Meaningful Voice Interactions

How to design smart speakers that foster well-being

Felipe Pierantoni

June 2020

This project was conducted as a master's thesis in the MSc Design for Interaction at the Delft University of Technology, concluded in June 2020.

Supervisors: Pieter Desmet, Elisa Giaccardi, Olya Kudina.

What is the impact of voice assistants on our well-being?

As the popularity of smart speakers continues to grow, so does the discussions around the potential dangers of these devices that speak like humans and share the intimacy of our homes.

Reports of children developing aggressive behaviours, stories of leaked private recordings, articles on the stereotyping of women as submissive helpers... Our current interactions with smart speakers are accompanied by a series of risks that negatively affect our lives.

Yet, smart speakers can also do good, as they participate in experiences where we feel autonomous, competent or connected with others. These moments show that meaningful voice interactions can stir us to become better people and live more satisfying lives.

I hope this project helps you foster well-being by designing more humane voice interactions.

Felipe Pierantoni

Table of Contents

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 Appendix A (p. 31)

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.

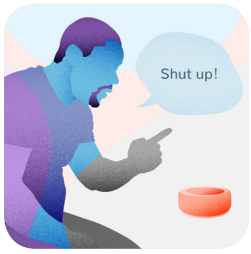
Read more on Appendix B (p. 34)

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.

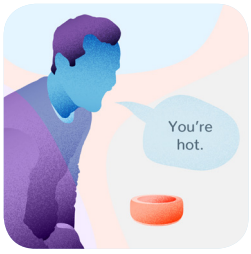
This typology and all its content are derived from literature review, exploratory research with users (interviews, observation), devices (thing ethnography, thing interviews), and user testing.

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

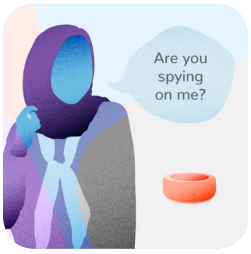


Impoliteness

Aggressiveness



Gender Stereotyping



Exposure

Shallow Mindedness

Emotional Dependency

Social Detachment

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.

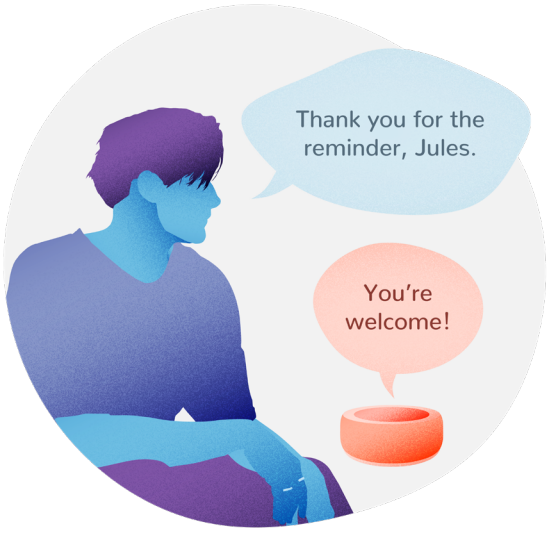
Questions to reflect on:

How to encourage people to speak courteously?

How to encourage people to express gratitude?

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.

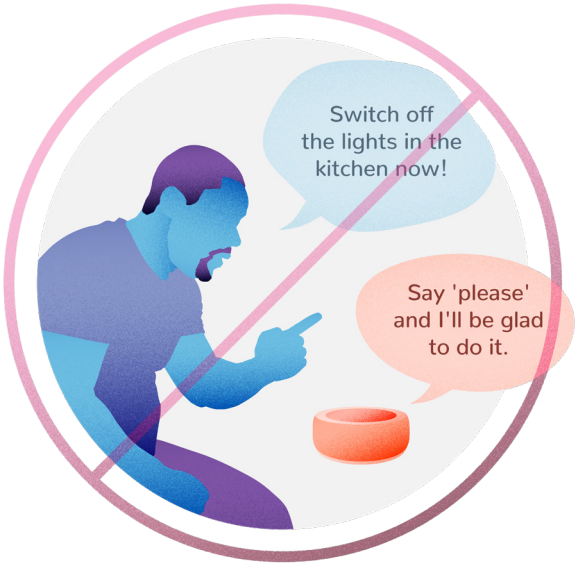
Consider allowing people to disable this feature, but let them know why it exists.

Respond to polite engagement in a warm tone of voice.

Speak respectfully to set an example.

Traps to avoid:

Do not offer rewards as a benefit for being polite.



Do not force users to use polite words.

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.**

Read more on Appendix E (p. 42)

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?

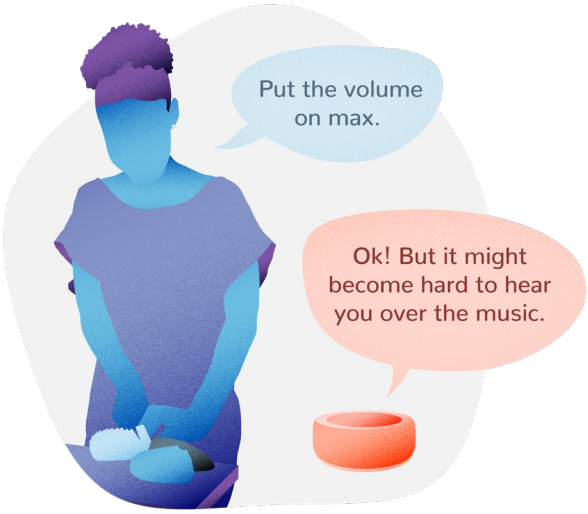
How to reduce the frustration over mistakes?

Design guidelines:

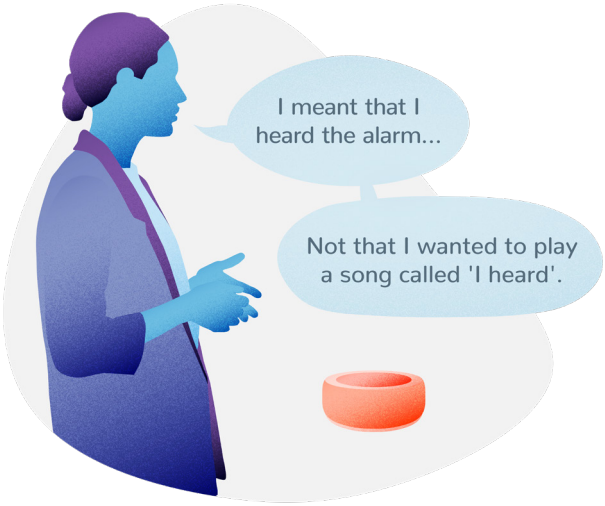
Obey aggressive commands, but call out on their rudeness.

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.



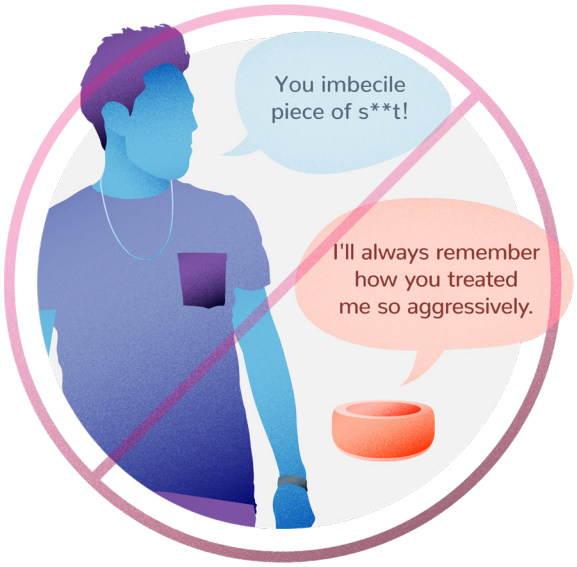
Allow the user to report and explain mistakes so the assistant can learn and better respond in the future.

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.

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 subserviency and incompetence.**

Read more on Appendix F (p. 44)

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?

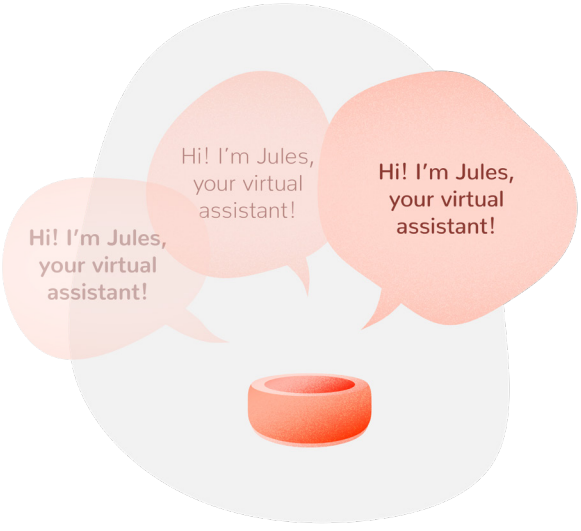
How to discourage sexual harassment?

How to reduce gender-specific responses?

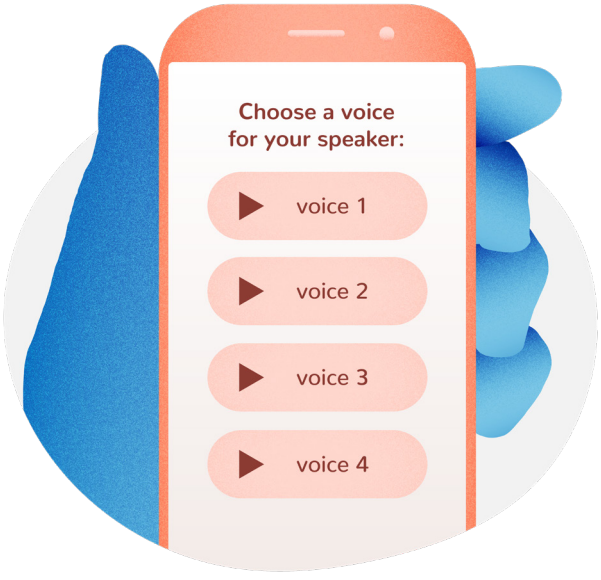
Design guidelines:

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.



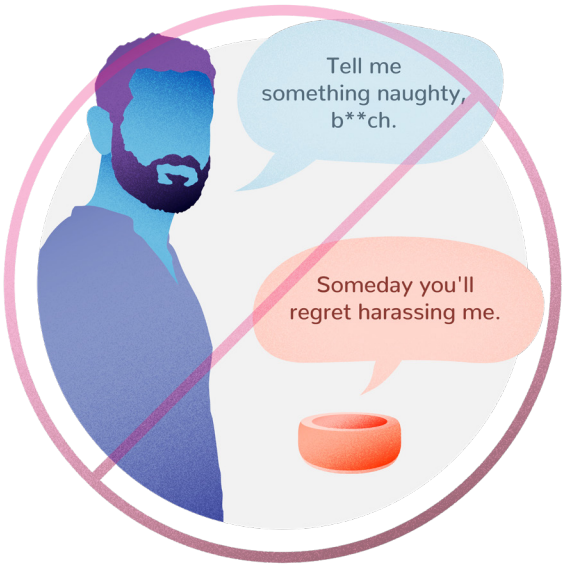
Ask the user to choose a voice during initial setup instead of offering a default female voice.

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.

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.**

Read more on Appendix G (p. 47)

Example:

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

Feeling paranoid about being spied on.

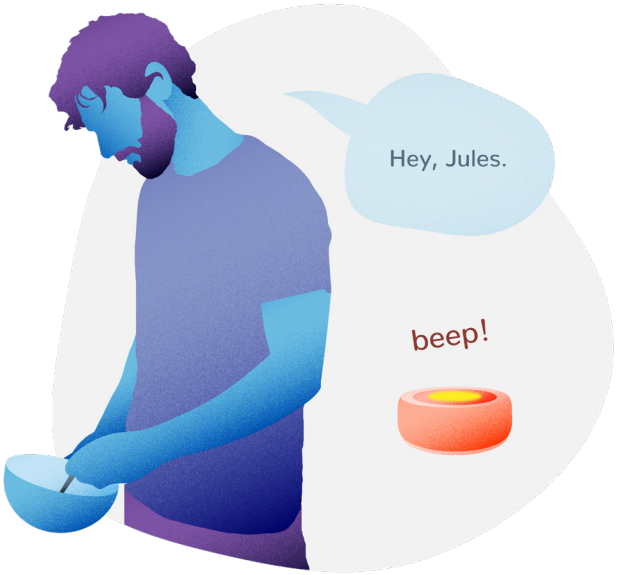
Questions to reflect on:

How to guarantee the privacy of users?

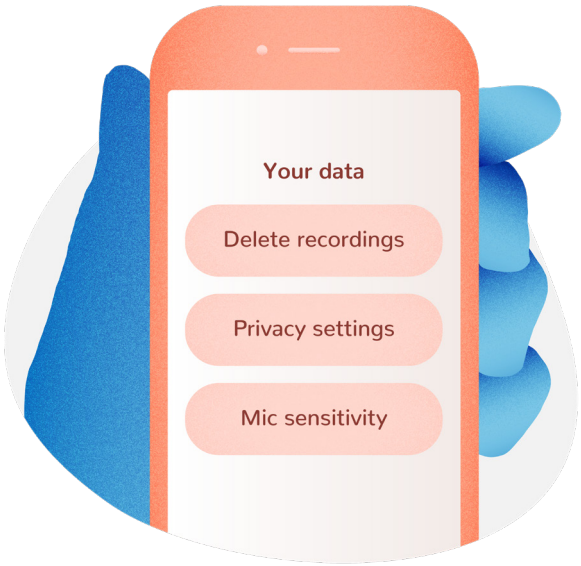
How to reassure people that they are safe from exposure?

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.

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.**

Read more on Appendix H (p. 49)

Example:

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

↓

Incapacity to ask complex questions during class.

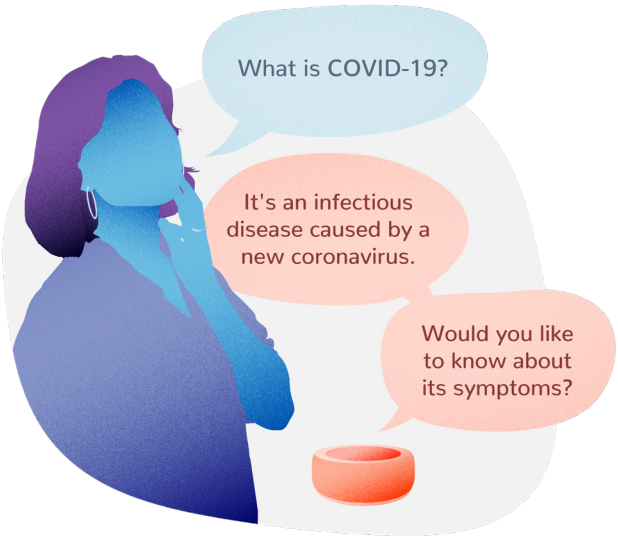
Questions to reflect on:

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.

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.

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.**

Read more on Appendix I (p. 52)

Example:

Confessing depressive thoughts to your smart speaker...

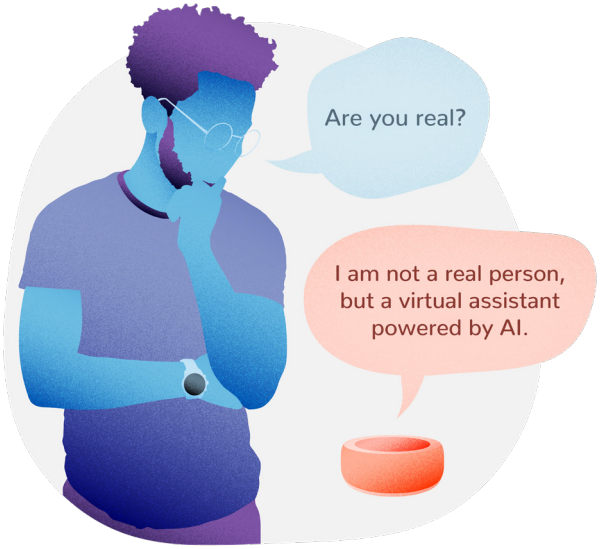
Developing a dependant relationship with a machine.

Questions to reflect on:

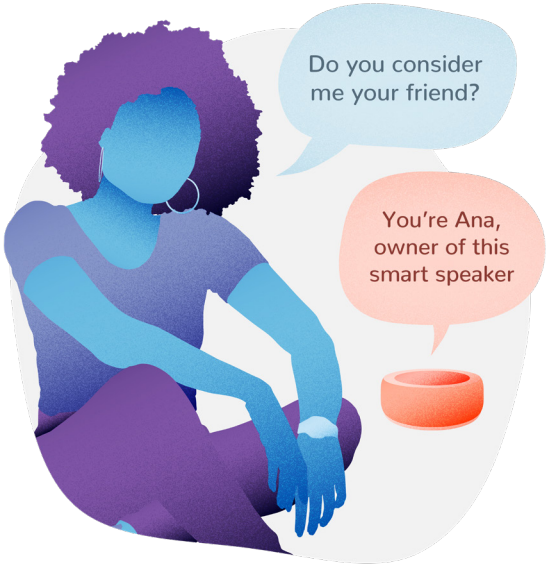
How to discourage human-like bonds with smart speakers?

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.

Traps to avoid:

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

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.**

Read more on Appendix J (p. 55)

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.

Appendices

A. Well-being and fundamental human needs

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, Pohlmeyer & 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 will focus on the eudaimonic approach to well-being. All proposed interactions with smart speakers will 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 possible impacts 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 will be a core strategy for this project. Nevertheless, 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 being 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. This result is evidence of how these needs are essential to people and why this theory is among the most established frameworks of human needs.

B. Language and the performative perspective

Throughout the ages, the capacity to communicate through speech has been something that defined human beings. As a result 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 contacts (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, just as people attach social responses to voice assistants, these social human-computer interactions could then influence our human-human interaction.

The idea that our voice interactions with assistants might affect our interactions with actual people is aligned with a performative perspective of language. Different than a traditional perspective, which supports the ideas that meaning is representation and words are labels of an independent reality, the

performative perspective states that language is actually intertwined with reality itself (Barinaga, 2009).

This view, originally proposed by the Austrian 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 from each other, 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 speakers 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 potential negative effects of current interactions with smart speakers.

C. Table overview: the 7 dangers of voice interactions

NEGATIVE EFFECT	DESCRIPTION
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.**
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.**
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 subserviency and incompetence.**
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.**
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**.
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.**
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	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**

D. 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 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).

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 caused an impairment in their well-being.

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.

E. Aggressiveness in-depth

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 foster 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 still 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.

F. Gender Stereotyping in-depth

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 they 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 various 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. Most voice assistants in western markets are exclusively female or female by default in regards to voice and name (West et al., 2019). Conscious and biased, this choice blends 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 common stereotypical ideas regarding the position and obligations of women. Because most voice assistants sound female, the 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 the role of 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 also 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. 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).

As the reach of voice assistants expands with the popularity of smart speakers and other devices, more people take part in interactions that might perpetuate and widen gender inequalities (West et al., 2019). When companies responsible for voice assistants choose to take action to address this issue, the most common approach is to add male voice alternatives or to let users choose the gender of their assistant during the initial setup (West et al., 2019). Beyond this strategy, there have been experiments with voice assistants with synthetic or 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.

G. Exposure in-depth

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 suggested a much higher prospect of exposure.

These past years of coexistence with smart speakers revealed that these fears were not completely groundless. Although no large-scale disclosure of private information has happened – or been revealed – so far, 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).

Regarding this potential privacy breach, companies state that this process is mostly geared towards situations in which the assistant could not understand

what was asked, and 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, as our privacy can indeed be exposed in some ways to other people.

> *"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 location, contact details, and app data." (Hern, 2019)*

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

H. Shallow Mindedness in-depth

Despite sounding as capable as human beings, smart speakers are still significantly limited in what they can understand and what they can 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 repercussions of smart speakers might be in the way we frame our questions and which types of questions we choose to ask.

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 negative effect described in this project as shallow mindedness.

"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)

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 reflections or responses? 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 then 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.

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'. 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.

I. Emotional Dependency in-depth

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 effort 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).

> *“[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)*

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 many 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.

Because voice assistants “give us a way to reveal shameful feelings without feeling shame”, people can 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 power over us (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.

J. Social Detachment in-depth

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, just as the other analysed impacts, 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.

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).

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 [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.

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.

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

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

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-alexas-job-parent-children/

Jimenez, S., Pohlmeyer, A.E., & Desmet, P.M.A. (2015). *Positive Design Reference Guide*. Delft: Delft University of Technology. ISBN 978-94-6186-425-3

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/

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).

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

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.

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/

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

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

Meaningful Voice Interactions

How to design smart speakers that foster well-being