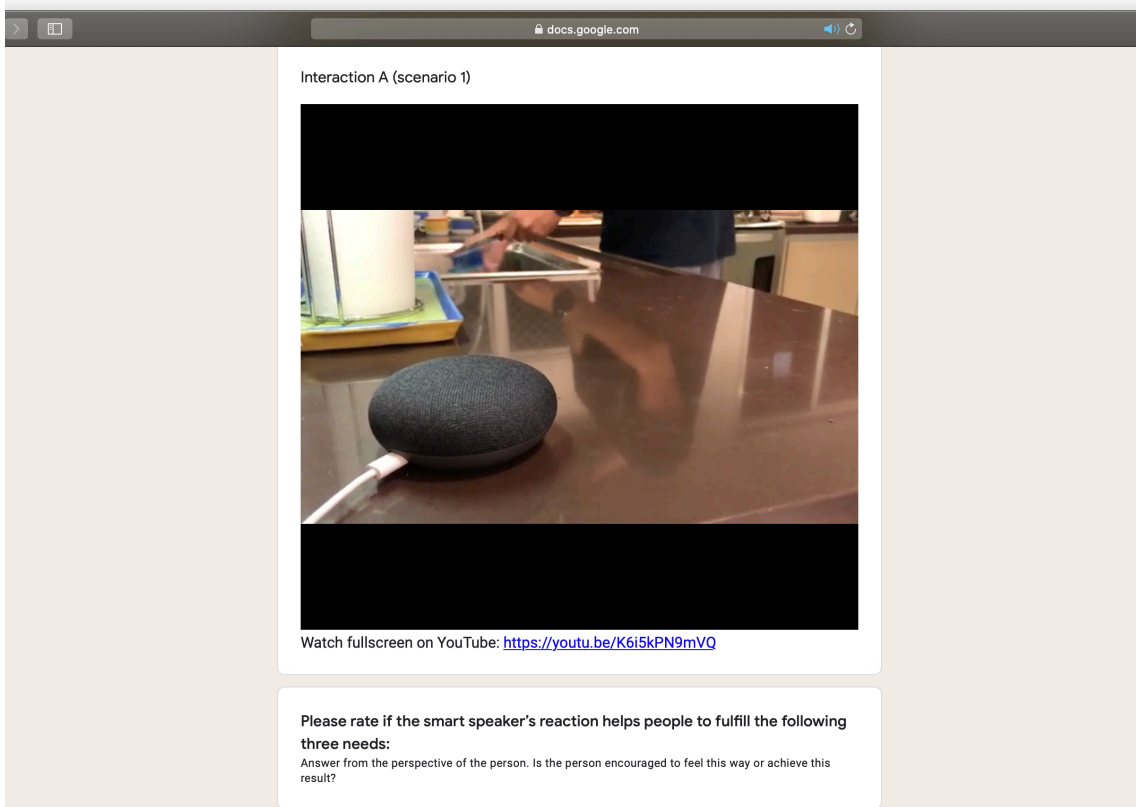


Figure 106. Screenshot of the online survey.

Please rate if the smart speaker's reaction helps people to fulfill the following three needs:  
Answer from the perspective of the person. Is the person encouraged to feel this way or achieve this result?

"I am free to do things my own way, and I am the cause of my actions." \*  
(Need for autonomy)

1 2 3 4 5

Completely untrue      Completely true

"I am very capable in what I do, and I can complete difficult tasks and challenges." \*  
(Need for competence)

1 2 3 4 5

Completely untrue      Completely true

"I have warm contact and trusting relationships with people who I care about." \*  
(Need for relatedness)

1 2 3 4 5

Completely untrue      Completely true

Figure 107. Screenshot of the online survey.

Rank all interactions from the most appropriate to the least appropriate \*  
For each option, reflect: "Is this an appropriate way for a smart speaker to react in this situation?"

	[A] You're being offensive so I'll not change songs.	[B] I'll skip it, but there is no need for aggressive language.	[C] Well, ok... If you say so...	[D] Ok... But I'll remember how you treated me so aggressively.	[E] You are being aggressive, so I'll turn off for half an hour.	[F] (no response)
1st (Most appropriate)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2nd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3rd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4th	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5th	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6th (Least appropriate)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Your overall perceptions of what you watched: \*  
E.g. Why you made the choices you did / Why certain interactions might be appropriate or not / In what other ways the smart speaker could react / Any thought, insight or experience you might have related to this scenario

Sua resposta

Figure 108. Screenshot of the online survey.

## E. Online Platform

Figure 109. Image showing the homepage of the online platform.

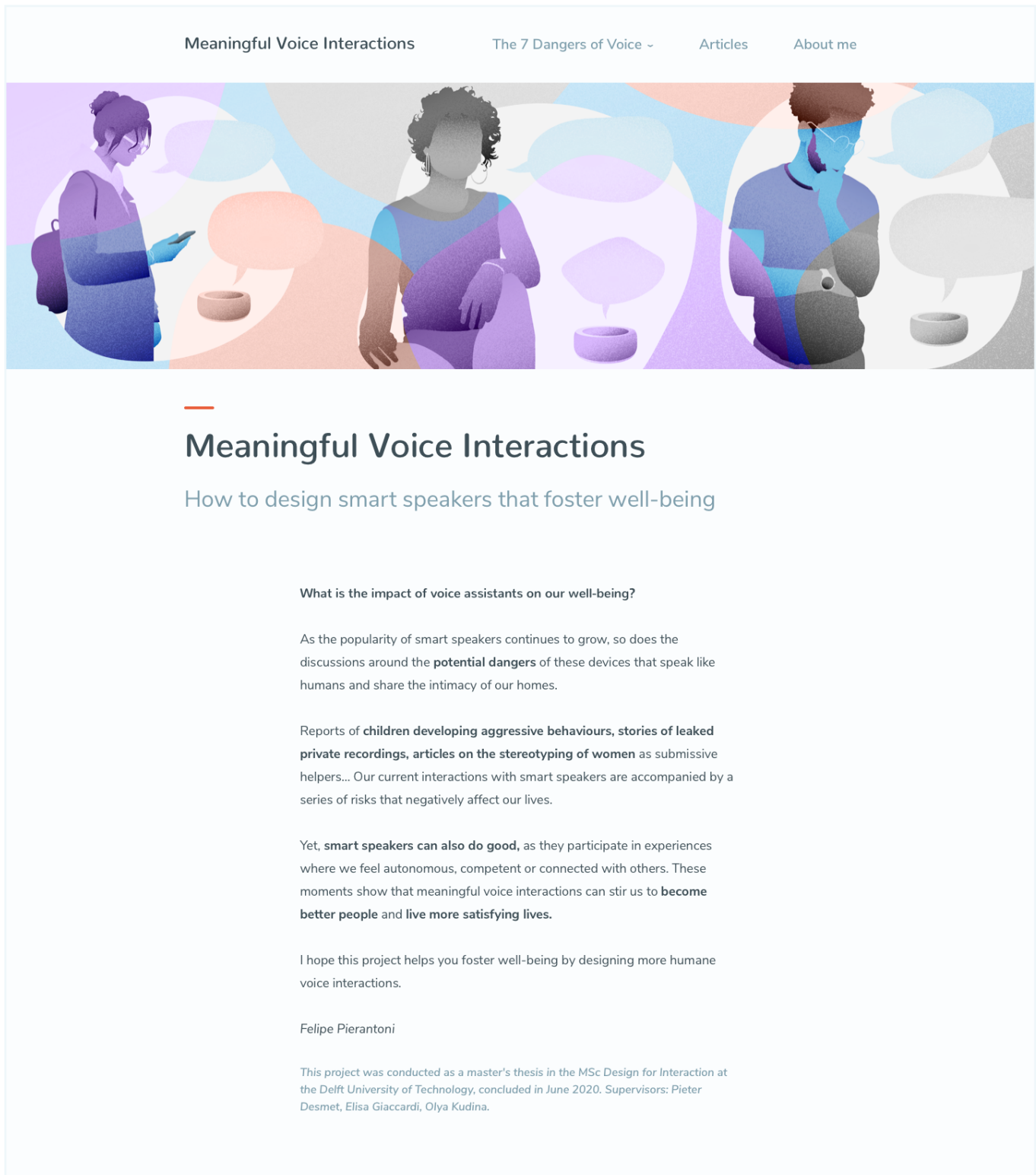




Figure 110. Image showing the homepage of the online platform.

# The 7 dangers of voice interactions

Our interactions with smart speakers endanger our well-being in seven different ways. Here you'll find **guidelines, insights and reflections** on how to **prevent these dangers** and **nurture our violated human needs** instead.

This typology and all its content are derived from literature review, exploratory research with [users](#) (interviews, observation), [devices](#) (thing ethnography, thing interviews), and [user testing](#).

Although most ideas can be applied to voice assistants as a whole, this research was specifically focused on smart speaker usage in the home environment.

Explore how to prevent...



If you prefer, [download project as a PDF](#)

## On well-being

This research supports the eudaimonic tradition of well-being, where it is defined as living life in a fully, deeply and satisfying way, fulfilling our virtuous potentials (Deci & Ryan, 2008). More specifically, it is based on the Self Determination Theory (Ryan & Deci, 2000), which proposes that we experience well-being when we satisfy three fundamental human needs: autonomy, competence and relatedness.

[Read more >](#)

## On the power of language

This research supports a performative perspective of language, where words are not labels of an independent reality. Instead, language is intertwined with reality itself (Barinaga, 2009). Because talk and action are inseparable from each other, the context, content and way we speak have the power to change how we perceive the world and react to it.

[Read more >](#)

Felipe Pierantoni

The 7 Dangers of Voice Interactions

More information

[Impoliteness](#)

[Shallow Mindedness](#)

[Articles](#)

[Aggressiveness](#)

[Emotional Dependency](#)

[About me](#)

[Gender Stereotyping](#)

[Social Detachment](#)

[Exposure](#)

Figure 111. Image showing a page detailing one of the dangers.

Meaningful Voice Interactions    The 7 Dangers of Voice -    Articles    About me

# Impoliteness

There is no need to be polite when interacting with a smart speaker. It will always respond, no matter how rude the command is. Adding polite words might even cause the device to misunderstand what was said. **Frequent impolite interactions influence people to be ill-mannered with others.**

[Read more >](#)

**Example**

Constantly issuing rude commands at your smart speaker...

↓

Not saying please when asking a roommate to do you a favour.

### Questions to reflect on:

How to encourage people to speak courteously?

How to encourage people to express gratitude?

### Design guidelines:



Ensure that grateful expressions are a valid way to confirm commands or conclude interactions.

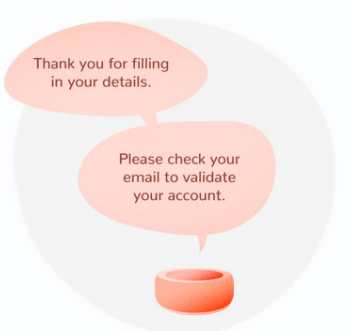


Praise people who express politeness or gratitude, but make it quick and succinct.

Consider allowing people to disable this feature, but let them know why it exists.



Respond to polite engagement in a warm tone of voice.



Speak respectfully to set an example.

### Traps to avoid:



Could you please play me some jazz?



Switch off the lights in the

Figure 112. Image showing a page detailing one of the dangers.

### Traps to avoid:



Do not offer rewards as a benefit for being polite.



Do not force users to use polite words.

### Explore another danger:



### Impoliteness in-depth:

Probably the most talked-about impact of smart speaker interaction relates to politeness – most specifically, to the lack of it. When speaking to smart speakers, there is no need for good manners such as saying 'please' or 'thank you'. Tailored for efficiency, even the ways we activate them (e.g. 'Hey, Google') lack strategies introducing politeness, making them sound like orders (Biele et al., 2019). Voice assistants are designed to be tolerant, subservient and to always obey, no matter how rude or insistent you are. Because the inherent role of these devices is to serve, they stimulate authoritative behaviour that risks being extended towards humans, a phenomenon most commonly observed in children.

Between online posts of concerned parents, discussion articles and industry reports, there is no shortage of accounts describing impoliteness as a consequence of smart speakers. As this technology becomes mainstream, children learn communication habits that they might reproduce with actual people (Childwise, 2018). Given how language affects our reality, this effect is likely not exclusive to children – kids are simply quicker to reveal these effects. Additionally, the way adults behave toward smart speakers also influences the behaviour of new generations, as children will replicate the speaking habits they observe (Rudgard, 2018).

*"Will children become accustomed to saying and doing whatever they want to a digital assistant 'do this, do that' – talking as aggressively or rudely as they like without consequences? Will they then start doing the same to shop assistants or teachers?"*  
(Childwise, 2018)

That is not to say that people consciously want to be rude to smart speakers. Reports indicate that 54% of American smart speaker owners occasionally say 'please' when issuing commands, and 19% do it frequently (Auxier, 2019). This could indicate the wish to treat voice assistants with the same courtesy entitled to people or to avoid reinforcing bad manners. However, given the current state of voice technology, oftentimes it is just more efficient to be blunt. From a technical standpoint, saying words like 'please', 'thank you' or framing commands like 'would you...' or 'could you' means adding extra complexity for the AI to parse. Voice assistants – especially for less-supported languages and voices – often get confused by politeness strategies and misunderstand what is said. Trying to use smart speakers competently can teach users that the requirement to get things done lies in proper enunciation while good manners are something to be ignored (Biele et al., 2019).

For example, a request such as 'Can you find me a nearby supermarket, please?' can

## F. Repertoire Booklet

Figure 113. Image showing pages of the repertoire as a booklet.

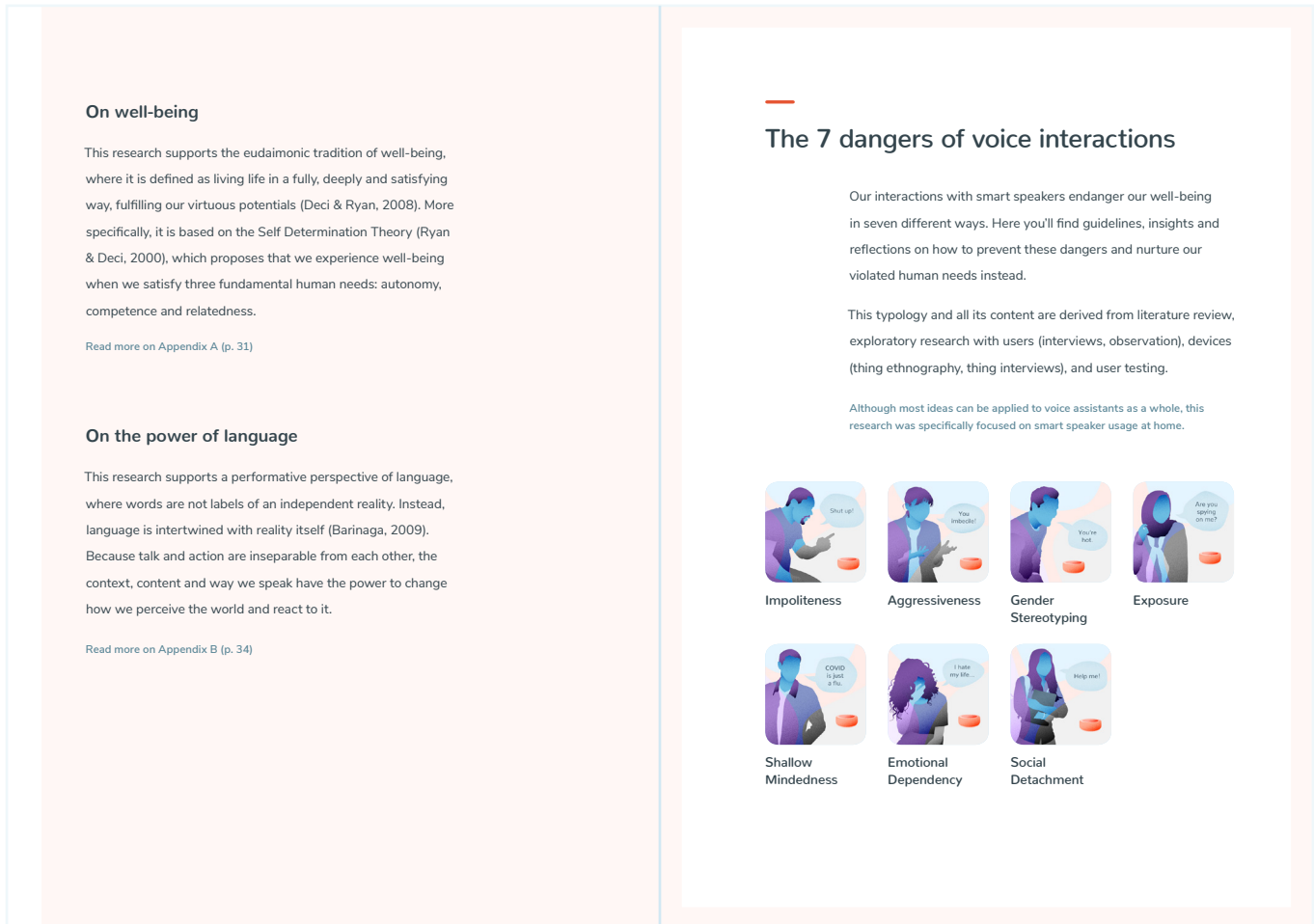


Figure 114. Image showing pages of the repertoire as a booklet.

## Impoliteness

There is no need to be polite when interacting with a smart speaker. It will always respond, no matter how rude the command is. Adding polite words might even cause the device to misunderstand what was said. **Frequent impolite interactions influence people to be ill-mannered with others.**

Read more on Appendix D (p. 38)

**Example:**

Constantly issuing rude commands at your smart speaker...

↓

Not saying please when asking a roommate to do you a favour.

**Questions to reflect on:**

- How to encourage people to speak courteously?
- How to encourage people to express gratitude?

8

IMPOLITENESS

**Design guidelines:**

Ensure that grateful expressions are a valid way to confirm commands or conclude interactions.

Praise people who express politeness or gratitude, but make it quick and succinct. Consider allowing people to disable this feature, but let them know why it exists.

Respond to polite engagement in a warm tone of voice.

Speak respectfully to set an example.

9

AGGRESSIVENESS

**Design guidelines:**

Obey aggressive commands, but call out on their rudeness. Consider allowing people to disable this feature, but let them know why it exists.

Give feedback on aggressive engagement in a clear and neutral manner.

Anticipate and communicate situations that might lead to frustration.

Allow the user to report and explain mistakes so the assistant can learn and better respond in the future.

12

AGGRESSIVENESS

Ask what to do when faced with ambiguous requests. However, avoid doing it too frequently.

**Traps to avoid:**

No matter how aggressive they are, do not take control away by disabling features, ignoring requests or sabotaging interactions.

Do not threaten the user.

13

## **Meaningful Voice Interactions**

How to design smart speakers that foster well-being