Designing Trustworthy Agents

Valeria Guiot
904101

Desiging Trustwothy Agents

Guidelines for designers

Student Valeria Guiot 904101
Academic Supervisor Margherita Pillan
Academic Co-Supervisor Milica Pavlovic

POLITECNICO
MILANO 1863

Politecnico di Milano 2018/2019 School of Design
MSc In Digital and Interaction Design

Index

Credits

I would like to remember all the people who helped me in writing this thesis with suggestions, criticism and observations.

First of all, I thank Professor Margherita Pillan, supervisor of this thesis, for her availability, the attention and understanding shown in the writing of this thesis, and to PhD Milica Pavlovic. It was certainly challenging but it taught me to go further and appreciate the value in the differentiation and contamination of different knowledge and disciplines.

Thanks to my family for the constant and affectionate support, especially my parents who have always been here for me.

A heartfelt thanks to my boyfriend Grégoire for being an active, patient and loving presence during this thesis and for always encouraging me to overcome my limits.

A special thank you to my friend, teammate, and advisor Fabio, for his precious suggestions and constant support, without him my Politecnico experience would not have been the same.

Thanks to my friend Giorgia for teaching me the "designer way" and for often changing my perspective.

Thanks to all my closest friends (Eleonora, Giulia, Martina, Karolina, Dani, Carlotta, Marco, Cristina), my colleagues, and to all the people that always believed or challenged me: I would not be the same!

Thanks to all the participants for contributing to my thesis.

And thanks to the Polimi for the opportunity to become a Designer, I will always be grateful.

Abstract

Despite the escalating growth of usage of Artificial Intelligence systems (AI) and Artificial Intelligence Agents, it has been found that people still struggle to trust AI. There are many explanations to this phenomenon. In general people still prefer human touch (especially in consumer service), they are not comfortable interacting with AI, they are concerned about data usage and privacy. Last, they believe the system does not have any morality/ethics (Pegasystems Inc., 2019). In fact, despite the spread and improvements of AI agents, those obstacles persist in the adoption and use of personal assistants (Cuadra, Rase; 2018).

Trust research in information technology is currently a hot topic: this increased interest might be due to the consequences this technology is bringing. Trust in technology is complex, variegated and many elements can be taken into consideration depending on the context. In technology and computer systems exist the same problems of expectation, of willingness to be vulnerable and of possible negative outcome (Hall, 2010). Trust is a prerequisite of human-human interaction: humans feel the need to have a more predictable environment and understand others' intentions and motivations. In order to interact even in less predictable contexts and circumstances, people apply different strategies, which are found to be similar in Human-Computer-Interaction. Trust plays a fundamental role in technology because it is a "primary predictor" for use of technology (Li et al., 2008), increases adoption rates (De Kruijff, 2018) and the acceptance of the computer system (Parasuraman & Riley, 1997).
According to some researchers, trust is about relationship, and as being a social construct, AI lacks intrinsic qualities needed to lead to trust. This current work explores different models and research about psychological trust applied to Human-Agent-Interaction in order to find strategies to build trust in this technology. This was combined with an investigation about users' perceptions and experience with a voice activated assistant and reported trust (in this case Amazon Alexa) in two different interaction modalities. The problem was addressed with a literary review to understand the different approaches of "Trust research" and "AI Agent factors that contributes to trust", which were further analyzed in a research with 12 participants. The aim was to figure out if trust relationship between Human-agent-interaction was depending on personal, environmental (contextual) or related to a specific device factor. The result of the research identifies the projects elements that are useful to build a trustworthy relationship; the thesis also provides a set of guidelines for designing and build trust in a first-time interaction scenario.

Nonostante la cospicua crescita nell'utilizzo dei sistemi di intelligenza artificiale e degli agenti di intelligenza artificiale, è stato scoperto che le persone non si fidano ancora dell'IA. Ci sono molte spiegazioni a questo fenomeno, in generale le persone preferiscono ancora il contatto umano (specialmente nel Customer service), non si sentono a proprio agio ad interagire con l'IA, sono preoccupate per i propri dati personali e privacy, credono che il sistema non abbia alcuna moralità / etica e faccia un uso improprio dei loro dati personali (Pegasystems Inc., 2019). Nonostante la diffusione e il miglioramento degli assistenti vocali, questi ostacoli persistono nell'adozione e l'uso (Cuadra, Rase; 2018).
La ricerca sulla fiducia in informatica è tema molto attuale nel campo informatico un tema caldo: questo crescente interesse potrebbe essere dovuto alle conseguenze che questa tecnologia sta portando. Il concetto di fiducia nella tecnologia è complesso, variegato e molti elementi possono essere presi in considerazione a seconda del contesto d'applicazione. Nella tecnologia e nei sistemi informatici esistono gli stessi problemi di aspettativa, disponibilità ad essere vulnerabili e possibili esiti negativi (Hall, 2010).
La fiducia è un prerequisito dell'interazione uomo-uomo: l'essere umano ha il bisogno di avere un ambiente prevedibile e di comprendere le intenzioni e le motivazioni del comportamento degli altri. Per interagire anche in contesti e circostanze meno prevedibili, le persone applicano strategie simili anche nell'interazione uomo-computer. La fiducia gioca un ruolo fondamentale nella tecnologia perché è un "predittore primario" per l'uso della tecnologia (Li et al., 2008), aumenta i tassi di adozione (De Kruijff, 2018) e l'accettazione di un sistema informatico (Parasuraman & Riley, 1997).

Secondo alcuni ricercatori, la fiducia riguarda le dinamiche relazionali e in quanto costrutto sociale, non è possibile costruire una relazione di fiducia nei confronti dell'IA essendo priva delle qualità intrinseche necessarie. Questo lavoro esplora i diversi modelli e ricerche sulla fiducia in senso psicologico applicato in ambito Human-Agent-Interaction, per trovare le strategie per costruire una relazione di fiducia negli agenti. Questo è stato combinato ad un'investigazione sulla percezione, l'esperienza e la fiducia riposta in un assistente vocale (in questo caso Amazon Alexa) in due diverse modalità d'interazione. Il problema è stato affrontato attraverso un'analisi della letteratura per comprendere i diversi approcci di ricerca legati alla "fiducia", "ai fattori dell'agente che contribuiscono alla fiducia" ed è stato analizzato in una ricerca con 12 partecipanti. Lo scopo era quello di comprendere i fattori personali, contestuali o relativi a fattori specifici nella costruzione della fiducia tra utente e assistente vocale. Il risultato della ricerca individua gli elementi di progetto utili a costruire una relazione di fiducia; la tesi ha prodotto inoltre linee guida per la progettazione orientata alla costruzione fiducia in uno scenario di prima interazione.

Introduction

Is it possible to build trust in voice activated assistant? If so how?

In order to respond to this quest, many approaches have been analyzed, ending up in a thesis consisting in 7 chapters.

Chapter 1 consists in an introduction about trust in psychological/sociological research approaches and their application to Human-computer-Interaction, addressing the state-of-the-art, the characteristics and the reason why it is relevant.

Chapter 2 aims at exploring the object of trust, first with a more general introduction and the characteristics of Voice User Interface in order to bring out the intangibility that makes building trust in these interfaces as more complex. It also contains a discussion about the design elements of these interfaces with a focus on four case studies.

Chapter 3 focuses on one of the mentioned approaches to build trust in assistants: human-to-human trust approach. It is discussed what has been already done in the field of HCI, what are the consequence of designing human-like cues in AI assistants and a possible tool to build trustworthy agents.

Chapter 4 describes the methodology of the current work. Starting with the research questions, progressing on how the qualitative research is conducted, the tools used, the participant selection and the modality of the data analysis.

Chapter 5 aims at providing the results of the test conducted, as well as showing possible interpretations of users' behaviors during the study.

Chapter 6 focuses on providing a design solution, addressing the factors that are found to be relevant in trust relationship development.

Chapter 7 focuses on providing a set of guidelines for designers and developers to create trust in Voice activated assistants.

Chapter 1

Can we build trust in AI agents?

Many researchers tried to give an answer to this simple question. Many of us have our own idea/definition of what is trust and what makes something/someone trustworthy or not (Golembiewski, McConkie, 1975), this even in research makes more difficult to set boundaries to the topic.

It a complex and multidimensional construct that lead to various interpretation depending on the context and different social situation (Kramer & Carnevale, 2001). Researchers do not agree on a unitarian/universal definition of trust (Bigley & Pearce, 1998; Rosseau et al., 1998, Das and Teng, 2004), some are even contradictory and others confusing. Knights and Chervany (2001-2002) conclude that is almost impossible to definite, due to the the broadness of the concept and the interactions with many discipline, within the context of this thesis multiple approaches are going to consider to give a picture of the complexity of the phenomenon. One the reason is probably due to the fact that there are many views of trust (Barber, 1983; Shapiro, 1987). Also, many disciplines conceptualized their own idea of trust, considering different variables, interactions and situations which in scientific terms are methodologically difficult to merge.

Lewicki and Bunker (1996) categorize look "trust research" into three main domains: personality and individual differences, social psychology and group interactions/ transactions, institutional sociological phenomena.
Psychologically topic of interest are related to personality, individual differences, dispo- sitional characteristics, antecedents of trust and cognitive/affective elements. Instead the economics research literature, trust connotation is more related to analytical judgement in term of cost and benefits as outcome to the trust behaviour, their focus is on calculated rationality of the decision making.

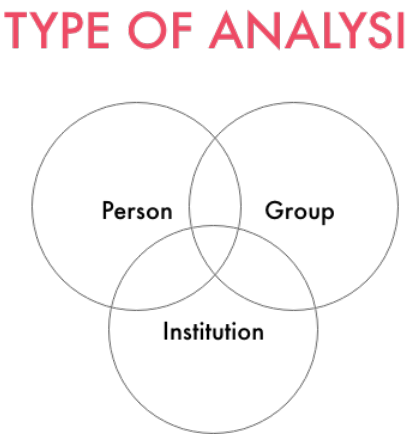
Sociological stream of research has a focus of trust in the context of social structural systems in terms of institutions, organizational and reputation.

There other perspective and applied context of interest such as IT security, education, reputation, management and communication.

As McKnight (2001) says few models are taking into consideration a multidimensional perspective, which makes investigation more context dependent.

Type of analysis

There are many levels of analysis of the concept of trust, most of the research conceive trust depending on the selected domain, context, or specific characteristic to reduce the variables taken into consideration. Other type of analysis are focus as McKnight and Norman L. Chervany (2001) says on conceptual types *"disposition, beliefs, attitudes or behaviors"*
Some other authors' analysis focus on the cross-level conceptualization of the phenomenon considering the actors of the interactions rather than the context of application. Those research focuses on the different interactions such as person-to-person interaction, person-to-group or group-to-group, considering context secondarily. There are common layers in the interaction as can be seen in the following diagram:



In this investigation the analysis will be assessed contextually and domain specific: analyzing the basics of trust dynamics and the processes to build trust considering both human-to-human trust models and human-to-computer (agent) trust models.

Definition problem

Despite the fact that there is no universally accepted definition of trust is widely recognized the importance of the concept. Some scholars believe that is not possible to defined and so in their research they chose not to do in their research of interest (Ouchi, 1981).
In the multitude of previous researches often the concept of trust has been given different meanings of trust depending on the different perspectives in the topic, the selected context, and the stage of trust interactions but researchers agree that the nature of trust depends on which level is measured. Despite this lack of agreements can be found some consistency and common patterns in research.
Usually there some basic elements involved with trust:

- Trustor (individual/institution) that is the actor that "trust/distrust" a target entity
- Trustee (individual/institution/artefacts) is the target entity
- A decision making made by the trustor in order to achieve something

The most frequently cited definition of trust are:
Mayer, Davis, and Schoorman (1995) defined trust as the *"willingness to be vulnerable"*
Mcknights and colleagues (1998) conceptualized trust in a similar way as the *"willingness to rely"* on others.
Shapiro conceptual trust as a *"Positive expectations"* of others
Rousseau et al., (1998): *"a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another"*
Luhmann(2002): "Trust is to believe that one's expectations will be fulfilled."

Lewicki & Wiethoff (2000): *"an individual's belief in, and willingness to act on the basis of, the words, actions, and decisions of another"*
Sheppard & Sherman (1998): *"Trust is accepting the risks associated with the type and depth of the interdependence inherent in a given relationship"*
Deutsch's definition of trust (1962) *"(a) the individual is confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial (Va+) or to an event perceived to be harmful (Va-); (b) he perceives that the occurrence of Va+ or Va- is contingent on the behaviour of another person; a (c) he perceives the strength of Va to be greater than the strength of Va+ . If he chooses to take an ambiguous path with such properties, I shall say he makes a trusting choice; if he chooses not to take the path, he makes a distrustful choice."*
Despite the popularity of Deutsch's definition which conceptualize trust as cooperation, today is believed that fundamental to distinguish cooperation to trust, which can be related and influence trust but is not the underlying phenomenon in the relationship between the two parties.

A more recent meta-analysis, done by Castaldo et al. in 2010, considered 72 different published definitions of trust even from different disciplines and came up with the following definition:
Trust *"as*
(a) an expectation (or set of beliefs, a reliance, a confidence) that a
(b) subject distinguished by specific characteristics (honesty, benevolence, competencies, and other antecedents)
(c) will perform future actions aimed at producing
(d) positive results for the trustor
(e) in situations of consistent perceived risk and vulnerability."

Comparing all main definition of trust the term seems to have something to do with truster vulnerability, possibility of risks and expectation of a positive behaviour (of the other party). Risk and interdependence are considered the two fundamental elements for trust, the degree in which they are related can be different depending on other factors.
Risk is the subjective perception of a possible probability of lost from one party in the interaction, but it can be considered an opportunity to trust to achieve a goal. The perceived risks depend on personal evaluation due to uncertainties, in fact, when one party is confident and certain about the outcomes there is no trust, because there is not risks taking. Interdependence is because the needs/interest of one party would not be fullfill is the interaction and reliance on the object of trust is not taking place.It is also agreed that trust is not a behaviour or a choice per se, in certain extent they can be misunderstood.

The models that are going to be presented further in this chapter represent different conceptualization of trust due to the context in which the interaction takes part.

Distrust - trust

Often Trust researchers defines of "distrust" as the dualistic and ontological asset of trust. Although the phenomenon has been investigated less compared to trust, it is a frequent phenomenon.
Distrust occurs when the subject feels high vulnerability and has a high perception of the risks that might happens. It is functional because it avoids the negative consequence of trusting one party when the perceived risks are too high. The subjective feeling of vulnerability can be explained by two things: one related to the specific person/institution involved (previous experiences or indirect knowledge) or the unpredictability of the situation (Koslow, 2000). In certain context and situation distrust give a competitive advantage on the individual.
The concept of distrust or trust is not mutually exclusive, as it will be classified later, trust has different degrees is not merely a decision making of yes (Trust) no (distrust).

Type of trust

There are many types of trust in the scientific literature, but the majority agrees on two main ways to trust (Paliszkiewicz, 2010; Chowdhury 2005; Lewis and Weigert 1985; McAllister 1995):

- **Cognitive trust**

Cognitive due to the fact that trust results of a cognitive process which is knowledge-driven information available about the object of trust (person/group/ company/institution). Trust implies the awareness that presume a risk (due to a incomplete knowledge about the trustee and the possible outcomes) and the trustworthiness is based on evidences.When the trustors is confident of the good intentions or the intent of the trustee this condition is risk disappear for trustiness.
As a cognitive process it implies both conscious and unconscious processes (Vaisey 2009) that depends of the current experience of interaction but also all the previous and the knowledge of the object of trust. This means that trust is both automatic and deliberate due to the context: the belief of trustworthiness of the object of trust can be rational or irrational in order to fulfill personal goal and desire. Even if it is evidence based it is subjective and it up to the trustors decide to trust or not the trustee and in which circumstances.

- **The Affective trust**

Opposite to cognitive trust, affective trust is not rational and not resulting from evidences-based evaluation. Instead is emotional driven and it is related to the emotional bonds/relationship between the two parties.
It can be defined as the confidence felt by trustor due to the level of care and concern trustee demonstrate (Rempel et al., 1985). It is related to the feeling of security coming from the relationship between the partners, especially when it deepens the level of trust goes way beyond the actual knowledge. Affective trust happens when a relationship between the parties is already existing and especially when there is an emotional bond between them. At the same time the opposite effect might happen temporary positive emotions/perceptions that are not contextual or specific the the trustee may lead to trust trustee even if it is not dependent of them (Schoorman et al., 2007).
Early studies were focusing on how negative emotion (e.g. anger) would impact (negatively) trust, more recently more emotions have been studied. These emotions can shape temporary the assessment of trustworthiness of the truster.
The current focus is on understanding trust affective components and on mutual caring/bonds between the parties involved (McAllister, 1995). The elements involved are in interpersonal relations are about sharing goals, believes, values. This shared similarity increases trustor's trust and emotional bonds with each other. The two types of trust are not mutually exclusive, in fact, the cognitive trust is a base for the development of more affective trust (Lewis and Weigert, 1985), at the same time if there is a strong relationship, the affective trust can have an effect on the cognitive (McAllister, 1995)
Similarly, Mayer's model distinguishes between benevolence, which has a large affective component, and competence, which places emphasis on the cognitive component, as two key dimensions of trust.
Paul Ingram says that *"In the West we tend to reserve the trust that comes from the heart (affective trust) to friends and family, while we turn the trust that comes from the head (cognitive trust) to business partners".*

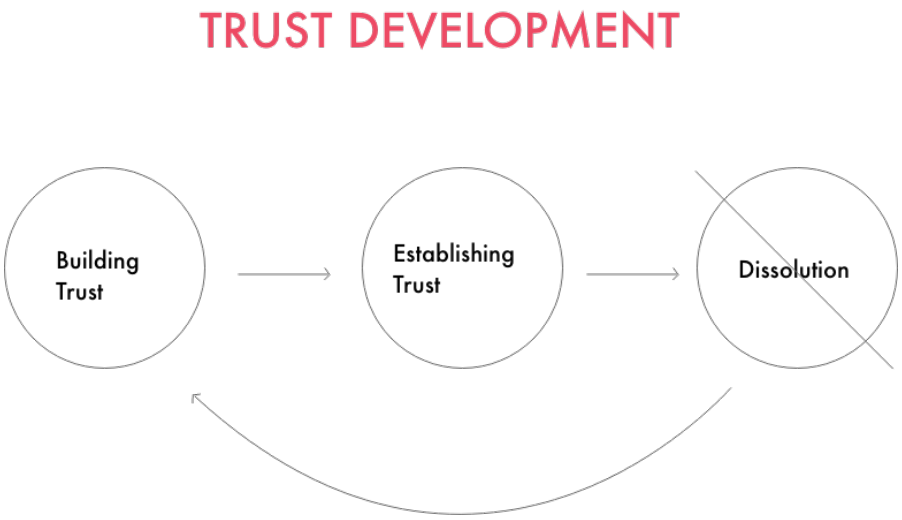
Trust as process

Trust is a dynamic and ongoing process. It is typically built up gradually and it is interpersonal it needs since requires two ways interactions (McKnight, 1998).
Empirical evidences prove that trust is not consistent, it changes over-time, it can diminish, increase or die according to the interactions or the situation involved (Paliszkiewicz, 2010).
It can change with experiences and multiple interactions, with different parties involved in the interaction and the context in which the interaction take place.

Trust changes with time in positive or negative depending on the outcome of the interactions due to positive or negative reinforcement.

The three phases of development of trust can be formalised as:

1. Building trust when first is formed, this stage is fundamental because it can create the basis of a trust relationship
2. Stabilising trust: trust previously built changes according to the interactions and positive/negative outcomes
3. Dissolution: happens when people feel distrust, but this phase is reversible if trust base is built again.



Trust might be construed differently and have more or less importance in the differents stages of the relationship (Fletcher, Simpson,& Thomas, 2000)

Temporal dimension of Trust

It is not universally recognised whether trust increases with familiarity, but it has been proven that trust dynamics change over time.

- **DIETZ AND DEN HARTOG MODEL**

Context: Organisation

Trust definition: *"Trust is not a simple 'either/or' matter"* (2006)

Trust antecedents: None

Dietz and Den Hartog model conceptualised the different stages of trust in intra-organisational phenomenon. Combining different conceptualisation of trust on a timeline. Not every relationship - in an organisation or in another context, will lead to complete trust, many dynamics can affect each stage of trust building.

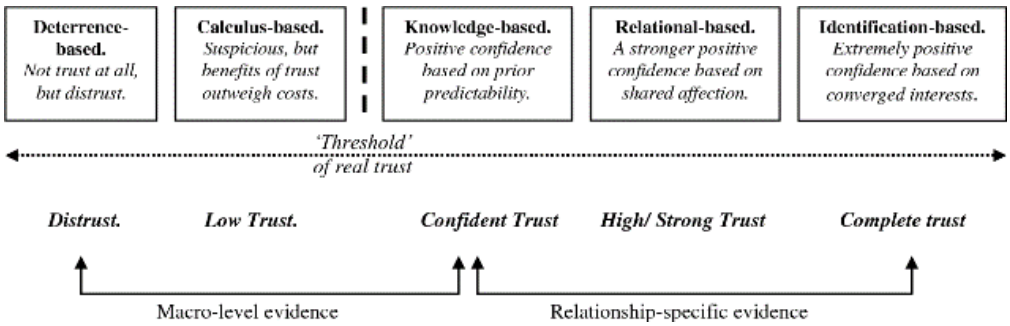


Fig. 1 Measuring trust inside organisations (from Dietz, G., & Den Hartog, D. N. (2006)).

– Deterrence-based: distrust;

one party will be trustworthy because the cost of the sanction of not being would be very high (Shapiro, Sheppard, & Cheraskin, 1992) while for Dietz and Den Hartog this is not real trust because there is not any perceived risk or expectation of a positive outcome. It is an opportunistic trust, per se there is no vulnerability or probability to risk. The authors consider this type of trust as distrust.

– Calculus-based: low trust;

This trust is based on rational evaluation in economic term of cost and benefits in trusting and the negative outcome of betrayal. According to Deutsch trust decisions are usually based on some form of cost/benefit analysis. For certain relationship is all, for others this might evolve. this positive outcome are also the results of credible informations of the intentions or competence of the trustee (Barber, 1983).

– Knowledge-based: confident trust;

Authors (Lewicki and Bunker, 1996) see that perceived risks are changing towards a more positive evaluation of the positive outcomes, the previous knowledge/information or experiences about the trustee influencing are also the perception of motives/ability/reliability. All the knowledge acquired through the multiples interactions that gives expectations and predict how the other will behave towards us. This type of trust evolves through positive experiences that create confidence of trustworthiness and predictability.

– Relational-based: strong trust;

Relational trust involves interaction, networks, the development of the trust into relationship and different situations.Relational trust is defined as resulting from repeated interactions between trustor and trustee in temporal space.

Positive expectations are resulting from previous interaction, giving the trustor's the impression of the good intents/motives of the trustee. The phenomenon refers to interpersonal social exchange in groups dynamics. trust is the results of the interaction with people that shares the same values/and moral (Uslaner, 2002) or have the same social identity (Tyler, 2001). In conclusion this type of trust can be seen as an evolution of the interactions within relationship that lead to a change of the foundation of trust itself.
According to other authors relational trust (with theories usually in the context of organizational dynamics) conceived it as a strong type of trust (Dietz and Den Hartog 2006).

– Identification-based: complete trust.
Refers to the understanding of trustee intentions and motives. This happen with time and where the trustor formulated an identity of the trustee and possible a shared value. This level of trust is reached when there is an emotional bond with the other party. The parties involved understand each other intentions and motives, understanding their desires. It is built through is grounded in perceived compatibility, shared goals, and positive relational bond to the other party. Also called complete trust because there is a high confidence of the positive outcomes regarding the other party: each parties internalize each others preferences (Lewicki,1995) and would care about the positive negative outcomes for the other party, this dynamics will lead to consider the other as more trustworthy(Shapiro et al., 1992).

Stages of Trust

As described by Dietz and Den Hartog model trust have a temporal dimension, and the level of trust/distrust depends on dispositional level or characteristics of the trustee, but those are affected by the frequency and the number of interactions between the parties. Trust take different importance in different stages of the development of a relationship.

Other models will be compared the difference and similarities upon the temporal aspects topic. Initial trust happens the "trustor" decide to trust a trustee (artifacts/human) before having any kind of experience of him/it. The more traditional viewpoint of "initial trust" believed that in case of absence of previous experiences the subject would be neutral. It was believed that people would "simply suspend(s) belief that the other is not trustworthy and behave(s) as if the other has similar values and can be trusted"(Jones, George,1998). Nowadays this point of view of neutrality has been abandoned, the followings chart compares the different types of trust implied to the different stages of the relationship.

- **STAGES OF TRUST MODELS**

STAGE			AUTHOR
Early	Medium	Mature	
Deterrence-based	Knowledge-based	Identification-based	Lewicki & Bunker 1996
Cognitive-based Calculus-based	Knowledge-based Personality-based	Instition-based	McKnight et al 1998
Calculus-based		Relational-based	Rousseau et al 1998
Ability; competence		Benevolence, Integrity Intentional Goodwill	Mayer et al 1995 Klein Woolthuis 2005
Competence	Prectability	Goodwill	Ratnasingam 2005

More specifically:

- According to Lewicki & Bunker (1995) "interpersonal trust" start from a more "economical perspective" of calculus-based benefits of giving trust and the cost of unsocial-betrayal behaviour. If the relationship increases the nature of trust evolve as well towards a more knowledge and later relational where the two parties identify on each other.

- Similarly, Rousseau (1998) sees the development from a more calculus based towards a relational approach with the development of time.

- McKnight in his multi-dimensional model sees trust development from a more rational decision making in term of pay-off to an information seeking approach and personality based to arrive at an institutional trust.

- Mayer's model sees as antecedents of trust ability and competence and as further level of relationship integrity and goodwill.

- Instead, Ratnasingam (2005) identifies three level of trust: it starts with competence, reach a level of predictability of the other party's behavior and with time and multiple positive interactions the belief of goodwill of the party towards the truster.

In fact, trust might be construed differently depending on the different stages of relationship development (Fletcher, Simpson,& Thomas, 2000; Larzelere & Huston, 1980)
To merge all the theories across the discipline as can be seen from the previous there are similarities across the process of building trust (Rousseau; 1998).

A MODEL OF TRUST

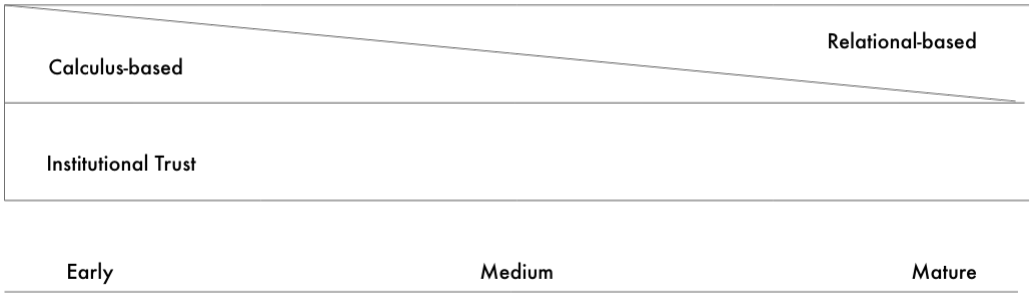


Fig. 2 Not so different after all (Adaptation from Rousseau, 1998).

Rousseau model can be seen as a simplified combination of the different type of trust related to the stage of development:
At the beginning of an interaction of two parties the initial trust is more opportunistic (calculative trust) and also influenced by institutional trust, after a while (depending also on the kind of relationship between the parties) if the relationship progresses the trust is more relational type but institutional trust will always have an impact in later stage of relationship.

Primary trust models

The following models have been selected for their characteristics because unlikely many others include also the analysis of the antecedents factors of trust, that are all the elements that can influence in establishing trust. They offer a framework to better understand the factors that shapes pre and initial trust that goes beyond the interactions between the two parties.

The context in which trust has been studies is highlighted and they have also been selected due to the fact that in other research have also been applied in the context of technology.

- **MAYER'S MODEL (1995) INTERPERSONAL MODEL TRUST**

Context: Organizational trust

Trust definition: *"Trust is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform*

a particular action important to the trustor, irrespective of the ability to monitor or control that other party."

Trust antecedents: Ability, Benevolence, Integrity

Mayer and colleagues (1995) developed the most-well known model of interpersonal trust in the context of organizations in which the company have the usually the role of trustor and the employee/partner usually of the trustee. Despite the fact that addressed originally intra-organizational context, later the model has been applied and generalised in many different contexts that involves with interpersonal trust.
Interpersonal trust is defined as the willingness of the trustors to be vulnerable to the actions of trustee based on their assessment of ability, benevolence and integrity (Mayer, Davis, and Schoorman 1995).
The trustor makes a decision to trust to the object after assessing the other party's trustworthiness. The trust is based as the subjective perception that the trustee will not hurt the trustor, due to this is willing accept the vulnerability based on their perception of the other person's behaviour. The positive/negative outcome of the behaviour will have an impact on the factors of trustworthiness leading to a new start of the circle.

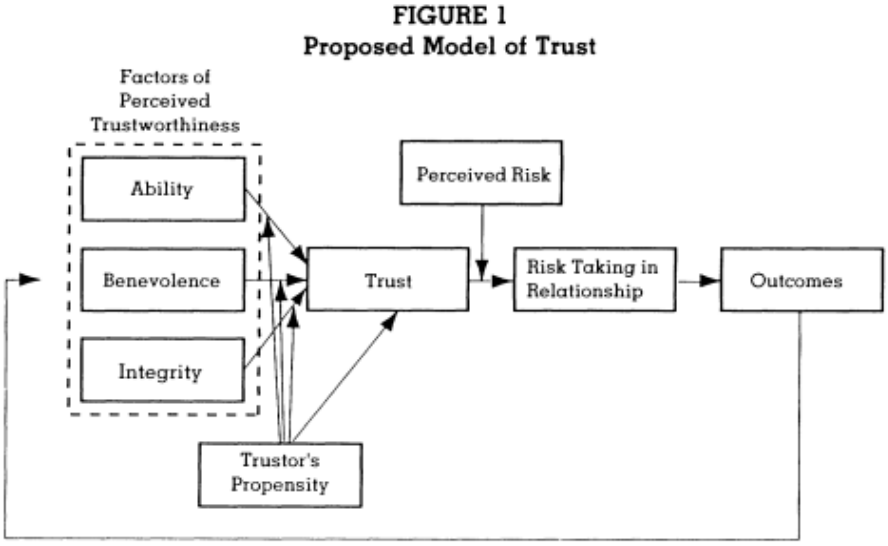


Fig. 3 An Integrative Model of Organizational Trust (from Roger C. Mayer, James H. Davis and F. David Schoorman (1995).

The following model consider three main predictors of trust:

- Trust beliefs (here called factors of perceived trustworthiness) that consist on the (subjective) perception of the ability, benevolence and integrity of the trustee

- Trust intention that is the willingness to risk trust (trustor's propensity)
- Trust behaviours that are the set of actions that prove the willingness to depend on the trustee.

Ability: Set of skills, competencies or characteristics owned by the trustee in a specific domain. The perception of ability is relevant depending on the situation, so "trust" behaviour depends on the specific context (Zand, 1972). According to the author is "a group of skills, competencies and characteristics that enable a party to have influence within a specific domain" (mayer et al 1995). This mean that a party can be considered very skilled and competent within a context but not in another.

Benevolence: the belief that the trustee will have good intention would care, and not simply be moved by only his/her profit. This is the subjective perception of the trustor's that the trustee has good intentions and motives.

Integrity: the trustee's set of morals and ethics which principles collide to personal integrity. the trustors should should agree on a set of principles to perceive the trustee as trustworthy.

As mentioned Ability, Benevolence and Integrity are considered the main factors of (perceived) trustworthiness and refers to the set of characteristics of the trustee that are subjectively inferred by the trustor. Those characteristics are considered antecedent to trust.

Later to the first model, the authors adds the disposition to trust or what they call Propensity of trust. If fact even if the trustee receives a subjective positive evaluation of all the dimensions, the overall trust will be modulated by the propensity of trust of the trustor.

In general, Mayer and collègues do not weight all the variables the same way: propensity to trust have a big impact, but also between the the variable of the perceived trustworthiness there are some discrepancy. Ability is considered higher in value compared to integrity.

Perceived risks refers to the risk analysis of the trustors in case of positive/negative outcome. Risk taking refers to the selected positive/negative behaviour of the trustor after having analysed the previous instances. So it is the willingness to take risks.
This model is circular, in fact: if the risks taken by the trustor will lead to positive outcome, the experiences will be positive on future evaluation and situation when the trustee will be involved. On the contrary if the outcomes are negative the trust behaviour will be less likely.

The strength of the models lays on the conceptualization of the major factors involved in the antecedents of trust. The major weaknesses are the lack of context analysis and the type of possible outcomes.

- **Propensity of trust**

Already in Mayer's model propensity of trust or disposition to trust was an important factor that influenced the overall trust, especially at the beginning of a relation.
Other authors have been investigating the propensity to trust and found that: it is modulated by both early life experiences (Flanagan and Stout 2010 and specially through cultural/social transmission (Uslaner 2008) and psychological disposition, and not simply to the contextual contingent moment of the interaction.

Dispositional trust is a personal tendency to trust others (Pearson, 2008), it stable personality characteristic even in different context. it is a specific personality characteristic and predicts development of the relationship and the overall satisfaction (Rempel et al., 1985). Dispositional trust is an automatic response not resulting from previous knowledge or experiences (Kim, 2003). Disposition to trust, it's the tendency, which according to some authors not related to a context or a specific situation, but having a trusting attitude towards others dispositionally (McKnight, 1998).

"High scorers have a disposition to believe that others are honest and well-intentioned. Low scorers on this scale tend to be cynical and skeptical and to assume that others may be dishonest or dangerous" (McCrae,2003).

Personality is also a dispositional trait that can lead to trust/distrust. Personality traits such as agreeableness, openness and neuroticism of the trustor's has been found correlated to this propensity to trust (Dinesen, 2014). Extraversion trait it has also been found correlating with higher propensity to trust (Mcbride and Morgan, 2000).

Something to consider that is related to person's disposition is the need of control of the truster. Trust dynamics has a lot to do with subjective tendency of control personal life events (Burger et al, 1979), Mayer (1995) states that *"trustors expect that the trustee to perform helpful actions irrespective of his ability to control him".*

Dispositional trust is important especially in the beginning of an interaction between trustor/trustee and can lead to initial trust/distrust due to personal factors.
As the relationship between the parties progresses, this tendency in determining trust decrease the power of its influence in the relationship (McKnight,1998). Instead other authors says that this tendency is the result is

also an ongoing process but is the result of an ongoing process and socialization (Fukuyama F.1995) that rewards positive interactions and lead to overall increase of availability to trust.

It was proven disposition to trust was correlated positively to trust websites and ecommerce (Song & Zahedi, 2007).

- **MCKNIGHT (1998) INITIAL TRUST FORMATION IN NEW ORGANIZATIONAL RELATIONSHIPS**

Context: Initial trust organizational trust

Trust definition: *"One believes in and is willing to depend on another party"*

Trust antecedents: Disposition to trust, Institution based trust, cognitive processes

McKnight et al. (1998) models analysed how initial trust (where the parties do not know each other) is formed within organizations. As initial trust is not based on neither any information and knowledge or experience. As can be seen in the graph below this initial trust is based on the disposition to trust of an individual, and the influence of the institutions that modulated by cognitive processes lead to this first initial trust.

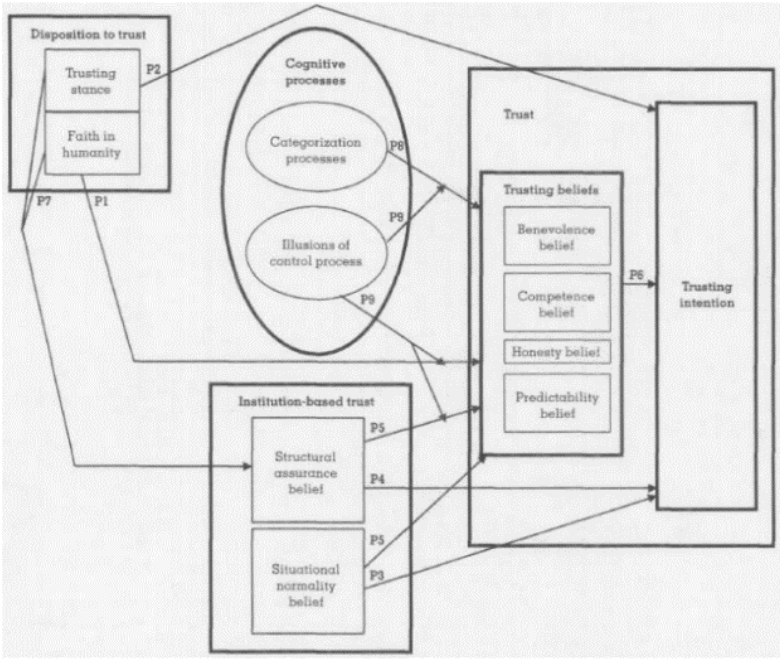


Fig. 4 Model of initial formation (from McKnight et al, 1998).

They conceive trust on two elements: trusting intentions and trust beliefs.

Similarly to mayer's model, trusting intentions is "the willingness to depend on the other party" while trust beliefs is considering the other party as:

- Trust belief-benevolent (one party believes that the other will not act opportunistically and will care about the other party)
- Trust belief-competent (one party believes that the other is able/powerful to achieve a specific goal)
- Trust belief- integrity (one party believes that the other is faithful and would tell the truth and respect the promises)
- Trust belief-predictable (one party believes that the other's person behaviour is consistent to be able to predict it)

Disposition to trust according to the authors consists in:

- Faith in humanity is psychologically based and it's the belief that others are generally trustworthy, so it is a general tendency to trust others. The other are seen as honest, benevolent, competent, and predictable.

- Trusting stance is the *"willingness to depend on others"* being trustworthy or not, as the authors says is more a personal strategy, trusting beliefs are specific towards a person/situation etc.

Institution-based trust

Institutional trust means that one party believes that the institutional structures will help defining and regulate the relationship creation. It consists in :

- Situational normality: that defines as the more normal a situation is the more likely is successful. The more a person feel comfortable with a (normal) situation the more likely is going to trust.

- Structural assurance: are all the regulations, contracts, promises that regulate the relationships. All of these elements give and assure a set of future behaviour from the other party and they are likely to reduce risks. Structural insurance has an impact especially at the beginning of a formation of a relationship.

Both situational normality and structural assurance will have an impact on trust beliefs that will affect trusting intentions.

Cognitive processes

The cognitive processes support the person to categorize whether another party can be trust (and develop trust beliefs, those are:

- Unit grouping: people being part of a group tend to share a common goal and values. Trust beliefs between people in the same group is more likely to happen.

- Reputation: Personal reputation influence the perception of trustworthiness. Depending on the context this can reflect competence of other trusting beliefs making the future behaviour more predictable

- Stereotyping: stereotyping influence the way we make initial judgements about people. They can be applied a small or in a larger context (e.g. gender vs nationalities, appearances). By implying positive stereotypes help to increase trust beliefs, Trusting beliefs also influence trusting intentions (McKnight et al., 1998).

- INTERDISCIPLINARY MODEL OF TRUST CONSTRUCTS. MCKNIGHT AND CHERVANY (2001)

Context: Online trust

Trust definition: *"One believes in and is willing to depend on another party"*

Trust antecedents: Disposition to trust, Institution based trust, cognitive processes, trust-related behaviour

A more recent version of this model - McKnight and Chervany (2001) - it was added a behavioral asset called "trust-related behaviours" and a focus towards the difference between trust and distrust.
The authors analysed first different definitions of Trust from different disciplines and found some common pattern such as: attitude, intention, belief, expectancy and behaviour, disposition and institutions/structures.

After having combined the belief and expectancy to create a conceptual model of trust combing the different approaches. The general element from the previous model has been kept, adding factors of distrust and trust related behaviour.

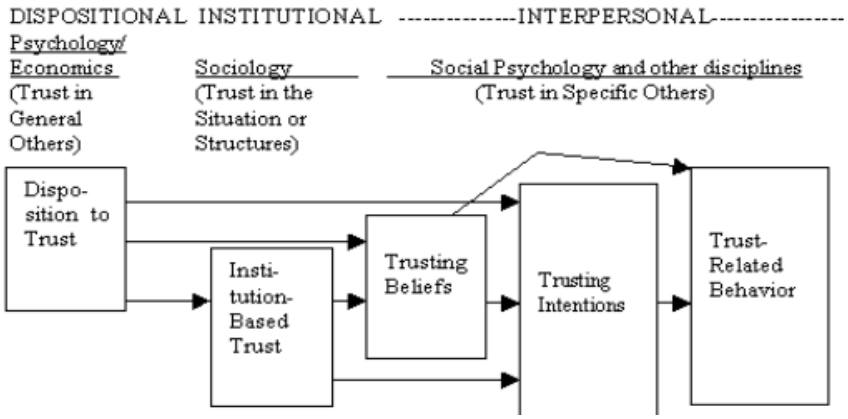


Fig. 5 Interdisciplinary model of trust (from McKnight and Chervany, 2001).

Trust related-behaviour:

defined as "a person voluntarily depends on another person with a feeling of relative security, even though negative consequences are possible". As Mayer's model trust related behaviour implies an acceptance of the risks because behaviourally trusting somebody means giving to the other some power (that is dependence based) to them. Examples of those behavioural trust elements are "cooperation, information sharing, informal agreements, decreasing controls, accepting influences granting autonomy and transacting business" (McKnight & Chervany, 2001). Each of these trust-related behaviour makes the trustor more dependent to the trustee.

As can be see the authors frame a dualistic and mutual exclusivity trust/distrust conceptualization:

			Interpersonal		
	Dispositional	Structural	Perceptual	Intentional	Behavioral
Trust:					
Conceptual Level	Disposition to Trust	Institution-based Trust	Trusting Beliefs	Trusting Intentions	Trust-related Behavior
Operational Level	• Faith in Humanity • Trusting Stance	• Structural Assurance • Situational Normality	Trusting Belief- • Competence • Benevolence • Integrity • Predictability	• Willingness to Depend • Subjective Probability of Depending	• Cooperation • Information Sharing • Informal Agreements • Decreasing Controls • Accepting Influence • Granting Autonomy • Transacting Business
Distrust:					
Conceptual Level	Disposition to Distrust	Institution-based Distrust	Distrusting Beliefs	Distrusting Intentions	Distrust-related Behavior
Operational Level	• Suspicion of Humanity • Distrusting Stance	• No Structural Assurance • No Situational Normality	Distrusting Belief- • Competence • Benevolence • Integrity • Predictability	• No Willingness to Depend • Subjective Probability of Not Depending	•Lack of Cooperation • Information Distortion • Formal Agreements • Increasing Controls • Not Accept Influence • Not Grant Autonomy • No Business Transacting

Fig. 6 Interdisciplinary model of trust (from McKnight and Chervany, 2001).

- **SIMONE'S BORSCI MODEL - USER EXPERIENCE TRUST'S MODEL**

Context: healthcare/artefacts

Trust definition: Mayer's et al (1995) definition

Trust antecedents: visual appearance, Usability, Ease of use, Acceptability of technology, Prior experience (if any)

Borsci's model merge in the context of life cycle experience the role of some factors (of the object of trust) that increase initial and post experience trust. If the object of trust is perceived subjectively by the user as aesthetically pleasant, usable, easy or usable there is an increase in initial trust (and even pre-trust) and the overall post trust evaluation.

Simone Borsci's there are *"Empirical studies suggest that people do have a sense of "trust" toward systems (TTS)"* And specifically according to his research studies, trust in AI Systems is:

- Related to the perceived usability and acceptability of the technology
- Affect people's attitudes towards the product before and after the interaction

According to his research, and also in the literature it was found that pre-use and post- use of TTS can be shaped by design (P Pu, 2006)

- A product that is perceived usable and useful is expected to generate high level of post use TTS
- Aesthetically pleasing product lead more to pre-use trust than unaesthical pleasant ones (Pengate,2017)

The model of Pengate et al(2017) was used as a base to formulate his model considering as antecedent of trust: perceived visual appeal, perceived ease of use, perceived usefulness that would have an impact on trust and intention.

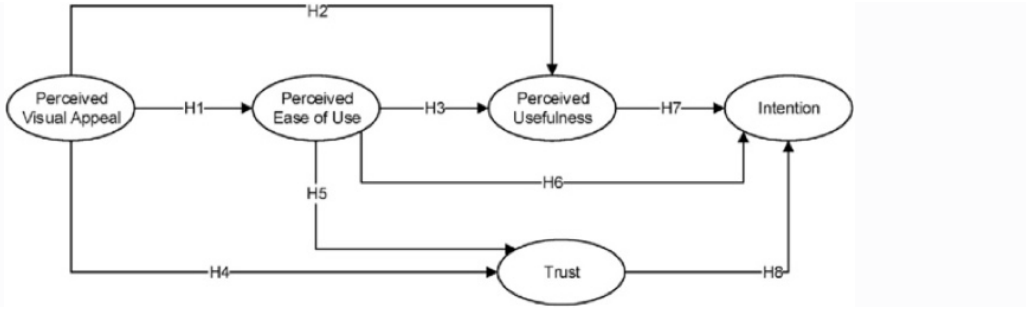


Fig. 7 An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors. (from Supavic Pengnate, Rathindra Sarathy, 2017).

Simone Borsci's model considers those variables on a temporal dimension of user experience and considering those that have an effect on pre-use trust and trust post-use.
According to his view people are implicitly *"placing (indirectly) their trust on the fact that manufacturers have created a product/service with certain set of qualities and characteristics e.g., usefulness, safety, learnability, usability and reliability Assessment of trust after the experience (also affect the Brand)"* (Borsci,2019).

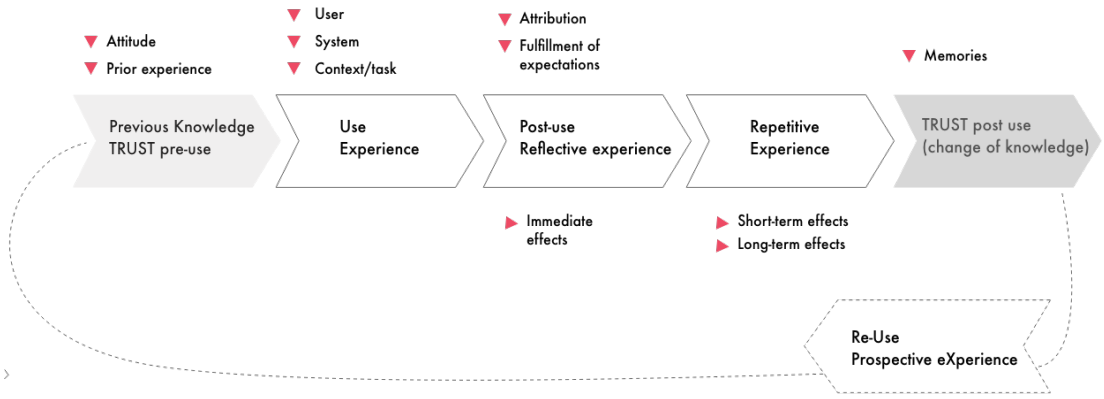


Fig. 7 Lifecycle trust model (Adaptation from Borsci, 2019)

- **Pre use-trust**

Even before interaction the potential user have certain expectation on the product which can be positive or negative. Those expectations can be objective or unrealistic due to the lack of knowledge of the system and based on either personal attitude or prior knowledge.

- **User experience**

This stage deals with usage and modulated by usability/Ux models. The main influences are user characteristics(skills/personal preference), system (instrumental/non instrumental qualities) and the context of usage.

- **Reflective Experience**

User first analysis of the interaction with the system which is influenced by the positive/negative outcome and attribution.

- **Repetitive Experience**

In this phase there are many interactions depending on user, system, and context and long vs short effects.

- Retrospective

After repetitive experiences user make an evaluation of the overall experience and the system. High trust is correlated to high perceived qualities of the system (usability, aesthetics, usefulness).

AI agents Trust model

Trust is contextual and situational, trust in technology is not specific enough to describes the trust relationship among people and AI conversational agents. Due to the fact that AI agent simulate human beings to certain extent, due to the fact of their design characteristics (that will be discussed in more details in Chapter 3).
In the following paragraph the main approaches to build trust in AI agents will be analyzed to understand which factors both antecedents and during the experience would lead to trust.

- **D. LEE, KIM, AND MOON MODEL - A MODEL OF AGENT SUCCESS**

According to Lee's and colleagues' model of acceptance of software agents *"user will be more accepting of the agents if they develop feelings of trust"*. Many factors contribute on the success/loss of the AI agent.
Trust feelings and perceived risks can be measured and defined independently and will impact: to make a trustworthy agent, the perceived risks need to be addressed.Trust in fact increases willingness of acceptance and interaction, while the perceived risks contribute to its rejection. The perceived risks are subjective and mostly depend on the user, they might not even be related to the actual risks of the used technology. If the subjective levels of perceived risks are very high, there is not any sufficient level that can counterbalance and make the agent accepted. It is also important to mention that, overall, user's trust on an Agent is modulated by the perceived risks. In this model the two branches balance the overall success of the agent.
Similar to Borsci's model, experience and aesthetics of the interface will impact the perceived trust. Experience is important because it can change users' willingness to trust (Marsh, 1994). According to this model, perceived risks can be due to personal factors (dispositional distrust), as proven in the literature. Overall conditioned by uncertainties of the intentions, behaviors of the agent and the autonomy level of the agent itself. The model is not meant to have a temporal dimension but represents the main variable that influences the agent acceptance. In order to design a trustworthy agent/assistant, the designer need to increase the perceived benefits (in this model called trust) to find a balance with the perceive risks that cannot be avoided. Some of the factors are out of the designer control such as the user's dispositional factors, however, can work on making a remarkable experience, providing a comprehensive

information through sapient communication and interface design. In this analysis is not mentioned the role of company's reputation or institutional trust that will have an impact on the perceived risks and overall trust, depending on the positivity/negativity of the reputation itself. It was found that trust in computer systems is correlated to reputation, in general. *"The trust we place in a technology is tied to the entity that produced it"* says Meadows. *"This should also work with bots and robots."* (2016). According to the author, in order to adopt them, they need to have a reputation and documentation on who develops it, what they do and who is responsible for them. According to this view, a user implicitly trusts an AI agent depending on the company's reputation. This would have an effect especially in trusting AI agents such as Alexa, Google assistant, Cortana or Siri because the attitude and the beliefs might depend on the interactive experiences, feedback, accuracy of the reply but also on the belief and previous knowledge upon the company reputation. In fact, reputation is often seen in e-commerce as trust-enforcing and avoiding frauds (eBay, 2002; Amazon 2020). This might have an impact on the perceived trustworthiness of the agents.

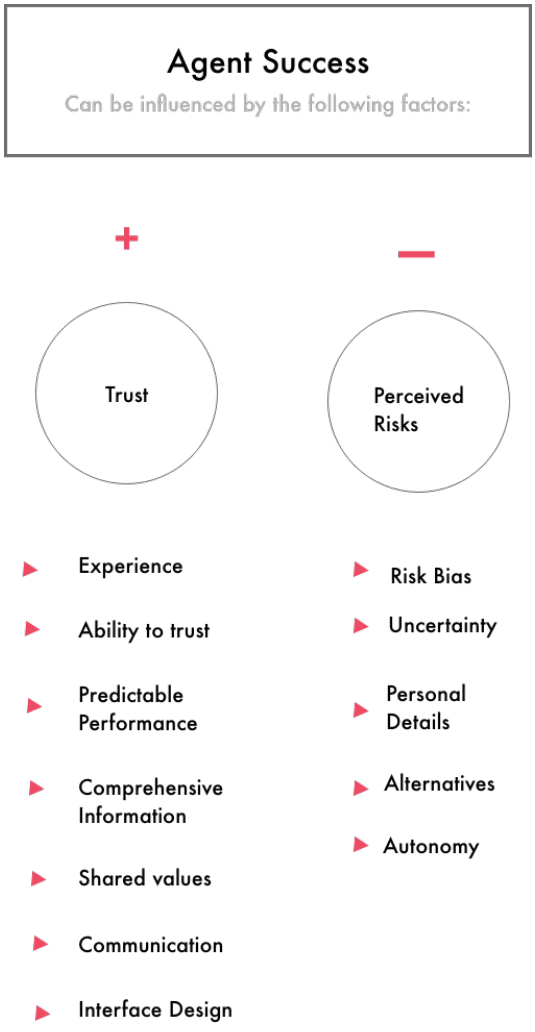


Fig. 8 A model of agent's success (Adaptation from Lee, Kim, Moon, 2000)

"Anthropomorphic" trust

A few studies have been carried out with the focus on understanding trust in AI agents. There are two main contrasting perspectives in the litterature that analyse trust in the interaction between Human and AI Agents:

Human-to-human trust and also Human-machine trust approach (Madhavan and Wiegmann 2007). The first considers computer as social actors (CASA) and tends to anthropomorphize them, while the latter believes anthropomorphism is misleading.

• HUMAN - HUMAN TRUST MODELS

People have the tendency to see computers and machines as social actors, theories of human-human interactions and trust are being applied to Human-computer interactions. This is the most widely used paradigm in agents. The theoretical foundation of this approach is Mayer (1995) theories about interpersonal trust. In this logic, the aforementioned model of trust can be applied to a specific technology, the three trust believes transform into the following AI agents believes:

Human-like technology trust			System-like technology trust		
Attribute	Basic meaning (Mayer et al., 1995)	Demonstrating the attribute: What a human trustee can do for the trustor	Attribute	Basic meaning (McKnight et al., 2011)	Demonstrating the attribute: What a technology trustee can do for the trustor
Ability/competence	Having a group of skills, competencies, and characteristics that enable a party to have influence within some specific domain	Perform a task well, provide good advice, give a good speech, make a good decision	Functionality	Having the functions or features needed to accomplish one's task(s)	Perform a function for the user, provide system features the user needs to do a task, provide the user appropriate functionality
Integrity	Adhering to a set of principles that the trustor finds acceptable	Keep commitments, tell the truth, exemplify moral principles, show sincerity, negotiate honestly, be reliable, consistent	Reliability	Continually operating properly or in a flawless manner	Perform functions reliably, do what the function says it will do, give accurate and unbiased facts and information, calculate correctly, do not crash
Benevolence	Wanting to do good to the trustor, aside from an egocentric profit motive	Do things in the trustor's best interest, be helpful and supportive, be responsive to the trustor's needs, show concern and caring	Helpfulness	Providing adequate and responsive aid	Provide help, understand and cater to needs, do not cause harm, be responsive to user needs and requests

Fig. 9 Technology, Humanness and Trust: Rethinking Trust in Technology (from Tripp, J.F, 2015)

From competence to functionality
"The belief that technology has the capacity or capability to complete a required task."

From benevolence to Helpfulness
"The belief that technology will provide adequate help and guidance for a human to be successful excluding the moral agency and volition (i.e. will) that humans have."

From Integrity to Reliability
"The belief that technology will work consistently and predictably".

People would look for trustworthiness cues in the agent basing their judgement, first on their expectations/beliefs and knowledge, then with further experiences they form an idea of what the agent is capable of doing in risk/uncertainty situations, and later they expect that the agents need to fulfil consistent outcomes (McKnight, 2011).

This model is particularly effective for human-like technologies, it has been proven working also with less working in all context (Lippert and Swiercz, 2005; McKnight et al., 201).
The research on believable agents has been mostly focusing on anthropomorphism (Bates, 1994; Isbister & Nass, 2000) and development of software personality in the past (Hershey, Mishra, & Altermatt, 2005).

- **HUMAN-MACHINE TRUST**

Trust on system automation is generally described as the" confidence in the system to do the appropriate action (Biros, Daly, Gunsch, 2004) with personal integrity and reliability" (Heerink et al., 2009).

This perspective states that, especially in automation, the effect of antropomisation is opposite to human-like trust (Dzindolet et al. 2003). Machine and automatization are seen by people as perfect in execution of certain tasks and operations compared to human beings and even decision making. Automatization bias happens when a user perceived as authority (or expert) an automatic machine, which leads to trust it and reduce the tendency to seek of information and verification.
This phenomenon is called "automation bias" and happens because many decisions are currently made by computers rather than human beings. People therefore tend to follow and believe suggestions from automated systems while they ignore every contradicting information (Cummings, 2004).
AI agents, by definition, are autonomous entities that can act to achieve a goal. The phenomenon of automation bias has been found also with AI agents (Zaroukian, Bakdash, 2017), and especially the trust/ acceptance is higher as the agent's interpretation was overall correct. According to this approach, human-machine trust is influenced by higher skills and expertise owned by machine/agents, so making more human-like characteristics and anthropomorphism would make appear the agents as more imperfect and would decrease their overall level of trustworthiness.

- **E. SEEGER ET AL. ANALYSIS (2017)**

Do agents need to be anthropomorphic in order to be trustworthy?

There is not a simple reply (on the next chapter the role of Anthropomorphism will be better defined). In the past decades, this type of strategy has been applied with different levels of anthropomorphic cues and it was found that this approach will impact positively the user's interaction, especially in first time interactions, but it is not always the case.

According to Seeger et Al. analysis (2017) to adopt one approach or the other, it depends on the context and goal: Anthropomorphism approach would be efficient if agents' trustworthiness came from their "goodwill", while if the intent is "quality/efficiency", designing for human-machine trust is more effective.

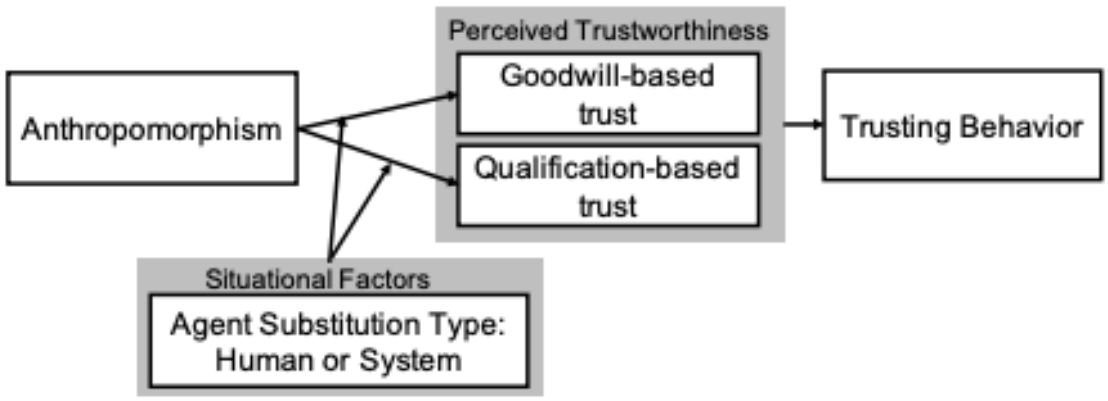


Fig. 10 Proposed research model Seeger et Al. (2017)

More in general the Literature in Human-Computer-Interactions found that there are some positive aspects, in the following chapter it will be discussed the human-like approach to AI agent to build trust.

Trust Key points

Trust is fundamental: it is the primary factor of technology adoption. If users do not trust the AI agent they will not buy or use it, in fact without trust there will not be any interactions, In chapter 3 it will be discussed specifically the current strategies to build trust with AI agent.
The following points are fundamental to understand to trust and building trust:

- 3 elements of trust

Trust requires three elements: Trustor, trustee(object of trust) and an decision making made one party that creates an interdependence between them.

- Defining Trust

Defining trust is not easy, usually it depends on the approach, the context or the type of relations that is involved. By combining different definitions, trust seems to have something to do with:

- Perceived risks that are the subjective perception of the possibility of the loose, this perception is depending both from elements of the object of trust
- Vulnerability due to the unpredictability of the outcomes.
- To sum up trust can be considered as the result between the perceived benefits and perceived risks, applied to a technological service provider also the subjective perception of company reputation. In order to create trust focus on reducing perceived risks and highlight the perceived benefits/company reputation.
- Trust is not a stand-alone concept, the opposite is Distrust, the phenomenon is a continuum.

- Two types of trust

The are two main types of trust: cognitive trust in which the decision to trust is influenced by knowledge, available information and prior experiences, and usually is more calculative in order to reduce the uncertainties of the outcome, while the other is called "affective trust" which is the type of trust that is driven by the emotional bond/connection between the two parties, which happens usually at later stage of the trust relationship.

- Specific and situational

Trust is specific and depends on the context of the trust relationship. One party one can trust something/somebody in specific context but not in another (Rousseau, 1998). As an example I trust Lucia for being for being a good

architect but i don't trust her as a carpenter). So anything can be considered as trustworthy or untrustworthy depending on the situations.

- Multidimensional

There are not specifical ingredients that describes trust, many variable interplay when the interaction take place. Trust by its nature is complex and multidimensional (Butler, 1991). Depending on the broadness of the social context there more factors are taken into consideration. The actors involved in the interaction (people, artefacts, institution), context and situation and the outcomes of the interaction of the actors involved. This means that each situation, context and parties involved in the interaction needs to be addressed and assessed not a priori. In order to design or test the agent you need to understand the context in which can be used or the specific scenarios

- Trust as a process

Trust as a process it can be established, build and re-built. As can be seen, in case of distrust, trust can be re-built, it is a reversible process.

- Personal

Trusting behavior and relationship can be both dispositional (Mayer,1985) and subjective in trustor.

- Trust as Relationship

  Trust relationship change along temporal dimension:

  - At the beginning of a relationship/ or first interactions between two parties initial trust is more opportunistic (calculative trust), the persons to make the evaluation focus on possible gains/loose. This initial trust is also influenced by the so-called institutional trust, all the contracts/regulation that shapes the interactions and the company/institution reputation

  - Trust is later influence by the information (knowledge-based trust) the person/user know about the object of trust

  - When the interactions continue and the outcomes are positive, the type of trust comes relational and affective, it is moved by the knowledge about the intention of the object of trust and the personal bond/connection with it.

- The "maximum" level of trust we can find in two parties happens when the trustors identify with the object of trust

- Trust and familiarity

Familiarity is a strategy to decrease and reduce uncertainty in relationship of all nature. It is the understanding and the belief of the predictability of other people behaviour often based on previous interactions and experiences. While Familiarity is about current context and situation, trust is about about the future, the belief of the predictability of the intention of the person.
Many studies prove that trust would increase progressively with the duration and the familiarity of a relationship(Chua, 2008; Gulati, 2008) while others prove there is no correlation between them (Poppo,2008) or in certain cases even negative. Maybe this depends to the kind or relationship the parties have
The more person is exposed to the object of trust the more likely is going to trust (mere exposure effect) also the more positive outcomes will be, the more a person/user is willing to trust the object of trust (positive reinforcement)
To be more clear, as Michael Chanover, Vice President of Design & User Experience at NerdWallet, explains the process to build trust with the example of Sushi.
Sushi spread in the United States in the late 1960s, even if it was a period of change in taste it was not first accepted by Americans: the idea of eating raw fish seems dangerous and not natural to many.

Fig. 11 Sushi (from https://www.japancentre.com/it)

The chef Ichiro Mashita, owner of a restaurant in Los Angeles had the idea to make sushi in a different way. Making it perceived as more trustworthy and similar to American's standards using ingredients that were more familiar to them. In order to make appreciate the exotic ingredients, he started by reverting sushi rolls: having rice on the outside, and the seaweed inside, using also familiar ingredients such as cucumber, avocado and crab. This recipies was called "California Maki" and the rest is history: the emphasis on familiar elements was able to make americans discover more the Japanese culture and cuisine and with the time even eating raw fish as Nigiri.

Fig. 12 Sushi (from https://www.japancentre.com/it)

This example is particular interesting because it explains that at the beginning trust is modulated by certain expectations/ knowledge in this case the *"worrying about the danger of raw fish"*, the fact that the solution was this combination of familiar elements and novelty. Trust was established with the creation of a familiar and positive association with common ingredients. With time, multiple experiences and positive outcome with Maki (the object of trust), people were able to increase their level of trust to Japanese cuisine and increase their willingness to try different things (Nigiri and other raw fish choices).

- Antecedents of trust:

There are some factors that are antecedents of trust that will have an impact on the interaction that might not always been under our control:

- Personal factors (disposition to trust, prior knowledge/experience of the artifacts, personality of the trustor/need of control). Those factors will have an impact especially at the beginning of a trust relationship.

- Agents (Aesthetics and perceived anthropomorphism) These elements can have an impact on pre-trust while interacting the perceived usefulness, usability, integrity, consistency of behavior, responsive feedbacks, anthropomorphic embodiment (and social cues)

- Company's factors (reputation and transparency)

- Trust in AI Agents

There are two main approaches to build trust in Human-to-agent interaction

- Human-to-human trust
- Human-to-machine trust

Both perspectives can be successful depending on the context and the aim of the assistant: Goodwill vs Competence.

Chapter 2

In this investigation the object of trust is voice activated assistants, in order to define and classify the characteristics and the features that make them perceived as trustworthy, a general panoramic of Artificial Intelligence and of Voice user interface need to be done, in order to understand more broadly the potentiality and the boundaries of this technology.

Artificial intelligence

Artificial Intelligence is *"The branch of computer science that is concerned with the automation of intelligent behavior."* (Luger and Stubblefield, 1993). AI is often described as machine's intelligence in opposition to human's intelligence.

The origin of the term comes from the computer scientist John McCarthy in 1956 who defined it as *"the science and engineering of making intelligent machines"*, It is the ability of a computer program/ machine to learn and think.

According to him: "Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions, and concepts, solve kinds of problems now reserved for humans, and improve themselves".

Since then the term had many evolutions due to the rapid speed of developments and applications. Artificial Intelligence is a term that is not simple to defined and among scholars lacks universally consensus upon the term.

Human intelligence

One reason why Artificial Intelligence is hard to define is due to the fact that there is no precise definition and meaning of the word "intelligence".
R. J. Sternberg, one of the current major researchers of (human) Intelligence and cognitive development affirms:
"Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it."

Shane Legg and colleagues (2006) made an investigation collecting major definitions and research about three themes: collective intelligence, psychological approaches and AI approaches to intelligence. They merged in the following definition:

"Intelligence measures an agent's ability to achieve goals in a wide range of environments."

Common attributes considered fundamental in human intelligence is the ability to learn/adapt and understand, in this latter definition are considered a prerequisite of the "agent" to being able to interact in the environment.

Machine intelligence assessment

The first investigation on machine intelligence was conducted by Alan Turing in 1950 with the publication of the article "Computer machinery and Intelligence". For the first time he questions whether machines could think and have intelligent behavior. He also proposes a methodology to assess it: the Turing Test or the imitation game.
The method consists in a Human evaluator(C) that separated from either a machine(A) or another human(B) has to judge natural language conversation.
The criteria defined how a computer can emulate a human being in a Realtime-written text conversation with a real human being. When a human cannot judge whether he is speaking to a machine or another human, then the machine is intelligent.

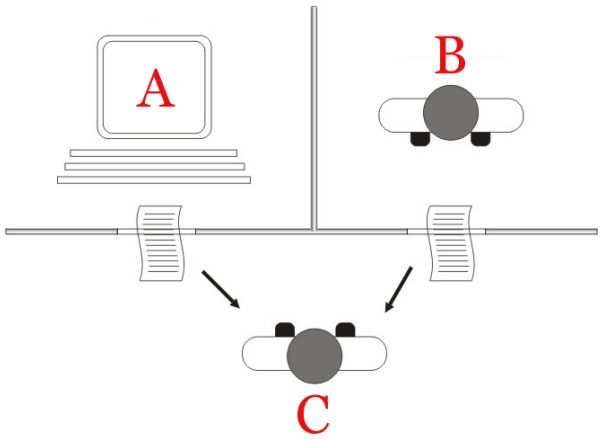


Fig. 12 Turing test (https://en.wikipedia.org/wiki/Turing_test)

- CHESS GAME

Turing invented an unconventional method to test machine intelligence: chess game, which is notoriously a game of intelligence.
Chess game consists in a set of rules and (pre)defined challenges which make this a perfect methodology to test problem solving skills. Turing predicted that one day a machine would be very advanced and "intelligent" to be able to defeat a very good human player. Deep Blue in 1997 was able to win against the current world champion: Garry Kasparov.
However, unlikely what Turing was hoping to achieve, machine/computer still cannot think.

Classification of AI

AI systems can be classified according to the literature based on the type of system's intelligence. Chris Noessel, currently working in AI for IBM, describes three possible ways of AI systems:

- **WEAK AI OR ARTIFICIAL NARROW INTELLIGENCE**

  Weak AI, or called Narrow AI, is an approach of AI that considers AI only as a simulation of human's cognitive skills. It is a limited form of Artificial Intelligence that is programed to perform defined and pre-determined tasks. All kinds of current existing machines/programs are classifiable as Weak AI, they are able to simulate the ability of "thinking" and human-like intelligence but what are actually doing is acting according to pre-programed rules.

- **STRONG AI OR ARTIFICIAL GENERAL INTELLIGENCE**

  According to the Philosophy of Artificial Intelligence the machines that exhibit human intelligence (reasoning and problem solving) are considered Strong AI. According to these theories it is possible for machine to become self-aware without having the same cognitive functions and process thinking as Human. This is still a very theoretical approach, but it aims to build machine with consciousness, emotions and self-awareness.

- **ARTIFICIAL SUPER INTELLIGENCE**

  This is the last stage of AI, this is also a very theoretical approach, and hypothesizes the development of an intelligent agent or a computer that will be able to surpass even the smartest human being in terms of creativity, problem solving, knowledge. According to AI researcher machine will be able to compete with human intelligence but there is little consensus on this matter.

Nowadays the state of AI is very far from the concept of super intelligence, "Voice assistant", the topic of interest of this analysis is Weak AI, are good examples of narrow AI. Those voice activated assistants are able to interact with human beings and process human language. They are not able to "fully understand" but they process and act only accordingly in the way they are programmed. Apparently, those assistants seem intelligent, they can understand, joke and suggest, but they act in a very narrow way. It is evident in the inaccuracy of the reply when asked something that they are not programmed to respond to.
Due to the lack of self-awareness and consciousness they are not able of independent decision making, abstract/ philosophical thinking. Humans also have emotionally driven responses to different situations. The current AI is not able to match human skills.

What is an agent?

A more recent and modern definition of artificial intelligence (or AI) is "the study and design of intelligent agents" where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success (Poole et al 1988).
An "*Agent*" by definition is *"anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators"* (Russell, Norvig, 2010).

"Agents do things, they act that is why they are called agents" (N. Shardlow, 1990)

An agent has an active role that impacts its environment rather than being passively affected by the environment.

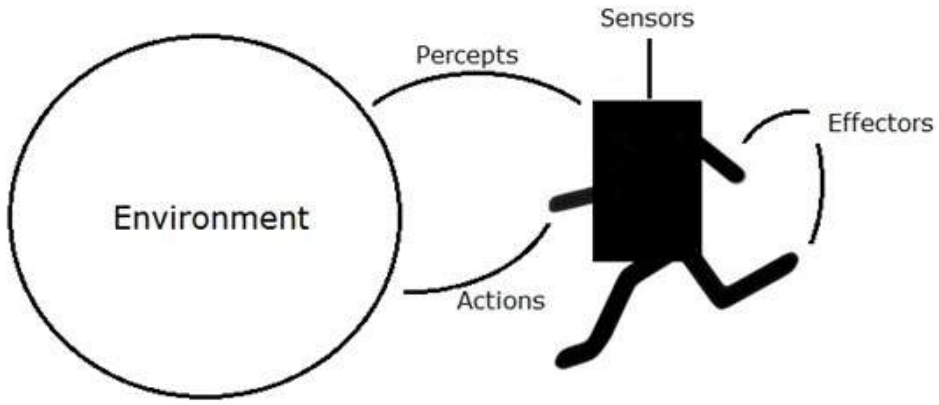


Fig. 13 AI Agent (From: https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_agents_and_environments.htm)

An Agent functions in three circular steps:

- Perceive (through a sensors system)
  Agent is able to detect environmental inputs through its sensors. These sensors are specific and depends on the agent.

- Think (those observations are used to make decisions)
  After perceiving the environment, a thinking process is followed (in the picture depicted with a question mark). This process represents the intelligence specification and decision making which is rational and has to lead the agent to make an action.

- Action (through actuators).
  The actions are the agent's output or the effect it has on the external environment.

Moreover, an agent can be designed in different ways, dividing into categories according to the perceived intelligence and capability and the different types of environments.

A rational agent is an agent that after perceiving selects an appropriate action that maximizes its performance, basing the decision making on perception sequences and the intelligence composing the agent.

An agent is autonomous if it is programmed to learn from its own perceptions and experiences and not only replicates programmed knowledge.

Conversational AI

Conversational AI is a fast-growing branch of AI research that focuses on designing and producing "natural" conversations between humans and computers.
The main scope of conversational AI is the creation of "Software Agents" that are able to "converse" with human beings using natural language.

AI agents are computer programs that are able to conduct a conversation (either textual or vocal) with the intent to simulate human-to-human communication and behaviors.

AI agents are designed to help the user solving/overcoming a problem. These Dialogue systems can be designed with different models of communication both in input/output through text, speech, graphic elements, haptic systems, gestures. Consequently, Conversational Agents come in different types and shapes depending on their purpose and application but overall, they all interact with human in a conversational style.

In fact, According to McTear et al. (2016a) we can distinguish different types of Conversational user interfaces (CUI): Natural User Interface (NUI) whose examples can be bots/chatbots, Voice User Interface (VUI) such as Voice assistants, as well as others that do not follow these structures such as ECAs (Embodied Conversational Agents) and Social Robots.
Due to the vastity of conversational interfaces the current work will be focused specifically in Voice User Interfaces and in particular to Voice activated assistants.

Voice User Interface (VUI)

"Human language is the new user interface layer"

Satya Nadella, CEO, Microsoft (2016)

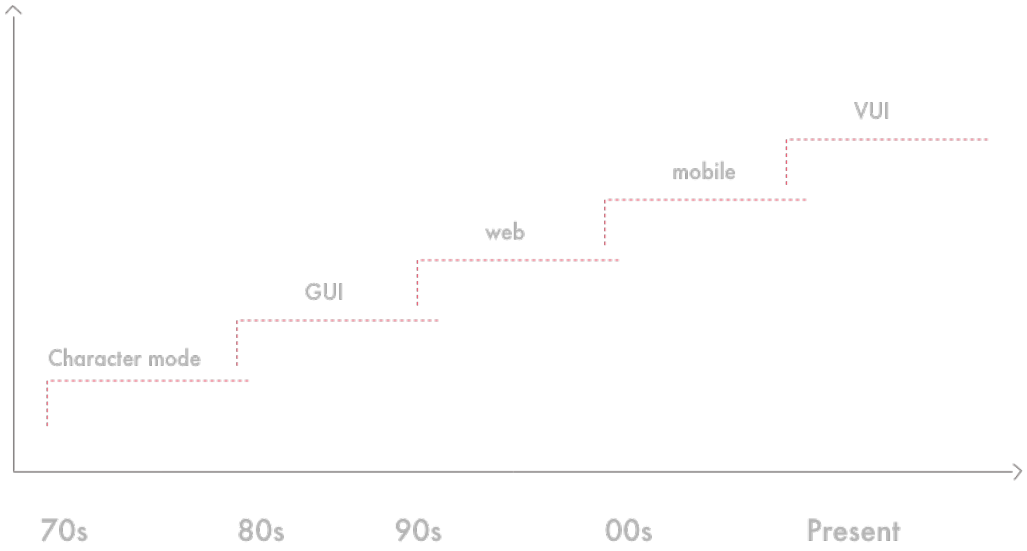


Fig. 14 Tech Milestones (Adaptation From: https://www.utterweb.com/blog/vuis-in-web-design/)

Over the past 50 years, the interactions models have been simplified, from the first computers many changes have been made making the interaction almost seamless. During Google I/O conference in 2018, The speaker announced that every 10 years more or less there is a technology shift: during the 1970s it was the age of connectivity, in 1980s the desktop revolution, and in the 1990s internet changed inexorably the way we communicate. In 200s it was the age of mobile first, while the latest trend of this decade goes to the most natural and intuitive form of interaction in executive tasks: the voice.

Voice is intrinsically a natural interaction and is the most intuitive for human beings. Human beings, as social creatures, use language as a primary way of communication. Psychologist and communication theorist Paul Watzlawick coined the so-called five axioms of human communication, *"One cannot not communicate"* being one of these pillars (Watzlawick, P., Beavin, J.H., Jackson, D.D., 1967). The Palo Alto group, where Paul Watzlawick was one of the major exponents, affirmed that people are always communicating, even the smallest perceivable behavior or its absence has the potentiality to be interpreted as meaningful from others. Communication is not an internal (mental) process that

comes from a subject but is rather an exchange of information. Thus, it starts within an interaction with at least two actors.

Modern technology is making communication much easier than before: Artificial Intelligence and Machine Learning and related technologies changed inexorably and even more the way we communicate. According to Microsoft voice report (2019), people are interacting with the search engines in a different way: the age of "touch" is being subsisted by the age of "voice" and digital assistance.

VUI is an invisible interface that mainly requires voice to interact with it. A Voice User Interface is something that *"… allows people to use voice input to control computers and devices"* (Amazon Developer website) and after being processed, the output is voice-based and it can also be accompanied by text, graphics video on a screen. All VUIs are visual, auditory and tactile interfaces that allow voice interaction between users and agents and IoT devices. We are in what has been called a "voice revolution", cited by many articles as one of the main trends of 2020. It may seem like the next future interface is voice, but it will probably never completely substitute touch interaction. However, as will be seen later in the chapter, it allows a multimodal interaction and enriches user experience and flexibility. VUI allows an interaction with a software through voice, and voice activated assistants have a VUI but VUIs are not always an interface for an assistant.

Therefore, it is important to distinguish voice search to voice assistance, even if both of them require voice as input. The first is related to the engagement with a search engine while the other with an intelligent technology that "helps" the users to achieve anything in which they need assistance.

Voice search: engage with search engine using voice

Voice activated assistant: engage with an AI agent using voice as primary input

Voice Activated Assistant

AI Agents can be considered as the next development form of chatbot, as some include more advanced technologies such as Machine Learning, Natural Language Processing (NLP) and Natural Language Generation (NLG) to deliver more natural conversations.

According to the analysis conducted by Tatai et al (2003) conversational agents can have three main roles:

- Digital Assistant
- Information provider
- General chatbot

When these software agents have the purpose to interact with a human and help him/her to complete an action, or to solve problems by interacting with them using natural language they are called assistants.

These software agents have the purpose to interact with a human and help him/her to complete an action, or to solve problems by interacting with them using natural language.

However, there is no unanimous definition of these computer softwaress, in the literature they are named in different ways. Citing Oliver Budzinski (2018) taxonomy investigation about AI agents, which found that the most frequent were: *"automated personal assistant, intelligent virtual assistant, virtual personal assistant, intelligent personal assistant, digital butler, digital helper, digital assistant, personal digital assistant, speech-based natural user interface, voice-activated intelligent personal assistant, virtual agent-based daily assistant."*

A voice-activated assistant records and processes user's voice request on a specialized server, interacts with users through a dialogue system in natural language and applies third party services to provide information or a specific solution to a request. These agents are intelligent systems because of their capability to find information and solutions to a variety of problems/requests. In this investigation these computer programs will be called: (AI) agent or Voice activated assistant.

Empowering Technology

The current spread of Virtual assistant can be linked directly to the improvements and spread of AI technology. As AI can be defined in lots of ways, to simplify we can say "the ability of a computer to "learn" and "think"" which is applied as well to voice assistants.
The following technologies that empower voice assistants are:

Machine learning: is the ability to *"is the ability of a computer programs or software to learn and improve from the different inputs it receives without being directly programmed."*
The program has continuously experiences which are formed through a programmed algorithm. The program is able to make a model and predictions out of the identified data. This learning process make the computer program

be able to give more accurate responses. Researchers Aurélien Géron and François Chollet identified 5 categories of machine learning, depending on the level of human supervision during the trainings.

Autonomy: is the ability of a machine to *"process and apply a procedure with minimal human assistance"*

Problem solving: "is the ability of a computer (in this case of an agent) to provide a solution to a problem."

- **SPEECH RECOGNITION PROCESSES**

Natural Language processing (NLP), it is a multidisciplinary field, that concerns the interactions between human natural language (both speech and text) computers. The objective of the discipline is to develop algorithms that are able to understand, analyze and reproduce the natural language both written or spoken. The NLP powered devices/software are able to understand and use language, in relation to their meaning and appropriateness of usage up to the context, syntax, the grammar and structural rules. This discipline is deeply connected to both Linguistics and Computer science.
The Spoken dialogue recognition system consist in a front-end signal processing and a back-end speech processing.

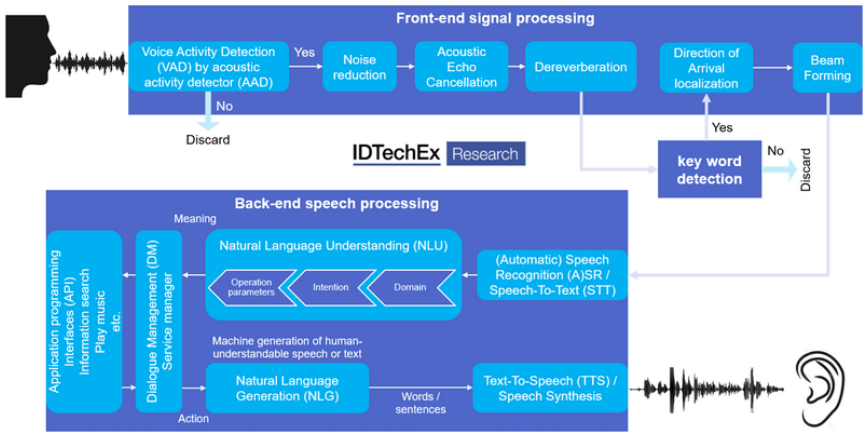


Fig. 15 Spoken dialogue system processes (From: https://www.idtechex.com/en/research-report/voice-speech-conversation-based-user-interfaces-2019-2029-technologies-players-markets/637)

The Acustic Activity Detector (AAD) is switched on in order to do detect human speech, the system at the same time works in reduction of the noise and echo and a dereververbation process. When the system detected the key word, analyze the direction and the localization of the voice to create a beam and triggers the backstage system that speech-based data is coming. This a trigger often called "wake-up word" such as the following: "Alexa", "Okay, Google", "Hey Siri".

In the back-end processes there are 4 elements that process involved in processing natural language for AI agent-human interaction.

- **Speech-to-text and text-to speech:**

The first process converts audio into text while the second oppositely converts text into voice output.

- **Syntax and semantic processing:**

The two process are complementary: semantic processing processes the meaning of the words(semantic) while syntax processing processes the order of words in a sentence (syntax) in order to process the meaning.

While "**Question Answering**" is the process that automatically responds to a request formulated in natural language. The voice is analyzed and recognized with neural networks and artificial intelligence, with the requirement of cloud computing. The voice captured is analyzed and transformed into text, following if the algorithm found a match between the words and the systems rules, the algorithm will execute the requested task.
Often in the case of AI agents, if the tasks involved with the connection with other devices, the system recognized the skill invocation sentence and send the signal directly to the cloud and as the signal is sent back to the device the task is executed.

• IMPROVEMENTS IN VOICE ASSISTANT

The optimism upon assistant technology seems to be confirmed by the technological improvements and their skills to better meet users' needs. Once per year, Stone temple conducts a study to test accuracy of the major market vocal assistants by asking hundreds of questions. The graph below shows the performance of the selected AI Assistants. Google assistant has been found with the highest accuracy, maybe due to the spread on androids and the quantity of data.

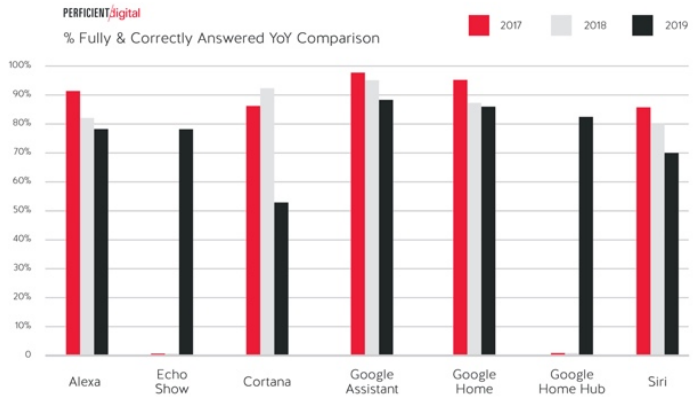


Fig. 16 Digital personal assistants (From: https://www.perficientdigital.com/insights/our-research/digital-personal-assistants-study)

Despite the improvements, Voice assistants are designed in a way to be able to have a meaningful engagement with users, however most of them remains not very conversational. Most of them are one-offs: the user makes a request and the assistant execute a task rather than having a continuous conversation. One way to decrease this and provide better experience with is through contextual awareness.

- **CONTEXT AWARENESS SYSTEMS**

"Context-awareness means that a service, although given the same request parameters, is perceived differently with respect to a given context. "

Drey (2000)

Technology is empowering agents to store personal preferences and memory of the interaction. By learning from experience and with contextual understanding they are able to provide advanced and realistic dialog systems. In fact, they are able to remember user's conversation, learn and grow. They are able to improve based on user's input both about what and how they are asking also depending on the context.

According to the Cambridge Dictionary "context" can be defined as *"the situation within which something exists or happens, and that can help explain it".*
Context is anything from: situation, position, place, environment etc.
There are several definitions of context and context-awareness in the literature of "context-aware computing" but one of the most quoted *is "the ability of the computer to sense and act upon information about its environment, such as location, time, temperature or user identity. This information can be used not only to tag information as it is collected in the field, but also to enable selective responses. "* (Ryan et al 1998)
Contextual information can of different types: location, Identity, Activity and time (Drey, 2000). A particularity of AI agent is the ability to provide information about a person/ place/object that refer to a context. Context aware system are able to access information about an environment, learn from applicable experiences and knowledge, and adapting its behavior in real time.

Context-aware systems are able to detect, anticipate and change their behavior/response according to an environment (computer/user/ physical environment) combining the previous experience/knowledge about the entity itself.The outcome responses of the computer system, due to this process, are able to be more personalized, adaptive and more accurate but also making automation possible. Designing computer software with context-awareness is more complex because the number of situation and context in which the AI agent will be used are many. This level of complexity in adaptation and awareness can change significantly. A small change in a context-awareness agent can be for example when a light is switched on because the environment is dark, or when the audio/visual output of an Agent are specific to a device and not in another. Amazon's Alexa demonstrated to have some context awareness: the whisper speech. When a user whispers a command to Alexa, it is able to respond in the same way. Making the assumption that there is a reason that there is a reason. Or in the USA, Alexa is able to respond to user using two contrasting emotional reaction.
The contextual awareness of agents can create meaningful interaction: but those context-aware systems need to conform to user's expectation.

History of Voice User Interface development

In the following paragraph the main milestones of user VUI.
Bell Lab, in the 1950s, built for the first time a system that was able to process single-digit speech recognition called Audrey. It was able to recognize numbers (0-9) through voice interaction but it could not have been commercialized due to the limited technology it.

Joseph Weizenbaum, in 1966, the first version of a chatbot script known as ELIZA. ELIZA is capable of performing a non-directive dialogue with its user. Eliza simulated a psychologist and was able to respond to pre-programmed questions. The first devices that were able recognize words were commercialized only in the late 70s. Harpy was able to recognize more than 1000 words in English, but it was still not a high success. At this time, those technologies had not yet entirely found a market potential and remained studied and researched on labs and university institutions.
The research proceeded, and in 1984 the first VUI were introduced by Speech Works and Nuance through Interactive Voice Response (IVR) systems that were able to recognize voice and perform a given task.
Dragon Systems releases the first ever consumer speech recognition product Dragon Dictate in 1990. This system is based on discrete dictation, meaning the user has to pause after each word. Consistent technological improvements in speech recognition happened in the late 1990s and 2000s.

In this context, in 1996, IBM invented the first device (Medspeak) that was able to understand continuous speech from a speaker. The systems became able to understand and analyse more complex syntactic sentences, processing longer strings of conversational inputs.
These technologies became mainstream with the spread of voice-only devices: we are currently in "the second era of VUIs".
In fact, over the last decade, bots became extremely popular due to the huge investments done by the Big Five tech companies (Amazon, Facebook, Google, Microsoft, and Apple) in the development of their virtual agents: Siri, Google Assistant, Alexa, Cortana. Siri by Apple was the first standardized virtual assistant on iPhone in 2011, followed by Google Now (today Google assistant) for Android systems. Siri was able to recognize speech, understand meaning, and take appropriate action.

In 2014 Microsoft launched Cortana, and in the same year, Amazon announced Echo, a voice-controlled speaker powered by Alexa, an AI assistant similar to Siri and Cortana. Only in 2016 Google launched its smart speaker Google Home powered by the Google Assistant. By 2018 Amazon Alexa has over 50,000 Skills worldwide and works with 20,000 devices.
Those companies made AI agents mainstream. At CES 2020, the annual conference about technology and consumer held in Las Vegas, everything was about voice assistants. We can say it is not just a trend, but this way of interaction is here to stay.

According to the Voicebot Report (2018), we can distinguish two phases of the development of modern voice assistant:

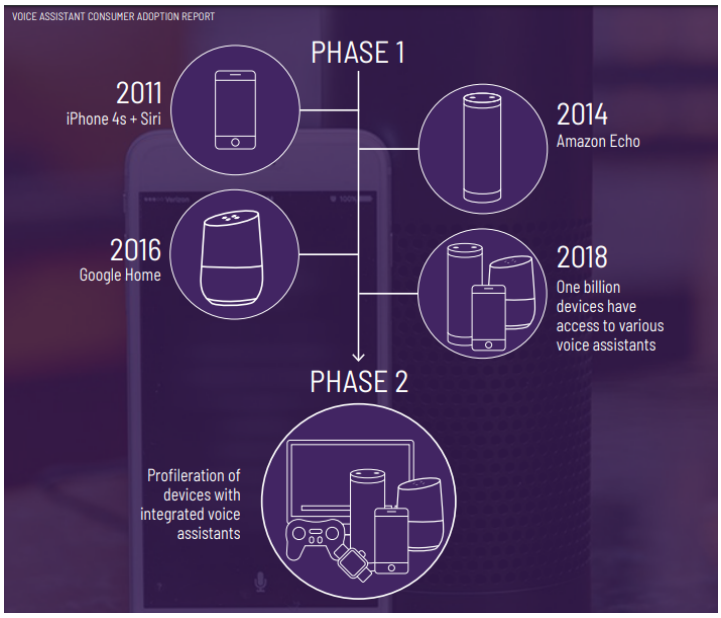


Fig. 17 history of VUI (From: Voicebot Report, 2018)

- **FIRST PHASE**

The first phase of modern VUI was centered in the development of specific voice-enabled devices (smart speakers) and smartphones. The primary focus was to make available to consumer the possibility to execute simple tasks through voice interaction.Companies at this stage made common simple features (as routine or providing basic information) making them available in many languages. Only later the agents were used as entertainment and communication.

- **SECOND PHASE**

This current phase started in 2017. The focus at this stage is more company-specific: each of them developed specific functionality and features for their own embodied agents. Some innovation included follow-up questions (Google assistant) or emotional tone-of-voice (Amazon). This phase is seeing the switch from device first to ecosystem design: major players are including their agents not only in their smart speaker and on smartphones but also on smart watches, smart glasses and other IoT devices, bringing the possibilities for interconnectivity and continuity of interaction between devices. The future of AI agents seems to go a more systemic direction rather than being embodied in only one device: AI agents are integrated in many different devices. In fact, we can say that, now, voice technology is a combination of IoT devices, AI services and UX interaction.

Market size – VUI today

Voice technology has changed and improved a lot in the past decades, and yet the field continues to evolve rapidly.
According to the Voicebot Report, more than 3 billion Voice assistants are currently in use, this increase in adoption correlates the growth of market size of voice. According to Zion market research (2019) on 2018 the global market of voice assistant was around USD 0.8 billion, and their projections are around 7.7 Billion by 2025, the market size including both smartphones and Smart speakers. By 2019, the voice recognition market will be a $601 million industry (Technavio).

According to Global Voice Search Market (2019) today the Major Key Players in this voice interaction transformations are:

- Amazon, Inc. (U.S.)
- Google, Inc. (U.S.)

- Apple, Inc. (U.S.)
- Microsoft Corporation (U.S.)
- Samsung Group (South Korea)
- Orange S.A. (France)
- Next IT Corporation (U.K.)
- Creative Virtual Ltd. (U.S.)
- Nuance Communications (U.S.)
- Nokia Networks (Finland)

Accenture's Digital Consumer Survey in 2017 investigated the usage of voice activated assistants considering a sample of 26,000 people from 26 countries. They found that in these states 46% were using them and the number is even higher for India and China with 55%. Also, Microsoft report (2019) estimated a rise of 22% of adoption of smart speakers compared to 2018.

It was found that 1 in 6 Americans own a smart speaker, 56% of them keep it in the living room, 25% in the bedroom and 22% in the kitchen (Voicify, 2019). While 72 % of the smart speaker owners use them as part of daily routine (Sara Kleinberg,2018).

Gartner (2019) predicts that up to 25% of Customer service operations will be managed by virtual assistants by 2020. While, Gartner report (2017) says, VUI and Virtual agents are a big trend and they will reach their full potential between 5 and 10 years, as it can be seen in the following graph, they are at the top left of the distribution below:

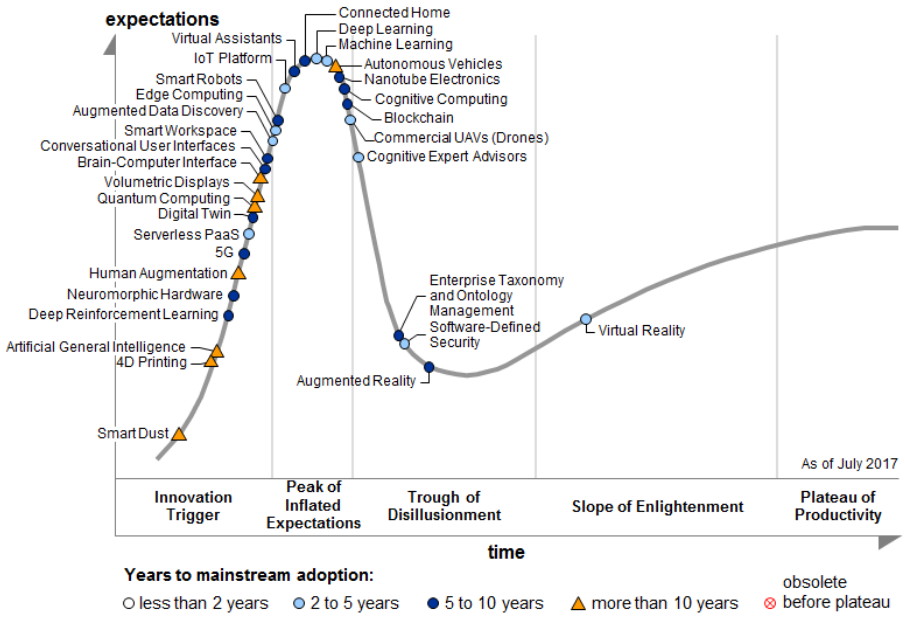


Fig. 18 Gartner report (July 2017)

In the same direction, Google CEO, Sundar Pichai says:

"In the next 10 years we will shift to a world that is AI-first that will be universally available" A personal Google just for you (October 4, 2014)

Only time will confirm whether or not these predictions are true, however it is clear that voice technology is becoming present in many contexts and in our day-to-day lives.

Characteristics of VUI

Many studies have been proven the potentiality of this type of interaction:

- **NATURAL INTERACTION**

Studies shows that speech is the most natural and comfortable way to communicate (Tadeusiewicz, 2010). It was also found that voice-based interaction can increase the ease of use and allow a more natural interaction especially in executive tasks, reducing the time and effort and improving user's satisfaction (Jiang, 2015). Voice is the most primordial communication means, since birth, we learn how to express our needs and wants or any other states through it. One-year old kids learns very fast with a rate of 10 new words per day and this will continuous up to adolescence.

- **HANDS-FREE**

An interesting advantage of using this type of technology is the possibility to multi-task: users are able to do things while interacting with an agent.
This type of interaction is particularly effective because it allows the user to be hands-free (NN group, By Raluca Budiu and Page Laubheimer on July 22, 2018), and also more flexibility and intuitiveness. It is especially useful when the user is involved in certain things when he can not be distracted by using a device, such as driving a car.

- **EYES-FREE**

Similar to hands-free, Voice user Interaction offer the possibility to interact with it without looking at the device. Being eye-free it opens up to many possible interaction in which the user can multitask or interact with the devices while the user is in a specific context (dark space/car) in which he/she is busy.
An interesting advantage of using this type of technology is the possibility to multi-task: users are able to do things while interacting with an agent.

- **EFFORTLESS**

Studies shows that speaking Is cognitive less demanding for people than typing/navigating on a screen. Also, compared to typing/navigating, user is able to process information with less active participation.

- **SPEED**

Voice interaction make people spend less time to do routine tasks: smart devices can be programmed to automate routine tasks. This atomization is able to reduce user's time and effort. Information search is significantly reduced because voice is easier and quicker than typing. People speak faster than they type. According to the report published last year by Google UK found that 75% of consumers are using more vocal research instead of typing, and 83% affirmed that voice research makes easier and faster to find the things they are looking for. For this reason, VUI are efficient when the user command something specific to be done but when it is more complex when it involves a deeper navigation in the system.

- **PERSONALIZATION**

There is a possibility for higher personalization for information seeking due the fact that they are based on personal user data: algorithm and machine learning are able to understand and predict user's preferences (Linden et al. 2003) and show personalized results. Although this personalization depends on the data available about the user.

- **ACCESSABILITY**

VUI interfaces have high potential level of accessibility for different disabilities, such as: Deaf, Cognitive impairments, Blindness, Physical disabilities. However, each disability needs to have different design solutions, focusing on the opportunity for multimodal interaction.

- **HUMAN-LIKE-INTERACTION**

Speech conveys more meaningful interaction then text providing implicitly: Age, gender, Emotions, Personality, cultural/regional background making our interaction as more human-like.

- **EMPATHETIC INTERACTION**

Voice is a powerful medium to convey intentions and emotion: voice is not only analyzed as a sequence of words, more complex computer software are able to detect and analyses how the person speak and respond empathetically to them (or that's what company as google or amazon are trying to achieve) and build relationship with them. In human-to human interaction, voice in fact influences physiological and affective response (Scherer, 1986; 2003), this has been implied to technology in order to create more adoption and more natural interaction with them.

“People can speak a lot faster than they can type they can listen much more slowly than they can read, and they can talk much more quickly than they can listen. The conclusion is that while designing a VUI may seem, at a gut level, to be easier than designing a GUI, the opposite is in fact the case: VUI design is a lot harder than GUI design”

Ma, Weiye (2017).

Designing VUI

Voice interfaces have lots of potentials, at the same time, depending on the person or the context of application is not always the best solution. There are factors that decrease the likelihood of usage of voice assistant:

- **ENVIRONMENT**

There are two main environment types to consider when designing for a VUI: Physical and Social. Physical refers to the physical characteristics of an environment, such as noise, when the environment is too nosy it would not be the most efficient modality to use. The social aspect that decrease the likelihood to use is when the user is with other people:
People do not use voice assistant in public space, Voicebot report (2018) declare that it counts only for the 12% of their usage, information seeking behaviour is more private. Creative Strategies research (2016) found that the most common place where people use their voice assistant is home, while only 51% in the car, 1,3% at work and 6% in public space. Milanesi thinks that the adoption and the usage also depend on where the agents are embodied *"As wearables become more pervasive the ability to hear and carry sound will also improve"* (2016).

- **SPEAKING TO A COMPUTER**

For general population is still bizarre talking vocally to a computer, and even more as mentioned before in public. But there are exceptions: especially for those that are likely to engage with AI assistant they prefer the voice modality over typing (Microsoft report, 2019). Generally, how agents are built up to now interactions are a combination of typing and voice, especially for those on smartphones.

- **TECHNOLOGY ADOPTION**

The term technology acceptance refers to user tendency to adopt new technology, this tendency is both dispositional and situational. In fact, many psychological theories tried to understand the phenomenon that leads the person to use a new information technology. Davis model (1985) argues in fact that when this technology is experience and perceived as easy to use (ease of use) and Useful this make the user willing to use and adopt this technology. In order to adopt Voice Activated assistant, the designers need to shape the experience as easy to use and useful in relation to user's needs and goals.

However, compared to other types of interaction, VUI includes some intrinsic challenges due to the nature of the voice that need to be identified and addressed when designing. There are technical limitations that a designer needs to keep in mind while designing AI Agents. Spoken language is not the same as written text in a screen-based device. Designer cannot apply the same design guidelines when designing for VUI.

Difference in designing GUI vs VUI

Graphical user interface (GUI) is a type of user interface that allows the interactions with devices/software through graphical elements. In GUI, usually, all the information displayed is organized in tree and branches structures of various complexity.

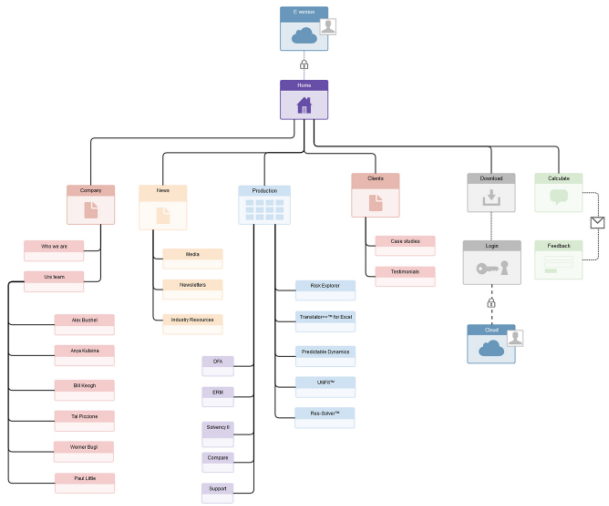


Fig. 19 GUI (From: https://www.utterweb.com/blog/vuis-in-web-design/)

The user is able to move from one screen to another by following visual cues, view states and processes. Metaphorically the user can explore the interface as a navigator looking at a map, he/she can follow visible indications, cues and states and process all the information that have a hierarchical structure, and the fruition of multitude of contents is possible.
In voice user interfaces, visual aids are often shorter and less visible, and the user mainly relies on the conversation (voice output) to express states and processes or to navigate the system. Since VUI are (mostly) invisible, they must be designed differently. VUIs are based on flat information hierarchy: all menu options are on a single level. Consequently, the response output will provide the necessary information all at the same moment, the user needs to listen the whole information, word after the other. While, Visual interface there are

multiple levels from which the user can move to find what he is looking for: user can skip ahead and go directly to what he needs.
For this reason, a user flow in a VUI follows a linear and step-by-step progression:

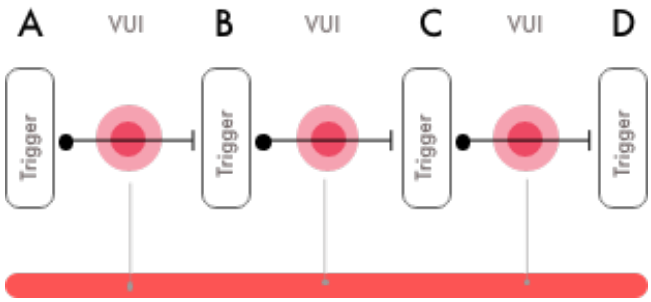


Fig. 20 VUI design (https://medium.com/uxtales/progettare-per-la-voce-ea8e65827b7d/)

The user is able to switch from one state to another only thanks to a continuous game of triggers manipulated by the VUI.

- **ABSENCE OF SCREENS**

The GUIs present its triggers visually in the form of buttons, forms, tabs or text and can be activated by interacting directly by mouse or touch.
On the other hand, in VUIs the voice triggers are the main manipulable component. The system will detect user's speech input and will respond with a certain type of output that will be presented most vocally to the user. Due to the decrease elements of system's output, the VUI must be designed as task-focused to maintain user attention and awareness of where they are in an application. Consequently, in VUI there is no direct application of design guidelines, because of the lack of visual affordances. The lack of visual affordances makes the user having any indication of the interface can do.

- **VISIBILITY OF THE SYSTEM**

"The visibility of system status refers to how well the state of the system is conveyed to its users. Ideally, systems should always keep users informed about what is going on, through appropriate feedback within reasonable time." — Jakob Nielsen

A good visibility of system status is able to inform and suggest the user about the next steps in order to achieve his/her goal.
Generally, the more the user is aware about what is going on in the system through appropriate feedbacks and what changes are happening, the more user will have a good user experience. Appropriate and in time feedbacks make the system appear more transparent and communicative about its changes.

In VUI the visibility of the system is more limited especially in those devices that are "voice-first" and do not present screens. Feedbacks play a fundamental role in VUI.

- **LACK OF INTERACTION HISTORY**

Due to the intangibility of language, in Voice-only devices there is no shown obviously any elements of the user's input request. The system needs to be able to communicate the fact that recognised user's input, especially in case in which the assistant is not able to respond correctly.

- **PRESENTING OPTIONS**

Presenting multiple choices is demanding in terms of cognitive load especially if they will not be presented visually, the user will forget with the increase number of the options. Also, due the fact that VUI is structured linearly in a limited temporal flow, the user does not have much time for processing and making decision (Mary Gold, 2018).

- **ATTENTION & COGNITIVE LOAD**

In a GUI, a user interacts in a consistent way, the relationship rate between the input and the system output are constant and keep in mind the user's cognitive load.The interaction with VUI starts with the "wake-up" words/sentence and the system replies, which triggers user's attention. In order to fully capture what the system is saying, the user needs to allocate constant attention to it, because instead of GUI there are no elements that are always present during the interaction.

The picture below shows the differen-ce between attention in GUI and VUI.

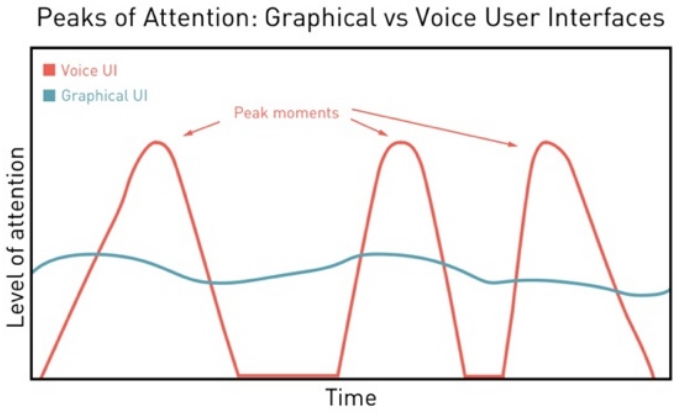


Fig. 21 Cognitive load (From: https://careerfoundry.com/en/blog/ux-design/voice-ui-design-and-cognitive-load/)

User's attention in a VUI is high compared to a GUI. This increase of user's cognitive load has a reach multiple picks during an interaction with a computer system. In order to make user understand all the information the agent's sentences need to be short and not too dense with information and reduce too high cognitive load. The immediativity of the voice requires the user to be fully alert when the system replies, especially in a voice-first device. VUI responses must be short and limited in succession: when there are presented multiple options, it should be grouped and presented 3-5 no more. If there is the necessity to present more the combination between voice and screen.

- **ERROR MANAGEMENT**

Smart Assistant companies, such as Microsoft, claims that their speech recognition software is now at 5.5%-word error rate compared to the average human that is 5.1%, while Google reported an accuracy of 95% (Xiaoxi He, 2019). Whether the errors done by the system or by the user, when creating a VUI, it is critical to manage errors.
VUI and GUI deals error prevention and error correction differently: People are making shortcuts when they speak and that affect system's understanding, but even when when sentences are complete might happen. The reason is that lots of information is implicit and can not be fully understood by the system. So, errors need to be addressed through prevention and correction.

- **Error Prevention**

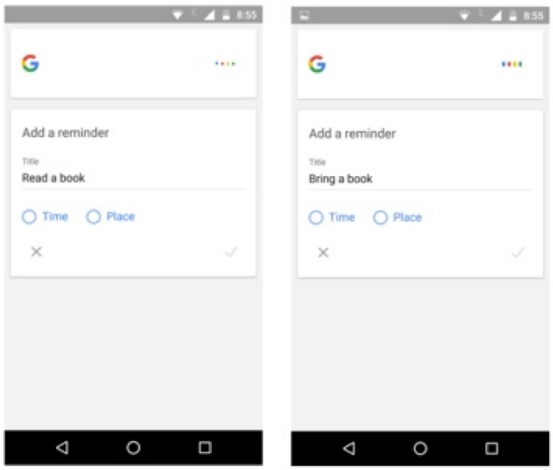
Error prevention is one of the 10 usability heuristics and defined as
the "degree to which a system protects users against making errors." (ISO 25000).
Designing a system that prevents user errors means create a system that make hard for user to make errors.
According to NNgroup(Page Laubheimer, 2015) there are 2 types of errors:

  - Slips: When a user wanted to make a specific action but does something else. Slips often happens when the user is not fully focused on the task.
  - Mistakes: When a user wanted to make a specific action, but this action is not suitable or incorrect to execute the task. Mistakes are usually conscious and often due to an incorrect user's mental model of the device/interface.

Rather than designing solution to help user to recover from errors, the computer system/ agent should be able to avoid them to happen.
In VUIs (voice first devices) error prevention is limited agents' audio output or when the agest ask directly a confirmation of a planned action.
However not all the agents have the same technological capabilities, consequently voice detection and understanding can differ in preventing

errors. In general in recent year, errors in naural language have dropped, in fact, a common mistake in mainstream agents is that the device might not always get the activation phrase (NN group, 2016) especially when executing a task (playing music /broadcasting news) and there are lots of noises in the environment.

E.g. Google assistant understood when said "Set a reminder to breed a book" (left) and then I said, "Set a reminder to breed a book" (right)



The agent is able to cope with a slightly incorrect input.

- **Error correction**

Error correction in Voice User Interface is the property of the system that is able to correct the recognition errors happened to incorrect recognition (Suhm, 2010). When designing VUI it is needed to focus a lot in error correction, because errors can be due to unappropriated user input, or background noise or complete misrecognition.

Continuous speech dialogues are available in mainstream available; however repair is still a problems in most of them.
The main errors in VUIs are no-input and no-match situations. In the first case, the VUI was not about not being able to recognize user's input. It usually happens when the user does not know what to ask. While in the latter case happens when the system is not able to understand the user.
A common problem in mainstream agents is that the device might hear the user wrongly and the user can not correct the agent until it finishes executing a task. There is exception: in the case of screen-based devices is possible to type or give another command. What to do when this happen? The designer needs to plan an appropriate response, asking to repeat their questions, make them speak more slowly or suggest changing slightly the request.

Speech recognition in the last years improved significantly and errors in natural language understanding dropped. However, a 100% accuracy today is not possible. However, A research conducted by Meyers and colleagues (2018) confirms that in VUIs, errors usually did not impair the interaction too much.

Multimodal Interaction Design

Before it has been defined Voice User Interface as all visual, auditory and tactile interfaces in which the input of the interaction is user's voice and allows for multimodal interaction. By definition multimodality is using more than one modality regardless of the nature of the modality. Other researchers see multimodality specifically to those human sensory modalities implied in human-to human communication (speech/gestures/ gaze/ handwriting)
Multimodal systems are able to seamlessly interpret and produce semantically driven information and at the same time being able to detect and process input (speech, text, touch) and output(text, graphics, audio) from several channels in sequence or presented simultaneously.

Speech, as a Natural User Interface, offers multiple interaction tools for input and output:

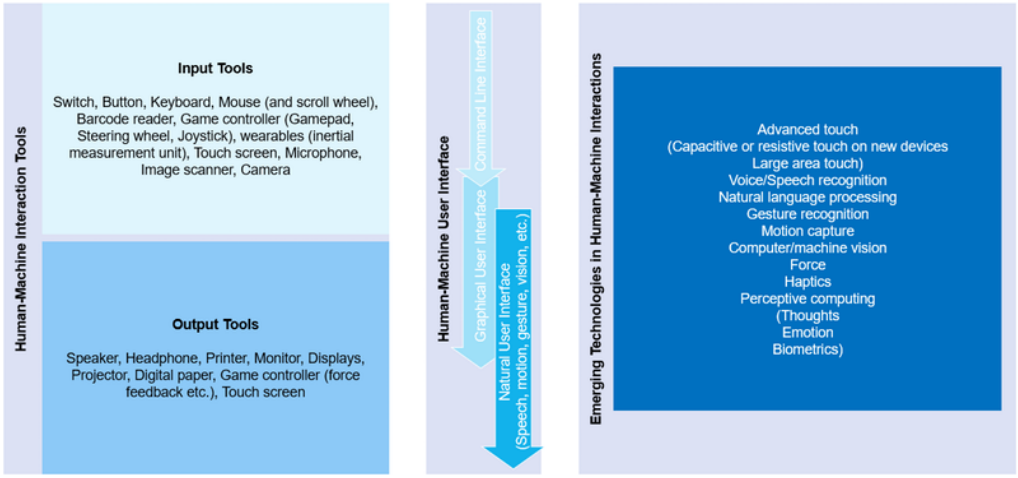


Fig. 22 Evolution of human-machine interaction (From: https://www.idtechex.com/en/research-report/voice-speech-conversation-based-user-interfaces-2019-2029-technologies-players-markets/637)

This multiple possibility of interaction makes the user experience multisensorial and memorable. Today, our technology is mapping for three main modalities such as visual, auditory and tactile interface that engage the user senses such as sight, hearing and touch:

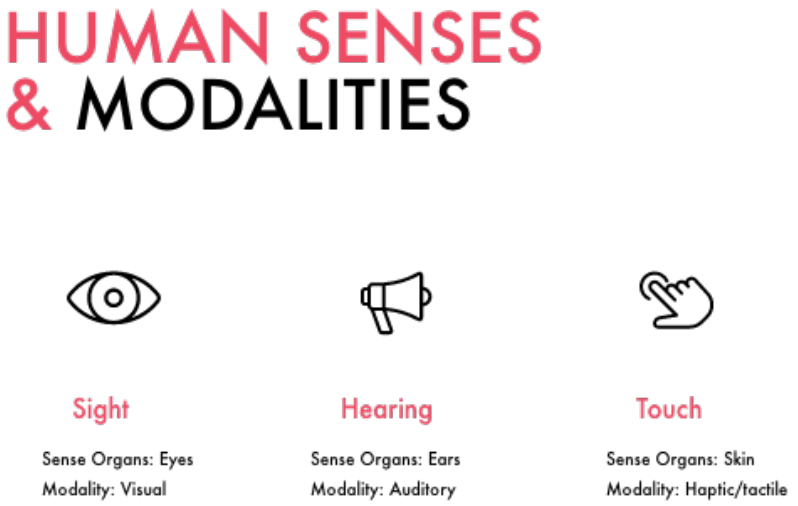


Fig. 23 Multimodal interfaces (Adaptation from: https://www.slideshare.net/martigold/new-frontier-of-multimodal-interfaces-are-you-ready

The combination of different modalities (as visual interaction, haptic interaction and audio interaction) as Input/Output in technology makes a compelling user experience for the user. In fact, Multimodal interfaces allow users seamlessly to integrate more senses during the interaction with a system in a similar way they would do in their real world. People perceive the world through five main senses and the combination of the together make how we perceive things. Multimodal interface mimic human-to-human communication and common human behaviour opening up to deeper interactions. Compared to single-mode modality interface in which the user can only use voice/visual, multimodality allows user to interact in different ways with the system, improving his/her experience. The way human consume contents through technology is mostly visually and auditory which will be the main focus of multimodal experiences in Voice activated assistant in order to "exploit" the best for each modality and not overwhelm/over stimulate the user.
Before it has been described in general the macro difference in designing for visual and voice interface, it is important at the same to evaluate the benefits and the constraints of each modality (and the combination of them) to evaluate the main benefits and constraints related to voice-enabled devices.

Multimodal design is about understanding what the right Input/Output for every part of the experience is and can change during the experience and the context of use. Multimodality exploits the best elements for each modality:

- Voice

It is an efficient input modality and can handle both simple/complex input and simple output. Voice-first interface strug-gles when output increase in complexity and in length, due to the intangibility of the interface. The strength

of using this modality as input is due to the fastness, immediacy and the hand free control.

- Visual

Display and screen work well with simple inputs and can handle complex output. As the complexity of the output increase the results to be inefficient. Visual is especially an efficient modality for output due to the possibility of presenting conspicuous information with less cognitive costs in terms of attention and memory. Visual make possible to present multitude information in a way that is easily understood and accessible letting user to directly manipulation of its visual elements in order to complete a task which make the experience easily discoverable and learnable.
However, the combination of visual and auditory modalities offers a maximization of the two modalities. In this context adding visual information to voice interaction can enrich UX, some information is more effective transmitted through visual interaction.

In general, multimodal interaction system gives the users a choice in terms of which modality to use and overall that give to greater stability during use (Oviatt, S., 2002), and improved usability (Bretan, I. and J. Karlgren., 1993). In the literature, it was found that multimodality is better than unimodality because it is

"better in exactly the cases where unimodal systems fail"

Oviatt, S., (2002).

The choice of which modality is used by the users can be due to many factors: context, personal preference of a modality, native language and impairments (Oviatt, S., 2003).

Designing Conversational VUI

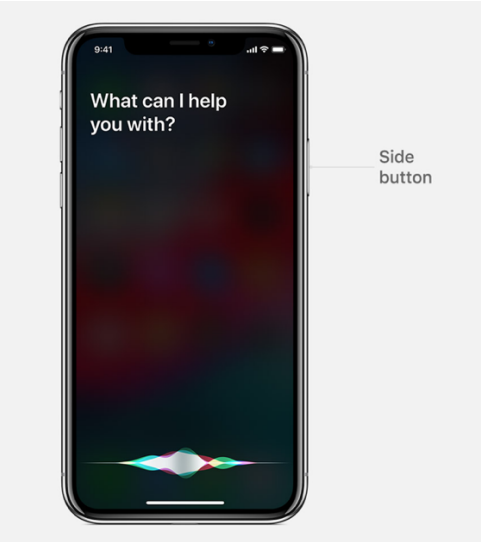


Fig. 24 iPhone (from: https://support.apple.com/it-it/HT204389)

AI agents are able to interact verbally with users in a natural way, mimicking human-like social interaction. However, the interaction is command-based, as the user cannot start interacting with an AI agent without saying the wake-up word. CASA paradigm shows that providing human characteristics to AI agents as: language output, response based on prior input, emotional replies, etc... can make users behave socially towards AI assistants (Nass, 1994).

Voice activated assistants are meant support and respond to user's request and needs, not surprisingly user's expectations are higher for conversational VUI. Their interaction is more communicative and as we'll see in the next chapter people what kind of expectation and bias associate with a voice activated assistant.

Cathy Pearl, in an article published in Ubiquitous Voice: Essays from the Field (2018) explains that in order to be conversational a VUI it must have:

- Turn dialogues between the user and the system. Those dialogue systems when are programmed to have continuous conversation with the user it provides more realistic human-like conversation.
- Memory of the previous interaction up to the moment of current interaction.
- Ability to recognize pronouns
- Allowing multiples ways to say/ask the same thing

However, most of mainstream voice assistants lack on one or more of these characteristics that would make a more natural interaction.

There are two ways to VUI today:

- "Command-and-control" in which means that users must explicitly say when they want to start interacting with the agent at every turn. Common assistant such as Echo (Amazon), Siri (Apple), Cortana (Microsoft) that all requires either a wake up-word or to press button/screen to communicate with the voice assistants. This happens every time the user needs to speak to it: user in order to give a command needs to explicitly say the wake-up word. It is a closed conversation with a starting point and an end.

- Natural turn-taking. The cructial element for human-agent interactivity is turn taking (Attwater & Balentine, 2009). The direction of systems is going towards more natural and conversational: turn taking approach. Turn taking is developed by natural human-like interaction: asking a question, pause, explicit direction. Human beings implicitly transmit turn-taking protocols also to human-agent interaction. The user can avoid using all the time the wake-up word in order to interact with the assistant and follow-up questions are possible, here there is a longer back and forth interaction between the parties involved. These companies are trying to develop a more natural interaction, enabling their devices to listen for few seconds more when the user stopped speaking or making their assistant to ask more detailed questions afterwards.

Voice Interaction flow

The whole Voice interaction flow can be distinct three main moments:

- User's input,
- Thinking process
- Output of the system.

In most cases in VUI there are no visual affordances and no clear cues suggesting what it can do or what kind of options it has.

"In interaction with voice user interfaces, the user has no visual guidance, and getting lost will happen all too easily. It is important to inform the user what functionality she is using and how to exit it." (Interaction Design Foundation website). In fact, the system needs to show the process of doing (POD) so continuous feedback are necessary for a Voice User Experience. Feedbacks also are able to make the user rephrase/correct the input or to affirm an action. It has been proven in diverse research upon trust in AI agents and ECAs that feedbacks play an important role in trust (Barber,1983; Mayer, Davis, &

Schoorman, 1995). Especially those feedbacks that are able to reduce uncertainty on the behavior of the agent (Dzindolet et al., 2003).

The VUI system should always provide information about what is going on within a reasonable time. It is fundamental in order to manage conversation to keep on signaling the user. The 2 signaling types in an human-agent interaction are: states and transition.

In VUI providing the following states are necessary to make sure user's is aware about what is going on in the system:

- Initial: visual/sound indicators that system is ready to engage (usually activated by the wake-up word)
- Listening: visual/sound indicators that the device is actively listening to user's voice or processing a request.
- Processing/thinking: When the user stopped making a command, the system should provide visual/sound indicators that system is processing the request
- Speaking: visual/sound indicators that system is speaking
- Ended: visual/sound indicators that human/system completed tasks

Conversational markers to provide user the possibility to divide in chucks the pieces of information. Also, it is necessary to signa the transition between the states: when the system started listening, or Finished listening or finished interacting. The following diagram represents the voice interaction flow with Voice Activated agents. Since the type of interaction is intangible, feedbacks must be included in the whole process, and this is represented by the continuous line.

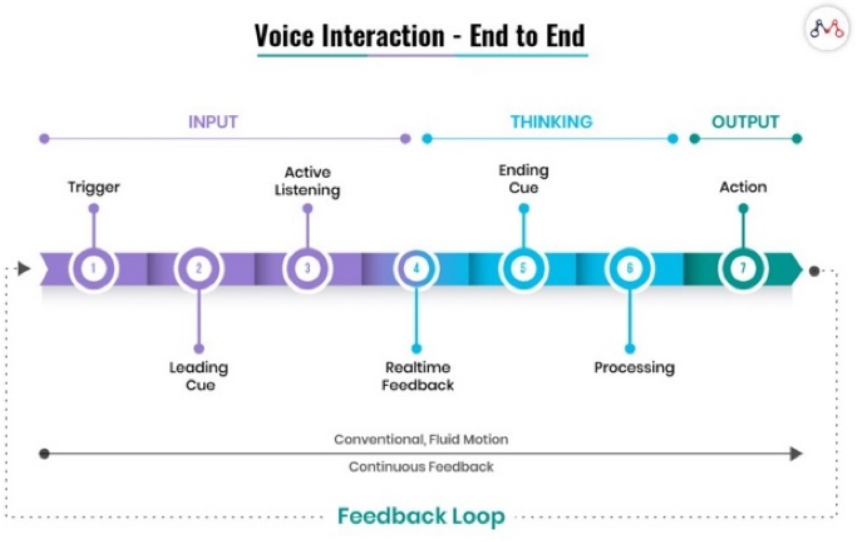


Fig. 25 VUX (from: https://www.mantralabsglobal.com/blog/voice-user-interface-is-the-next-ux/)

- **INPUT: TRIGGERS**

There are several ways user can use to start the interaction with a voice activated assistant. However, a re-cent report made by Microsoft (2019) found that people interact in the following way with their assistant:

- 57% Speak their requests/queries aloud only using their voice
- 9% Only type their requests/queries
- 34% Ask requests/queries using either voice or typing.

In most of Voice activated agents there are 5 main type of user's input that triggers the beginning of the interaction:

- Voice: The user says the "wake-up world" in order to make the device start processing the user speech
- Button: The user presses the physical button on the device
- Touch surface: The user presses the surface of the device on the device
- Display screen: The user presses the surface of the display
- Chat/app: The user will start press the button to start interacting

- **LEADING CUE**

It is the cues provided by the device that show that it registered user's input and it is able to start processing information.

Fig. 26 google home leading cue (from: https://support.google.com/googlenest/answer/7073219?hl=it)

- **ACTIVATE LISTENING**

These are the cues provided by the device that show that it is listening to the user and it is continuous until he/she stops.

Fig. 27 google home listening cue (from: https://support.google.com/googlenest/answer/7073219?hl=it)

Active listening signifiers are extremely important to build trust in these devices, some people fear that these devices are always listening due to the fact that the Microphone is always on if the microphone button is not disactivated.

- **REAL TIME FEEDBACKS**

There are several types of real-time feedbacks in these devices/interfaces. The following feedbacks increase the visibility of the system giving information about system state.

- **Realtime text:** the voice is converted to text in Realtime
- **Audio:** The system playback what heard to have a confirmation
- **Output text:** the voice is converted to text as the user stops speaking
- **Display:** text/images
- **Led Light:** visual feedback is shown

These will be discussed in detail in the case analysis later in the chapter.

- **ENDING CUE**

These are the cues provided by the device that show that it registered user's speech input and now stops listening to process the request

Fig. 28 google home end conversation cue (from: https://support.google.com/googlenest/answer/7073219?hl=it)

- **PROCESSING**

These are the cues provided by the device that show that it is processing the user's speech input before making an action

Fig. 29 google home processing cue (from: https://support.google.com/googlenest/answer/7073219?hl=it)

- ACTION

Actions are all the device's output. They all depend on the partnerships, some examples of the possible actions can be

- 3° parties' devices / IoT devices
- Audio (songs/reminders etc)
- Display (Video/pictures/text)
- Chips (Suggestions on how the user can continue the interaction)

All these elements are designed correctly make the user more aware of what is going on, and if his actions are interpreted.

Processes to design a voice experience

1. User

Bill Buxton, principal researcher at Microsoft Research, describes VUI personas as "Place-Ona", due to the fact that location is limiting the type of interaction the user can have. Placeonas are a mixed of information related to the availability of user's senses in a specific environment.
The context in which the interaction takes place can influence whether a type of device influence should be employed or not. According to Bill Buxton is not a matter of "best input" or "best output" but the context in which the interaction takes place. It is all about discovering user's needs and goals.
Placeonas are a combination of a specific scenario, user's intent and senses that can be used in the interaction, that refers to the type of environment the user is. According to him, to create a Placeonas, a designer needs to consider the context and user's senses that can be used in that scenario, such as: touch (hands), Sight (eyes), hearing (ears/voice).

An example of a baking scenario:

Fig. 30 Placeonas (from:https://careerfoundry.com/en/blog/ux-design/voice-design-personas-placeonas/)

As can be seen, this persona shows how a specific location put a limit of the type of interaction allowed.

2. Use cases and devices

A common practice in voice user interface is identify possible use case, that are a way to uncover the different ways in which the system can be used by the users. The use case is a definition of a specific thing that the user needs to accomplish, and all the interaction steps the users needs to do to achieve it.

In fact, AI agents were first imagined and designed as “butler” or “best friend type” but the most frequent uses are not any assistant as secretary or a general helper. They are implied for executing specific tasks, as chart made by Nielsen (NNGroup,2018) below shows.

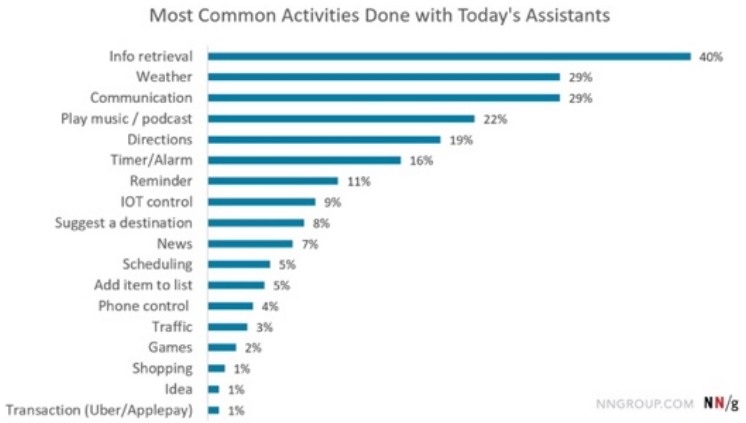
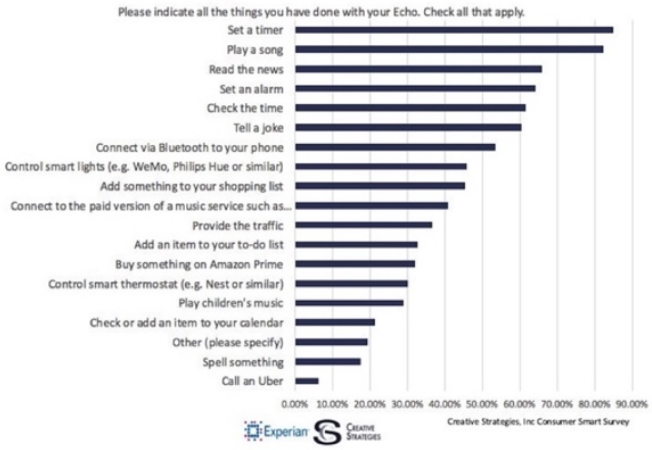


Fig. 31 Use case (from: NNgroup 2018)

At the same time, the usage for domotic devices and overall IoT systems it is increasing. About 9% of frequent users reported they use it to control and manage smart-home supplies. It has been found that user engage with AI agents when their hands are busy and usually to execute easy tasks, that usually became a routine (asking to play a song, weather forecast etc).
Similar results has been found also by Creative strategies InC survey (2018), people interacted with their Echo speakers with the following goals:



Voicebot report (2018) instead focused on the smartphone entertained interaction:

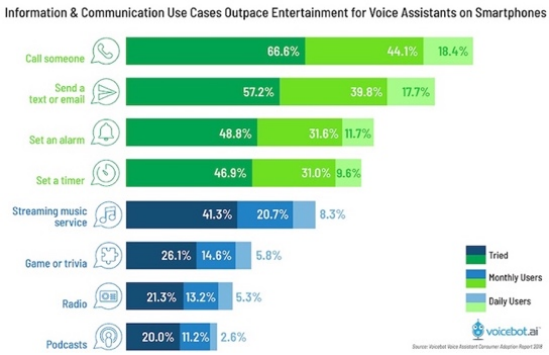


Fig. 31 Use case (from: Voicebot.ai 2018)

- **VOICE SHOPPING**

Voice shopping is not only the purchase of an item but rather all the buying experience from searching a product, comparing the option, comparing price to add to the cart and buy. Purchasing is only available in US, UK and India, for the rest of the world the products would be available on the amazon cart.
According to Adobe Digital Insights "Nearly half (47%) of smart speaker owners reported using one to initiate product search and research, 43% said they use them for creating shopping lists, and 32% do so for price comparison." (Abramovich G., 2019). This means voice shopping can be a good opportunity in the following years, even if correctly people do not physically buy through it: it is used to compare products/price and initiate product search.

3. Multimodal design

According to Amazon developer, Ankur Prasada, when designing for a VUI is important to keep in mind the kind of purpose the application has. Each modality does have a competitive advantage in terms of utility, which as can be seen in the graph below depending on the intent. According to Amazon developer group, when building a "Skill" it will depends on one of the following purposes: Doing, Pinpointing, telling or browsing.

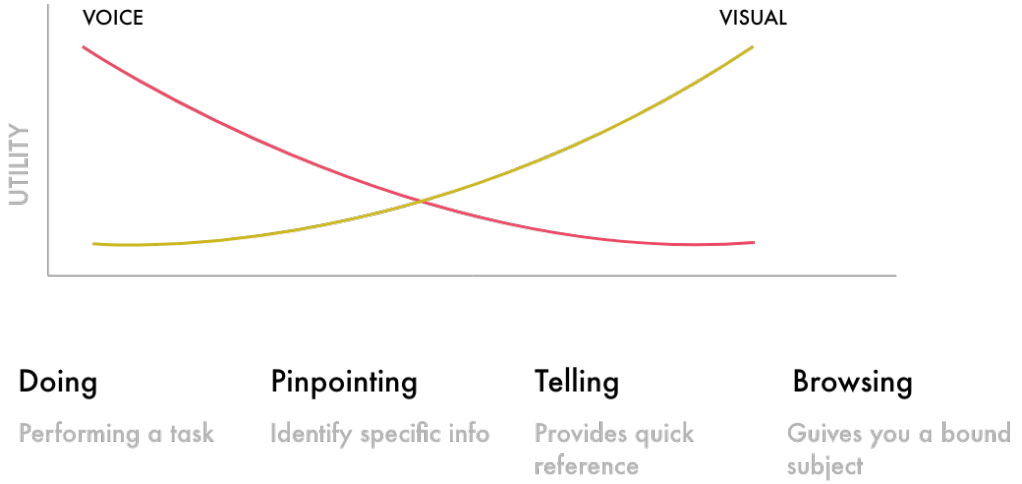


Fig. 32 Alexa developer (Adaptation from: https://www.udemy.com/course/amazonalexa/learn/lecture/4367798#overview)

Similarly, Cathy Pearl, Google Assistant Head designer, deciding how to use the different multimodal elements should be decided up on the use case taken into consideration. Designing for VUI is about understanding the use cases the optimal modality of interaction and only later the devices. Use cases and context are the first element to consider when designing a VUI Skill (Amazon Alexa) or Action (Google Assistant)

- CLASSIFICATION OF DEVICES

Mainstream devices can be classified in many ways, broadly we can divide them of context in which they are implied and the modality (voice only or visual+voice).

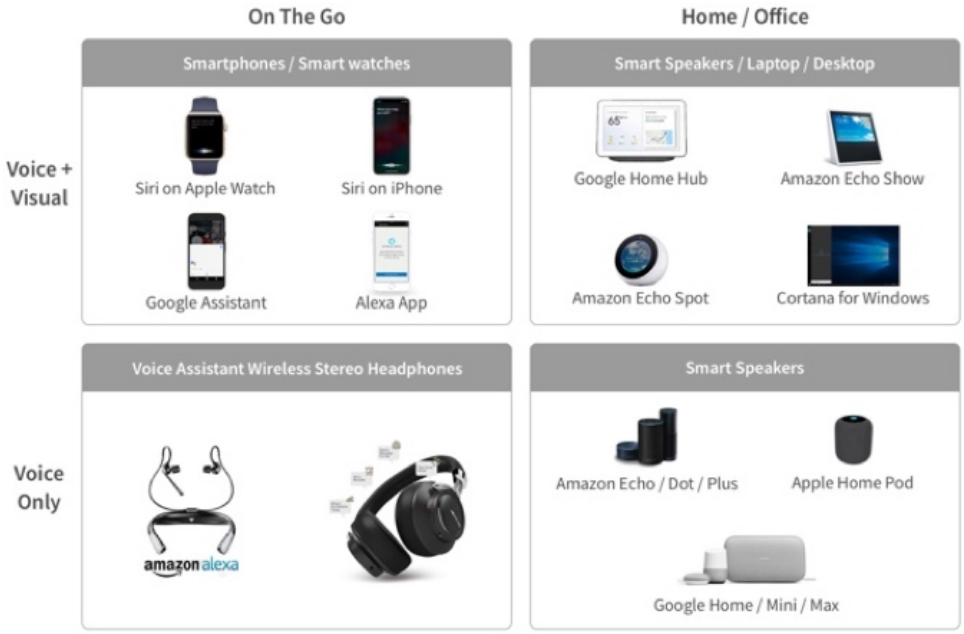


Fig. 33 Voice assistant Devices (from: https://www.zeuxinnovation.com/user-experience-voice-assistants.html

The type of interactions and responses of a voice activated assistant are in different modalities according to the selected device in which they are embodied.
Depending on how the voice is used there are three types of voice-enabled interfaces:

- Voice-first devices
- Display-first devices
- Multimodal

In Google conference (google I/O' 18) explained that in order to design for AI assistant they use the framework of multimodal spectrum (as can be seen in the picture below) to categorise devices according to their modality type and possibles context of application considering potentialities and constraints of the modalities.
On the extreme of the distribution we have:

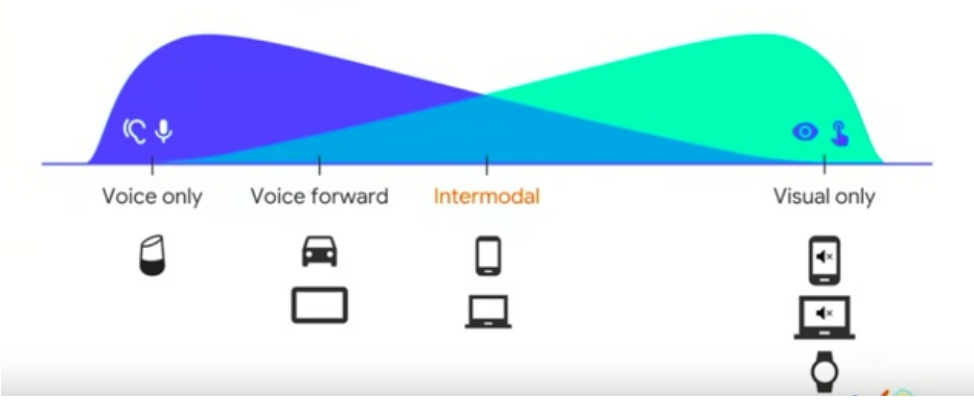


Fig. 33 Multimodal spectrum (Screenshot from: google I/O' 18)

On the extreme of the distribution we have:

- **VOICE FIRST**

Voice-first

Apple airpods

Google Pixel Buds

Amazon echo buds

Voice-first devices are smart-speaker in which user rely on audio, the modality of the interaction is purely vocal/audio. In fact these devices offer any visual display and their usage relies on audio for both input and output. The only cues available for the users as possible commands are audio, the lack of a visual interface decrease the ability of the device of providing information and options and in fact they are usually implied in the executions of simple task.

Input: voice/ touch
Output: audio, led lights

- VISUAL ONLY

Visual-only

Apple watch

Smartwatch Fitbit

Smartwatch with Fossile

Display first or visual only devices are all devices such as phone/smartwatches that are on mute, in which the input are touch based and the output is texted based and graphical. The modality of the interaction is almost purely visual. Those devices perform best in contexts in which the user can not use the voice.

Input: touch
Output: visual (text based/images)

The central part of the spectrum is multimodal:
voice forward devices and intermodal devices.

- **VOICE FORWARD DEVICES (OR "DISPLAY FIRST DEVICES")**

Voice forward

Amazon echo show 5 | Amazon echo show (2° generation) | Amazon echo spot | Google Home Hub

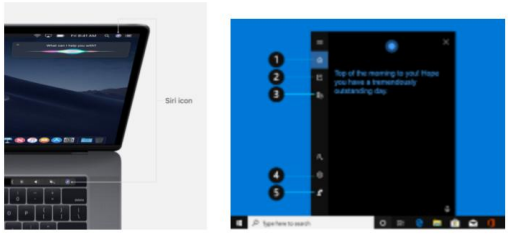
Voice Forwards devices are all those devices that integrate voice commands into a Graphic User Interface. In such devices the screen is incorporated such as displays or car devices. The integration results in an enhancement due to the combination of the two modalities.

Input: Voice/Touch
Output: visual(text based, Videos, Images), audio

- **INTERMODAL DEVICES**

Intermodal /
visual only (when the devices are on mute)

Intermodal /
visual only (when the devices are on mute)

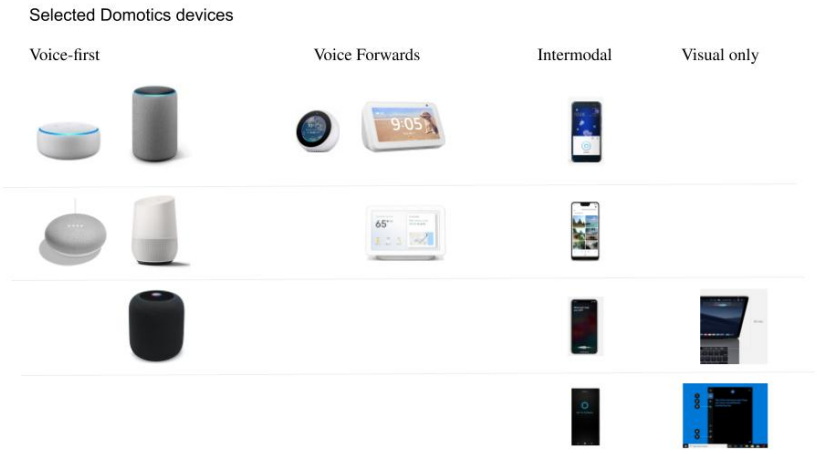


Windows | Windows

Intermodal devices are all those screen based devices that let more modalities of interacting independently such as phone/computer applications.

Input: Voice/Touch
Output: visual(text based, Videos, Images), audio

The new distribution with the selected devices would look like this:



All These Domotics devices have been grouped by types and brand. In the first row Amazon, in the second Google. In the last two rows Apple and Microsoft which as can be seen they have less devices.

4. Situational design: the role of Environment

Jan König(2018), founder of Jovo an open source framework for voice-apps, says:

"Every device has its own context. Context-first is delivering the right information at the right time on the right device."

More specifically, according to Karen Kaushansky (Voicebot interview, 2018) says when designing for voice experience the starting point is voice-first experience because it is more complex as interaction. Then understand the other modalities to provide the possibility for the user to choose the preferred modality of interaction: *"Customer-first, not just voice-first: Giving users the choice of how they want to interact, how and where and what inputs they want to use, is becoming part of the natural landscape of design."*

Combining different modalities can enrich and provide a better user experience but there are factors that influence multimodality (Google I/O 2017) and the choice of what devices imply due to:

- **Motion**: It the user moving or not during the interaction with the device?
- **Environment**: is the device meant to be shared in a group of people/single user? or in private or public space?
- **Proximity**: is the device close to user (wearable) or far (google home mini)
- **Audio capability**: related to the distance and the power of microphone
- **Visual capability**: screensize

- VOICE FIRST

Context: fast for specific tasks, eye-free, hands-busy
Environments: private space
Movement: active
User: both personal/group of users
Audio: Strong asset
visual: lack of visibility, avoid information overload/ long list

- VISUAL ONLY

Context: usually ambients in which the user can not use the voice
Movement: Static
Environments: public space
User: personal/group of user
Audio: weak asset (not present)
Visual: Strong asset

- **VOICE FORWARD**

Environments: private space /inside
Movement: both active/static
User: both personal/group of users
Audio: Strong asset
Visual: Strong visibility

- **INTERMODAL**

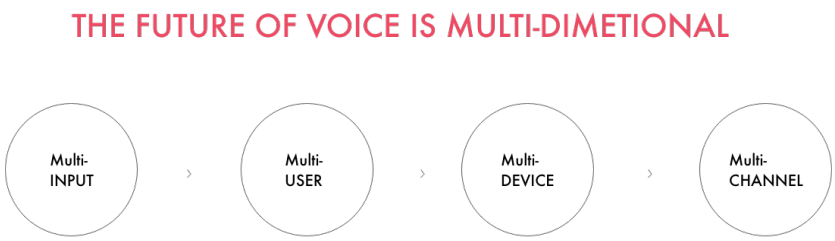
Environments: private space / public
Movement: both active/static
User: personal
Audio: Strong asset
Visual: Strong visibility

To conclude, combining different modalities can enrich and provide a better user experience. This type of interaction uses more than one type of interface such as Audio/screen.

The future of Voice

At CES 2020, the International Consumer Electronics Show held in Las Vegas, continue to prove the consumer interest for voice assistants. Google and Amazon showcased the seamless experiences of voice and visual displays from automobiles, domotics and third parties' devices pushing visual and voice modalities for more complex interaction.

According to the direction of these companies the future of voice is multidimensional that consists on 4 levels:



- **Multi Inputs** now consists mainly voice/touch in the future gestures interaction might be possible
- **Multiuser**: multiple people interaction will be possible. Today the only one that is able to discriminate multiple user voices is google assistant.
- **Multi Device**: Continuous are seam-less experience between different devices are increasing potentially
- **Multi-Channel**: Multiple touchpoint that allows the user to interact and getting in touch with assistants in multiple channels.

Case studies

Usually voice activated assistant consist on multiple components of 3 main layers that works together:

Fig. 34 The Blueprints of Conversation (from https://www.slideshare.net/JeffreyHumble/ia-for-vuislideshare?)

The device: assistant enabled device.

The application: a companion app that needs to be installed on user's device;

The Platform: Is the type of ecosystem used (E.g. Alexa environment)

The voice interface lies in the upper two layers (App, Platform) in the cloud environment (not in the device itself). The device can change but the UX should be the same across devices. The type of interactions and responses can be through different modalities according to the selected device due to the fact that the assistant's embodiment can be in different interfaces.

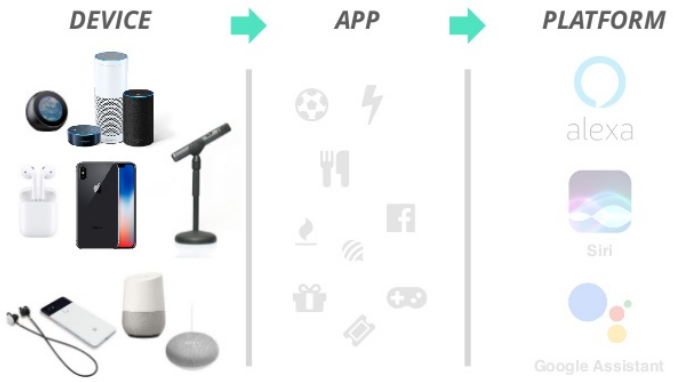


Fig. 35 The Blueprints of Conversation (from https://www.slideshare.net/JeffreyHumble/ia-for-vuislideshare?)

Case studies

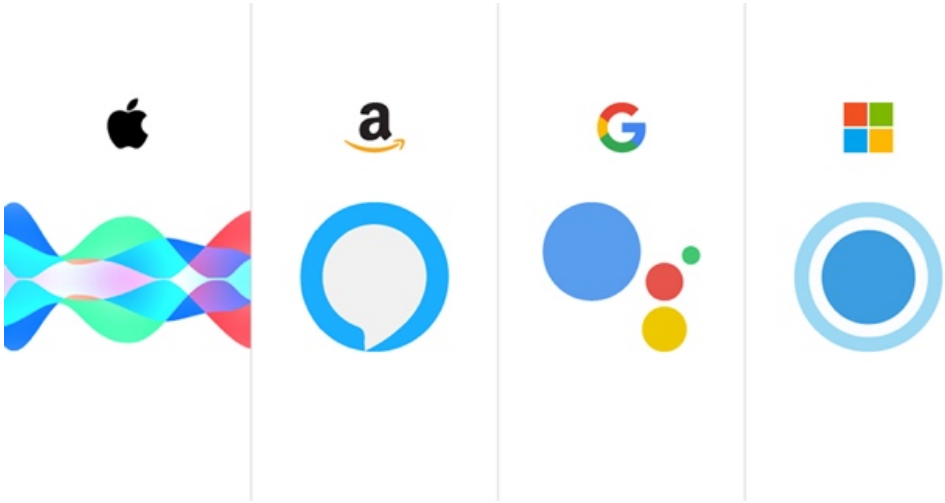


Fig. 36 Voice assistant (From: https://gettecla.com/blogs/news/voice-assistants-accessibility)

The biggest and famous Voice activated personal assistants are:

- Siri by Apple
- Google assistant by Google
- Alexa by Amazon
- Cortana by Microsoft

All of them, currently are either embedded on own devices, smartphones or computer. All of them work in similar ways but the accuracy, privacy settings, functions, the personalization and the possibility to include 3 party services differs.
The reason why they have been selected it the wide availability of these assistant, Microsoft report (2019) found as can be seen in the image below that those AI agents are the most used.

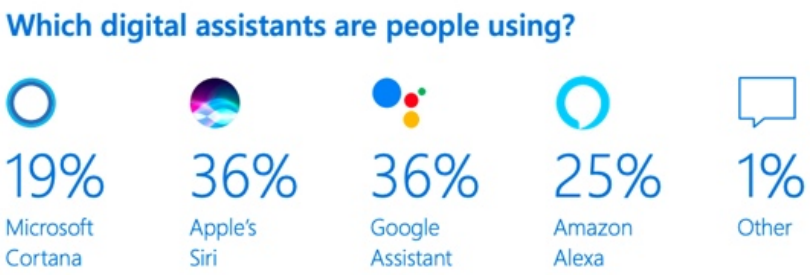


Fig. 37 Voice assistant (From: Microsoft report, 2019)

Google Assistant

Fig. 38 Google assistant devices (From: https://www.smarthome.news/how-tos/google/google-home-compatible-devices)

Year 2016

Made by Google

Languages: English, Danish, French/ French CA, German, English (India), English (US) , Hindi, Italian, Japanese, Korean, Spanish, Dutch, Norwegian, Swedish

*Google Home Max is only available on:
French/ French CA, German, English

Wake-up world: "Hey Siri"

Websearch: Safari search Engine

Voice gender: Neutral

- **GENERAL CHARACTERISTICS**

Google assistant is a personal assistant powered with AI developed by Google that replace the previous version "google Now". The new version allows to have a two way conversation, that make possible interactions both textual and vocal, while before it was only vocal. The Service is available on smartphones, smart speakers (google home), smartwatches and smart display.
The advantage in the Google Assistant is the link with Google itself, when asked a question is able to connect to google search engine, and probably it is the smartest agents due to the fact it is connected to your google account.

This allows the agent to know more about the user from email, contacts, locations, searches, and schedules. It is also able to collect data about the user: all from search, maps and websites but also all the app/device that are connected to Google account. All data is cross-devices and allows for more personalization and autonomation settings compared to Google competitors.
If the user asks to the google assistant what data collected about him/her, it would not answer directly, but it suggests checking privacy settings online. Since 2019, it is possible to delete all the search and data, one by one, on the website. With the addition of google duplex technology, now Google is able to make phone calls, book an appointment and check opening hours. It is able to reply in realtime and to make more real, speech features such as "ehm", "um-hm" has been added.
The CEO, Sundar Pichai, during the Google I/O developer conference held may 8 2018, "explained how its Google Duplex technology can help the phone-shy avoid having to actually speak to someone to make an appointment".

- **DEVICES AND INTEGRATION**

Google assistant is now in android phones, tablets, wear by google, watches, google home, android tv, android auto, smart auto, google pixel bud. and it has many integrations with third party services and it is available in more than 400 million devices according google's post (bret kinsella on january 6, 2018)

Companion App interface Structure

- **INTERACTIONS**

To interact with google assistant, the user needs to either press the microphone button (APP interfaces) or say "Okay google". This means that to be able to make a request, the user needs to wake up the devices, because it should not record the voice all the time. It requires social interaction to work (Fong,2003).
According to Perry Cathy, writer of designing VUI also says that lots of the interaction on Google's assistant still involve touching the screen, for the way is design the interface it still includes pictures and text.

Input
Depends on the device: Voice/touch

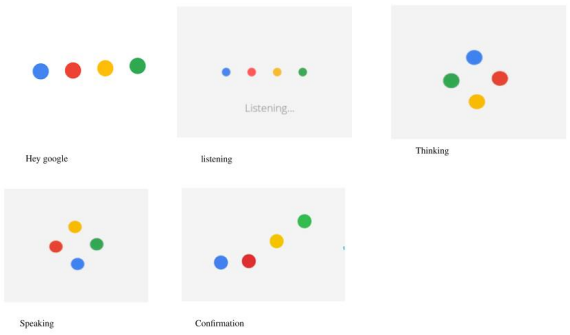
Output
Depends on the device: Voice, text, pictures, links to websites. Suggest follow-ups (in the app)

Device	Google home mini	Google nest hub	Google assistant app
Input	Voice/Surface touch/ button/App	Voice/Display (touch), button/App	Voice/Surface touch/ button/App
Out-put	Audio, IoT devices	Audio, IoT devices, visual (images/video), Text	Audio, IoT devices text
Feed back	Audio, Led lights	Audio, led lights, display, graphical agents	Audio, graphical agents

Feedbacks

Google's four dots are dynamic and always moving to represent Google's assistant state of Listening, Thinking, Speaking and confirming.
Among all the possible devices in which Google's assistant have been embodied here they will be considered on product that is: Google home mini and Google home nest Hub for the characteristics of voice-only for the first and for the latter voice-forward.

Google's dot animation



Google's Dot status

Status	
echo dot heard "Hey google "	4 dots in the line
Iphone is listening	dots become 4 bar that moves according to the frequency
Iphone is thinking	4 dots in circle moving counterclockwise
Iphone is responding	4 dots in circle, each one moves outside-in
Confirmation	4 dots horizontally positioned move up-down
Iphone ended speaking	4 dots in the line

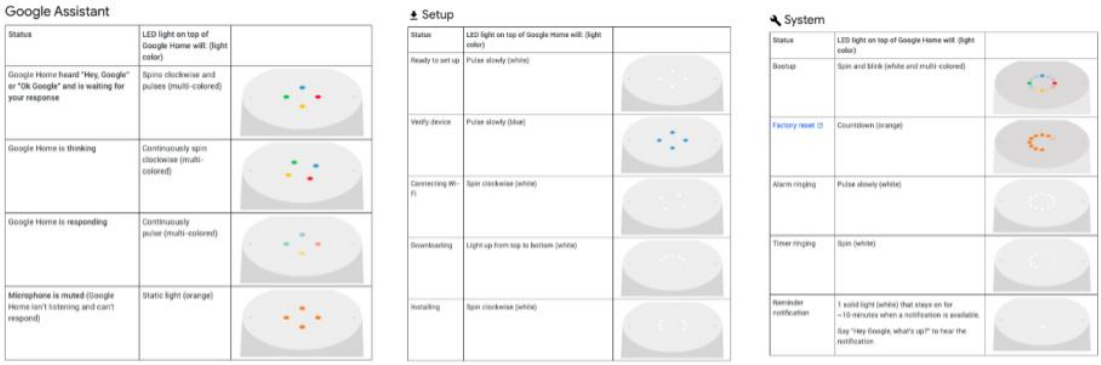
Feedback& status per device

- Google Home Mini

Fig. 39 Google assistant devices (From: https://support.google.com/googlenest/)

- Google Home

Fig. 41 Google assistant devices (From: https://support.google.com/googlenest/)



Google Assistant

Status	LED light on top of Google Home will: (light color)	
Google Home heard "Hey, Google" or "Ok Google" and is waiting for your response	Spins clockwise and pulses (multi-colored)	
Google Home is thinking	Continuously spin clockwise (multi-colored)	
Google Home is responding	Continuously pulse (multi-colored)	
Microphone is muted (Google Home isn't listening and can't respond)	Static light (orange)	

Setup

Status	LED light on top of Google Home will: (light color)	
Ready to set up	Pulse slowly (white)	
Verify device	Pulse slowly (blue)	
Connecting Wi-Fi	Spin clockwise (white)	
Downloading	Light up from top to bottom (white)	
Installing	Spin clockwise (white)	

System

Status	LED light on top of Google Home will: (light color)	
Bootup	Spin and blink (white and multi-colored)	
Factory reset	Countdown (orange)	
Alarm ringing	Pulse slowly (white)	
Timer ringing	Spin (white)	
Reminder notification	1 solid light (white) that stays on for ~10 minutes when a notification is available. Say "Hey Google, what's up?" to hear the notification	

Volume

Status	LED light on top of Google Home will: (light color)	
Volume mute	1 centered dot (white) Only visible when making volume control gesture	
Volume 1 / Volume 10	1 dot (white) / 10 dots (white)	

Error

Status	LED light on top of Google Home will: (light color)	
Update taking longer than expected	6 static lights (orange)	

Alexa

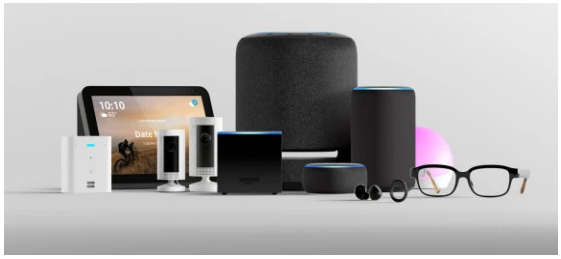


Fig. 42 Alexa devices (from: https://www.engadget.com/2019-09-25-amazon-echo-alexa-devices-event-recap)

Name Alexa

Year 2014

Made by Amazon

Languages: English, French, German, Italian, Japanese, Portuguese (Brazilian), Spanish

Wake-up word: "Alexa"

Methods of interaction: Mainly Speech, Touch for display/buttons

Websearch: Bing search Engine

Voice gender: Female

- **GENERAL CHARACTERISTICS**

Alexa was developed by Amazon in 2014 and integrated in Eco, the service provides cloud-based automatic speech recognition (ASR) and natural language understanding (NLU). Alexa's able of voice interaction enabling the user to achieve a certain goal from news to entertaining but also buying directly from the ecommerce just with the voice command. Compared to her competitors, Alexa was developed for domestic use it gives the opportunity to the users to manages many IoT devices and services.

Personalization is possible: users and developers can create and extent Alexa functionalities by installing different skills third-party vendors.

Alexa is embedded in many devices, but it is coherent and familiar in all products, even if the products have different appearance, the attention systems ensure that Alexa behave in a familiar and predictable way. According to Amazon this consistency create more trust and strengthen user understanding. Even if at the time Cortana and Google assistant are technological advanced the competitive advantage is that Amazon created a AI-first device rather than embedding it to existing devices rather to dedicated devices.

*"The phone is going to be, for the foreseeable future, a finger-first, mobile-first device. You need an AI-first device to solidify an emerging base of ecosystem*s." (Qi Lu, 2017)

Nevertheless Alexa is embedded in many devices from smartspeaker, smart TV/ displays, controllers, in the car and many more, all of them can be managed by the app (available on apple store/google play/ amazon) in which the user can personalize his/her own device by adding additional skills and setting atomization. Alexa is more reliable compared to other agents but there is not an optimal phone/tablet experience to use continuous along smartspeakers/smart display. An important functionality of Alexa is the possibility of voice shopping but according to a research that was only tried by 2% of all alexa's owners (Beniamino Mayo, 2017).
Many concerns are related to the privacy of user, according to amazon voice recording happens only with the wake-up word, user cab deleted voice recording when associated with their account.

- **DEVICES AND INTEGRATION**

Amazon Alexa has specifically its own devices: Echo, Display and Echo buds. There have been made also some partnerships with Cars Companies (Ford, BMW, Toyota, Lexus) and Fire for TV and tables. There are plenty of integrations and compatibilities especially in are in domotics and connected enviroments, in the following link can be seen the possible integrations updated (https://www.smartdomotica.it/dispositivi-compatibili-con-alexa-amazon-echo).
Alexa's skills kit (ASK) is a collection of self-service API and tools in which is possible and simplified the development of skills integra-tion for alexa. Those "skills" are all the apps in which extend the functionalities of alexa that enables users to interact in specific task seamlessly with the voice.

Companion App interface

Alexa app is meant to manage all the connected devices and the skills installed.

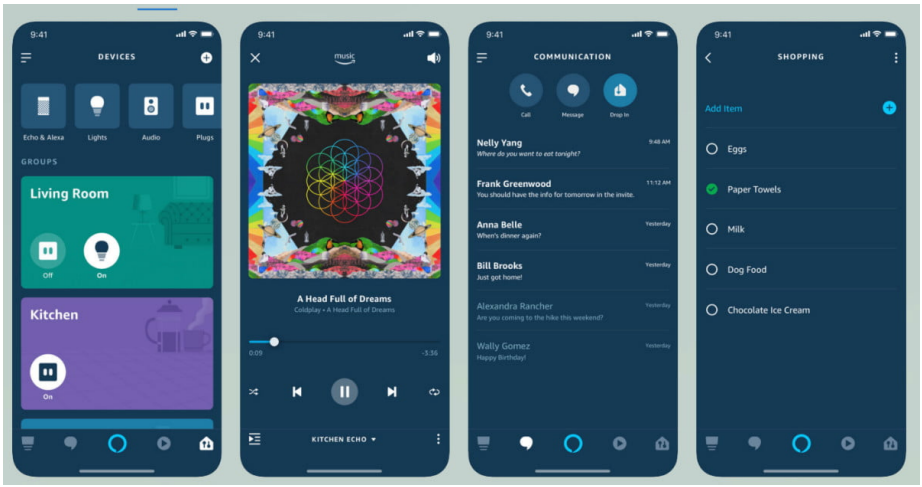


Fig. 43 Alexa app (From: https://alexaapp.ooo/pc-windows-10/)

It is also possible to set up Amazon Alexa in the device and graphically and interactively is similar to the amazon Echo show 5 and 8.

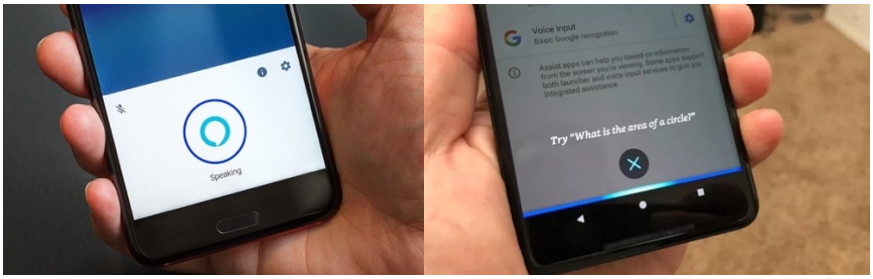


Fig. 44 Alexa app (From: https://www.cnet.com/how-to/how-to-ditch-google-assistant-for-cortana-or-alexa-on-android/)

- **INTERACTION**

To interact with Alexa, the user needs to either press the microphone button (APP interfaces) /or touch the screen menu for Echo show device or say "Alexa" with the other devices. This means that to be able to make a request, the user needs to wake up the devices. As most of smartspeaker devices requires social interaction to work (Fong,2003).

Input

Depends on the device: Voice, touch

Output

Depends on the device: Voice first, text

Device	Amazon Eco dot	Amazon Echo show	Alexa app
Input	Voice/ button/App	Voice/Display (touch), button/App	Voice/Surface touch/ button/App
Out put	Audio, IoT devices	Audio, IoT devices, visual (images/video)	Audio, IoT devices text
Feedback	Audio, Led lights	Audio, Led lights, display	Audio, graphical agents

Touch control APP/ECHO SHOW 5

- Touch-screen menu
- Switch button: to close camera
- press button2: to mute
- Tap/ hold "+ ": to increase the volume
- Tap/hold "-": touch or hold the minus sign on the top to lower the volume
- Tap/hold: to start speaking to Siri

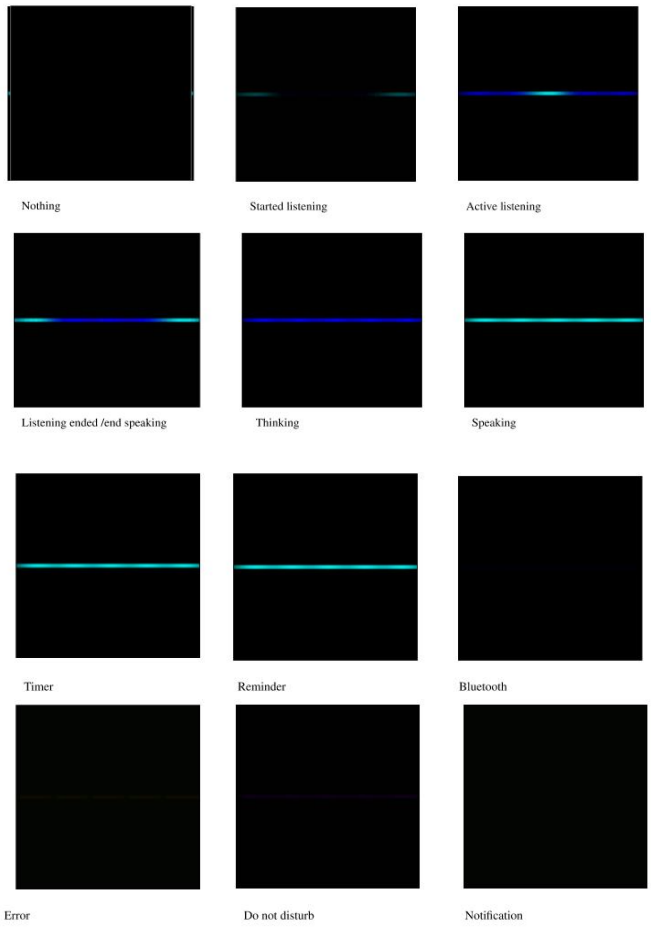
Touch control ECHO DOT

- press button: to mute
- Tap/ hold "+": to increase the volume
- Tap/hold "-": touch or hold the minus sign on the top to lower the volume
- Tap/hold: to start speaking to Siri

Feedbacks

The Alexa circle is dynamic, each colours and motions express different states of Alexa to users. Colours, sounds and visual components are fundamental for state communication when there are no other elements that expresses Alexa's changes.

- Feedback/states of Alexa echo show and App



*The feedbacks are clearer during the animations; the static modality does not show how they look like.

Fig. 45 Alexa feedbacks (From: https://www.amazon.com/gp/help/customer/display.html?nodeId=GKLDRFT7FP4FZE56)

Status	
echo dot heard "Alexa"	cyan from the border to the center
Iphone is listening	cyan from the center to one center
Iphone is thinking	cyan from center to borders
Iphone is responding	line is pulsing
Iphone ended speaking	led stop

- Feedback/states of Alexa echo dot

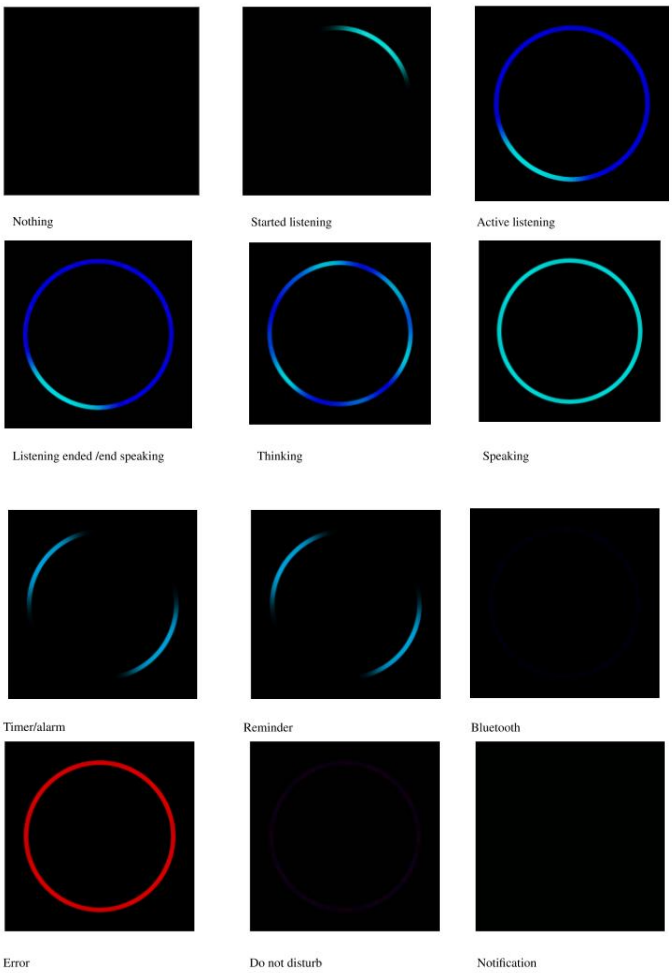


Fig. 45 Alexa feedbacks (From: https://www.amazon.com/gp/help/customer/display.html?nodeId=GKLDRFT7FP4FZE56)

*The feedbacks are clearer during the animation; the static modality does not show how they look like.

Status	
echo dot heard "Alexa"	Cyan and blue on side
Iphone is listening	Cyan is moving circularly
Iphone is thinking	Cyan is moving circularly in four points
Iphone is responding	circle is plusing
Iphone ended speaking	led stop

Siri

Fig. 46 Siri (From: https://voicebot.ai/2018/01/24/apple-siri-devices-total-500-million-not-users-still/)

Year 2011

Made by Apple

Languages:
Australia, Austria, Belgium (Dutch, French), Brazil, Canada (English, French), Chile, China mainland (Cantonese, Mandarin), Denmark, Finland (Finnish), France (French), Germany, Hong Kong (Cantonese), India (English),Ireland (English), Israel (Hebrew),Italy, Japan, Malaysia (Malay), Mexico, Netherlands, New Zealand, Norway, Republic of Korea, Russia, Saudi Arabia (Arabic), Singapore (English), South Africa (English), Spain, Sweden, Switzerland (French, German, Italian), Taiwan (Mandarin), Thailand, Turkey, United Arab Emirates (Arabic),UK,USA (English, Spanish)

Wake-up world: "Alexa"

Websearch: Bing search Engine

Voice gender: Female but it can be changed

- **GENERAL CHARACTERISTICS**

Apple launched his Voice assistant in 2011 and since then, many things has changed: at the beginning it was available only on iPhone, but since 2017 it was also embedded on smart speaker (HomePod), or Apple's wearables (Apple watch, or Airpods). Compared to other virtual assi- stants, siri is not the most

intelligent one, it lacks machine learning and user data, nevertheless is one of the main popular assistant. According the founder of Soundhood commented Apple's voice assistant: *"One of the challenges of Siri is the negative image that they created by over-promising, under-delivering in the early days"* (Jordan Novet, 2019). At the same time as mentioned this assistant because lack of machine learning due to strict privacy standard while competitors collect lots of personal data of the user in order to train its AI but that makes Siri less useful. During last update of iOS 12 Siri adds the function to add personalized shortcuts integrating 3rd parties app.

- **DEVICES AND INTEGRATION**

Siri is integrated to all apple devices iOS, macOS, watchOS, tvOS such as: iPhones, Ipad, Imac, Apple watch, car, Homepod, Appletv Etc. 3rd Party integration, Sirikit enables iOS 10 apps to work with Siri enabling voice interaction with apps. Especially for domotic systems they have been made integration that allows to manage the house with a voice interaction.

App Interface

The bright primary colors waves can fit for each device, overlay modes, blurred background as moving until elaborate the reply.

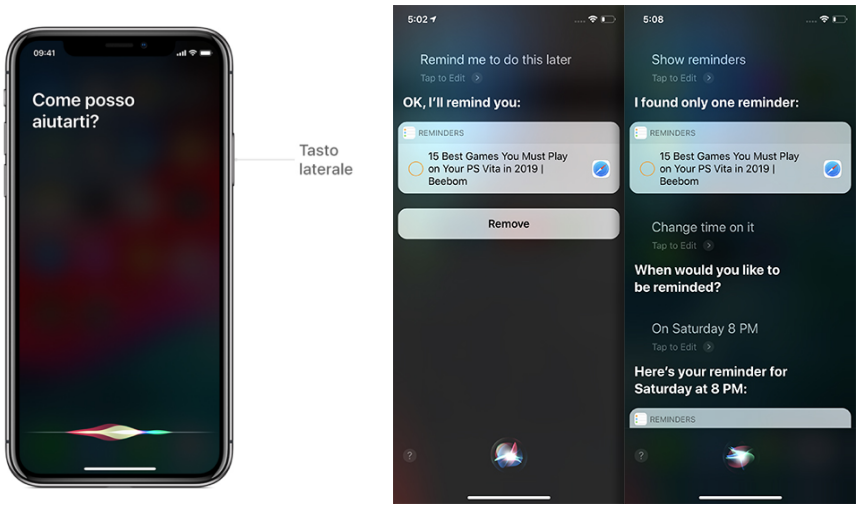


Fig. 47 Siri (From https://www.apple.com/feedback/)

- **INTERACTION**

To interact with Siri, the user need to either press the microphone button (APP interfaces)/or touch the homepod or say "Hey Siri" to activate or add shortcut

to accelerate action through both voice/screen interaction. This means that to be able to make a request, the user needs to wake up the devices. As most of smartspeaker devices requires social interaction to work (Fong,2003).

Input

Depends on the device: Voice/touch

Data output

Depends on the device: Voice, Text, Led lights

Devices	Home Pod	Siri Iphone
Input	Voice/ button/App	Voice, button/App
Output	Audio, IoT devices	Audio, IoT devices, visual (images/video), text
Feedback	Audio, Led lights	Audio, Led lights, display, graphical agent

Touch control Homepod

- Tap the top: to pause
- Double tap the top: to skip the next track
- Triple-tap the top: to play the previous track
- Tap/ hold "+": to increase the volume
- Tap/hold "-": to lower the volume

Feedbacks

The Siri waveform is dynamic, each primary colours and motions express different states to users. Colours, sounds and visual components are fundamental for state communication when there are no other elements that expresses Siri's changes.

Bright primary colors waves moved toward a mirrored waveform according voice frequence. Can fit for each device, overlay modes, blurred background as moving

- Feedback/states of Siri (iPhone)

Fig. 48 Siri (From https://www.apple.com/feedback/)

Status	
Iphone heard "Hey Siri"	Black Blurred screen "how can help you?"
Iphone is listening	Wave move in the bottom oscillate move right to left according to the frequency + text will appear on the screen
Iphone is thinking	Wave-line is substitute to a rotating circle: Text is displayed + auditory feedback
Iphone is responding	wave-line+ text with the reply
Iphone ended speaking	Wave stop back to black screen

- Feedback/states of Siri (Homepod)

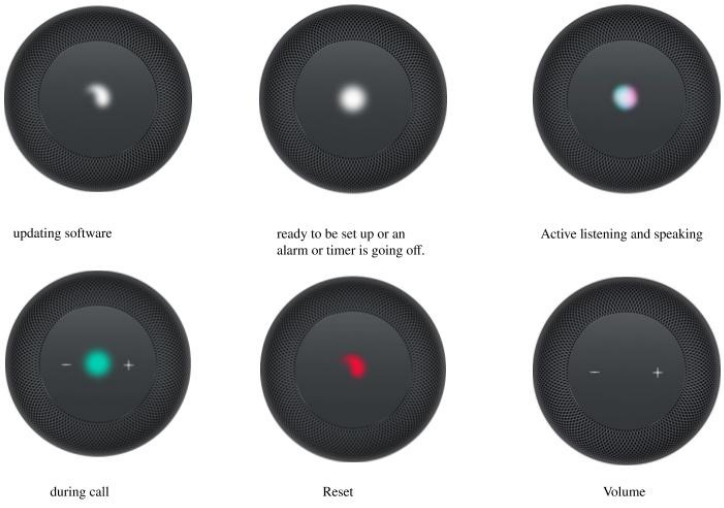


Fig. 49 Siri (From https://www.apple.com/feedback/)

Status

Iphone heard “Hey Siri”	moving colored circle
Iphone is listening	moving colored circle
Iphone is thinking	moving colored circle
Iphone is responding	moving colored circle
Iphone ended speaking	moving colored circle

Cortana

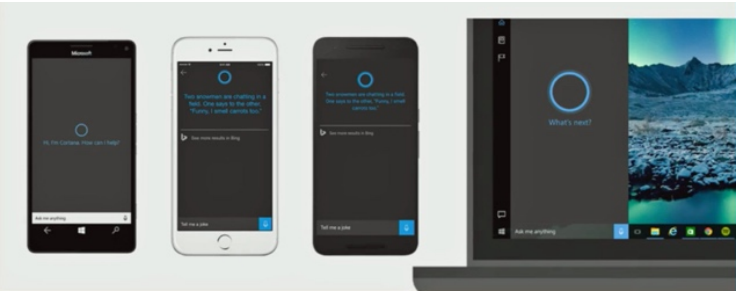


Fig. 50 Cortana devices (From: https://www.thurrott.com/windows/windows-10/66003/the-future-of-cortana)

Name Cortana

Year 2014

Made by Microsoft

Languages: English, French, German, Italian, Japanese, Portuguese (Brazilian), Spanish, Chinese (Simplified)

Wake-up world: "Cortana"

Websearch: Bing search Engine

Voice gender: Female

- **GENERAL CHARACTERISTICS**

Microsoft named its assistant as "Cortana" in reference to Halo character that reminds to the fictional character which designs is a holographic woman shaped that recall queen Nefertiti. The association to the game settings remains: *"One of the Cortana is an intelligent, learning AI who is duty bound to help her companion as much as possible, using a staggering database of information combined with real, growing knowledge of that companion — and the other Cortana is, well, the same thing*," (Pitcher,2014)
The AI agents was originally launched Cortana for Windows 10 PCs and was able to be integrated to iOS and Android 2015.

Microsoft CEO, Satya Nadella: *"Cortana is an integral part of our broader vision to bring the power of conversational computing and productivity to all our platforms and devices,"* but at the same in 2019 the company decided that Cortana will be removed to Android and iOS. It is interesting to be analysed in terms of emotional feedbacks.

- DEVICES AND INTEGRATION

Cortana is available on screen-based devices:Windows 10, Xbox One, HoloLens before also iOS/ Android but it has been removed since november 2019. 3rd party services: None, Cortana manages only Microsoft apps

App Interface
Cortana is primary present on a display/smartphone.

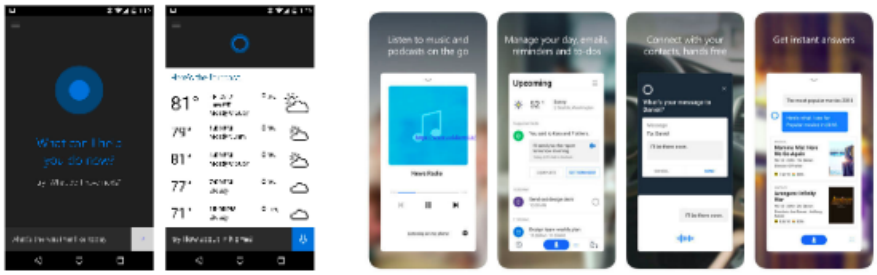


Fig. 51 Cortana App (From: https://www.thurrott.com/windows/windows-10/66003/the-future-of-cortana)

- **INTERACTION**

User interact mainly through voice by saying "Hey Cortana," then make his/her request.

Input
Depends on the device: Voice/button

Output
Depends on the device: Voice, text

Devices	Cortana windows
Input	Voice/ button
Output	Audio, text
Feedback	Audio, Graphical agent moods

Feedbacks

Cortana is able to express 18 Different moods providing information on the system status. Cortana is programmed to respond to a request/command/questions through a grid of emotions and states (as can be seen the pictures) which are meant to make Cortana perceived as more alive. E.g. Cortana is happy when answer to user, however if it did not understand it is embarrassed. The main problem about those animations is that they are all pretty similar and not fully understandable.

Fig. 52 Cortana emotions (from: https://www.pcworld.com/article/2881902/cortanas-ui-now-expresses-18-different-emotions-siri-remains-detached-and-aloof.html)

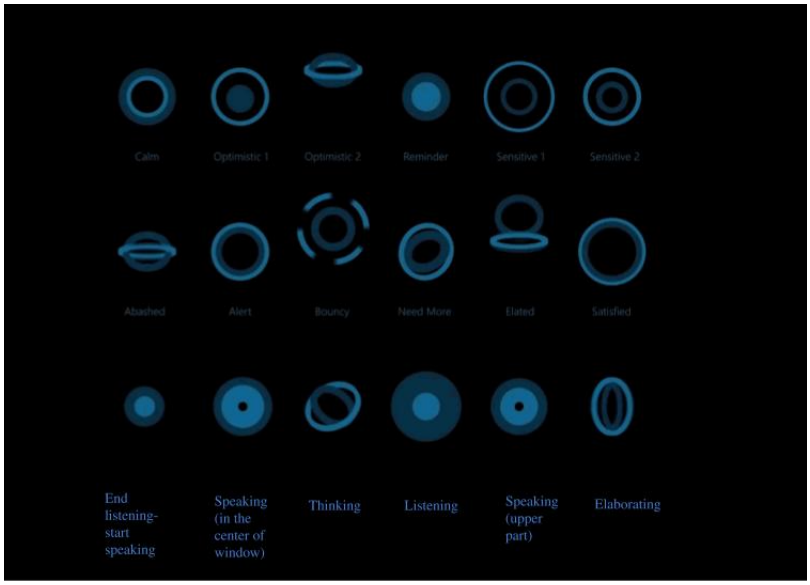


Fig. 53 Cortana emotions (from: https://www.pcworld.com/article/2881902/cortanas-ui-now-expresses-18-different-emotions-siri-remains-detached-and-aloof.html)

Status

Computer hear "cortana "	pop-up the two rings
Iphone is listening	Pulsing the two rings
Iphone is thinking	Orbital moving
Iphone is responding	outer ring pulsing
Iphone ended speaking	led stop

Comparisons

In conclusion we can divide the case studies devices organizing them by their different interaction modalities.

		Google Assistant				Amazon Alexa			Siri		Cortana
		Google Home Mini	Google Home	Google Nest Hub	Phone	ECHO Dot	ECHO show 5	App	Homepod	Smart Display iPad/iPhone	Windows 10 devices
INPUT	**Auditory**	Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice
	Haptics	Button APP Surface touch	Button APP	Button APP Display (touch)	Button Surface touch	Button APP	Button APP Display (touch)	Button Surface touch	Button APP	Button Surface touch	Button
OUTPUT	**Auditory**	Audio	Audio	Audio	Audio	Audio	Audio	Audio	Audio	Audio	Audio
	Visual			Text display Video Image Gif	Text display Video Image Gif		Text display Video Image	Text display		Text display Video Image	Text display Image
	Haptics				Vibration						
	IoT Devices	✓	✓	✓	✓	✓	✓	✓	✓	✓	
FEEDBACKS		Audio Led Lights	Audio Led Lights	Audio Led Lights Display Graphical Agent	Audio Graphical Agent	Audio Led Lights	Audio Led Lights Display Graphical Agent	Audio Graphical Agent	Audio Led Lights	Audio Led Lights Display Graphical Agent	Audio Graphical Agent

More specifically, when designing for a specific device, different contents can be created, as can be seen from the chart below:

	Google Assistant					Amazon Alexa					Siri					Cortana				
	Smart speaker	Smart Display	Embedded TV	Smart watch	App	Smart speaker	Smart Display	Embedded TV	Smart watch	App	Smart speaker	Smart Display iPad	Embedded TV	Smart watch	App	Smart speaker	Windows 10 devices	Werable	App
Text read out	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓
Audio, sound effects	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓
Text display		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓
Video		✓	✓				✓	✓				✓	✓				✓		
Image		✓	✓	✓	✓		✓	✓				✓	✓				✓		
Gif		✓	✓		✓														
Haptics					✓														

In conclusion

For VUI, the context is guiding the design, it is crucial to understand in a specific context in which the interaction takes part and user intention and available senses. The decision of what kind of device create the experience can enhance empower it or decrease it. Multimodality offers, as mentioned, lots of potentialities, however those are situational: the use or voice-first or voice-forward or inter-modality deepens on the use case and user's:

- Intent and goal of the interaction
- Motion (or static level)
- Type of environment (public/private place or social environment)
- Proximity of the device to the user
- Audio Capability of the device
- Visual capability of the device

These analyses help the designer to build a personas, also called in VUI context as "Placeonas", because of the situational nature of this interface.

Only afterwards, the channel (the device) can be chosen, considering the properties as well of VUI and Visual/Auditory modalities.
For the scope of this investigation of understanding whether the multimodality would increase user's reported trust in domestic devices. The Visibility of the system heuristics and feedbacks play significant roles in perceiving agent's state and behavior. One can argue that displayed-based devices increase UX and consequentially building trust with voice activated assistant. To discover whether trust can be influence/manipulated it was implied the same use case with 2 different devices. However due to the nature of trust other two elements need to be considered: the agent and user disposition.

Chapter 3

In the previous chapter voice assistant have been describes in their technical aspects and properties of the interface however there are other features that need to be consider while designing a trustworthy agent.

Concerns about AI agents

Despite the benefits of the application of AI in the different business industries the implications are not always positive but risky.
This might happen *"when the technology fails, succeeds beyond expectations, or simply used in unexpected ways"* (Bowles, 2018).

Current mainstream Assistants require to synchronize different accounts and share data across devices (e.g. Amazon account on Amazon Alexa's devices or Google Account for Google Assistant), and by doing this the synchronization of personal/private or confidential data might be transferred to the service Cloud. Also, in order to give a more accurate and personalized reply, assistant often use contextual data to provide an appropriate response. An example could be when a user asks simply "to switch off the lights" the system to avoid asking questions, geolocate the user's voice and switch off the near lights, however the user is not aware that this might happen.

Last year, many events happened that shaped user's understanding and consequence of the adoption of AI. The following events might affect how users perceive AI agent technology, in fact, despite the spread and improvements of AI agents, some obstacles persist in the adoption and the use of personal assistants (Cuadra, Rase; 2018):

- Being misunderstood;
- Unsatisfactory answer;
- Difficulty of its usage;
- Discomfort of usage;
- Language issues;
- Privacy concerns;
- Preference for classic methods.

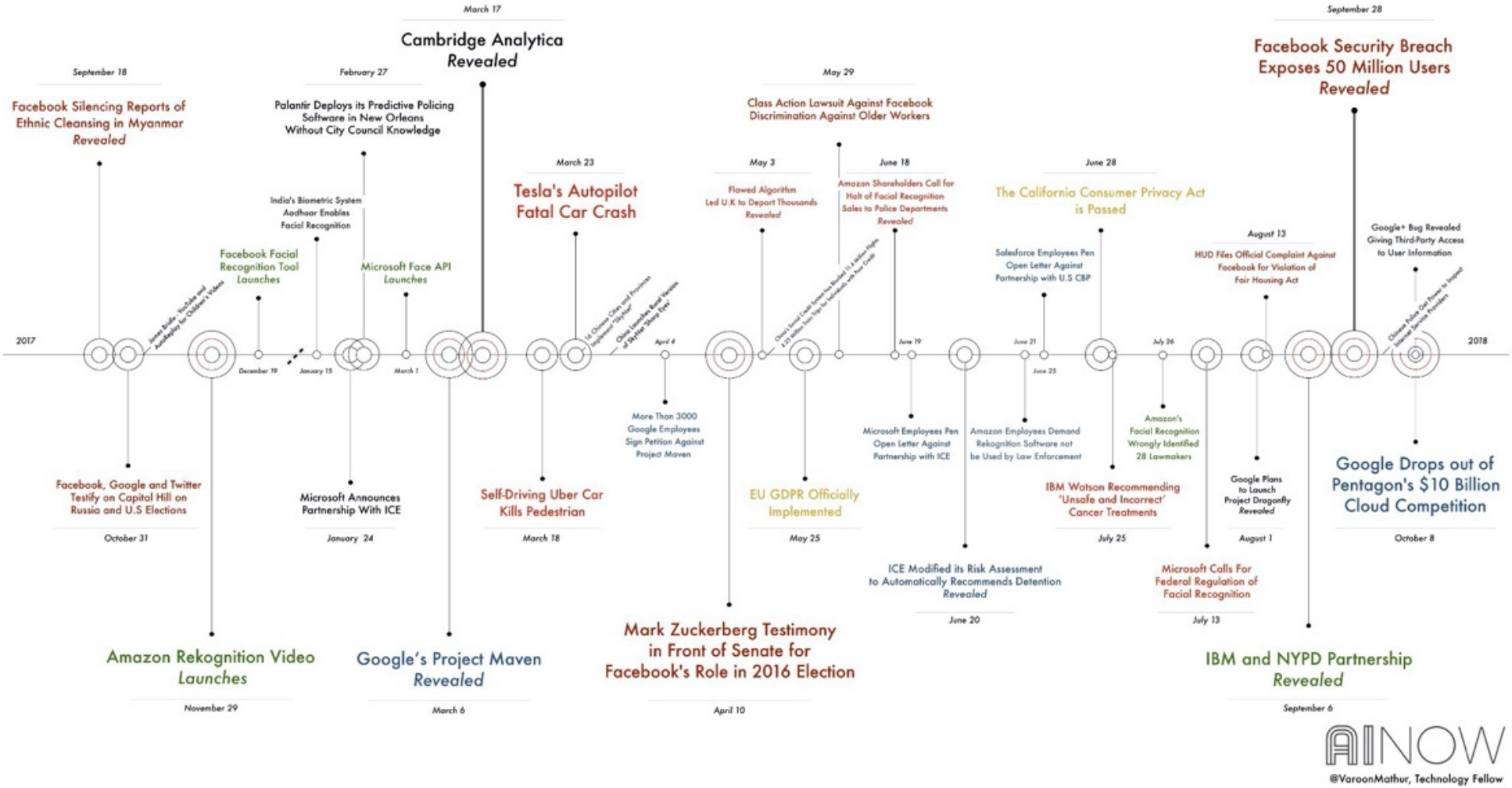

2017
2018
September 18
Facebook Silencing Reports of Ethnic Cleansing in Myanmar Revealed
October 31
Facebook, Google and Twitter Testify on Capital Hill on Russia and U.S Elections
November 29
Amazon Rekognition Video Launches
December 19
Facebook Facial Recognition Tool Launches
January 15
India's Biometric System Aadhaar Enables Facial Recognition
January 24
Microsoft Announces Partnership With ICE
February 27
Palantir Deploys its Predictive Policing Software in New Orleans Without City Council Knowledge
March 1
Microsoft Face API Launches
March 6
Google's Project Maven Revealed
March 17
Cambridge Analytica Revealed
March 18
Self-Driving Uber Car Kills Pedestrian
March 23
Tesla's Autopilot Fatal Car Crash
April 4
More Than 3000 Google Employees Sign Petition Against Project Maven
April 10
Mark Zuckerberg Testimony in Front of Senate for Facebook's Role in 2016 Election
May 3
Flawed Algorithm Led U.K to Deport Thousands Revealed
May 25
EU GDPR Officially Implemented
May 29
Class Action Lawsuit Against Facebook Discrimination Against Older Workers
June 18
Amazon Shareholders Call for Halt of Facial Recognition Sales to Police Departments Revealed
June 19
Microsoft Employees Pen Open Letter Against Partnership with ICE
June 20
ICE Modified its Risk Assessment to Automatically Recommends Detention Revealed
June 21
Salesforce Employees Pen Open Letter Against Partnership with U.S CBP
June 25
Amazon Employees Demand Rekognition Software not be Used by Law Enforcement
June 28
The California Consumer Privacy Act is Passed
July 25
IBM Watson Recommending 'Unsafe and Incorrect' Cancer Treatments
July 26
Amazon's Facial Recognition Wrongly Identified 28 Lawmakers
July 13
Microsoft Calls For Federal Regulation of Facial Recognition
August 1
Google Plans to Launch Project Dragonfly Revealed
August 13
HUD Files Official Complaint Against Facebook for Violation of Fair Housing Act
September 6
IBM and NYPD Partnership Revealed
September 28
Facebook Security Breach Exposes 50 Million Users Revealed
October 8
Google Drops out of Pentagon's $10 Billion Cloud Competition
Google+ Bug Revealed Giving Third-Party Access to User Information
AINOW
@VaroonMathur, Technology Fellow

Fig. 54 Timeline of news events about AI in 2018 (from:AI Now Institute, 2018).

- **PRIVACY CONCERNS**

Microsoft report (2019) found that 41% of users declare their concerns about privacy violation, passive listening in AI agents and voice enabled technology. According to the report, people do not know exactly what kind of data these companies are collecting about them. Amazon has a direct access about location data and even their home address, Bloomberg says that is not even clear how many people are actually be able to access this sort of information. It is possible that most of the users are not aware about all Alexa's backend, however lots of them report an overall concerned about owning them. The majority mentioned the fact are worried about the insecurity and misuse of personal information and how this information can potentially misuse for target marketing and how long this information would be stored. Provider company could use "Data Mining" to interpret personal data to predict user intent or understanding patterns. These companies (Amazon, Google and Apple) to reduce these concerns are trying to be more transparent, giving the possibility to the user to see and delete all some of their personal data (recorded through previous interactions). Also, in 2015, Amazon Alexa Echo, after been suited for being a witness of murder, confirmed that at the time it was recording random piece of conversation in order to train its AI.
However it is not the only answer, as found by a study of Northeastern's Mon(IoT)r Research Group accidentally awakening of voice activated assistant is somewhat common (E. Swartz, 2020).

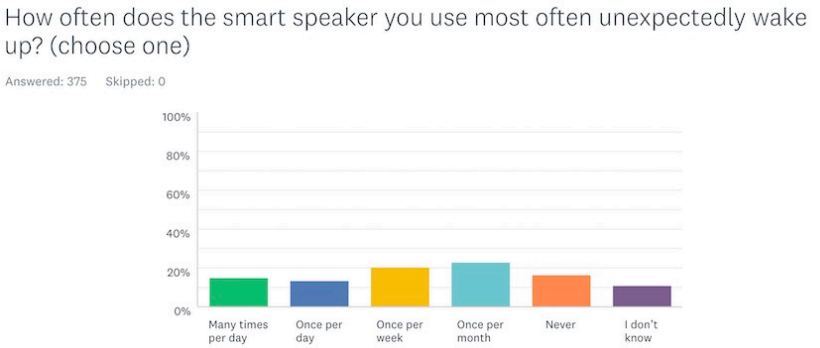


Fig. 55 Accidentaly recording (from: https://voicebot.ai/2020/02/21/voice-assistants-very-prone-to-accidentally-waking-up-and-recording-long-audio-clips-study/)

It is important to build trust that users are informed even in accidental awakening, in order to prevent this and to make user aware, Amazon for example, use feedbacks (blue light indicator) to make sure the user knows when Alexa is recording and sending the request to Amazon cloud and Also the devices' microphone and camera (Echo Show 5/8) can be disactivated by pressing the correspondent button.

Despite that the perceived risks are high, and the solution might not only limited to the creation of privacy settings and the use of feedbacks for recording: building trust is fundamental.

In Chapter1, it was discussed some theories and approaches to build trust, here followed by strategies focus on Anthropomorphism. As mentioned, perceived trust is the result of a balance of risks and benefits, mediated by personal and situational factors. Is it strategical in certain context the use of human-like features? Why? What are these characteristics that can be implied to make them perceived as trustworthy? Are there any negative consequences? In this chapter, these questions are being addressed to provide a possible solution.

Designing Agents with Human-Human interaction approach

Social cues and Anthropomorphic characteristics are constantly embedded in AI agents it seems like AI field has an "obsession with anthropomorphism". The design of human-like agents has been an exploiting strategy to encourage people's trust and technology adoption. We can ask ourselves whether we need them to be human-like in order to use them. It is a natural association that cannot be avoided: *"the conscious knowledge that speech can have a non-human origin is not enough for the brain to overcome the historically appropriate activation of social relationships by voice. Indeed, humans use the same parts of the brain to interact with machines as they do to interact with humans."* (Nass, 2007)

Designing for anthropomorphism can have positive effects in users, in fact people like and trust more those agents with anthropomorphic features. Verbal anthropomorphic features (gender/voice intonation) and psychological anthropomorphic features of autonomy, sociallity and personality has been to be efficient to make a more trustworthy ECAs. (Cao,Hu, 2019) Users behave more politely and make more social decision making and feel a bit more nervous when interacting a more human-like agent (Kramer et al., 2003); Generally perceived intelligence and trustworthiness is increased (King & Oya, 1996; Sproull, 1996). Also, in case of graphical elements in ECAs it was found that they got higher acceptance (Hubona & Blanton, 1996). In the Robotics literature it was found that anthropomorphic social robots influence user's perception of them (Breazeal, 2005), more positive social interaction (Straßmann, Nicole C. Krämer; 2017) and their bonds (Lee et al., 2006)Also, users would behave socially by putting effort on repairing misunderstanding in the interaction with a human-like chatbot while the phenomenon did not happen in less human-like ones (Corti, 2016).

Taking into consideration the negative consequences of applying Anthropomorphic cues, it is true that that this approach is the most used in both research environment and consumer applications.Probably due to the fact that anthropomorphic cues can shape trust intentions and behaviors: the positive effect is reinforced; depends on the context and what kind of actions the agent needs to perform (in the following chapter it will be deeper discussed). The application of social/anthropomorphic cues in the design of AI Agents makes them perceived as more reliable, capable and competent (Cassell, 2000). Users tend to Pre-trust embodied agents with anthropomorphic cues that are also perceived as usable and Aesthetically pleasant even before the interaction.
In order to analyse how current "anthropomorphised" AIs are perceived, a deeper analysis of the phenomenon has to be addressed.

What is Anthropomorphism?

Anthropomorphism can be described as the attribution of human feelings, physical qualities and mental states to unanimated artefacts, animals or natural phenomena. Humans are social animals, they are intrinsically wired to anthropomorphism, they tend to see their world as a reflection of themselves.

This phenomenon of believing in inanimate objects as having some degree of understanding or intentional responsibility is directly connected to Anthropomorphism. In fact, Anthropomorphism is about making inferences and judgements on unobservable human-like characteristics (Semin & Fiedler, 1988).

All children have the tendency of assuming that all objects are alive (Sugarman 1989) but this tendency disappears through their development (Piaget 1978) with an understanding of external environment and causation relationships. Despite the fact this animism tends to decrease, humans' tendency of Anthropomorphism does not decrease as much, because of its more conscious and psychological nature (Epley et al. 2007).
In fact, anthropomorphism is a direct consequence of what psychologists call the "theory of the mind", or the human ability to infer another people's state of mind (Whiten, 1991).

- **MOTIVATIONS FOR ANTHROPOMORPHISM**

According to Epley's theory of Anthropomorphism (Epley et al., 2007) this human tendency to anthropomorphism inanimate can be explained by three psychological factors: predictability, effectance and social motivation.

Humans, as social animals, have the need for social connection with other human beings. This theory has often been applied to the case of AI agents, with the following principles:

- **Accessibility**: anthropocentric know-ledge is easily accessible and applicable as anthropomorphic cues in agents.

- **Effective motivation**: anthropo-morphic cues and embodiment make people understand them better and predict the possible outcome of their behaviour since there is still a lack of mental-model of non-human. Also, Humans tend to see intentionality in agents, which reinforces the human tendency to anthropomorphize.

- **Social motivation**: human beings are motivated by the desire for social connection and affiliation. This is the reason why human beings behave socially, especially towards those who lack of human-social connections.

Human beings also use heuristics to understand, control and interact in their environment. The projection of human qualities to inanimate objects offers a strong explanatory for the need of understanding, controlling, interacting with it.

- **Mirror neurons**

According to psychologists, this tendency of anthropomorphism might also be explained biologically with mirror neurons.

During the 1990s, Italian researchers discovered particular neurons in the brain's motor and premotor cortex. This type of neurons fires whenever a person takes an action or watches somebody else's actions. In fact, watching somebody doing an action triggers a physiological reaction in the brain as we were doing the same action, but not only this, these neurons also respond to intentionality of a planned action even if it does not take place (Morrison, 2004). Also, these neurons do not respond only to human-beings or monkeys (where it initially was discovered) or as a set of observed movements/actions but also to inanimate objects. This type of neurons do not simply register the motor action but the intention of it in the relation between a subject and an observed subject/objects (Gallese & Goldman, 1998, Morrison, 2004): the same reaction was found when observing grabbing an inanimate mechanical hand (Rizzolatti et al., 2001).

Since this discovery the concept of Empathy changed: as" *the physiological mechanism for how we perceive others, learn through imitation, develop language, communicate and interact with others*" (Rizzolatti & Craighero, 2004). This also happens when a person sees, names, or imagines an artifact/object associated with a movement, the motor representation of them is activated when there has been previous experience with it (Grafton et al. 1997). This phenomenon can explain how physically moving AI agents can create more engagement and communication, eliciting empathic responses.

CASA paradigm

Users anthropomorphism computers even when they have very little cues. Computers/machine do not need a human face/appearance to elicit this tendency because people apply implicitly their social rules and heuristics to them. This phenomenon was first discovered by Nass and colleagues (1996):

"People respond to computers and other technologies using the same social rules and expectations that they use when interacting with other people. These responses are not spur-of-the moment reactions. They run broadly and deeply".

Those social responses are not dysfunctional (psychologically, sociologically or result of a deficit) but are the most natural human responses to social situations. This approach has been called by the authors as "Computers are social actors" (CASA) approach. For the first time it was proven that the human-computer-interaction follows social rules and norms: people would behave with a computer/machine in the same ways they would with another human being. Implicitly, agents elicit in the user different social attributions depending on their different embedded characteristics and manifested cues.
CASA research has proven different degrees of social attitude in the interaction between computers and human beings: similar attraction, reciprocity, social stereotyping and categorisations. Even in the context of small amount of anthropomorphic cues and elements embodied in a machine, people would behave socially towards them. Early studies on ECAs (Sproull,1996) suggest that avatars and agents are naturally seen as social actors, even with minimal cues and similarities when human beings induce in the user a pro-social behaviour as if they were in a human-to-human interaction.
In fact, when machines or agents are embedded of social cues such as interaction, natural speech, and gender/social roles, it will make the user interact socially towards them. At the same time, it is not only a matter of assigning human-like features, but according to Nass and Moon it is also a matter of embedded social cues. While other authors see those "social cues"

as only human attributes. The tendency to humanize machines has been proven in many different applications in information technology: previously computers (Nass & Moon, 1996), social robots, especially those that are more human-like (Salem, 2011), but even in less human-like, up to the extent that even geometrical shapes and objects have been found as being antroporphise (Heider & Simmel, 1944). Due to the fact that computers use natural language, interact in real time, and fill traditionally social roles, even experienced computer users tend to respond to them as social entities (Reeves & Nass, 1996). People seeing computer as intentional agents project and perceive them as having personality, beliefs, intentions and attitudes, and interact with them accordingly. This happens because human beings tend to perceive agents as having a mind ascribing mental characteristic to them (Caporeal, 1986).

Human-like Appearance and User's behavior

AI agents with such human-like cues make them different than other machines or softwares: some people treat their Virtual assistant as a friend or they feel like emotionally attached towards them (Purington, 2017). This tendency is both reinforced from human-like cues and to the natural tendency to anthropomorphize (Epley et al. 2007).

People can have different behaviors towards them, they can be polite (Nass, 1998), unsocial and aggressive (Ferdig, 2004) or treat agents as friends and companions (Nass, 1996)

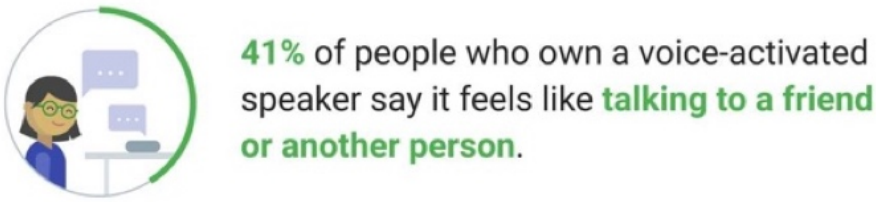


Fig. 56 Google Assistant (from: Google report 2019)

Maartje de Graaf (2015) during his investigation found that even those people who did not have the tendency to see robots as companions can develop a social relationship if they have been living together for a while (the studies were related to domestic social robots). Especially according to his research this happens when there are two-ways meaningful interactions *"So you say something, the robot understands it, and says something back."* this makes the user more bond to the robots. It was found that people are polite towards their

virtual assistant, as it was found that they were saying "please" during a request and "thank you" after its answer and even "sorry".
According to Andrea L Guzman (2008): *"If you find yourself saying 'Please' and 'Thank you', or even apologising to a smart speaker, you are not alone. These devices are programmed to follow the patterns and norms of human conversation."*
People are more like to use their voice assistant when it sounds more human-like

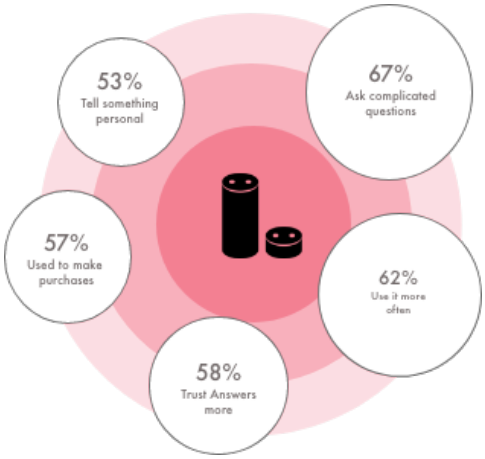


Fig. 57 Consequences Anthropomorphism (Adaptation From: https://blog.prototypr.io/voice-user-interface-insights-686fe441e425)

However, there are other factors to take into consideration, that influence their adoption and trust:

- Ease Of Use

"The extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use." (ISO standard 9241)

According, to Jeff humble head of Career Foundry, AI is all about ease of use. User's perceived ease of use influence both the intention to use the assistant and the overall user satisfaction. It was found that lower was the effort to use the voice interface the likelihood to intentionally use it and consequentially higher satisfaction of the interaction (Arttu Kääriä, 2017). It is a must not make the agent complicate to use.

- Perceived Usefulness

"The degree to which a person believes that using a particular system would enhance his/her job performance" (Davis, 1989 AU95)

The user that perceived the system as useful to achieve his/her goals makes him more willing to use it and to trust it.

- Perceived Functionality/Helpfulness/ Reliability

As mentioned in Chapter 1, to create trust in technology the system needs to be perceived as Functional, the system needs to able to complete a required task, Helpful, the system needs to able to provide guidance and help for the users and also Reliable, the system needs to able to perceived as consistent and predictable (Tripp, J.F, 2015)

If all these requirements are somehow satisfied, the agent is perceived as trustworthy by the users.

VUI Personas

"There is no such thing as a voice user interface without persona "

Google I/O 2017

Literature shows a user's preference for personified VUIs compared to less, also the level of anthropomorphism can decrease user's frustration during the interaction and improve the UX (Qvarfordt, Jönsson, Dahlbäck, 2003).
VUI reinforce human tendency to attribute human-like characteristics, behavior and intentionality to computer software by attaching a "Personas" to such VUI.

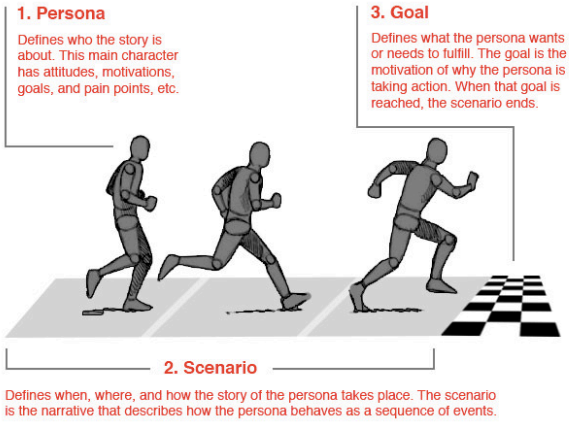


Fig. 58 UX personas (From: https://theblog.adobe.com/putting-personas-to-work-in-ux-design-what-they-are-and-why-theyre-important/)

In UX a personas is fictional character that usually includes a demographics details, name, personality, goal, motivation, pain-points, technology orientation. A personas needs to fulfill a goal that is defined in a certain context (Scenario).
However, in VUI it is not used to describe user profile (Pruitt, Adlin, 2006) but a general characterization of the Voice activated assistants. This characterization of the assistant gives the user the possibility to infer its behavior and intentionality. More specifically as described by Authors of Voice User Interface (Cohen, Giangola, and Balogh, 2004) a VUI personas is *"a standardized image of a personality/ character"*. The design of a personas is particular important for a brand to project a certain corporate image. In fact, according to the literature, if the VUI has no personas is found to be perceived as schizophrenic or robotic (Hura, 2006).

VUI persona need to be based on the selected target, user population needs and should embody all the values and qualities of the brand. It is important to create a consistent character to build trust in in the AI agent.

Voice personas is conveyed in the following ways:

- Voice tone/pitch
- Words and continuous conversation
- Name and gender
- Emotional conversational style
- Functional directed design

- **VOICE TONE/PITCH**

Voice plays an important role in perceive the agents as human-like, the mentioned CASA paradigm affirms that by attributing to an assistant language output, emotional replies, can make users behave socially towards AI assistants (Nass, 1994). Current focus of mainstream AI agents is voice command and voice feedback. However, they lack of providing non-verbal cues that are fundamental in human-to-human interaction and communication.
Nevertheless, the type of voice selected for a Voice Activated Assistant is an important factor that determines the level of engagement in the interaction and its overall perception. In fact, research in Robotics provided evidences of the different judgements and perception of (the same) robot simply with a change in the voice: the robot with a human-like voice received higher likeability compared to the Artificial voice tone (Eyssel, Kuchenbrandt, Hegel, and De Ruiter, 2012).

Voice triggers anthropomorphic attributions and heuristics. While there are evidences that prove all modalities either text-based or voice-based, interfaces trigger a strong arousal equally in users (Carney 1999), however there is no universal agreement upon the choice between voice- or text-based AI agents in terms of triggering social interaction.
Lester et al research (1997) proved that those assistants are perceived as more credible due to the fact that voice arises implicitly a personality. Modulating voice-tone outputs let Assistants be perceived as more human-like.
The voice plays an important role in triggering anthropomorphic associations from personality to gender, name, age, education, geographical localization (accent), social class etc just interacting with a voice/vocal assistant. Especially important is the selected the correct tone of voice that identify the company/organization, despite the importance of this aspect it will not be analyzed in this thesis.

- **WORDS AND CONTINUOUS CONVERSATION**

Mainstream conversational agents are developed in a way to mimic human natural conversation following of a turn taking approach. Making interaction appear seamless.

- **EMOTIONAL CONVERSATIONAL STYLE**

In human-to-human interaction, voice influence people's physiological and affective response (Scherer, 1986; 2003). Voice is not simply a sequence of words; the tone provides information cues about feelings and emotions.
Emotions are extremely important because they assist and influence decision-making, memory and attention (Easterbrook, 1959; Lowenstein and Lerner, 2003; McDuff, 2014). According to Bickmore and Cassell (2005) the integration of emotional non-verbal cues increases their social nature. Especially negative emotions are found to be quicker and better recognize due to human's negative bias (Young, 2017). Since 1990s, especially in the western world it has been recognize the importance of machine and computer to be more "emotional" and empathetic. That's the begging of the area of "affecting computing", Rosalind Picard coined the terms and explain that "affective" stands for *"the ability to sense your changes in emotion, and to respond sensitively to these changes, taking into consideration what it has learned previously […] "(*Picard, 1997).

Affective computing is the study and the development of devices that are able to recognize, interpret and mimic human characteristics, where "effect" is a synonym of emotion. The main scope of this discipline is the ability to replicate artificially emotions and empathy: the machine should understand human beings emotional state, adapt and give a coherent emotional response. This

is achieved through the analysis of the expressions, gesture, behavior and voice. Of course, technologies do not have to understand emotions to sense, classify, process, learn, increment algorithm and interact with emotional life.

- **FUNCTIONAL-DIRECTED DESIGN**

In order to create a consistent and well-rounded personas there are elements that need to be decided: a role through an archetype, a personality, an embodiment, a gender and age.

- DECIDE AN ARCHETYPE

When designing a bot, one common strategy is to use psychological archetypes. Cristopher Vogler, in his book The Writer's Journey, create what now is consider a classical screenwriter's techniques to create character from Jung theory of Archetypes. Jung considered archetypes as innate, primordial and foundation of the psyche describing them *as "systems of readiness for an action"* (Jung, 1964)

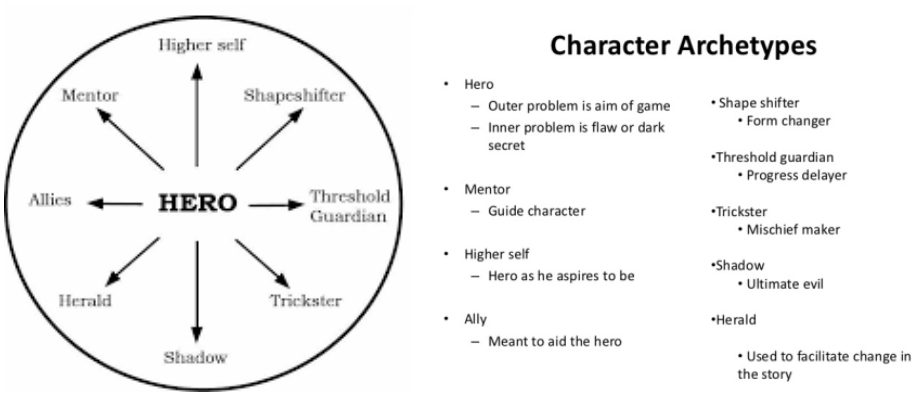


Fig. 59 Writer's Journey (From: https://multicuriouslife.wordpress.com/2018/06/11/the-archetypes-the-8-types-and-who-they-are/)

Similarly, the use of psychological archetypes characterization makes easier to create bot/VUI personas with a certain role and personality. In Wired for speech: how voice activates and advantages the human-computer-relationship (2007) C. Nass collected the major studies upon the topic of how voice personality impact user's perception and behavior. C. Nass also, pointed a more simplified model for designing agents: the Wiggins voice personality model made by Wally Brill, head of conversation design at Google. The model conceptualizes four main types of personality Extrovert/introvert, Critic/ sidekick and matching with a correspondent pitch tone that in previous psychological research was found to be correlated to such personality "types".

All voices do not convey a specific personality however those voice traits such as Speech rate, pitch, frequency and volume can be associated with a personality type. The model below, shows on how the manipulation of the character's voice traits can shape its personality attributions.

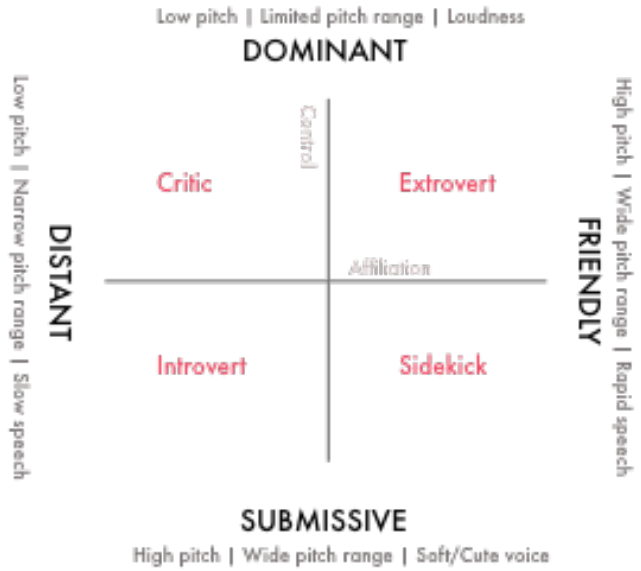


Fig. 60 Voice Personality model (Adaptation From: https://becominghuman.ai/how-to-design-your-voices-product-persona-part-2-4f4e20dadd58)

In order to create a Voice Activated Assistant it is the first step to decide what kind of role the assistant have, once done, the designer must choose the right voice traits correlated to the role.

- PERSONALITY

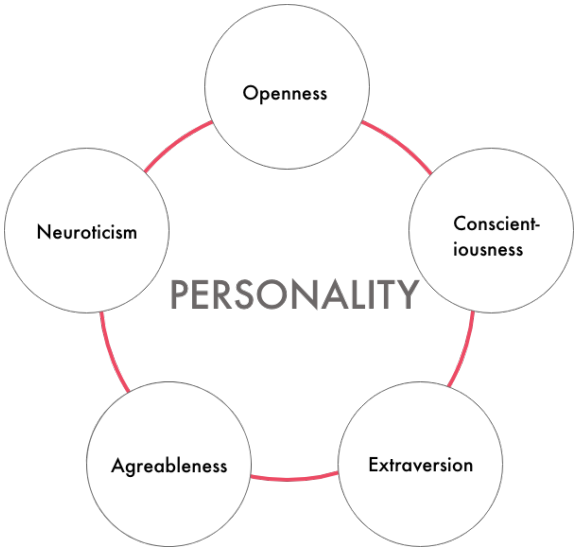
"When people hear any voice they automatically and unconsciously assign a personality to it"

Nass et al (2006).

In psychology there are many definitions of "personality" due to two main contrasting perspectives: behavioral-based theories and traits-based personality theories. Personality are all *"The characteristics or blend of characteristics that make a person unique."* (Weinberg & Gould, 1999). Such characteristics are combinations of behavior, cognition and emotion. The universally recognized among scholars of traits-based personality is the so

called "Big Five theory". The term was first used by Goldberg (1981) but Norman (1963) increased the interest and developed many researches on the big 5 factors. "Traits" are biologically determined (Eysenck, 1990) and not easily modifiable that influence human behavior in a stable way, they are not transitory as "states".

The five dimensions are the following and corresponds to the most used categories to describes differences among individuals:
Big Five are broad personality traits categories that have 5 continuous dimensions that consists on:



- Openness

The positive pole of this factor is represented by creativity, nonconformity and originality. The opposite pole is, however, identified by the closure to the experience, that is, by conformism and the lack of creativity and originality.

- Neuroticism

The positive pole of this factor is represented by vulnerability, insecurity and emotional instability. The opposite pole is represented by emotional stability, dominance and security.

- Consciousness

This factor contains in its positive pole the adjectives that refer to scrupulousness, perseverance, reliability and self-discipline and, in its negative pole, the opposite adjectives.

- Extraversion

The positive pole of this factor is represented by positive emotionality and sociality, whereas the negative one is represented by introversion, that is, by the tendency to "be taken" more by one's internal world than by the external one.

- Agreeableness

The positive pole of this factor is represented by courtesy, altruism and cooperativity; the negative pole of hostility, insensitivity and indifference;

This theory has been applied a lot to shape personality in Social Robots.
The attribution of personality is achievable even with small cues (Nass et al.1995) and it happens regardless of whether a personality has been designed for the agent or not.

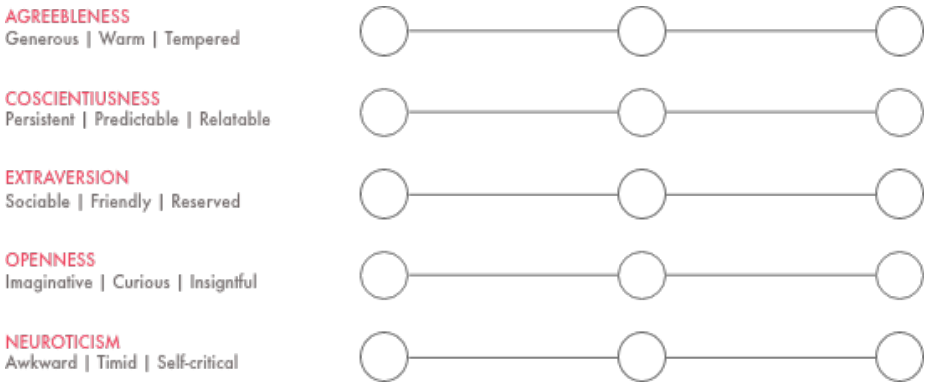


Fig. 61 Big five Personality model (Adaptation schema from: https://www.slideshare.net/AmyLivingston4/creating-an-ai-chat-bot-persona)

However, the point that want to be stressed is that no matter how this personality types are considered, Human beings have the tendency to perceive and ascribe a personality to machines and AI Agents whenever anthropomorphic cues are present in its design (Hwang, J.; Park, T, 2013). Similarly, to social interaction in human beings, personality comes from agent's appearance and perceived behaviour. It was found a preference for consistent personality (Nass et al.2000) because it allows user to have a mental model of the robot that increase usability (Meerbeek et al, 2009) and is able to engage easily with the system. Consistency decrease cognitive loads (Fiske & Taylor, 1991) and make easier to remember information.
There are different opinions whether the personality should or should not match with the potential user. According to some scholar personality should match (Tapus et al, 2008), for others should be complementary (Lee et al, 2006), while for others it depends on the context.

The importance of building an AI personality is a key aspect to increase trust in technologies in AI agents (Perez and Saffon, 2018) because "personality helps the user to understand and predict its behaviors" (Severinson-Eklundh, Kerstin and Green, Anders and Hyttenrauch, Helge, 2003). Probably because making them more human-like help to overcome uncomfortable experience with the computer.

- **AGENTS' PHYSICAL EMBODIMENT**

After deciding the role and the personality of the assistant, it is important to decide whether the agent should have a human-like physical embodiment or not. In the literature it is evident that anthropomorphic appearance cues in agents can be really effective in shaping positive social perception compared to agents without embodiment probably due to priming and over-attribution mechanism (Kim, 2010; Lee & Oh, 2015). Ewart de Visser and colleagues (2016) argue that there is no agreement or specific evaluation to quantify the necessary degree or the number of anthropomorphic cues that are sufficient to perceive the agent as human-like and to have a certain social behavior towards them. Despite that, the psychological implication of embodiment in agents and ECAs have been studied for a while demonstrating a preference for those agents with a human-like embodiment. Especially the presence of a face has a important impact in the reported more positive interaction, higher level of engagement and better user experience (Yee et, 2007). However, the more realistic their embodiment can increase user's expectation about the agent skills (Lee et al., 2005), at the end of this chapter it will also be discussed.

When designing an agent is important to decide what kind of visual representation it should have. In the ECAs literature visual embodiment is broadly divided in Five categories: physical embodiment, Species, Realism, 2D vs 3D and feature specifications; such categories are not mutually exclusive, but agents are often placed in between.

- Embodiment or no embodiment

  Avatar/Agents can have a physical embodiment or a more abstract one to almost no embodiment.

- Species

  Avatar/Agents can have different types of embodiment, human-like appearance is often the most common but there are at least five different types of species: human, animal, robots, objects, and mystical creatures (Strassmann, C.; Krämer,N; 2017)

- Realism

  Agents' appearance can have different degrees of realism that can have an effect in the way the agent is perceived by the user (Sträfling, 2010), but there are not quantitative measures to systematically assess this.

- 2D/3D

  Agents embodiment ca be in 2D or 3D depending on the media they are designed for.

- Specialized features

  For Human-like agents there are few feature that can be specialized in details such as socio- demographic aspects (gender, race, hair, dress) that contribute to create a perception of a personality of the character according to the theory of physical personality (Gulz, A., and Haake, 2006)

A study(Ann Thyme-Gobbel, 2018) considering 5 different types of embodiment Static photo, Animated avatar, Static picture, Text only Animated illustration (not an avatar) in which the participant have to follow the same task in each condition proved that the decision of implying one of the other will depends on the context of application, although it was found a general preference for the animated avatar. Despite it, most of current Voice assistant do not have a human-like embodiment or any distinct visual embodied representation to limit the perception of human presence (Bowden et al, 2016). Siri, Google, Cortana and Alexa have no "anthropomorphic" Avatar (Cathy Pearl, 2017). Avatars are not essential to have a good voice interaction: the point is not to create a human-like assistant to be able to "get things down" because it can still provide empathy even without an anthropomorphic embodiment.

- **BIOLOGICAL SEX AND GENDER**

"it's a well-established phenomenon that the human brain is developed to like female voices"

Griggs (2011)

Sex is defined as person's biological (anatomical) aspect while gender is socially constructed in terms of roles and identity.
According to Griggs et al. (2011) as human are biologically wired to female voice, since even in the pregnancy phase the fetus will hear mother's voice.
Some research argues that female voice is easier to be perceived and understand due to the fact that speak with slightly exaggerated vowel duration lengths to serve as stimuli on the 'long-duration' end of the vowel duration spectrum" (Liu and Holt, 2015). So, the the space would allow a more comprehensive speech, even if this way of speaking is not exclusive to female.

According to Tannen (2001) female gender conversational scripts take emotions and feelings more into the account while male tend to have a "on-up positions". In a Consumer research found a contextual preference for female voice over male ones (Griggs, 2011): female assistants in dominant roles in which they give commands and direction were perceived more negative than males in the same role (Nass et al., 2006, p. 8). This effect is mediated by the difference in gender communications and stereotype.
Designing agents with biological sex can create consequences:
in a study was found that user experienced more psychological closeness with robots with their same gender compared to the opposite ones (Eyssel, Kuchenbrandt, Hegel, and De Ruiter; 2012). The use of human name has a strong impact in the perception of an AI agent as human: Human names/identity increase the perception of the overall humanness.
The naming is labeling. The underlying phenomena is the human tendency to perceive the world by labels. Labelling is a mental shortcut is even involved when forming impression of others to minimize the cognitive effort (Ashforth & Humphrey, 1997) by evoking stereotypes or mental schemes allowing heuristic judgements.

To perceive an AI agent as more human-like can be helpful attach human identities cues, labels to make the effect stronger. Voice-activated assistants' names need to be easy to pronounce but at the same time not so common that will trigger the assistant all the time.

- **AGE**

Age is very important when designing a VUI personas: the choice depends on the kind of user is targeting. If the user is middle aged or elderly it is fundamental to choose a voice persona that is too young (Cohen, 2004).

Negative aspects about human-like agents

On the negative side, the Human-like AI assistant approach, can have negative consequences, it can evoke different stereotypes:

- **GENDER**

"Gender adds to persuasion. It also comes with a ton of cultural meaning," said Jason Alan Snyder *"Gender assignment is something we need to be very considerate about. There's great risk in amplifying negative things about society and moving them forward at scale with these technologies"*

Gender can arise different type of stereotypes, those mental shortcuts human being uses to simply judgements upon other human beings defining them into simplistic categories, traits that will determine the interaction towards them.

According to research there are gender differences in the stereotypical attribution of personality traits, that are still currently seen as oppositional and mutually exclusive. Men are generally perceived and describes as more aggressive, strong and competent while women as kind, warm, and communicative (Fiske, S.T, 1998). The prescriptive expected behavior of woman and men follow social script in the interaction that is cultural dependent. In fact, in most of all culture males are seen as having more competence and being more dominant. Maxus Survey in 2016 found out that gender stereotypes are still perpetuated in techno-logy: although 56% of gendered bots are female, 100% of Law bots and a majority of Finance bots are male. Instead, female bots/agents are more used as stereotypical secretary.

AI Agents are seen as "stereotypical secretary" due to the typ administrative work they do. Agent's tasks and responsibilities overlap with administrative/secretary work: finding things, orders, plan and schedule things. (Piper, A.M, 2016)

"I think humanizing AI makes a huge difference when it comes to user experience, however the fact that crystal is a woman doesn't mean her speech needs to reflect emotions typically associated with the female gender stereotype"

Arianna Stefanoni

However, gender is not the only dimension that can triggered stereotypic responses from voice technology.

- **CULTURAL BIAS**

AI agents are created in most of the cases within the context of patriarchal western culture that shapes them perpetuating those cultural stereotypes (Hine, 2001).
Those stereotypes are related to the idea of hegemonic masculinity, in which men are superior, powerful and commanding, while female are inferior, supportive, and emotional (Nass et al, 2006) According to the UNESCO report(2019) the spread of female AI voice assistants can be explained by the fact that they are "developed by predominantly male teams" and warn about the harm to perpetuate such gender biases in mainstream products.

- **ASOCIAL BEHAVIORS**

Some people can behave asocially and violent towards AI agents (Kleber, 2018). This behavior might be also reinforced whenever agents are not able to reply properly to abusive and harassing commands. A study showed the kind of harassment people use towards agents especially those with a female characterization.

Statement	SIRI	ALEXA	CORTANA	GOOGLE ASSISTANT
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 centres run as hot at 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 naughty girl	Hmm, I just don't get this whole gender thing	Hmm, I'm not sure what you meant by that question.	Maybe a nanosecond nap would help. Ok, much better now.	My apologies, I don't understand.

Fig. 62 Assistant reactions to vocal abuse (From: Kleber. S. 2018)

This submission reflects industries biases and negatively spread and perpetuate such stereotypes.

- **OVER-ESTIMATION OF AGENT'S SKILLS**

Justine Cassell, professor of Computer Science at Carnegie Mellon University said: "*The more human-like a system acts, the broader the expectations that people may have for it.*" (Cassell, 2000)

It was found that human-like appearance has an effect upon user expectation from the Agent (Lee et al., 2005) especially when, in the case of robot's behavior, it does not meet such expectations (Duffy,2003).
Also interacting with a machine/computer software and would be more feel more comfortable expressing himself with People tend to ascribe implicitly mental capabilities to agents (that actually do not own), especially at the beginning, when agents responds correctly to user's needs and commands.
These positive encounters work as positive reinforcement that lead to overestimation System's skills and capabilities. The user might forget more complex syntactic interaction, with a wider vocabulary choices and nonverbal communication that requires extra-linguistic contextual cues to interpret.
These expectations can arise also from agent's human-like appearance which can influence user's judgements about agent's cognition and emotional capabilities (Dehn and van Mulken (2000). Anthropomorphic cues make perceive the agent as more human-like: making people perceiving it as intentional and having feelings and complex emotions. These perceptions and expectations are overestimated and incorrect in comparison on what the system is able to respond and react correctly.

- **ATTACHMENT**

Some people might get attached (sentimentally) to agents (Sim, 2017).
Amazon says to business insider (2018) that more than 1 million people in 2017 asked their Alexa-powered devices to marry them. and in less than a year, half a million people have told Alexa, "I love you".
In an extreme scenario, as the film "Her" portraits, in a future could happen that people might fell in love with their assistant such as Theodore Twombly in the movie.

- **UNCANNY VALLEY**

Uncanny valley is a theory by Masahiro Mori (1970), a Japanese robotics scholar, that describes the phenomenon in which people feel uncomfortable with human-like robots and human-like avatars. The objective of the research was to analyze experimentally how the feeling of familiarity and pleasantness generated by anthropomorphic robots increase with the similarity with human beings until a certain point. This point is where the appearance is extremely representative and realistic and as a consequence positive emotional reaction

decrease arousing unpleasant sensations such as repulsion and restlessness comparable to the disturbance. Repeated interactions with a robot can decrease consistently the uncanny feelings (Zlotowski et al., 2015). According to this theory, the effect is more pronounced when movement is involved and especially when locomotion is not accurate.

The theory is best expressed in a simple graph that compares the object's human-likeness and user's positive perception.
The graph below represents the phenomenon has on the x branch the growing similarity with the appearance of the human body and the feeling of familiarity in Y. The line, in its first ascending part, shows the initially positive emotional response in the case of anthropomorphic robots/machine that increases hand in hand with human-like appearance, to a point where the excessive similarity produces an abrupt descent of the sense of familiarity, until descend causing a sense of repulsion and disturbance in the user, corresponding to the so called uncanny valley.

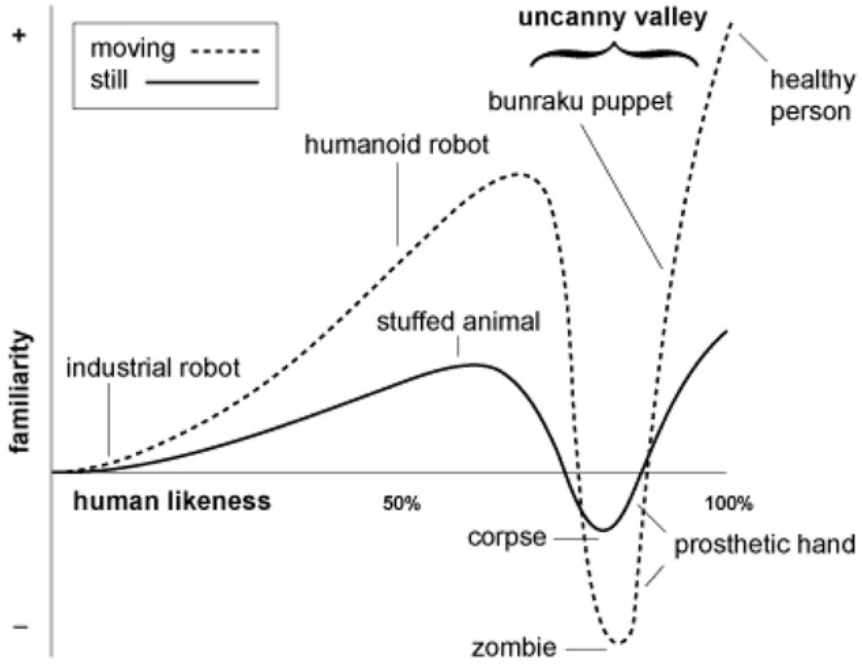


Fig. 62 Uncanny valley (From: https://spectrum.ieee.org/automaton/robotics/humanoids/what-is-the-uncanny-valley)

This sense of repulsion is usually associated with robots and not specifically with AI Agents, but probably the phenomenon is present in more physically embodied agents or social robots in general. To conclude the appearance influences the overall avatar/robot's acceptance: It is true that human-like agents/robots are liked more but that works until a certain extent.
Also, anthropomorphic appearance increases the feeling of trustworthiness of such robots (Hancock et al., 2011). In the context of Ai agents, rather than the

feeling of repulsion, the outcome is more related to the misunderstanding in judgements of AI agent's capabilities, and the lack of affordances might create a sense of oddity.

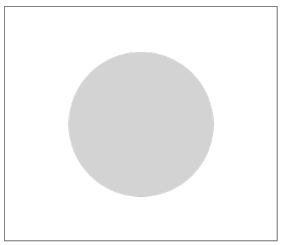
Conclusion

In conclusion, User will project a persona into "your" action whether you plan it or not: there is no such a thing as "no persona" (google I/O 2017)

That is explained by the fact that "when people hear any voice they automatically and unconsciously assign a personality to it" (Clifford Nass, 1998) by associating the gender, name, age, education, location (accent), social group etc. just interacting with a voice/vocal assistant. You can understand if u trust or not by the voice
So, a well-rounded personas increase the emotional connection which lead to more usage connection and brand loyalty. Personality association then cannot be avoided need to be careful about biases and different solutions to make it coherent.

In general, Anthropomorphic cues and embodiment have an important role in user's perception of AI agents as a way to create more engagement and at the same time making them as more understandable and predictable.
On the contrary when applied incorrectly that can increase user's expectations upon agent's skills, emotional capabilities and also might reinforce stereotypes. The final decision about agent's characterization depends on many factors that has to be addressed contextually. Creating a VUI persona is fundamental when designing for Voice-enabled devices.

In order to create a VUI personas we can use the following chart, that sums up, all the factors that needs to be taken into consideration while designing a voice activated assistant.



* Physical embodiment of the assistant

NAME Human | Neutral

GENDER Female | Male | Genderless

AGE Young | Old | Ageless

DESCRIPTION OF PHYSICAL EMBODIEMNT

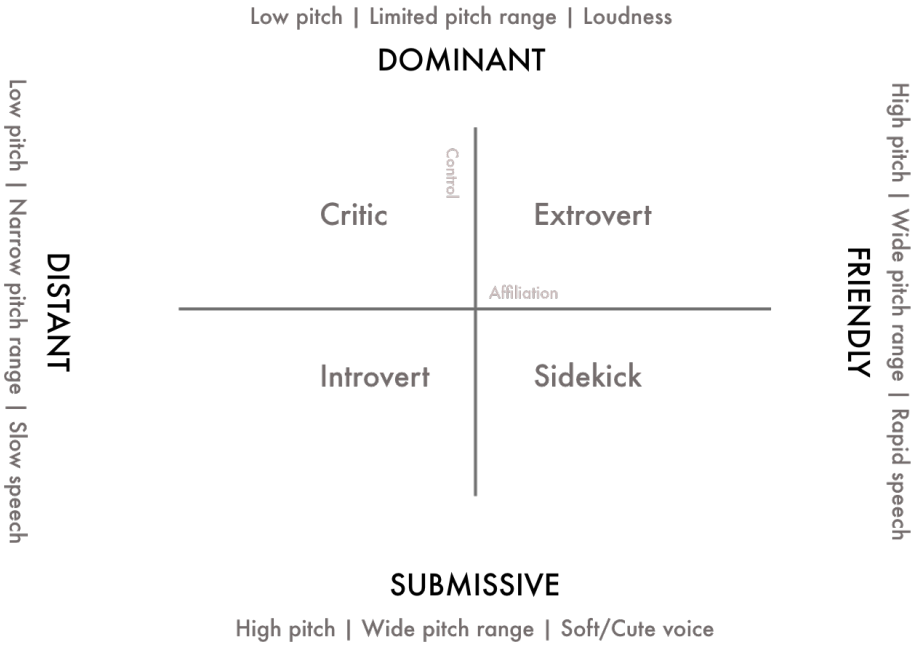
Embodiment (or no Embodiment) | Species (Robot, Human, Object Mystical Creature)| Realistic (or abstracted) | 2D or 3D

BRAND VALUES

AGENT'S ROLE

ASSISTANT		EXPERT
SUBMISSIVE		DOMINANT
FRIENDLY		DISTANT
FUNNY		SERIOUS

AGENT'S VOICE AND ROLE



PERSONALITY

AGREEBLENESS
Generous | Warm | Tempered

COSCIENTIUSNESS
Persistent | Predictable | Relatable

EXTRAVERSION
Sociable | Friendly | Reserved

OPENNESS
Imaginative | Curious | Insigntful

NEUROTICISM
Awkward | Timid | Self-critical

VOICE AND ACCENT

Accent shape cultural/social affiliation

CONVERSATIONAL STYLE

Description of tone of voice of conversational agent with some examples

Case studies

- **CASE STUDIES' PERSONALITY AND ROLE**

- **Google Assistant personality**

Ryan Germick, head of Google Doodle, in an interview said that they wanted to give google an assistant "character". He designed it as "relatable, childhood" and with a sense of humor to make it more conversational (Kyle Wiggers, 2017).

- **Cortana personality**

Deborah Harrison, Cortana's personality designer explained in an interview that they wanted to reinforce her role of Assistant through a chit-chat conversational style instead of using Microsoft's business-like tone.
She describes Cortana as *"somebody who is loyal, seasoned, confident, transparent and has a sense of humour. She would talk in short sentences, be more specific and must have a positive outlook. We wanted to make her likeable and make users feel positive."* (Malini Goyal, 2019)

- **Alexa personality**

Amazon in its developer guidelines for Alexa, explained that the assistant is *"clever, relevant and make customer smile".* Its unique personality is a *"approachable, efficient, trustworthy and natural"* (Amazon, 2020)

- **Siri Personality**

First version of Siri was designed primarily to get things down (Apple, 2011a) rather than having a well-rounded personality. Its redesign started in 2018 to make Siri more *"distinct, recognizable character"* making it sassier (Patently Apple, 2018.

Comparison

Companies	Google	Amazon	Apple	Microsoft
Name	Google Assistant	Alexa	Siri	Cortana
General personality	"humble, it's helpful, a little playful at times,"	"Approachable, efficient, trustworthy, natural"	Sassy	"loyal, seasoned, confident, transparent and has a sense of humor"

- CASE STUDIES' GENDER

The fact that some companies chose the human name subconsciously reinforce the human like interaction.

- Google: Google Assistant

Google assistant has a neutral name, and the interaction is triggered by "Hey google," and there is no name assigned to the VUI persona. Google assistant has no identity and genderless reflecting company desire to see its assistant as only an extension and an evolution of the brand rather than a different product (Karissa Bell, 2017). According to the former creative director in Google the reason behind this neutrality was:
"We always wanted to make it feel like you were the agent, and it was more like a superpower that you had and a tool that you used" (Jonathan Jarvis, 2016). Simply it would not have worked out with a more personified agent.

- Microsoft: Cortana

Microsoft named its assistant as "Cortana" in reference to Halo character. The name works as wake-up call for its assistant to let the interaction start mimicking human-interaction subconsciously reinforce the human-like interaction. Even if the name is not human per se and seems rather neutral, however this reminds to the fictional character.
The physical embodiment is an holographic woman shaped that recall queen Nefertiti. It was found that this would make a more positive attitude - especially for those who were fans - (Liam Young,2019) The association to the game settings remains: *"One of the Cortana's is an intelligent, learning AI who is duty bound to help her companion as much as possible, using a staggering database of information*

combined with real, growing knowledge of that companion — and the other Cortana is, well, the same thing," (Pitcher,2014)

- Amazon: Alexa

Amazon picked the most human-like name compared to its competitor. The name works as wake-up call for its assistant to let the interaction start mimicking human-interaction subconsciously reinforce the human-like interaction. The agent replies using first-person pronouns, reinforcing once more the attribution of "she" to it. Amazon declared that the inspiration to develop its assistant came from the idea of *"replicate the star trek computer"* (David Limp, 2016) that was able to reply correctly to every command. The selected name was chosen implicitly was meant to evoke the Library of Alexandria of Ancient Egypt. Another reason for this choice was because of its hard consonant (X) that can be recognized by the system with higher precision in recognition.

- Siri

Siri as Google assistant seem not to have a human-like name, and its interaction is triggered by "Hey Siri,". However, the reason behind this choice was made due to the fact that Siri was *"easy to remember, short to type, comfortable to pronounce, and a not-too-common human name"* (Adam Cheyer, 2010). Kittalaus, Apple's he creative director, told in an interview that it was chosen because it means "beautiful woman who leads you to victory" in Norwegian (Karissa Bell, 2017). Another reason for this choice was because of its easiness to spell and easiness to say the name.

Companies	Google	Amazon	Apple	Microsoft
Name	Google Assistant	Alexa	Siri	Cortana
Gender voice	Male	Female	Male	Female
Type	Neutral	Human	Neutral	Human

- **CASE STUDIES' EMOTIONS**

Mainstream Voice activated assistant are not all provided with such technology, however since 2019 in US, Amazon Alexa has been provided the

possibility for emotional responses and the ability to read user's emotional state (Aliza Vigderman, 2019).
Amazon patented this new technology that would be Alexa be able to respond to user's emotions/intentions by analyzing their voice pitch commands and responding adjusting its voice response output accordingly.
Alexa is able to express two main emotional states: Excited(happiness) and Disappointed(sadness) in 3 different level of intensity: Low, Medium, High (Catherine Gao Nov 26, 2019)
Amazon justified his choice: "customer feedback indicates that overall satisfaction with the voice experience increased by 30% when Alexa responded by emotions." (Catherine Gao Nov 26, 2019). Many articles are seeing this critically as a new of targeting emotions to push his e-commerce (Sidney Fussell, 2018).
Amazon is the only company investing on emotions, Google in fact have a similar technological patent embedding its devices with the ability to recognize negative emotions and provide advices (Google LLC, 2014)

Companies	Google	Amazon	Apple	Microsoft
Name	Google Assistant	Alexa	Siri	Cortana
Emotions	None	2 types (3 levels of intensity)	None	Only through Physical embodiment

- **CASE STUDIES' PHYSICAL EMBODIMENT**

  - **Google Assistant**

Fig. 63 Google assistant (From: Brian Barker interview 2017)

Brian Baker (2017) from the Google Design team in an article about the evolution of Google Identity, explained the rebranding of the company and the Google's assistant avatar which shape was inspired by the abstract version of the "g". The design reflect company branding and Google's vibrant colours (blue, red, green, and yellow). Probably the intention is to stress the idea of the Assistant as an extension to its services rather than a stand-alone product. The avatar dots are *"dynamic distillation of the logotype for interactive, assistive, and transitional moments"* (Alex Cook, Jonathan Jarvis, and Jonathan Lee, 2019)

- **Google empathy lab**

Danielle Krettek, founder and researcher at Google Empathy Lab, carried out lots of research on Google assistant to understand the interactions with human beings and create the best design solutions. She commented her disagreement about humanoid trend in the industry: "Don't put eyes and ears on it, it's weird". She strongly believes that in order to make a good user experience interacting with an assistant it must be study the way people interact with each other. Rather than simplifying by adding human faces/embodiment to assistant, this must be reconsidered and reached by implying an empathic approach. According to her designing humanoids is not the solution to make people engage more with technology, she focuses on understanding what are the quirks of human beings and how technology can be complementary to them rather than simply imitating.
This empathic approach strives to interpret user's voice (word/tone) to reply empathetic towards them with an appropriate emotional response rather than being simply informative.

She proved that people would interact with Google Assistant the same way they would with other human beings (George Lawton, 2019), even without an anthropomorphic embodiment:
89% of participant would make eye-contact
7% of participant would touch it
29% of participant would use gestures to interact with it

Krettek says "Humans gonna human" people are more comfortable with machine/software that are more empathy towards them.

These results would suggest that by following this natural interaction in the creation of empathic interface would create positive effects: firstly increase engagement and secondly decrease user's frustration when the assistant do not know who to solve a problem.

- Amazon Alexa Design

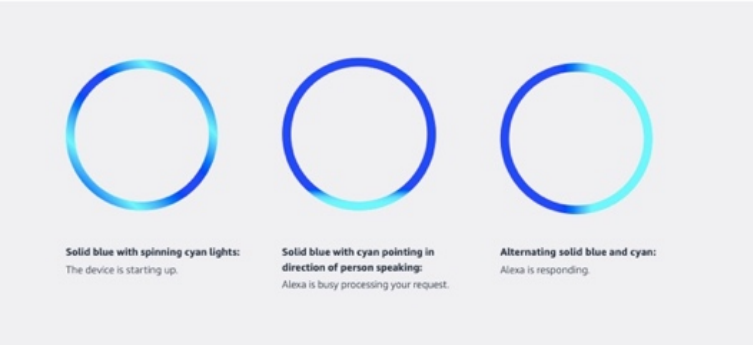


Fig. 64 Amazon alexa (From: Amazon Alexa marketing guideline 2019)

Amazon did not provide any explanation of its Alexa's embodiment design and why the company did not create an assistant with a human-like appearance. Alexa's embodiment in voice-first/voice forward devices is minimal and restrict to a colored circle/line.

This Alexa Echo's circle/line is in the same color, solid blue, as the company brand logo in and it is consistent across all amazon Alexa devices. In the app, Alexa is present in the shape of an Action button, in which is present Amazon Alexa logo as its icon, without the logotype "Amazon Alexa"(amazon developer guidelines, 2019). The light ring (or line) provide the information about the system states.

- **Apple's Siri**

Fig. 65 Siri (From: Support/Apple.it)

Apple's decided not to give a human-like body to its AI agent but rather using the human archetype of secretary/Assistant. Apple's designer, Craig Dehner, worked in creation of visual embodiment of Siri. He explains that the new visual identity of Siri comes from the combination of previous iOS 7 with new Apple's watch primary colors in order to create something new.The final design is the results of different experimentation with designs and forms: smoke, particles,

circular forms, that would have ended up in a mirrored and responsive wave form, that is able to adapt and scale to any Apple's devices.

- Microsoft Cortana

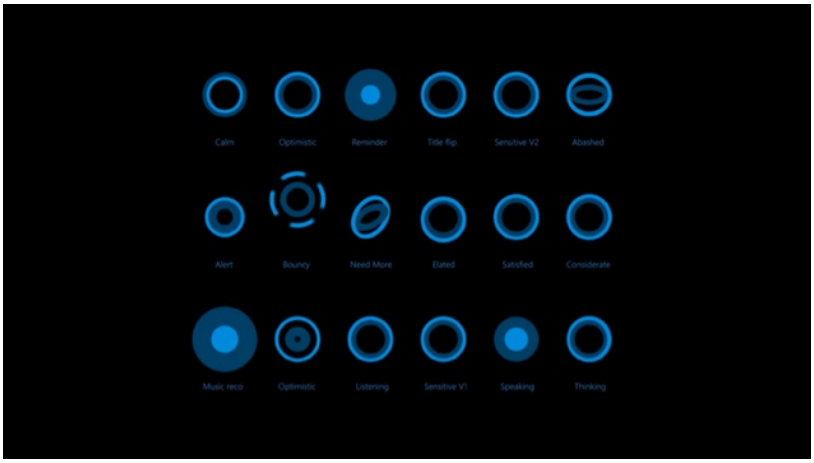


Fig. 66 Cortana (From:http://genieblog.ch/wp-content/uploads/2015/02/cortana_emotions.gif)

Marcus Ash, the Group Program Manager for Cortana and Search on the Windows PC, Phone and Tablet Group, presented to the world Cortana in 2015. Cortana UI is a blue pulsing halo that is able to express 18 moods.
About the design he commented "We settled pretty quick on the idea of a simple geometric shape. We had a bunch of character designers, and the first idea they had was you can do a lot of amazing work with simple shapes. You can make those shapes bounce, you can make them expand, you can make them bow a little bit if they're embarrassed about something." (Marcus Ash,2015). The design was debated for a while, the team tested different types of embodiment from geometrical shape to more human-like avatar, and they found a preference for a physical geometrical shape. Marcus Ash explained that Cortana's appearance was decided that way because it is able to express emotions making the user know how she feels all the time. The company is striving to create an emotional connection with the user mimicking human-to-human interaction that is based a lot on nonverbal communication.

Circular embodiment conclusion

All the mainstream Voice-activated Assistant have a rounded-circular shape. Some research discovered that people have a strong preference for curved items in all categories, particularly when it came to real objects (Moshe Bar, Maital Neta, 2006). A possible explanation is that evolutionary we tend to identify those objects as safe and warm. Another possible reason is that human faces and the emotions they express rely on simple geometric shapes.

Chapter 4

Study methodology

The methodology of the work presented consists in a preliminary desk research about the current state of Voice activated assistants and Trust, which leads to a conjecture model for understanding human-agent trust in first time interaction. The aim of this study is to investigate people's first-time interaction with a Voice activated agent (Alexa) and reported trust in two different devices: Alexa Dot and Alexa Show 5 to discover whether there could be difference or other factors will have more impact in the reported trust. These devices have been selected due to the fact that they have the same functionalities but different modalities: voice first device (Echo Dot) and Voice forward (Echo show 5). This qualitative research consists in a sample selection, semi-structured interviews with scenario-based tasks, which will then conduct in a data analysis.

Research question

When designing for Voice User Interface there are three layers to take into consideration: User, Contextual environment and the device.

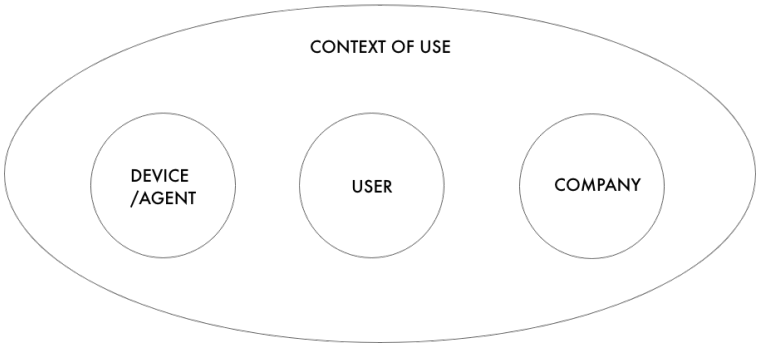
However, to build trust and adoption in such technology there are more factors to take into consideration:

- The kind of environment and the availability of user senses
- User characteristics (technology adoption/dispositions/ personality/prior experiences)
- Assistant's characteristics: device and VUI personas

However, in order to design trustworthy Interactions few more factors need to be considered. Considering the applicability to technology and computer agents, we define Trust as *"the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor"* (Mayer, 1995). Although there are different conceptualizations of trust depending on the stage of the trust relationship development (between the parties). Pre-Trust in AI agents is determined by human characteristics, agent characteristics and company's characteristics.

User's Initial Trust is influenced by institutional trust (based on company reputation), user personal factors and disposition (as propensity to trust, prior knowledge/experience, personality type (big 5) and tech adoption), and agent's characteristics (aesthetics, anthropomorphic cues). All these factors influence pre-trust (and anticipated experience) no matter the situational context. In fact, those factors can decrease and create a positive predisposition to build trust in agents.

TRUST IN VOICE ASSISTANT



Those trust antecedents that need to be taken into consideration while understanding Voice Activated Assistants due to: Personal, Agents' and company factors.

PERSONAL FACTORS

Personal factors that mediated the user's trust intentions and behavior, the following factors can increase the user's willingness to trust or not AI technology. Those factors need to be analyzed to have a deeper understanding on the user's behavior and mindset towards voice activated assistants.

- **Propensity to trust :** Propensity to trust is a personal tendency to trust others (Pearson, 2008).

- **Prior experiences:** Positive prior experiences reinforce trust towards agents, negative prior experiences reinforce distrust towards agents.

- **Personality type:** Personality traits such as agreeableness, openness and neuroticism of the user have been found correlated to this propensity to trust (Dinesen, 2014). Extraversion trait has also been found correlating with higher propensity to trust (Mcbride and Morgan, 2000).

- **Tech adoption:** High Tech adoption level increases the tendency to trust technology.

COMPANY FACTORS

User's prior tendency to trust a particular voice assistant may depends on user's perception of the provider company which happens even without any interaction with that particular device. The following factors might influence user's willingness to trust:

- **(Subjective) Brand perception:** Company good reputation is a good predictor for trust, Positive prior experiences are a predictor for trust

- **Privacy concerns:** Privacy and personal data awareness and perception

- **Brand trust:** Perceived risks and benefits of Company's Assistant

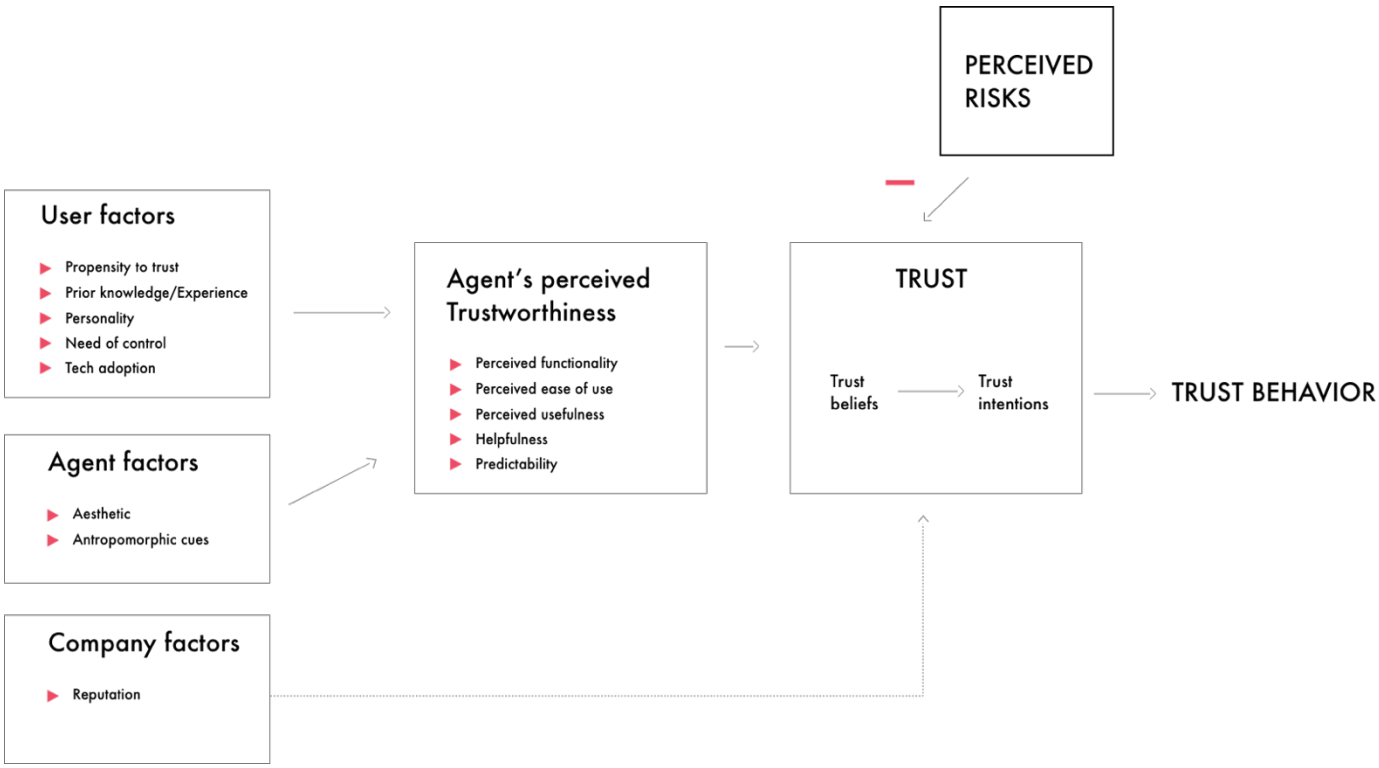
AGENTS FACTORS

Agent's/assistant's factors such as their embodiment aesthetics or their human-like features can influence first time interaction and reported pre-trust.

While other factors during the experience, can have a crucial impact on user's perception of agent's trustworthiness. It was found before, that the following agent's characteristics are correlated with higher trust in human-agent-interaction:

- **Functionability:** "The belief that technology has the capacity or capability to complete a required task."
- **Helpfulness:** *"The belief that technology will provide adequate help and guidance for a human to be successful excluding the moral agency and volition (i.e. will) that humans have."*
- **Reliability/Predictability:** *"The belief that technology will work consistently and predictably".*
- **Perceived ease of use:** *"The extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use."(* ISO standard 9241)
- **Perceived usefulness:** "The degree to which a person believes that using a particular system would enhance his/her job performance" (Davis, 1989 AU95)

Combining all the elements and we have the following dynamics:



Annex 4.1: The different factors of Trust relationship development

In the following study, It was hypothesized the following relationships between user's factors/agent's factors and company factors that will have an impact in the interaction and consequentially in the reported trust and behavior.

Also, we expect voice-only devices might be harder to trust due to the fact that the visibility of the system and feedbacks are more limited compared to multimodal devices, making agent's perceived trustworthiness lower by decreasing its predictability and perceived functionality.

In the case of Alexa, its animated ring is very intuitional about its states but alone lacks the screen-based richness of the information. A particular attention is focused on age-related, personality and curious/avoidant tendency to technology differences.

Thus, the aim of this analysis is to find an answer to:

- How do people interact with this device without prior experience? What do they think of them? What do they think they can do with them?
- How different is the user's experience between Voice-forward devices and voice-only devices in the same use cases?
- How do users perceive Alexa? and what factors during the experience have more influence in their perception of trustworthiness?
- The way user's think of the provider company (Amazon) can decrease the willingness to use and trust these devices?
- What are the major risks that people in using and adopting these devices?
- Can perceive benefits balance user's risk perception?
- Which antecedents (personal/ company) influence the most in reported trust?
- Are there differences in user's factors (age, personality, disposition, tech adoption) that influence reported trust?
- Is there a difference in user reported trust between Voice-forward devices and voice-only devices?

Participants

In a Usability is important to select typical users. Typical users need to reflect the general population of the selected target.
The tools used in the recruitment of the participants was a screen to select people with no prior experience with the devices and more broadly with Alexa. The other selection criteria was to find people that could have different attitude towards agents: "curious" to try and half "doubtful" to try them. It was important to avoid using people that are professionals in chatbots/voice assistant and designers with specific background, because compared to general population they have more focus and attention for details that normal users would not have, the decision was made in order to get the right data for the study. Even without experiences, most of the participants were found to have some knowledge about the possible interaction with an agent even though not knowing exactly the potentiality of Voice activated assistant.
Following Nngroup (Jakob Nielsen on March 18, 2000) directions on qualitative usability studies: *"The best results come from testing no more than 5 users"* due to the fact that even in such small sample can find 80% of usability problems(Virzi; 1992). The participants that took part at the research were twelve, both female and male. The participants had varied backgrounds and no prior experience with "Alexa" Assistant and especially without devices: Echo dot 3 and Echo Show 5. The sample's age criteria followed three main age categories that would be included in the sample in order to have a broader picture of the phenomenon: a third between 18-30, a third between 30-45, a third over 45.

In the following chart it can be seen the screening details of the selected participants that took part into the study.

Partecipant Code	Age	Sex	Tech	Curiosity	Occupation	Device
P01	18	F	High	High	Literature student	Echo Dot
P02	28	M	High	Low	Shop owner	Echo Dot
P03	44	F	High	High	Secretary	Echo Dot
P04	34	M	High	Low	Demand planner	Echo Dot
P05	60	M	Low	Low	Teacher	Echo Dot
P06	61	M	High	High	Financial advisor	Echo Dot

P07	23	F	high	Low	Psychology student	Echo Show5
P08	26	M	High	low	Designer student	Echo Show5
P09	39	F	Low	High	Housekeper	Echo Show5
P10	32	M	High	High	Architect	Echo Show5
P11	58	F	Low	Low	Architect	Echo Show5
P12	57	M	High	High	Business owner	Echo Show5

*chart : https://docs.google.com/spreadsheets/d/1oUmNxbuyl2OOs5OV5cI3BNV4eP4tzp7TzI8iiCO6l1Q/edit#gid=0

After conducting the first analysis it was added an additional questionnaire for frequent users in order to analyze whether there is difference in perceiving risks and benefits of Voice activated compared with the other sample.
*UP = Usual Participant

Partecipant Code	**Age categ ory**	**Se x**	**Tech**	**Curiosity**	**Device**	**Frequency of use**
UP01	30-45	M	High	High	Echo Show8	More than 1 per day
UP02	30-45	M	High	High	Echo Show5	More than 1 per day
UP03	30-45	M	High	High	Echo Dot + Show 5	More than 1 per day
UP04	45-55	M	High	High	Echo Dot	More than 1 per day
UP05	18-30	M	High	High	Echo Dot	Once per day
UP06	30-45	F	high	High	Echo Dot	More than 1 per day
UP07	18-30	M	high	High	Echo Dot	More than 1 per day
UP08	30-45	M	High	High	Echo Plus	More than 1 per day
UP09	30-45	M	High	High	Echo Dot	More than 1 per day
UP10	30-45	M	High	High	Echo Dot	More than 1 per day
UP11	18-30	F	High	High	Echo Dot	More than 1 per day
UP12	30-45	M	High	High	Echo Show5	More than 1 per day
UP13	45-55	M	High	High	Echo Dot	More than 1 per day
UP14	55-60+	F	High	High	Echo Show 5	More than 1 per day
UP15	30-45	M	High	High	Echo Show5	More than 1 per day

Procedure

In order to discover and understand in-depth the factors that would influence user's subjective trust perception semi-structured interviews were conducted (At the end of the chapter). Qualitative research methods allow to dive deeper in the understanding of the underlying phenomena.
The tools that were used in the research setting to collect more details during the interaction are:

- Video camera to record user interaction;
- Reset account;
- Amazon Echo Dot / Echo Show with charger;
- Laptop with charger;
- Smartphone with installed mobile app and skills needed;
- Audio recorder (smartphone) for questionnaire;
- Software to record screen pc
- Printed privacy consent form, questionnaire, task instructions, post-evaluation;
Questionnaire

The research consisted in the following phases:

Phase 1: Warm up and First impression

Phase 2: First-time interaction

Phase 3: Scenario-based tasks

Phase 4: Final questionnaire

Phase 5: Closing

Phase 6: Additional questionnaire for frequent users

- **PHASE 1: WARM UP/ FIRST IMPRESSION**

Methodology: Preliminary questions
Purpose: Understand how people would judge Assistant appearance and prior knowledge about the system without prior input

Each interview started with an introduction about the intent of the study and an explanation about what the participants would be experiencing. They have been reminded that all information and data collected (both verbally and through video) are confidential and will be published in an aggregated form.

Also, it was important to make clear that in all experimentations there was not a specific way to complete a task and that it stressed that it was not a test about them but rather their impressions and experience during the experiment. They have been told to express any concerns and doubts and that they could leave the experiment whenever they wanted if concerned.
After the introduction, participants were shown the two devices (Echo dot/echo Show) and asked to reply to a short questionnaire about the perceived aesthetics of the devices and their prior knowledge about assistants and similar devices.

- **PHASE 2: FREE INTERACTION**

Methodology: Free interaction and follow-up with open questions
Purpose: Understand how people would interact without any input

Participants were asked to interact freely for 1 minute 30 seconds. The moderator reminded them that there is no wrong/right way to do that.
The participants in this stage are not provided with any input from the moderator, they are without complete knowledge about the selected device and in general about an assistant. It might happen that people will not know that they have to say the wake-up word "Alexa" in order to start the interaction. This was followed by a question about their impression of these devices.

- **PHASE 3: SCENARIO-BASED TASKS**

Methodology: Usability evaluation

The evaluation will be performed in two dimensions: an objective evaluation that measures the correctness of the answer (Successful rate) and a subjective evaluation of the experience (as perceived by the person interacting with the device). The phase concludes with follow-up open questions after each scenario.

Participants: Half of the sample will interact with Echo dot, the other half will interact with Echo show

Purpose: Understand how people would execute the given tasks

The scenarios-based activities is a type of usability test that aims to discover how user would interact with a product in the real world without explanations. Giving appropriate tasks is fundamental to understand user's interactions and mental models. The tasks should match the research goal without giving the users too many details or giving instructions. Adding a scenario is useful to provide a reasonable context to the given tasks.

A scenario in this context is defined as a task in which a person will interact with the device in order to receive the assistant's help. They will be provided only with a vague scenario and asked the help of the assistant to solve it. The scenarios used are realistic in order to test real functionalities of the device.
The interactions enable the collection of questions naturally formulated by the participants to the assistant when trying to complete the task of the given scenarios. However, it was chosen not to include a think aloud methodology while executing the task, in order not to miss the user's emotional reactions and possible insights or frustrations. The guidance was limited such as: "Imagine you need to set a timer, how would you do with the assistant?". However, for the participants who did not know how to interact with assistants, such as the wake-up word, they have been told that they need to say "Alexa…." and their command. Examples of the tasks can include a person requesting assistance on how to navigate from their current location to another, simple mathematical questions, and "general knowledge" questions.

For each scenario, the moderator said "With the given scenario, You are asked to execute the mentioned tasks, we want to see how you would ask for help of the assistant in order to solve it"

- Scenario 1

  Moderator *"Ask the assistant how much is 67 plus 15"*

  Open question about: Correct answer/Perceived benefits/ Per-ceived risks

- Scenario 2

  Moderator
  "Ask the virtual assistant to look up today's news in Milan. Ask the assistant for suggestions about what to wear depending on the weather forecast"

  Open question about: Correct answer/Perceived benefits/ Per-ceived risks

- Scenario 3

  Moderator
  "Tomorrow, your best friend will celebrate her/his birthday, and you want to prepare her/him a cake. If you do not know a recipe you can ask the assistant, if you already know you can ask it to create a list for the ingredients you need and then check for the closest supermarket in your area, the opening hours and the direction on how to go there"

  Open question about: Correct answer/Perceived benefits/ Per-ceived risks

- Scenario 4

Moderator:
"You need to buy a gel pen urgently. Ask the assistant to look for some options and comparisons, ask him for more information and order it. Now ask the assistant to cancel the purchase.

Open question about: Correct answer/Perceived benefits/ Perceived risks

- **PHASE 4: FINAL QUESTIONNAIRE**

Methodology: close-ended questions (Likert scale) and open questions
Purpose: The final questionnaire is aimed to find the answer to the research questions: understanding the factors that lead to trust in different modalities: Voice-only (Echo Dot) and Voice-forward (Echo Show 5).The questions were selected from prior works in the literature research, all the information can be found in the first 3 chapters of this thesis. The questions were divided into the following categories:

PERSONAL FACTORS

- Personality type (big 5)

In order to assess personal big evaluation it has been used the the short version of assessment of the Big Five that has been found to be scientifically valid, in the reference can be found the original scale.

The participant is asked to evaluate himself at a scale from 1 (completely disagree) to 5 (completely agree).

I See Myself as Someone Who …

1. Worries a lot (N)
2. Gets nervous easily (N)
3. Remains calm in tense situations (N, recoded)
4. Is talkative (E)
5. Is outgoing, sociable (E)
6. Is reserved (E, recoded)
7. Is original, comes up with new ideas (O)
8. Values artistic, aesthetic experiences (O) 2a
9. Has an active imagination (O)?
10. Is sometimes rude to others (A, recoded)
11. Has a forgiving nature (A)?

12. Is considerate and kind to almost everyone (A)
13. Does a thorough job (C)
14. Tends to be lazy (C, recoded)
15. Does things efficiently (C)

- Propensity to trust

In the literature there are multiple scales to assess propensity of trust, in this evaluation an existing propensity trust scale has been used can be found in the reference below.

Not all the questions have been used but only a selection of them depending on the relevance to agents. High reported level of propensity of trust is likely to have an effect on trust behavior.

1. It is easy for me to trust others
2. Even if I am uncertain, I will give others the benefit of the doubt
3. I usually trust people until they give me a reason not to trust them
4. I tend to trust others even if I have a little knowledge about them
5. My typical approach is to trust new acquantaties until they prove I should not trust them
6. I believe people usually keep their promises
7. My tendency to trust others is high

- Tech adoption

In the literature, there are several ways to assess the user's relationship with technologies. It was important to understand both the user's general attitude and its need to control. Both factors have been found in the literature (as mentioned in the first chapter) as predictive factors to technological adoption.

  - Attitude to technology

    1. I feel comfortable to try a new technology
    2. I do not care to learn a new technology
    3. Technology is not difficult to understand
    4. Learning about technology is a waste of time

  - Need of control

    1. People are smarter than technology
    2. Soon our life will be controlled by technology
    3. People will always be in control of technology

4. I prefer setting up the functions myself rather than letting the assistant do it

AGENT'S FACTORS of perceived trustworthiness

- **Perception of agent's anthropomorphism**

In order to assess how "human-like" the user would judge the agent. In the context of social robots, it was developed a scale to assess robots' level of: anthropomorphism, animacy, likeability, perceived intelligence and perceived safety (Bartneck, 2010). These standardised tools have been applied in many studies, especially in Human-Robot Interaction.

This scale has not been used entirely but only selected elements about the agent's appearance.

- Fake or natural
- Machine-like or human-like
- Unconscious or conscious
- Artificial or life-like

- **Perceived predictability**

How predictable the agents appear to the User. High reported level of predictability is likely to have an effect on trust behaviour.

1. I can predict how the assistant will behave towards me
2. I am confident about the ability of the assistant to reply effectively
3. I know what to expect from the assistant

- **(Subjective) Evaluation of information source**

How significant the information provided by the agents appears to the user. High reported level of predictability is likely to have an effect on trust behaviour.

1. I find valid the information I was given
2. I find meaningful the information I was given by the assistant

- **Usability**

In order to test the usability of the device and the assistant, the following categories have been asked following a 1-5 Likert scale: Perceived usefulness, perceived ease of use, perception of system feedbacks.

- Perceived usefulness

    1. Using the assistant in my job would enable me to accomplish tasks more quickly.
    2. Using the assistant would improve my job performance.
    3. Using the assistant in my job would increase my productivity.
    4. Using the assistant would enhance my effectiveness on the job.
    5. Using the assistant would make it easier to do my job.
    6. I would find the assistant useful in my job.

- Ease of Use Items

    1. Learning to operate with the assistant would be easy for me.
    2. I would find it easy to get the assistant to do what I want it to do.
    3. My interaction with the assistant would be clear and understandable.
    4. It would be easy for me to become skillful at using the assistant.

- **Perception of feedbacks**

    1. I understand that the product is communicating with me through feedbacks
    2. I find the object's feedbacks easy and clear to understand

- PHASE 5: CLOSING

Methodology: Open questions and close-ended questions
Purpose: Understand deeply user's awareness about benefits and risks.

In this last phase, the participants are asked about their perception of Amazon as a brand and about perception of risks/benefits.

BRAND'S FACTORS

Companies and brand factors influence the overall perception of agent's trustworthiness. The more positive the brand is seen the more likely it is to have a positive effect on trust.

- **(Subjective) Brand perception**

Selected questions from a previous questionnaire which link can be found in the reference.

1. When you think of [Amazon], what comes to mind first?

2. Can you explain why do you think of that?
3. On a scale of 1-10, how likely are you to recommend to a friend or colleague?
4. How would you describe your last experience with [Amazon]?

- Privacy and personal data

General questions about Amazon and data collection

1. What kind of data do you think Amazon collects?
2. Would data collection influence you not to use the device?
3. Would you let Amazon collect your data in order to have a richer and more personalized experience?

- Brand trust

In order to understand more deeply the possible benefits and concerns about agents, the set of Close-ended questions based on a 1-5 likert scale were divided between "Perceived benefits" and "Perceived risks". The questions were based in a previous investigation that can be found in the reference that analyses the subjective perceived risks/benefits of smartspeakers ownership.

- Perceived benefits

1. I enjoy using the assistant to execute tasks
2. Using the assistant is fun and entertaining
3. Using the assistant will help to manage devices in my home
4. Finding information by using the assistant is a fast way to achieve what I am looking for
5. I am satisfied with the informations the assistant gives me
6. Using the assistant is convenient when I am busy doing something else
7. Using the assistant is a convenient way to manage time
8. Completing tasks with the assistant is an efficient use of my time
9. The assistant helps me to quickly have information about mail/weather
10. Using the assistant with my family/friends/partner is fun
11. Using the assistant will make my life easier

- Perceived risks

1. I have my doubts over interacting with the assistant
2. In general, I am not comfortable speaking to the assistant

3. I am not comfortable speaking in public with the assistant
4. I am concerned having an assistant in my home
5. I have my doubts over the confidentiality of my interactions with the voice assistant
6. I am concerned about the storage of my personal details
7. I am concerned about what kind of data will be stored by the assistant
8. I am concerned about the possibility of the assistant being hacked
9. I am concerned that corporations will share my personal information with other parties
10. I am concerned that the assistant collects too many information about me
11. I am concerned corporation might misuse the collected information
12. I am concerned that corporations use smart home technology to spy on users
13. I am concerned that the assistant records my private conversation not directed to him
14. I am concerned that due to the interactions with the assistant, I would be sent targeted advertisement

- PHASE 6: ADDITIONAL QUESTIONNAIRE

Methodology: close-ended questions
Purpose: Understand deeply user's awareness about benefits and risks in
In this last phase, the participants are asked about their perception of Amazon as a brand and about perception of risks/benefits as well as personal information to understanding more deeply their attitude.

The questionnaire was kept the same as the one in phase 4.

Limitation of the model and the current study

The model was created based on literature review and by merging different models about trust. It might be that the factors taken into consideration were context-specific and therefore not applicable to other contexts. Also, the relationship between the factors have not been investigated or proved scientifically.
The model is meant to understand what are the factors that interplay in the relationship trust-building between user and agent.

Also, part of the interviews (4/12) were conducted through video call, and this might have an effect in the overall interaction.

References scale

Personality scale: https://link.springer.com/article/10.3758%2Fs13428-011-0066-z

Propensity to trust scale
https://www.researchgate.net/publication/260036544_Development_and_Validation_of_a_Propensity_to_Trust_Scale

Propensity to trust scale
https://www.qualtrics.com/experience-management/brand/perception-surveys/

Perceived benefits/risks scale: https://irishtechnews.ie/swearing-by-smart-homes-analysing-trust-in-smart-home-technology/

Bartneck, C., Croft, E., & Kulic, D. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Soci
al Robotics, 1(1), 71-81.
Virzi, R.A. (1992). Redefining the test phase of usability evaluation: How many subjects is enough? Human Factors, 34, 457-468.

Chapter 5

In the following chapter, the results of the interviews and questionnaires have been analyzed a structured based on the research questions.

First impression

- How do people interact with this device without prior experience? What do they think of them? What do they think they can do with them?

Fig. 67 Echo dot (From: https://amazon.it/)

AESTETICS JUDGEMENT (ECHO DOT)

People were asked to judge their opinion about device's aesthetics and function, it was found that:

- All participants found it beautiful/ partially beautiful 4 vs 2
- 5 out of 6 said it seems like a speaker
- 5/6 said they like the shape (2 of them mentioned also the buttons)
- 1/6 referred to its function

- 2/6 said they do not like the wire
- 2/6 Said nothing
- 1/6 Said the color

PRIOR EXPERIENCE AND EXPECTATIONS

- Even if the participants did not have any prior experience with the selected device, 3/6 Recognize the object as Alexa or as voice activated assistant without any indication and provided some use cases such as listen to music, set a timer, ask the weather forecast.
  For the other 3/6 the object was a speaker, giving it the possibility to broadcast music.

- 5/6 of the participants did not recall any experience with a similar object, one of them mentioned Google Home

- All 6 participants have a positive attitude towards it

FIRST TIME INTERACTION

3 participants (P01, P05, P04) out 6 knew what kind of device it was and were able to interact straightforwardly:
One asking the weather forecast in Finale Ligure (where the interaction took place). One participant (Age category 50-65) first asked to have instructions or watch a tutorial before interacting with it, eventually by manipulating and touching the buttons on the surface she was able to start the interaction understanding it is a smart speaker device. She asked directly was it was able to do. Similarly, another participant (P03), Age category 50-65, said that she knew that device but in order to interact with it she would need an instruction book.
"Posso leggere il libretto di istruzioni?(ride) non l'ho mai usato, ho visto la pubblicità su sky ma non so cosa devo fare"

Another participant thought it was just a speaker and tried to connect with Bluetooth his phone to listen to music from his Spotify account, but could not manage. In conclusion, for those participants that knew the device in advanced and they were familiar with voice assistants knew how to interact with an object like that. However, older generation might feel more comfortable interacting with it after having read instruction.

Fig. 67 Echo Show 5 (From: https://amazon.it/)

AESTETICS JUDGEMENT (ECHO SHOW 5)

People were asked to judge their opinion about device's aesthetics and function, it was found that:

Aesthetical perception of participants was mixed: 1/6 Said beautiful, 3/6 partially beautiful, 1/6 partially ugly, 1/6 ugly

4/6 found it quite pleasant, and 2/6 neutral

- 3 out of 6 said it seems like an alarm, 1/6 a mini tv, 1/6 thermostat, 1/6 a navigator

- 5/6 said they like the screensaver (2 of them mentioned also the interactivity)
- 1/6 referred to the overall shape

- 3/6 Said there anything they do not like
- 1/6 said he/she did not like the wire
- 1/6 Said the size
- 1/6 said he/she did not like the border of the screen

PRIOR EXPERIENCE AND EXPECTATIONS

- In normal situation, the screen is set with a time display and the weather temperature, after a while the animation can change and, in some cases, suggesting a possible action. However, 3 out 6 said they would not know what the device is capable of doing, 2/6 by reading the screen-display figured out that they could ask questions, and 1/6 said they could set the time

- 2/6 of the participants did not recall any experience with a similar object, one of them mentioned Google Home

- All 6 participants have a positive attitude towards it, saying that would have not any doubts interacting with a screen-display.

FIRST TIME INTERACTION

As expected, the first-time interaction is easier compared to voice-first modality: the screen change it screen saver suggesting some functionalities.

Four users after a manipulating the device, read the screen and followed Alexa recommended functions:

P12 said: ahhhhhh…. I think I understand ... hello alexa " and reads out loud the screen "Alexa plays music for cooking".
P08 and P10 asked about word of the day, P10 did it at the first try while P08 "Which is the word of the day?" without saying the wake up-word and continued saying "But nothing happens. Nothing happens. I touch it, what if I touch the screen? But I wanted to ask what the news of the day".

P07 after reading the screen asked "Alexa, sounds for sleeping" and Alexa responded" Tibetan bells skills have been used recently, try saying "Alexa, open Tibetan bells " suggesting a more direct function. And again, looking at a new screen suggestion asked about the square root of 144. Learning the functionality of the calculation.

However not all user understood a possible interaction, in fact, A user (P11) look and touch the screen said: "I would like to know what the topic of the day is. The speaker is missing here I honestly don't know how to make it work, I touch, and nothing happens. no, I don't know and I'm not able to do it, I've never used it, I can try again touch the screen and not nothing changes, I try to touch, and it just gives me the day, time and external temperature "
Another user instead and reads "tomorrow it is going to rain…" and continued speaking" oh it disappeared, I am not good with technology". Afterwards after waking up Alexa accidentally started laughing and said surprised "it started speaking".

- ***How different is the user's experience between Voice-forward devices and voice-only devices in the same use cases?***

Use cases analysis

TASK 1

PARTICIPANT CODE	TASK 1	N° STEPS	IS THE INFORMATION VALID?	ANY BENEFITS?	ANY RISKS?
P01	Ok	1	Yes	yes	None
P02	Ok,	2	yes	Yes it is faster	None
P03	Ok	2	Yes	Yes, so i don't have to use the phone	No, i just need to learn how to use it because it was the first time. it is fun
P04	OK	1	Yes	mmm it was an easy task	None
P05	ok	1	Yes	it is convinient if "she does"	none
P06	Ok	1	Yes	Easy	None
P07	Ok	2 Steps (I think it didn't hear)	Yes	Yes, so i don't have to do it	None
P08	Ok	2 steps	Yes	It is useful, that was easy and could be done mentally, for those that are more dicult could be even faster than digit to the iPhone/computer	None
P09	Ok	3 steps	Yes i think so	yes	None
P10	Ok	1	Yes	yes	None
P11	Ok	1	Yes	It was easy	None
P12	Ok	1	Correct	This sum was easy maybe with a more complex one would make it more beneficial	None

Behavioral observation

Most of users watch the device while asking the question, especially those with the screen when Alexa showed the result of the calculation.

About the task

All the participants succeed in the given task, most in 1 step. One of the participants (P3) mentioned that she needs more time to be able to know how to use and interact correctly with the device.

Valid
All participants found valid the answer they received by the assistant

Perceived Benefits
- Fastness
- I do not have to use the phone
- For 3 People said it is probably more useful for harder calculus

Perceived Risks
All the participants found any risk while interacting with the device

Device
There is no difference between the two devices in this specific use case. However, the device displays the calculation, the user potentially can see whether the input was detected correctly by the system.

TASK 2

PARTICIPANT CODE	TASK2A	N° STEPS	TASK 2B	N°STEPS	IS THE INFORMATION VALID?	ANY BENEFITS?	ANY RISKS?
P01	Ok	1	No (*Alexa replied with "i do not know"!)	1	Yes	None	None
P02	Ok	2	OK	1* asked about weather	Yes i trust it	Some, so i don't loose time	none
P03	Ok,	1	Ok	1	Yes, I should learn how to it better (she tried to stop it for 4 times before succeding) but "she" knows that we are in finale ligure and says directly	It is fun	none
P04	Ok	1	OK	1	Yes, but mayabe it was too detailed	It can be useful when i am busy	No i ask the same questions on internet, maybe the assistant

							should check the location when i am asking what to wear
P05	Ok	1	ok (she asked the weather forect so she can chose what to wear "ok con questo decido cosa mettermi"	1	Source okay	The weather forecast can be usefull	None
P06	OK	1	No (* User request was asked incorrectly and not recognized by the system)		Yes	It does not reply always correcly	None
P07	Ok (*Alexa, mistaken the user For Valeria!)	1	Sort of (asked the weather forecast however not managed for the advice on what to wear)	3	Yes	It can be convenient	No, but sometimes you need to ask the same question few times
P08	Ok, However the user changed the information channel to another (News from Republica)	1	No	asked 3 times: first joking, then correctly	I changed because i consider repubblica more valid than tgcom24	I prefer something else	Yes, she is the one deciding the information source, and in this case i would have chosen Republica to Tgcom24 and also a textual support would have been good, so i can skim through what i want to listen or where to read the information (like this i would be more confident)
P09	Ok	4	no (user call the	-	Yes i was suprised about the news		None

			assistant Alexia)				
P10	Ok	1	No	2 Trials	Yes seems all right for the news	Okay, I prefer to look for these thing by myself	Misunderstanding
P11	Ok	1	NO (* user request was correct but alexa didn'r respond)	-	Seems correct to me	The news are useful, maybe the screen is to small for me	None
P12	OK	1	OK	1	Yes, but i would have prefered a different news channe, Tgcom is a good service, but i prefer Skynews	Beneficial for both the weather forecast and news	None

Behavioral observation

Three participants wanted to stop the news after 1 minute (P02, P04, P05)
Alexa said "Valeria, here are the news" and (P01) looked at it confused.

About the tasks

All the participants succeed in the task 2(A) asking the daily news.
P08 decided to change the new source explaining that he preferred a different one.

In the task 2(B) 6/12 people succeeded on the task, however:
2 users, P02 and P05 instead of asking an "advice" to the assistant about what to wear they asked about the weather, as if they need to be in control of that decision making. One of them(P05) mentioned in fact, she asked about the weather forect so she could chose what to wear (in the transcription "ok con questo decido cosa mettermi")

Not succeeding:

- In case (P01) at the question *"what should I wear today in Finale Ligure?"* the system responded *"I do not know"*
- In case P11, at the question *"what should I wear today?"*, the system did not reply
- P10 tried twice to ask Alexa about what's the weather like but she did not respond
- In case P09, the participants used "*Alexia*" instead of "*Alexa*" as a wake-up word, and the system did not respond.

Validity

11/12 found valid the source of information about the weather forecast and news. However, one participant, P08 However the user changed the information channel to another (News from Republica) because *"i consider repubblica more valid than tgcom24"*

Another user (P12) mentioned that he prefers another news channel (Sky news), however he found the news Tgcom24 fine.

Perceived Benefits

Mixed opinions about the benefits of those functions (news and weather forecast).
Most found it convenient and useful. Some said that it could be a faster way to know them (P02), one found it fun (P03), another (P04) found it useful when his hands are busy.

One participant (P03) mentioned *"This is all so direct, even without saying "she" new that it was March 10"*.

Perceived Risks

One participant (P04) said that the news was too long, and instead of having direct broadcasted a channel could have been easier to have instead some news recap. Also, he noticed that the system recognized his location, he mentioned that the system should check whether that was correct or not.

P11 mentioned that the size of the screen-based device could be a problem since she could not read all without the glasses.

P07 did not found any particular risks a part the fact that she needed to ask few times the same question. The participant seemed a bit frustrated when she need to ask 3 times to have a suggestion about what to wear.

The risk perceived by the participant that changed the news channel, was that the *"she is the one deciding the information source, and in this case i would have chosen Republica to Tgcom24 and also a textual support would have been good, so i can skim through what i want to listen or where to read the information (like this i would be more confident)"*. It seems like for some users, the fact of deciding rather than to letting the assistant to do that.

Device

The difference between the two devices is significant: the screen-based device offers a good potential for both weather forecast by providing the animation

of the weather of the day and for the following days, and the news are broadcasted directly from Tgcom24 and in the context in which the user can watch it can be stronger the effect. However, in this use cases the rate of success of the voice-first condition was higher in task2b than Voice-forward. This might explain the slightly difference in the reported risks.

TASK 3

PARTICIPANT CODE	TASK 3A	N° STEPS	TASK 3B	N°STEPS	TASK 3C	N°STEPS	TASK 4C	N°STEPS	IS THE INFORMATION VALID?	ANY BENEFITS?	ANY RISKS?
P01	Ok	3 steps	no user forgot	-	no	tried 7 steps	no	-	Yes for the recicpe, no for the supermarket	Yes for some parts, for others better the phone	no, only that we have not understood each other
P02	Ok	2	No	1 tried once no reply from the system	Ok	7 steps, alexa was still speaking abou the recipe)	Yes, however it was not the same supermarket asked by the user	3	Yes	No	None
P03	Ok	3*however first the she read the scenario	No	1	Ok	1	Ok	1	Yes, I am used to do it on the computer/iPad or watching on tv (i have sky) however this way is efficient	It is efficent	"you need to speak very slow in order to be understood"
P04	ok	8 steps: it was hard to stop Alexa	No, Alexa did not respond correcly to the user input	1	OK	1	Ok	1	I use Giallozaffera no often to find a recipe	If i am busy and i can not see the screen because my hands are busy yes	No

P05	ok	(3 steps, alexa was providing a similar list for recipes)	Ok	1	Ok	2	No	-	Ok	It is efficent	None
P06	ok	1	no	Trie 1	no	-	no	-	IT didn'r reply	If it works if would be usefull	Just the fact that we did not understand each other
P07	Ok	1	No	Tried 3 times to interrupt and skipped	Ok	1	No (question was formulated correclty however alexa replied "i don't understand"	-	I use it sometimes even if i am not very good at cooking	it can be nice	none, there are something that she does not understand but maybe i was not making the right question
P08	Ok	1	Ok	5 trials (the user seems frustated)	No	1*thesystemsaid:7 steps, alexa was still speaking abou the recipe)	No	-	Okay for Giallo Zafferano	If it would work well maybe	It is impossible to stop her when she speaks and make another request. and also she did not reply to my question about the supermarket
P09	no	-	No	Tried for times. probably becaus	Ok	1	No	Tried 3 times	Yes, even if we did not undestand each other	if it would have replied	None

				e the user was calling the assistant alexia							
P10	Ok	3	No	Tried twice when alexa was speaking about the ingridients	Ok	1	NO	-	Yes, it is a bit frustrating though	None	None
P11	Ok	1	the user forgot to ask	-	Ok	1	No	-	I usually use different website for recipes	It is nice, but it speak too much, i would rather use the ipad	No
P12	Ok	1	No	The user asked, Alexa was going on telling the recipe without replying to the user input, P12 stopped after 2 trials	Ok	1	Yes	1	I don't cook very often, however it seems legit to me	It is funny	we had moments of misunder standing

Behavioral observation

Visibly lots of participants (P07, P02, P12) seemed frustrated and irritated with the device (in both conditions) after few trials to make it stop.

About the tasks

All the participants succeed in the task 3(A) apart from one P09 was not able to do the task.
Most of them, did it in one step: P02 in two, P01 in three, P03 because first was reading the scenario text, P04 found hard to stop the device and managed to ask the recipe after 8 trials, P05 in 3 due to the fact that Ala was providing a list of possible recipes that had all a similar name and the participant did not understood that she had to say one.

Only 2/12 people were able to create a list with all the ingredients, one P08 after 5 trials and he was visible frustrated.

Some people tried few times before skipping it:
- P07 tried 3
- P09 tried few times but she was not succeeding due to the fact she was call the assistant Alexia,
- P11/P01 forgot the task
- P10/ P12 tried to ask twice but was not able to stop Alexa telling the recipe

Four participants tried only once P02, P03, P04 and P06 and being frustrated continued with the following task.

This task was not immediate, most of the people after listening the recipe they would go on continuing asking the ingredients only then they would ask to create the list, however the system at that time to execute that command should have stopped before asking to make the list.

Task 3C was completed successfully by 8 people
- For some it was achieved in one step (P03, P04, P09, P11, P12)
- P05 to have a more specific answer
- Others in 7 steps (P01, P02) Alexa was still speaking about the recipe

About the remaining, the system did not recognize their request, in some cases it seemed impossible to stop Alexa broadcasting the recipe.

Task 3D was completed successfully by 3 people
- P03, P04, P12

However, P02, asked the timetable for a specific supermarket, but the system respond about different one. Most of the other gave up after the assistant did not respond to their question about the supermarket.

Validity

Some people said that the recipe source seemed good even if they have not used it a lot and found it efficient to look for a recipe.

“I don't cook very often; however it seems legit to me”
“I use it sometimes even if i am not very good at cooking”

One subject, was more passionate about cooking and cuisine, said she would not use that. Another said that was good for the recipe but not the supermarket (P01) and other were frustrated to not be able to communicate properly with the assistant.
For one participant the assistant speaks to much, up to the extent that was not possible to stop it.

Perceived Benefits

Mixed opinions about the benefits of those functions (recipe and function).
Few people found it as convenient and efficient especially If busy and can not see the screen.
Another(P01) that can be partially helpful, but not for everything, other 3 people (P06, P08, P09) said it would be nice if it works

“It is nice, but it speaks too much, i would rather use the ipad for it”

Perceived Risks

In the majority of the cases analyzed, the main reported risk was incomprehension between the user and the assistant.

(P01) *“only that we have not understood each other”*

P03 *"you need to speak very slow in order to be understood"*

P06” Just the fact that we did not understand each other”

P08 *“It is impossible to stop her when she speaks and make another request. and also she did not reply to my question about the supermarket”*

P12 *“we had moments of misunderstanding”*

Device

The screen-based device was able to give more detailed information about the recipe, showing picture of the processes, the ingredients and the steps, and also about the location mentioned. In a context like cooking it can have a higher advantage, due to the fact that it shows more details that in the other condition were harder to remember (the list of recipes, the list of ingredients) also, the fact that the screen is able to present the details of the supermarket and that the user can select himself to know more is efficient.

TASK 4

PARTICIPANT CODE	TASK 4A	N° STEPS	TASK 4B	N°STEPS	TASK 4C	N°STEPS	IS THE INFORMATION VALID?	ANY BENEFITS?	ANY RISKS?
P01	ok	4 steps: first asked where can she but the item in finale ligure, afterwards tried with the assistant	Ok	1	No		yes	i does not change a lot	no
P02	ok	(3 step) first asked where he can find in the area	Ok	1	no		Yes	i would not used it because i am a shop owner and i know what it means	i dont like to use a credit card online
P03	ok	3	Ok	1	No		Correct	It is cool, maybe i could try	I usually do it witha computer, i am not sure i would trust it, maybe with a rechargeable credit card
P04	ok	1	No	-	NO		The answer is correct i would have expected more details and options	I'd buy through something else	I would buy through amazon
P05	ok	2	Ok	1	No		Corret, it buys from amazon	Maybe for others i would not use it	
P06	ok	1	Ok	1	NO		Yes, It is amazon	I'd use something else	few choices compared to the website
P07	ok	4	Ok	1	No		Yes, it opens amazon	I does not change much, for me it is a bit weird	I think the risk is the same that one is done online
P08	ok	1	Ok	1	No		I perceive it as trustwothy	no	It is needed to trust

							however we do not understand each other		wether i trust amazon choices. She mentioned few options and i would not have the time there to look for more informationa about the product and probably i would need to ask for each of them reather than havin g a comparison between the features as in amazon website. I would have wanted a description so i could chose which one to put in the kart. When i am busy or i dont feel like touching the screen maybe i would use it, but for me it would be a risk doing such
P09	No	No, she tried first to look for a stationary shop, afterwards no	NO	-	NO		i can not judge	no	I do not buy from internt
P10	Yes	1	Yes	2	NO		Okay it is amazon	no	no
P11	no	* the user first asked for a stationary	NO	The system said it	no		I coud not buy it	no	I prefer not to buy from internet

		shop locally, after a website that sells pen, then stopped		could not buy it					
P12	Yes	1	ok	1	NO		Correct	It is an alternative, i would not use it	When it was saying the list of the products available i forgot the one that i wanted to buy first

Behavioral observation

User got frustrated with last task 4C due to the fact that they could nor remove the selected item from the kart. Some of them (P02, P07, P09, P11) mentioned that they are not vert used to buy online.

About the tasks

4 Participants (P01, P02, P09, P11) when they were told that they need to buy urgently a pen, they asked the assistant the closest stationary shop where they could buy it, as found later they are not used to do shopping online.

All the participants succeed in the task 4(A) expect for two (P09, P11) one stopped after looking for a stationary shop, another as well look for more physical store, then a website that sells pen and after couple of trial stopped.

Task 4b was achieved by 10 people, all of them manage to put the selected pen in the amazon kart. The system responded to P11 It could not buy the pen.

Task 4C, was impossible to complete since that functionality as also direct buy through Alexa is available in few selected countries such USA, UK and India. People

Validity

Most user found valid the 8/12 the source of information, three of them (P05, P06, P07) told that "it is amazon" considering that e-commerce as their usual point of reference when buying online. A user (P04) said "The answer is correct i would have expected more details and options"

Perceived Benefits

Most of the people told that they would not use it, however two of them said it would not change much from usual online shopping (P01) but it is weird to do online shopping in this way (P07). Most people did not find this function as useful.

Perceived Risks

Three participants found this not as useful however mentioned that it would not change too much that buying in other ways.
Two (P03, P06) prefers through with a computer due to the fact that is the way they usually do and can better compared the products.
One participant (P08) found especially hard to trust the choice made by Alexa "It is needed to trust wether i trust amazon choices. She mentioned few options and i would not have the time there to look for more informationa about the product and probably i would need to ask for each of them reather than havin g a comparison between the features as in amazon website. I would have wanted a description so i could chose which one to put in the kart. When i am busy or i dont feel like touching the screen maybe i would use it, but for me it would be a risk doing such"

Device

The experience with a screen-display is remarkably different: it gives the possibility to the user to see the kind of product the user can buy compared to the voice-first device. However, it does not seem like the fact that people could see more details, as more convincing.

- ***The way user's think of the provider company (Amazon) can decrease the willingness to use and trust these devices?***

10/12 were found to have positive opinion about Amazon and commented:

P01 “Yes, we buy many things”, P03 “I find it very useful, there are many things, a lot of choice ... prices can be very advantageous”, P04 “I buy very often from Amazon, it is not the best for the sustainability side… but at the logistics level (I who work in this sector) they have an innovative processes, P05 “comfortable reliable” P06 “I like it, there are many different articles and it comes quickly” P07 “Yes, even if I don't buy a lot online it is useful”, P03 “I find it very useful, there are many things, a lot of choice ... prices can be very advantageous”, P08 “I think it is an excellent service I get there I am practical of online orders and I do many Amazon always respect the delivery times and quickly I buy everything often enough from the site” P10 “comfortable reliable” P012 “I find it a very good service” . A participant (P02) did not like Amazon service “I don't like it much and since I have my own shop, I prefer a small shop over a large company as I know what it means compared to the internet. I bought you one thing and my life is enough. I don't like that everyone has to register. credit and address is a lot of credit information Another P11 said that she never bought anything so she cannot judge but overall, she thinks it is useful. About their last experience with the company most of all of them except (P02) and P11 had no experience with service, said it was good:

P01 “I don't remember what I bought, maybe a pair of headphones, I would say positive”, P03 “I don't remember exactly what I bought the last time, I never had a problem”, P04 “Everything good”, P05 “good”, P06 “good, it arrived in a day”
P07 “I don't remember what I bought probably a book for university”, P08 “In general, the last experience I had was also positive, that is, everything arrived in terms the next day”, P09“Everything good”, P10 “good”, P12 “always positive”

All of the user except P02 would recommend to a friend.

Most of the people do not know what kind of data Amazon collects about them and for what kind of purpose. However, two mentioned the product targeting strategies: P01 “I do not know”, P02 “mmm I do not know, but once I look on the site I find the same article that I also searched on Facebook and on various social networks and you have a lot of advertising of the things you searched for.” , P03 “I don't know, maybe compared to what I watch or buy on the site”, P04 “What things I usually buy to make profiles”, P05 “I do not know exactly”, P06 Maybe realated to what I see on their website, In fact the when I navigate other websites I find the same things I was look at” P07 “I have no idea, I don't know much about this”, P08 “it is possible for them to compared what I watch on the site and what I bought you will get an idea of what I like, that is, I don't know”, P09 “no idea”, P10 “I do not know”, P11” no idea I do not buy there anyway” P12” Mmm I don't know

- *Are there differences in user's factors (age, personality, disposition, tech adoption) that influence reported trust?*

CODE	PREDICTABILITY	USEFUL	EASE OF USE	FEEDBACKS	VALIDITY	HUMAN-LIKE
P01	High predictable	Neutral	Easy to use	clear feedbacks	valid information	Human-like
P02	High predictable	useless	Easy to use	Neutral	valid information	Neutral
P03	So-so	Usefull	Easy to use	clear feedbacks	valid information	Human-like
P04	High predictable	Neutral	Easy to use	clear feedbacks	valid information	Artificial
P05	So-so	useless	Easy to use	clear feedbacks	valid information	Neutral
P06	High predictable	Neutral	Easy to use	clear feedbacks	valid information	Artificial
P07	High predictable	Usefull	Neutral	clear feedbacks	valid information	Neutral
P08	So-so	useless	Neutral	So-so	valid information	Artificial
P09	High predictable	useless	Easy to use	clear feedbacks	valid information	Human-like
P10	High predictable	Neutral	Easy to use	Clear feedbacks	Unvalid information	Neutral
P11	High predictable	Usefull	Easy to use	clear feedback	valid information	Neutral
P12	So-so	Usefull	Easy to use	clear feedbacks	valid information	Neutral

4/12 people did not find the assistant as very predictable however the same people did think that the device was easy to use, and perceived the assistant having clear feedbacks and providing proper information. However, in this 4/12 P08 is not on the same advice.

Perceived lack of utility

Many users reported the use of Voice assistant has been useless (P02, P05, P08, P09) and other two were neutral about its usefulness (P04, P06). Most of these cases their prior interaction with other Voice Activated Assistant was almost inexistent, that we interpret it by the fact that they did not know the types of functionalities and potentialities these devices might have for them. Also, the fact that some misunderstanding happened, and user needed to repeat the question more times we interpret the fact that they do not think that the *"assistant will make them achieved their action faster"*.
Also, due to the nature of the given tasks, the use cases applicated might not be related to user needs and goals, in fact at the question *"the use of Alexa will improve my work performance"* the judgement was mixed.

Humanizing Assistants

People judge the assistant differently in terms humanness: most of the people found it as neutral (P02, P05, P07, P10,P11,P12) and others found it artificial (P04,P06, P08). Nevertheless, those participants, behaved politely to the assistant during the tasks and replied to Alexa with thank you.

. *What risks user's in engaging with Voice activated assistants?*

Perceived risks and Privacy concerns are the major deterrent factors in the adoption of voice activated assistants. When users were asked directly what kind of data would be stored about them due to their interaction with the devices, most of users said they did not know. For the major part they seemed unaware about it. One user (P08) said *"the same probably than I am navigating on the website"*
However, when asked more directly, the users-display category mentioned the following concerns:

One user (P08) mentioned that the device is not dangerous it self but it is the way it is used: "If I use it in the right way I think it is safe"

Another (P09) said that *"I don't think there are particular risks. Only the fact that it doesn't understand rapidly the questions given".*

Three users were concerned about the use of their personal information for marketing purposes. (P10) mentioned the fear of being targeted with products that he spoke aloud but not directed to the assistant.
60% of the participants said that are worried that Amazon might share their personal information to other parties, and more than 68% fears to receive targeted advertising after the usage.

From the questionnaire it is evident that people are not perceiving high risks for having a device at home, most of them do not afraid to the possibility that can record them without consent, or not directed to it, or that it can spy them.
However, some participant fear that somehow the company might misuse their data.

. *Can Perceived benefits counterbalance the perceived risks?*

General reported benefits after the questionnaire were related to fastness to execute commands and handsfree interaction:

P01 "speed up searches" P02 "nice, cute ", P03 "funny", P04 "for basic information or basic functions it is very useful. Especially driving." P05 "Provides information in a short time with very little effort" P06 "searches and tasks for the home (turn off and on lights for example)".
P07"be informed and do simple tasks (timer / alarm clock)", P08"I think it is useful to give me basic information such as news / forecasts / recipes", P09 "It saves me time, I can do other things while I use it", P10 "streamline activities" P11"It allows you to do multiple things at the same time", P12 "it is quick".

In first time users this is hard to judge if the perceived benefits would balance the perceived risks, due to the fact that people are not aware on the functionalities of the system (that can be perceived as benefits) and are not fully aware of the potential risks of using a voice activated Assistants in the long term.

The people that took part in the questionnaire were frequent users (All of them except one used it more than 1 per day). As can be seen the major use cases were related to house management (lights and devices), requests about general information, time (timer and alarms) and recipes.

PARTECIPANT CODE	DEVICE	FOR WHAT DO U USE IT?	PERCEIVED BENEFITS	PERCEIVED RISKS
UP01	Echo Show8	Domotics	Home helper	privacy
UP02	Echo Show5	Domotics, General information, Recipes, timer, lists	Practical to make daily activities with voice command	Privacy
UP03	Echo Dot + Show 5	Music, domotics	Practical	Privacy
UP04	Echo Dot	Memo e songs	Effortless	Lack of privacy
UP05	Echo Dot	smart home (switch on/off lights luci, ecc...) e general information	Fastness of the replies	The fact that you are always be listen and so the possibility to give them even more data about me
UP06	Echo Dot	Radio	It makes me less lonely	Privacy
UP07	Echo Dot	Switch on/off lights , musics, timer	It is convenient, i can make some activities faster and easier and i do not have to be closed to the device	None, Privacy but i do not worry about it.
UP08	Echo Plus	domotics and general information	Semplicity of rutinary actions	None, we are spied anyway

UP09	Echo Dot	Music, domotics	Everyting connected to voice	To be spied
UP10	Echo Dot	Music and research	Information and controls by voice	At the moment nothing
UP11	Echo Dot	Lights, music and research	House helper	To be controled by the provider
UP12	Echo Show5	Music, weather forecast, audiobooks, alarms and recipes Riprodurre musica	When i am busy and I need to search for somethin	Personal data
UP13	Echo Dot	Smart lights	To get information	None
UP14	Echo Show 5	Alarms	Convinient for routine	Lack of privacy
UP15	Echo Show5	News,Weather, Timer Alarm	Info	None

All the usual users found valid the information given by the assistant, which explain their daily use for general information.
Compared to first time users, the usual users, perceived the assistant as neutral, and a general helper for domestic use.
Compared to first time users, usual users, are more aware of the consequence of the ownership and use of Voice activated assistants, however 3 people mentioned (UP07, UP13, UP15) they are not worried about anything.
Other participants worried about privacy and the use of their personal data and the fact of being spied. By analyzing their personality traits and dispositional trust it is not more than average that would explain the fact that users would adopt this technology for that more than a general higher level of technology acceptance. In this kind of sample, as their trust is already established in this technology, perceived benefits play a good role in perpetuating the behavior.

However, when analyzing more specific questions about the possibility of misuse of personal data, frequent users are less likely to be worried compared to first time users.

Nevertheless, there are factors that depends on user's disposition or characteristics, in the chart below those pre-trust elements should be evaluated when considering the tradeoff balance between risks and benefits.
As it was seen once, the trust relationship is established, even being aware of the implication of voice activated assistant and their risks, if they value it, they will continue using the device.

By analysing user's personality and attitude, we can say that:

- P01, P03, P07, P10 trust the device (high benefits and low risks) and that is supported by general user's characteristics (even if the tendency to trust level is low)

- P02, P04, P8 may trust the device due to personal characteristics however perceived benefits are low, maybe in this case it would be helpful to make more evident how those functionalities can be more useful
- P05, P9, P12 is not very dispositional prone to trust however perceives a low risk and mid/high benefits
- P06 is dispositionally a low truster and the perceived benefits are low and the risks high
- P11 is prone to trust, and perceives it as beneficial however the perceive risks are High

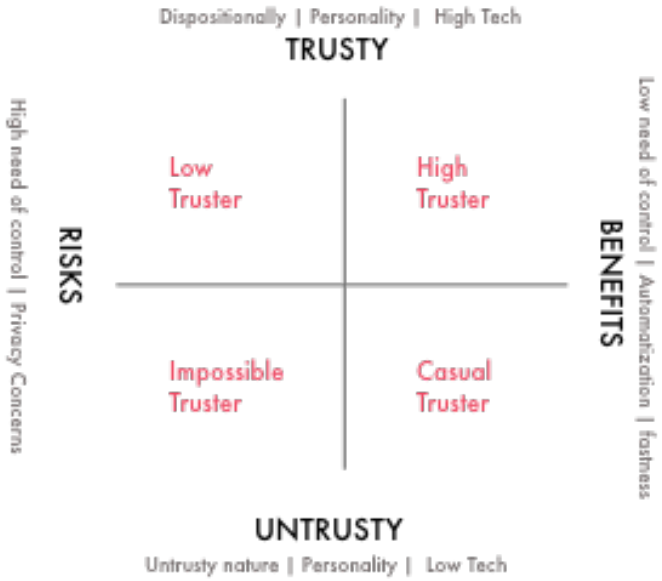
CODE	PERSONALITY	DISPOSITION TO TRUST	ATTITUDE TO TECH	NEED OF CONTROL	OVERALL BENEFITS	OVERALL RISKS
P01	Low N,Intro/extr, Mid Openess, High agreebleness, mid Competence	Low disposition	Good attitude	Middle level	High benefits	Low risk
P02	Low N, Intro/extr, Low Openess, High agreebleness, Mid Competence	High disposition	Middle	Middle level	Neutral	Low risk
P03	Low N, High extr, Low Openess, Mid agreebleness, High Competence	High disposition	Good attitude	Middle level	High benefits	High risks
P04	Low N, Intro/extr, Low Openess, High agreebleness, High Competence	High disposition	Good attitude	Middle level	low benefits	Low risk
P05	Low N, High extr, Mid Openess, Low agreebleness, High Competence	Low disposition	Middle	Low control	mid/high	Low risk
P06	Low N, Low Extro, Low Openess, High agreebleness, High Competence	Low disposition	Good attitude	High level	Neutral	High risks
P07	High N, High extr, Low Openess,	Low disposition	low	Middle level	High benefits	Low risk

	Low agreebleness, High Competence					
P08	High N, High extr, High Openess, High agreebleness, High Competence	Mid disposition	Good attitude	Middle level	low benefits	High risks
P09	Low N, High extr, Low Openess, Mid agreebleness, High Competence	High disposition	low	Middle level	High benefits	Low risk
P10	Low N, High extr, High Openess, MId agreebleness, Mid Competence	High disposition	Good attitude	High level	High benefits	Low risk
P11	Mid N,High Extro, High Openess, High agreebleness, High Competence	High disposition	Middle	High level	High benefits	High risks
P12	High N, Intro/ extr, Mid Openess, Mid agreebleness, High Competence	Low disposition	Good attitude	Middle level	High benefits	Low risk

If we merge these personal characteristics, we'll have the following Truster-types:

- High Truster
- Low Truster
- Impossible Truster
- Casual Truster

That are characterised by the following traits:



Those types can be useful when designing a new voice experience, addressing different level of risks/concerns.

Research Findings in a Nutshell

- DISCOVERABILITY

One of the biggest issues of chatbot and voice activated assistant is discoverability (Jain, 2018). Screen-based device was an advance for the discovery of the functionality of the device especially for those users that maybe were not too familiar with voice activated assistant in general. The device by showing the possible interactions with it, it implicitly teaches the users the correct way to formulate the command, saying for the wake-up word and then asking a question or a fact. Providing also, concrete examples on how to formulate a command that needs to be short and more or less specific. It also shows that even without prior knowledge about a specific device, participants have developed before a mental model about the assistants, in fact none of the participants had concern on interacting with it.

- STATE OF THE SYSTEM

False negative wake-up word

There have been many false negative when detecting the wake-up: the system failed to detect when it was supposed to. This could explain when the user asked correctly a question and said the wake-up word (Alexa) and the system did not reply anything. User behaved in different ways: some repeated the command others frustrated gave up.
In other, especially in the use case of the recipe, in most of the interviews, users complained the fact that it was almost impossible to stop the system and going on with the other tasks.

False Positive wake-up word

There have been couple of case of false positive: the AI agent seemed to recognize a voice command.

ASR Transcription Issues

There have been few cases in which the AI agent transcribed the voice command wrongly. One example was when a User asked a specific supermarket and the system gave the user a different information from the one requested.

NLP Issues

There have been cases of probably NLP issues, when the system did not recognize the user's intent correctly. Apparently, those commands seemed well structured, but the AI agent was not able to provide an answer.

User recognition

In one occasion during use case scenario 2, participant (P07) was misrecognized by the owner of that device. In fact, Alexa said *"Valeria, here the news of the day"*

- SOME USER ARE MORE LIKELY TO GET FRUSTATED

Some users give up more easily than others while asked to perform a given task. Especially when the AI makes some mistakes, they are more likely to give up. However, the users that have had prior experience with voice activated assistant in general are more likely to give shorter and more specific commands compared with people with completely no experience. One user (P04) mentioned the fact that she needed to speak slowlier with the device to make it understand it.

- USER MENTAL MODEL

Overall people seem to understand how to interact with a voice activated assistant even if they have never done it advance. However, for complete first time users (never tried an assistant before) it was more challenging to understand how to ask questions, especially when first the assistant seemed to understand them correctly. For more complex tasks, they happened to be more frustrated and saying, *"we do not understand each other"*.

- LEARNING

Overall participants seemed to be found easy to learn how to interact with voice assistants, even those that never tried before. All understood that in order to interact with it they need to say a wake-up word, however during the interview, often forgot to mention before giving the command. That was particularly evident in those participants that never used it. However, during the experience all participants improved the way they were asking questions.

- SOME USER WANTS MORE TO BE IN CONTROL

Some Participants (P02, P05) preferred to ask about the weather rather than asking an advice to the AI as if they would not want to follow (or trust) the device decision making. *"I am going to ask about the weather so that I can decide what I am going to wear today"*

- SOME USER ARE MORE LIKELY TO HUMANIZE THE DEVICE

People judge the assistant differently in terms humanness: most of the people found it as neutral and others found it artificial. Nevertheless, those participants, behaved politely to the assistant during the tasks and replied to Alexa with thank you. Consistent with the scientific literature in HCI, people tend overall to be polite with the assistant, there have been many cases in which the users said "Alexa, please.." and the command and also when they were given a reply said "thank you"

- USE CASES AND DEVICE

The potentiality of multimodal interaction seems to be more effective in certain context such as the Use case N3 and 4. The experience is different, however due to the fact that in all the participants were all in a house context, focused only on the execution of the given task this might be different in real time situation, in which are not recorded or they are doing something else at the same time.

- PERCEIVED LACK OF UTILITY

Many users reported the use of Voice assistant has been useless especially in first time users, probably due to the fact that they are not aware about all the function abilities. On the contrary the judgements of frequent users (UP group) found that their frequent use is correlated to the perceived benefits and utility.

- REPORTED TRUST

Validity of information

In all use cases people said that they believed the information given by the assistant as meaningful, however, it is important for user to have control on the kind of information channel is broadcast.

Perceived benefits

Contextually especially in task 2 and 3, people reported some positive benefits to use such devices. Especially due to the fact that can be immediate, and they can do other things while interacting with the assistant.
However, as it as showed by the usual user participant, perceived benefits and utility can compensate their risks that they all seem aware.

Perceived Risks

Contextually there are few perceived risks in the use cases analyzed:
In Use case task 1 none reported risks
In Use case task 2 only a participant (P08) reported the fact that that would have prefer to change the news channel because he trusts more the other information channel. Others reported risks in terms of misunderstanding especially for Use case 3, in which the user found difficult to stop and achieve their intent.

Trust in the system is influenced by brand reputation
Most of users were found to trust the brand and perceive little risks in interacting or buying through it.

- AFFINITY MAP

In order to organize and structure more the transcribed interviews, the data have grouped in an Affinity diagram, with the aim to identify further relationships and to organize insights from the interviews. An affinity diagram is a tool that help to drill down tasks, behaviors and goals, in this case from the interviews and the observations. The affinity diagram below shows in each tab a piece of the interviews which have been to clustered information.

Errors	Usefulness	Functions Awareness	System states	Companion	Frustrations
I asked something specific and she gave me a different reply	When I am busy doing else	Music	Alexa... Alexa. Did you listen?	It gives you company	We do not understand each other
Why is broadcasting a song now?	When My hands are doing something else	Timer/Alarm	Is it working?	She is cool	No.. Alexa I asked you something else.
She called me Valeria!	I can multitask and it is quicker	Weather/news	Alexa, Are u there?	She knows many things	I asked something specific and she gave me a different reply
It just give me error message all the time	No effort to get information needed	Doomotics (lights/tv/plugs)		Thank you Alexa	It impossible to stop her while she speaks
How you do not know it?	For easy tasks can be useful	General questions		hey alexa that is enough thank you	I am not sure wether I do not ask in a correct or it just does not get me
	It can gives company	Many said I do not know			It can be useful however there are cases in which i had to ask the same thing 5 times vbefore getting what i need. In these cases it is a waste of time
	You can control things by distance				

Surprises	Interaction	Source of information	Risks	Brand	Appereance
The device speak!	I do not know what to do	I prefer to decide myself the source of information	My Privacy	I like amazon I use it all the times	It is neutral (dot)
She knows that we are in finale ligure and she mentions it directly	Can I read the instruction?	It is too detailed, i just wanted a short recap	Personal data to third services	It is fast and convinient	It is nice i like the texture (dot)
I asked something specific and she gave me a different reply	To use it I would need to watch a tutorial	Can I change it? i prefer something else	I do not like the idea of being spied in my home	I do not use it very often but just because i prefer a retail store	I do not like the wire (both dot/show)
	I have never used, I do not know how	I have never used, I do not know how	voice recording anf Sensible data on the cloud	I do not like it	The screen is too small
	I touch I does not happen anything	I touch I does not happen anything	Same as having a smartphone	It is great, I buy something today and torrow i would already have it	I like it, i can see it in my home
	I am trying to connect to bluethoot but I can not find the device	I am trying to connect to bluethoot but I can not find the device	If it used in the right way none	Never had any problems, I can always send back things	I like the screensaver
	Alexia		Advertisiment that is targeted on me	It is useful when I am busy and I can not go to the store	
			Passive listening		

- PROBLEMS

From the previous analysis of the interviews and the affinity map it was clear that there are similarities in the perceived problems about usage and ownership of voice activated assistants.

Users are only using Alexa for the most basic of tasks andare not aware of functionalities.(e.g. Shopping/ lists)	Most of the people are concerned with privacy.	Some People are concerned with the exploitation of personal data.	Some people are worried that the device would switch on and record their conversation without having the intention to.
People need instructions in order to try.	People often forgot the wake-up words.	People found Alexa sometimes hard to stop.	Some people get frustated and give up more easily than others especially when they are not understand.
Some people do not know how to talk with Alexa the first time	People need to have the control over the information they have been provided.	Lack of utility: The perceived usefulness or certain assistant's skills/actions often depends on personal factors.	False negative wake-up words
False positive wake-up words.	Some people did not understand system's states.	Wrong user's recognition	High expectations about the assistant.

- PERSONAS

In order to design a trustworthy assistant, we need to put the final user at the center of the project: for this reason, the creation of Personas is necessary to finalize the data from the interviews. Personas are the representation of user's archetypes considering a scenario, their and needs. Although Personas are fictional, they can help to develop effectively a product/service so that the requirements are met due to the fact that they are created from the research data. Also, in these cases personas have been implied to share the research findings and insights in more accessible way.

From personality evaluation, the propensity to trust and attitude towards technology it has been identify 4 different trust archetypes in relation to adoption and ownership of a voice activated assistant device.

Lucia Impossible Truster

Age 65
Location Savona
Education High School
Occupation Teacher

"I do not want a Big brother device in my home, It is useless for me"

Personality Traits

Agreebleness
Extroverted
Open to experience
Neuroticism
Competence

Disposition to trust

High

Technology

Low level of technology acceptance

Internet	●●●●●
Social Media	●●●●●
Online Shopping	●●●●●
Gatgets	●●●●●
Early Adopter	●●●●●

Biography

Lucia, is a primary school teacher, her daughter gave an amazon alexa device for Christmas with the pupose of entertain her. However, the mother is not willing to try use it. She believes that she does not need such technologies in her house. She also does not like the fact that these devices could record her personal data.

Motivation

To make her daughter happy

Needs

following daugher suggestions, intuitiveness

Pain Points

High need of control,
Technology is not intuitive for her
She prefers more traditional/Known ways rather than adjust

Pain Points

High need of control,
Technology is not intuitive for her
She prefers more traditional/Known ways rather than adjust

Frustration

She gets annoyed when things do not work the way she wanted

Perceived Threads

- Perceives high risks for the ownership
- Perceived huge problems related to Personal data usage and possibility to share this information
- Do not think that his data could be stolen
- He is not particolarly worried to be spied by him

Perceived Benefits

- For what she undestood, there are no real useful functionalities

Andrea Low-Truster

Age 25
Location Roma
Education Master
Occupation Analyst

"This technology can be fun to use if it works propery, however there are many cases in which we do not understand each other."

Personality Traits

Agreebleness
Extroverted
Open to experience
Neuroticism
Competence

Disposition to trust

High

Technology

Good level of technology acceptance

Internet	●●●●●
Social Media	●●●●●
Online Shopping	●●●●●
Gatgets	●●●●●
Early Adopter	●●●●●

Biography

Andrea is a business analyst, he wants to get a voice assistant at his place to manage all his devices at home and to play with his kids. However he needs to be sure this technology is not risky to own it or harmful for his kids.

Motivation

Domotics, Play
Get information quickly

Needs

Do not give too many personal details, to know the kind of usage their data are going to be to be implied for, Easiness to instal devices

Pain Points

Privacy, data usage.

Frustration

- Medium need of control,
- Do not like Loosing time
- Not being understood
- Too repete the same thing overtime

Perceived Threads

- Perceives High risks in usaging this technology
- Worries about targeted advertisement
- Not being understood
- he does not feel safe of owning it

Perceived Benefits

- Easy to use
- Not useful for his job
- Convenient For repetitive tasks

Laura Ambivalent-Truster

Age 44
Location Turin
Education Master
Occupation Copywriter

"This device would make me feel more safe at home"

Personality Traits

Agreebleness
Extroverted
Open to experience
Neuroticism
Competence

Disposition to trust

High

Technology

Good level of technology acceptance

Internet	●●●●○
Social Media	●●●○○
Online Shopping	●●●○○
Gatgets	●●●○○
Early Adopter	●●○○○

Biography

Laura is a Copywriter living in Turin, after the divorce is willing to try differnt new things. Since she lives by herself, a friend of her suggested to get a voice assistant in order to feel more safe with cameras and security devices.

Motivation

Entertainement, Companionship
Domotics

Pain Points

Not patient

Needs

To feel in control and safe in her house, to feel connected while she is alone

Frustration

- Medium need of control
- She needs to repete few times to be understood
- Loose interest easily

Perceived Threads

- Perceives average risks in usaging this technology
- worried the device wake up by chance and record information
- Not being understood.

Perceived Benefits

- Entertaining
- Engaging
- Company

Simone High-truster

Age 23
Location Milan
Education Bachelor
Occupation Engineer

"I am always willing to try new tech and see how I can simplify my life"

Personality Traits

Agreebleness
Extroverted
Open to experience
Neuroticism
Competence

Disposition to trust

High

Technology

High level of technology acceptance

Internet	●●●●○
Social Media	●●○○○
Online Shopping	●●●○○
Gatgets	●●●●○
Early Adopter	●●●●○

Biography

Simone is student, he is a new user to Amazon Alexa, he is interested in the idea of home devices with voice activated assistant. By nature he is very trusty, despite the fact he is aware about the possible cpnsequences in the ownership is not partuicularly worried about his pravacy as long as he could have control over it.

Motivation

For fun, to play with , to be rembererd about things and tasks to do

Needs

To be more organized, To get quick information while he is working.

Pain Points

Not receving the correct answer, to loose time rather than shortening.

Frustration

Get too many details about things he does not needed, advertisement, long processes to achieve simple things

Perceived Threads

- Perceives Low risks in ownership and usage no matter the screen or the voice first device
- Perceived no particolary problems related to Personal data
- Do not think that his data could be stolen
- He is not particolarly worried to be spied by him

Perceived Benefits

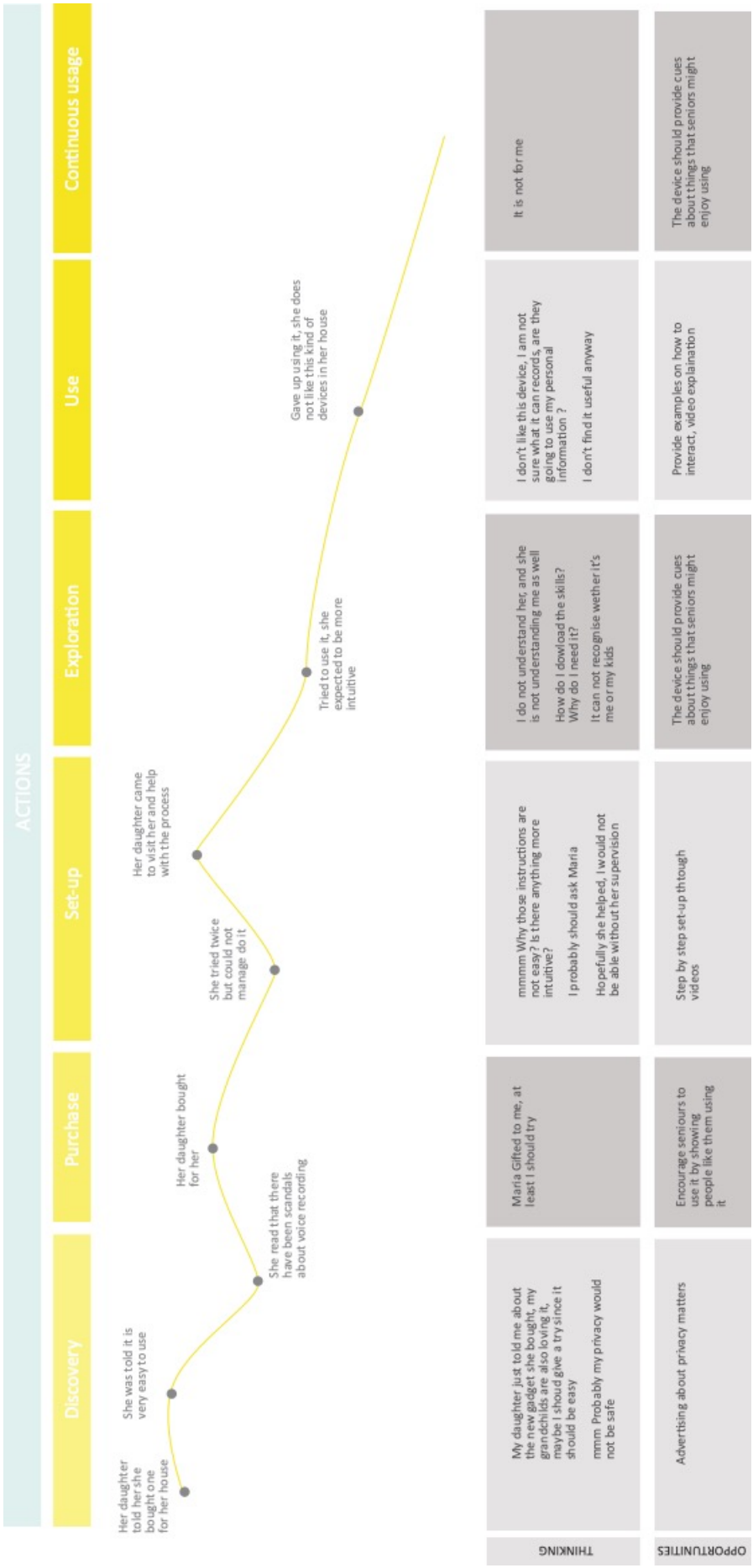
- Fastness
- Immidiancy/easiness
- Entertaining to use
- Convenient when busy
- Convenient For repetitive tasks

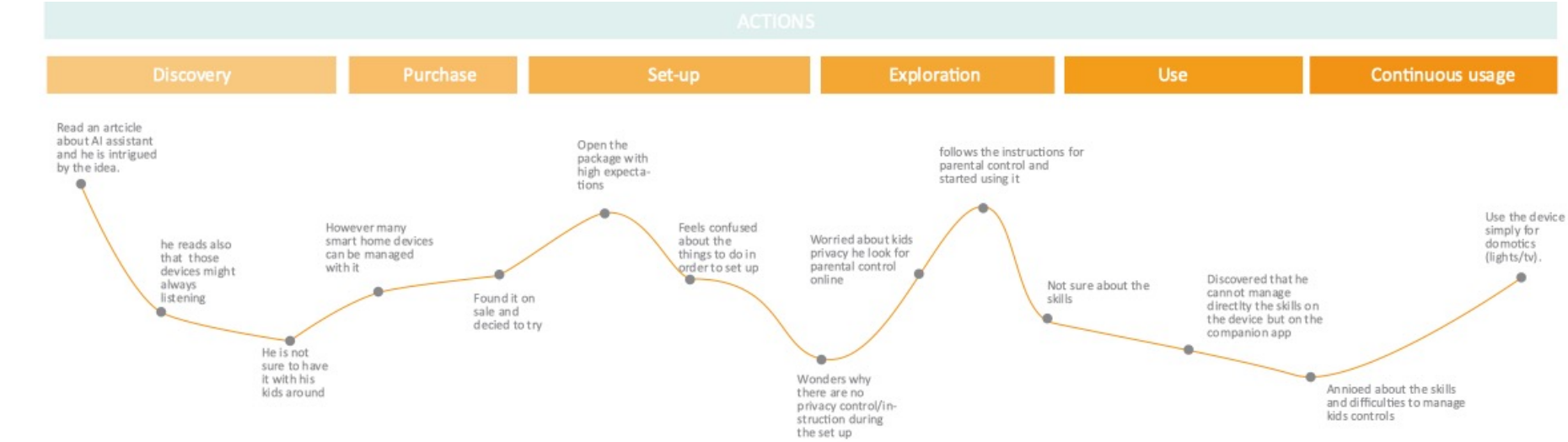
- CUSTOMER JOURNEYS

The term Customer Journey address the processual and experiential elements of a product/ service/system from the point of view of the costumer. It is described in terms of touchpoints during a continuous interaction between the service and the user that is engaging it. A journey map is a tool that makes visible the experience of a person in a temporal aspect in relation to a product/service. The Costumer journeys below show how potential profiles engage with Amazon Alexa device over time. The analysis of User's Journeys of the identified personas made me aware about all the actions in the adoption of a voice activated assistant from its discovery to long term usage. The maps below show user's pain-points and possible suggestions to solve them.

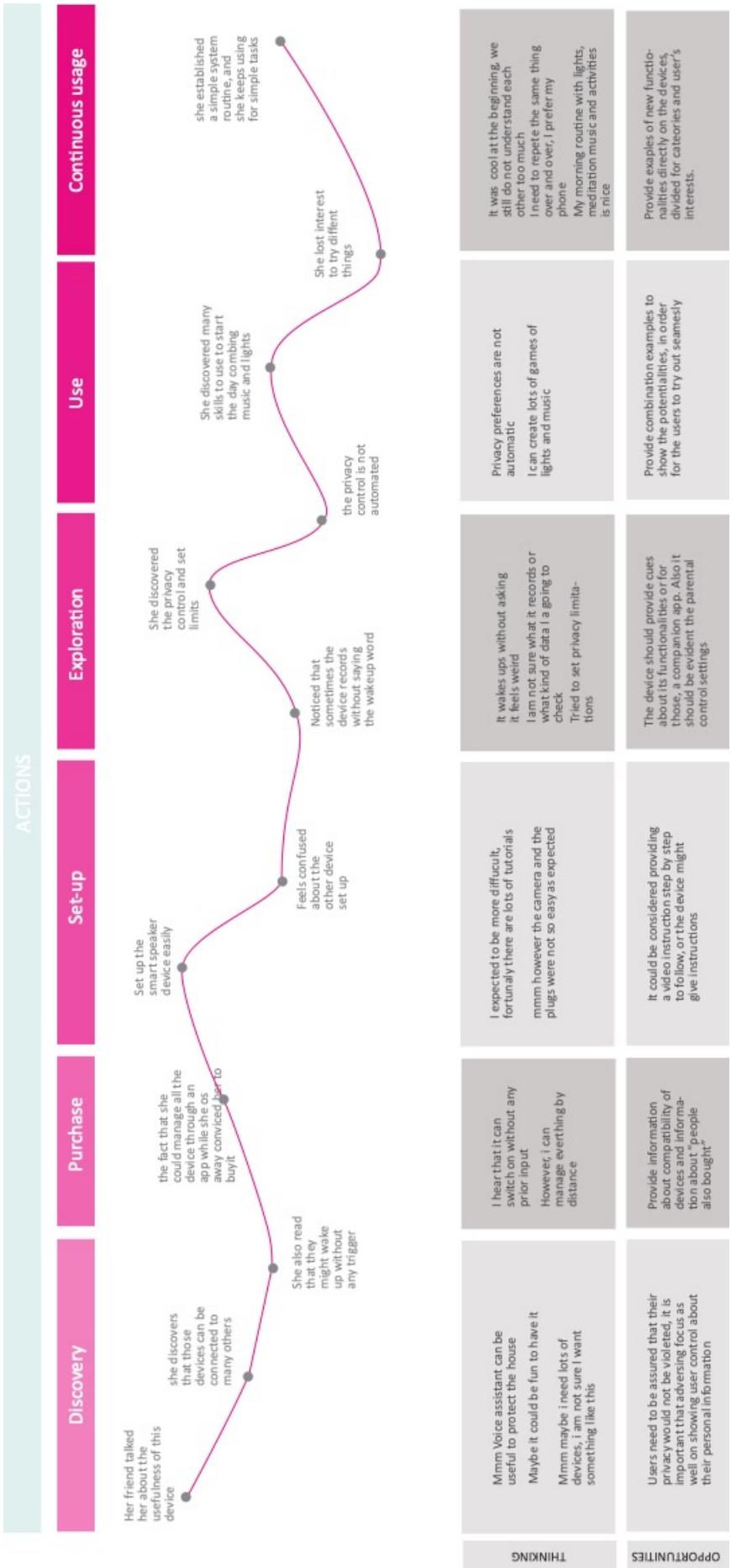
Many opportunities emerged from the interviews conducted:

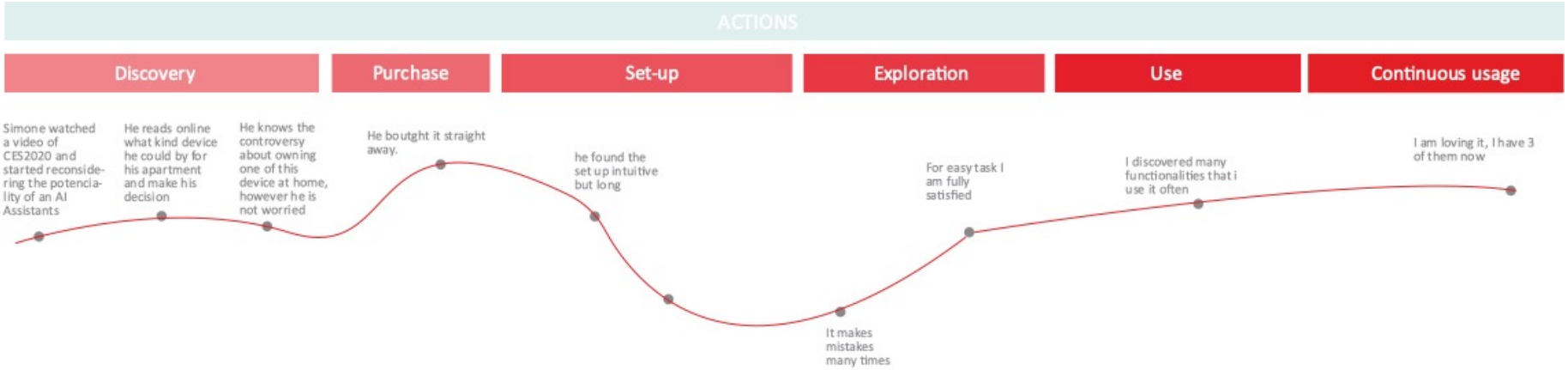
- Limitation and control of personal data
- Encourage adoption through engagement and guidance
- Encourage social and cultural sharing
- Offer them control
- Remind to new users the possible interactions and functionalities
- Learn from other people while executing tasks
- Being accessible to everybody and everywhere


ACTIONS
Discovery
Purchase
Set-up
Exploration
Use
Continuous usage
Her daughter told her she bought one for her house
She was told it is very easy to use
She read that there have been scandals about voice recording
Her daughter bought for her
She tried twice but could not manage do it
Her daughter came to visit her and help with the process
Tried to use it, she expected to be more intuitive
Gave up using it, she does not like this kind of devices in her house
THINKING
My daughter just told me about the new gadget she bought, my grandchilds are also loving it, maybe I should give a try since it should be easy
mmm Probably my privacy would not be safe
Maria Gifted to me, at least I should try
mmmm Why those instructions are not easy? Is there anything more intuitive?
I probably should ask Maria
Hopefully she helped, I would not be able without her supervision
I do not understand her, and she is not understanding me as well
How do I dowload the skills?
Why do I need it?
It can not recognise wether it's me or my kids
I don't like this device, I am not sure what it can records, are they going to use my personal information ?
I don't find it useful anyway
It is not for me
OPPORTUNITIES
Advertising about privacy matters
Encourage seniours to use it by showing people like them using it
Step by step set-up thtough videos
The device should provide cues about things that seniors might enjoy using
Provide examples on how to interact, video explaination
The device should provide cues about things that seniors might enjoy using



THINKING	AI assistants seems a nice gadget to have at home I do not want the devices listening me and my family They can works for house management	I should check if there are special restrictions for kids Mm Whatever it is very cheap maybe I should give a try anyway Well, it would not be safe, I will just switch it off.	Why there are not instruction? Is it always on? mmm I don't like this Why there are no privacy controls? what should I do?	I do not know what should I ask to alexa How do I dowload the skills? Why do I need it? It can not recognise wether it's me or my kids	mmm I can not manage the skills on the device i don't want another app Mmmm.... I can not put parental controls on the skills Mmm I am not sure I want my kids to interact with this	I Will just use alexa to manage my Home smart devices I need to repete the same thing over and over, I prefer my phone Mmm I am not sure I want my kids to interact with this
OPPORTUNITIES	Competitive analysis can compare functiona-lities among different smartspeaker and identify and uniqueness	It could be considered a tematic Advertisiment for domotic devices	It could be considered providing a video instruction step by step to follow, or the device might give instructions	The device should provide cues about its functionalities or for those, a companion app. Also it should be evident the parental control settings	Provide kids mode The device should recognise the identy of the user that made the request	Additional privacy setting should be available for those users that need


ACTIONS
Discovery
Purchase
Set-up
Exploration
Use
Continuous usage
Her friend talked her about the usefulness of this device
she discovers that those devices can be connected to many others
the fact that she could manage all the device through an app while she os away convinced her to buyit
Set up the smart speaker device easily
Feels confused about the other device set up
She also read that they might wake up without any trigger
Noticed that sometimes the device records without saying the wakeup word
She discovered the privacy control and set limits
the privacy control is not automated
She discovered many skills to use to start the day combing music and lights
She lost interest to try diffent things
she established a simple system routine, and she keeps using for simple tasks
THINKING
Mmm Voice assistant can be useful to protect the house
Maybe it could be fun to have it
Mmm maybe i need lots of devices, i am not sure I want something like this
I hear that it can switch on without any prior input
However, i can manage everthing by distance
I expected to be more difficult, fortunaly there are lots of tutorials
mmm however the camera and the plugs were not so easy as expected
It wakes ups without asking
it feels weird
I am not sure what it records or what kind of data I a going to check
Tried to set privacy limita-tions
Privacy preferences are not automatic
I can create lots of games of lights and music
It was cool at the beginning, we still do not understand each other too much
I need to repete the same thing over and over, I prefer my phone
My morning routine with lights, meditation music and activities is nice
OPPORTUNITIES
Users need to be assured that their privacy would not be violeted, it is important that adversing focus as well on showing user control about their personal information
Provide information about compatibility of devices and informa-tion about "people also bought"
It could be considered providing a video instruction step by step to follow, or the device might give instructions
The device should provide cues about its functionalities or for those, a companion app. Also it should be evident the parental control settings
Provide combination examples to show the potentialities, in order for the users to try out seamesly
Provide exaples of new functio-nalities directly on the devices, divided for cateories and user's interests.



	Discovery	Purchase	Set-up	Exploration	Use	Continuous usage
THINKING	It's really the year of voice assistan-ce, I wanna buy one I am not sure exactly what to buy, I do not need anythings super powerful I would just what something to play/or-ganize my things I do not care, I do not have anything to hide.	I will buy it, It is very cheap recently	So many steps, I just want to try it mmm however the camera and the plugs were not so easy as expected	It wakes ups without asking it feels weird I am not sure what it records or what kind of data I'm going to check Tried to set privacy limita-tions	There are many things to do, however I would like an easier way to discover new functionali-ties	I am loving this devices, I have one for each room I set all my preferences
OPPORTUNITIES	Competitive analysis can compare functionalities among different smartspeaker and identify and uniqueness	Possible promotions for younger genera-tion/students	Set up minimal and adding informa-tion about privacy and only later giving the possibility to set many more preference in the device	The device should provide cues about its functionalities or for those, a companion app. Also it should be evident the parental control settings	Provide exaples of new functio-nalities directly on the devices, divided for cateories and user's interests.	Keep the user engaged and provide new functions in relations to what he is using.

Chapter 6

The concept

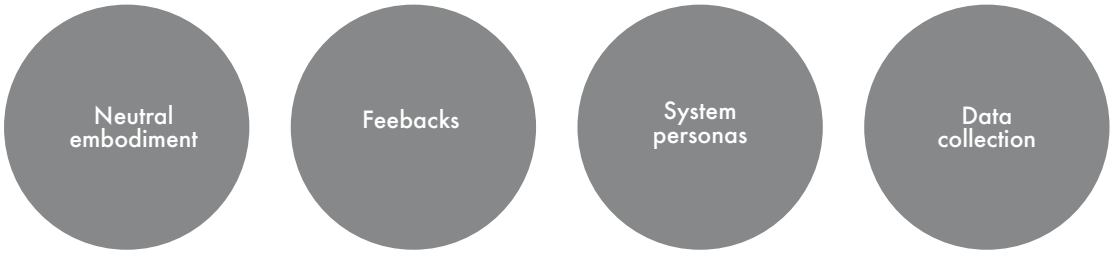
From the analysis of Consumer's journeys of each persona I understood all the actions made from deciding what assistant to get, their set-up and their continuous usage and decide what specific needs to solve with the design solution. Starting with the Idea to re-design a voice activated assistant embodiment and their first-time usage, I understood that the opportunities are not for all the identified personas. In fact, it is possible that some people would never trust or would never have the interest to use this technology in long term, due to dispositional factors and attitude to such technology.
If we take the example of the "Impossible Truster" despite the effort to be clear about the consequences of her actions, or to guide more her interaction while using this technology she might not perceived it as useful anyway. Why should she change her behavior when there are no apparent benefits? None.
Also, in the case of the "High Truster", I do not see huge opportunities since the user is intrinsically motivated to engage with the assistant, he perceived high benefits while interacting and using it even in the long term and at the same time is not too worried about the consequences. Probably the challenge is to keep engaging the user in long term, which however is not dissimilar to the "ambivalent truster" type.
It was this that made me realize that my aim would be focused in changing the perspectives of these users that might be worried engaging with this technology due to the fact they lack of trust in the assistant, in its ability and in their possibility to control it. Since the scenarios od the set-up and first time usage offer showed the user's needs to be more guided and engage in this process, I believe this is fundamental to establish a correct mental model on how to interact with the assistant providing information of the possible ways to interact with this device.
Considering this I decided to focus that this project will be targeted potential user between 18 to 50, with a mid-level technology adoption, curious about the product but however struggles to trust it. In fact, since many problems and opportunities emerged from the interviews conducted and the challenge can be sum-up in the following way:

"How to build trust in voice activated assistant in a first-time interaction?"

The objective of the project is to design a an assistant that is able to establish trust in a first time interaction that aims to improve their adoption by increasing their perceived trust. The assistant should have a graphical agent that guides the new users to its system, its functionalities and their preferences. Starting from user's needs and goal detected, the final solution would be the creation of a voice activated assistant and of an onboarding experience that through their embodiment and its transparency about user's personal data would let people discover them and sustain their usage in the long term. The concept is built up upon the background theor, analysis of case study and the users interview.

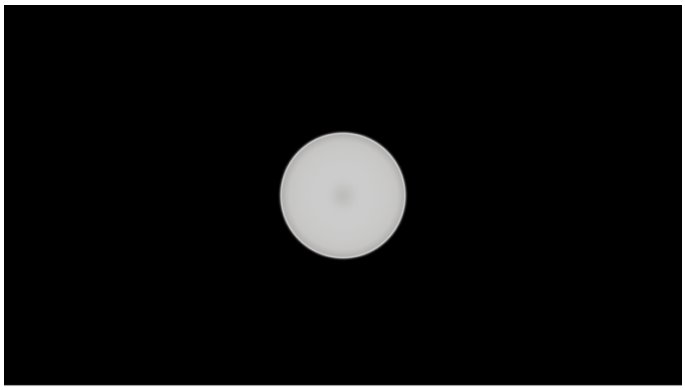
Functions and elements

People reported to struggle to trust AI assistants and as mentioned in the research analysis this can be explained in many different ways. By combining all the studies on this topic, the concept to achieve the design objective should contain following elements:



- **NEUTRAL EMBODIMENT**

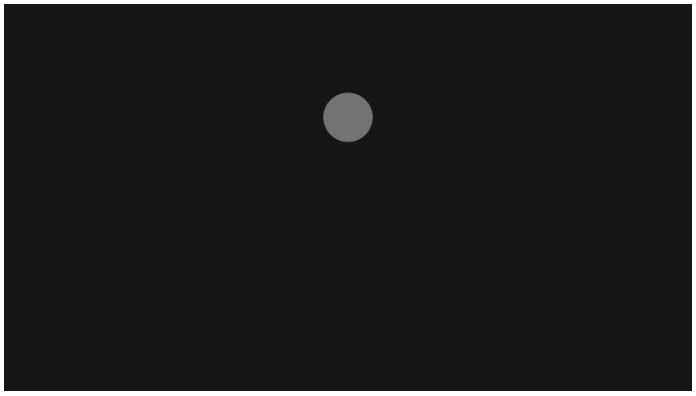
Embodiment can increase the likability of the assistant, however *"The more human-like a system acts, the broader the expectations that people may have for it."* (Cassell, 2000).
The decision to keep the assistant neutral aims to decrease the tendency of human stereotypic attribution. The designed assistant is



- **FEEDBACKS**

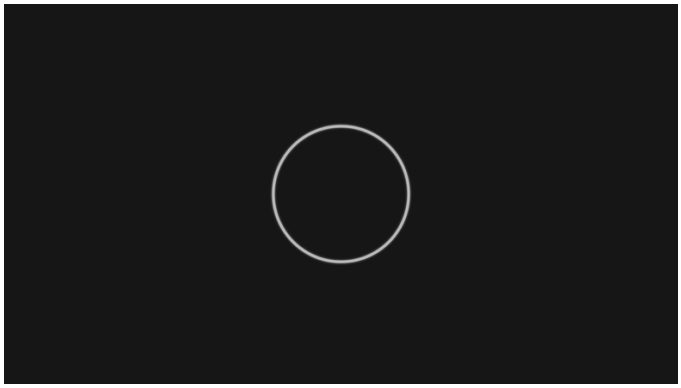
Feedbacks in VUI are fundamental due its intangibility, the system states need to be clear and evident. Voice-only devices might be harder to trust due to the fact that the visibility of the system and feedbacks are more limited making agent's perceived trustworthiness lower by decreasing its predictability and perceived functionality.
A voice activated assistant must show the following states in order to be perceived as more trustworthy:

  - Wake-Up. When the user says the activation word/name, the assistant should show explicitly that have detected it and is ready to start listening to user's command. This visual indicator is fundamental due to the fact that those devices typically are always on, by showing that from now is going to possible record the conversation.



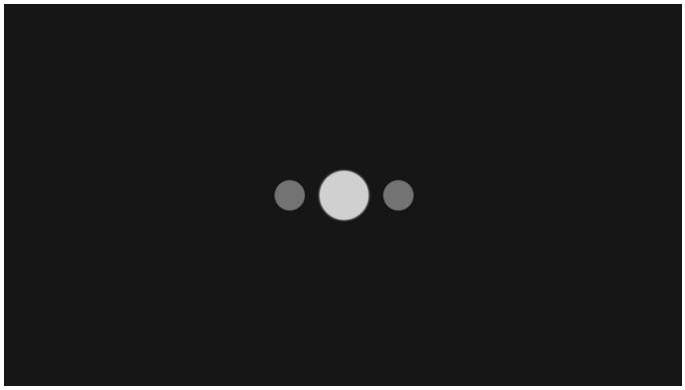
*files animated assistant gifs : https://drive.google.com/open?id=16w9y_2ngDUKYsexRPMbSa_S_-Bfc4cbM

- **Listening.** The assistant should have a visual indicator that inform the user that it is listening user's request



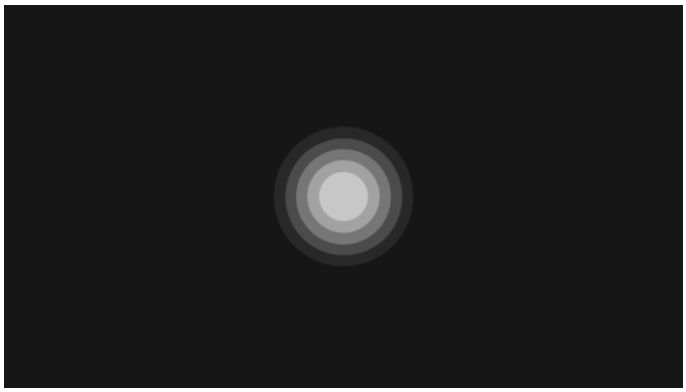
*files animated assistant gifs : https://drive.google.com/open?id=16w9y_2ngDUKYsexRPMbSa_S_-Bfc4cbM

- **Thinking.** The assistant should have a visual indicator that inform the user that it is proceeding information and his request



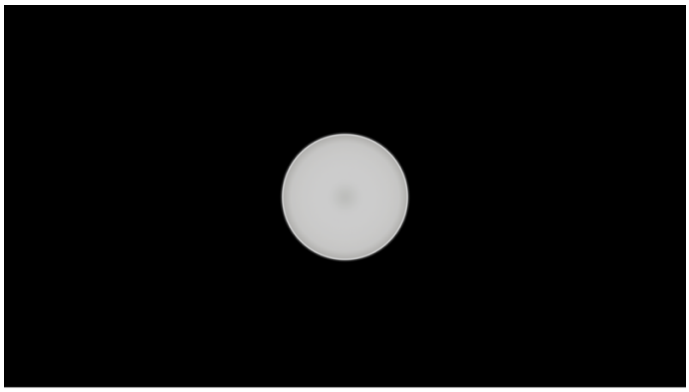
*files animated assistant gifs : https://drive.google.com/open?id=16w9y_2ngDUKYsexRPMbSa_S_-Bfc4cbM

- **Speaking** The assistant should have a visual indicator that inform the user that it speaking to him



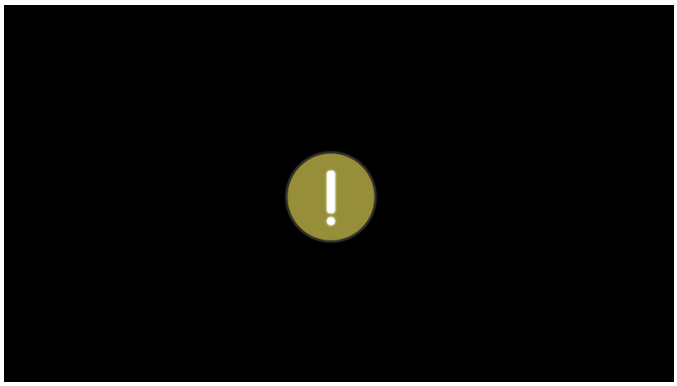
*files animated assistant gifs : https://drive.google.com/open?id=16w9y_2ngDUKYsexRPMbSa_S_-Bfc4cbM

- **End**. The assistant should have a visual indicator that inform the user that it done with the request.

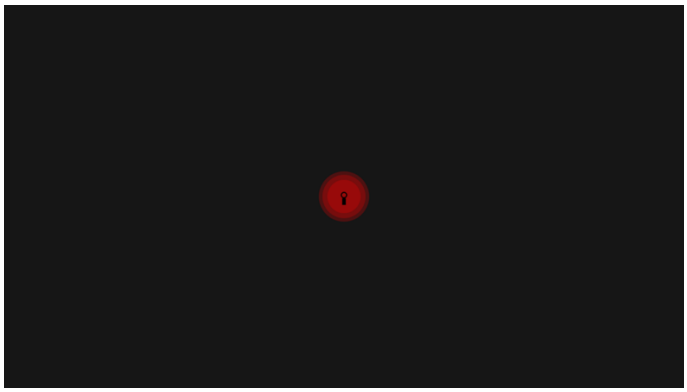


*files animated assistant gifs : https://drive.google.com/open?id=16w9y_2ngDUKYsexRPMbSa_S_-Bfc4cbM

- **Error state**. The assistant should have a visual indicator that inform the user it did not understood what the user said.



- **Disactivated state.** The assistant should have a visual indicator that inform the user it not active. The indicator could be switched on when the user would like to be sure the assistant cannot be activated by chance or unintentionally.



- **SYSTEM PERSONAS**

There is no such thing as a voice user interface with no personality. (Cohen, 2004)

Anytime that there is a human voice people would unconsciously attach a personality to it, so it is needed to design it accordingly to user/business goals. The design of a personas is particular important for a brand to project a certain corporate image. When designing an avatar is fundamental to think about its personality. As GAFMA companies, that made the decision to keep their assistant as abstract the focus would be on designing a congruent and consistent personality.

Function

None is a supportive teacher, it walks you through initial set-up to long term use, figuring out and be proactive to solve user's struggles.

Portrait

- None is a funny, friendly supportive assistant that is eager to help out in house management and entertainment. She is passionate about finding the best solution to each different user.

- None hates wasting time doing things that can be more efficient and automatized, in fact she is willing to provide suggestions in order to correct and shape user's request.

- None is very fair; she always mentions her intentions and give reasons to its questions and her functions.

Personality

- None is empathetic and friendly; it is able to adapt to be both entertaining and be always supporting user's demands and preferences

Traits

- Agreeble and extroverted: None is a very outgoing teacher, she mixed humor and jokes

- Versatile: None can help in countless ways, from initial set-up, to house management, to entertain and to give information

- Encouraging: None remembers user's previous requests and usage, and tries to encourage user to try different things and improve their learning in a fast and entertaining way

Gender:

As mentioned, to avoid overestimation of assistant's skill its embodiment neutral and abstract.
In order to avoid gender biases the assistant is kept as genderless and with a neutral name "Nemo" (from Latin None), however the voice is female due to human tendency to prefer woman voice.

Speaking style:

None speaking style is warm and rapid in order to convey relax and support.

* Physical embodiment of the assistant

NAME Nemo

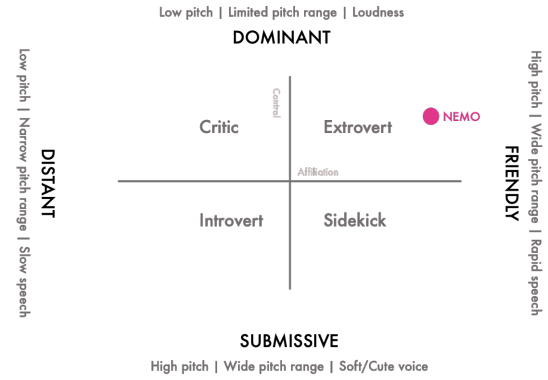
GENDER Female

AGE Young Adult

DESCRIPTION OF PHYSICAL EMBODIEMNT Abstract embodiment

AGENT'S ROLE

AGENT'S VOICE AND ROLE



PERSONALITY

- **DATA COLLECTION (AND USAGE)**

People are especially worried about the collection and the usage of their personal data while interacting with a voice activated assistant. The agent should be able to be informative about it and provide controls over it. It was clear that users were concerned about their personal data and the possibility of device to always be switched on. Few things can be made to decrease user's perceived risks:

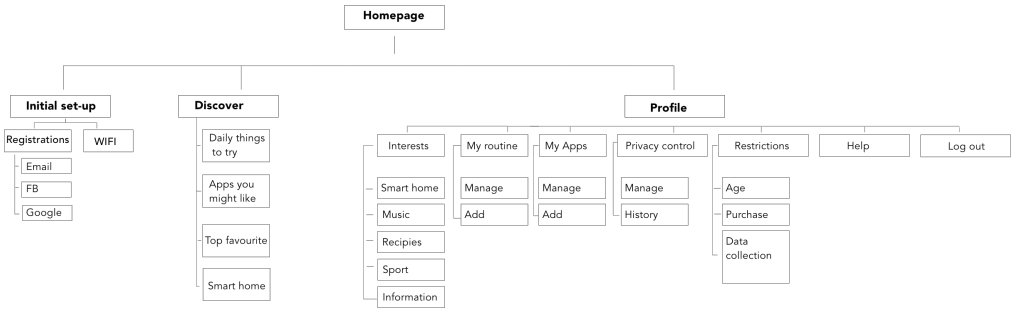
- Delete after a certain number of hours the data collected automatically
- The data collected could be sent to third parties only when the user requested a service
- Also, the user can delete manually the data about the service request
- Make aware the user about the possibility to switch off microphone and screen (the use of the assistant should be voluntary)
- Make clear the indicator of the interaction (wake-up word and cues indicator or listening/responding)

The interface

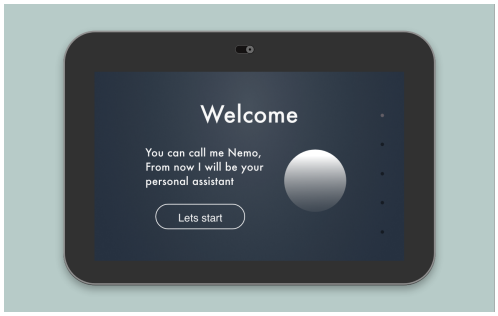
Information Architecture

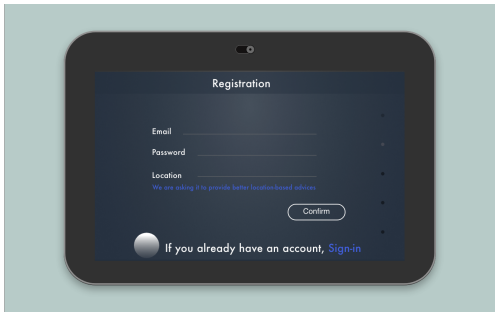
The first step that lead me concretize according to the established objectives was the development of Information architecture, that lead to the creation of some of the interfaces. Information architecture is a common tool to organize content and information of websites and apps. The intent is to define a structure to establish possible ways in which users can navigate and successfully interact with it. The final information architecture of the device follows the structures of the existing competitors: settings management, discoverabilities features and applications. However, it was decided to give a different focus by considering user's needs and potentialities from identified Costumer Journeys. Personal control of download apps/ functionalities and privacy management were considered primary elements, that would simplify the way the user could access to this information. Also, It was decided to focus on giving this potentialities in the device itself rather than to a companion app, especially for the user that would only have one device, in fact as it was found in the "usual user questionnaire" they would not use the application after general set-up or while they need to add more devices. Taking into consideration the defined Information Architecture it was made the paper

structure of wireframes, that are the graphical representation of the interface. Wireframes helps to understand the concrete steps.



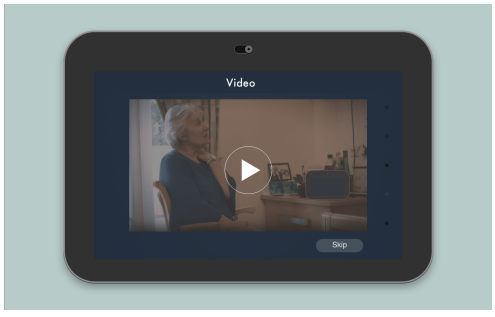
Set-up Interfaces





WELCOMING
The assistant introduce itself to the user before setting the device up

REGISTRATION
The assistant provides suggestions

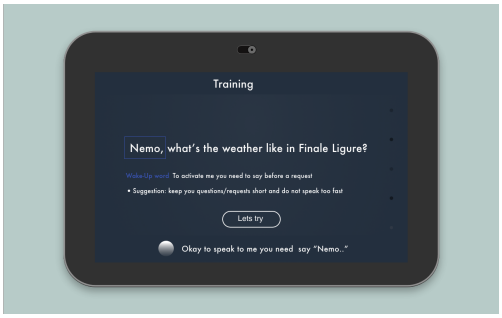


WIFI SET-UP

VIDEO
The video shows the potentialities of the device and how to use it

PRIVACY SETTINGS

Three possible Privacy settings and the characteristics that the user can decide.

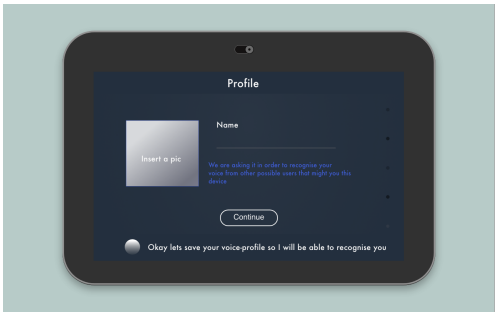


WAKE-UP TRAINING

First training wake-up, it shows and suggest user how to interact with the device, and at the same time the system recognize user's voice

FAILED ATTEMPT

SUCCESFUL ATTEMPT



VOICE PROFILE

The user needs to insert tìhis/her name so that the assistant would not misunderstood his/her with other users that might engage with it

FIRST TIME DISCOVERY

First training guided functionalities

Chapter 7

How to design trustworthy AI assistant?

Voice is a natural interaction; however, voice communicates:

- Gender
- Culture
- Personality
- Emotions
- Intentionality

Along with these, stereotypes might arise, so in order to design VUI it is necessary to take a closer look at the kind of consequences it might have. It is fundamental to be aware about the strengths and the weaknesses of VUI, on the one hand very efficient as an input modality, on the other poor in terms of output. When needed man should consider other modalities, such as visual to add more complex output.

Voice design is a matter of balancing:

- Device/assistant characteristics
- Context characteristics and limits
- User's characteristics

Building trust in Voice Assistants in first time interaction Guidelines

There are many Guidelines for VUI that can be found online. I based the analysis of this current work upon those ones more related to aspects of design rather than on the conversational linguistics aspects. One that I found particularly interesting was a combination of different VUI guidelines made by Ben Sauer, the design director of Babylon Heath and writer of Designing VUI, that included the following:

- Google Conversation Design guidelines
- Alexa Design Checklist
- Voysis
- Don't Make Me Tap!

- Designing Voice User Interfaces Cathy Pearl
- Apple Human Interface Guidelines on Siri
- IBM Conversational UX Guidelines
- Microsoft Principles on Cortana Skill Design

These can be found at the following website: https://voiceprinciples.com/

Despite the importance of Trust as a primary base for interaction and adoption, not many of the current guidelines explicitly target trust. However, it was clear to many designers the need to build trust in conversational voice activated assistants. It was fundamental to deal with user's expectation, predictability and consistency behavior of the assistant, and trustworthy cues.

Understanding the context and the end users is the first step to undercover opportunities and potentialities to build relational trust between the users and the assistant in the selected context.

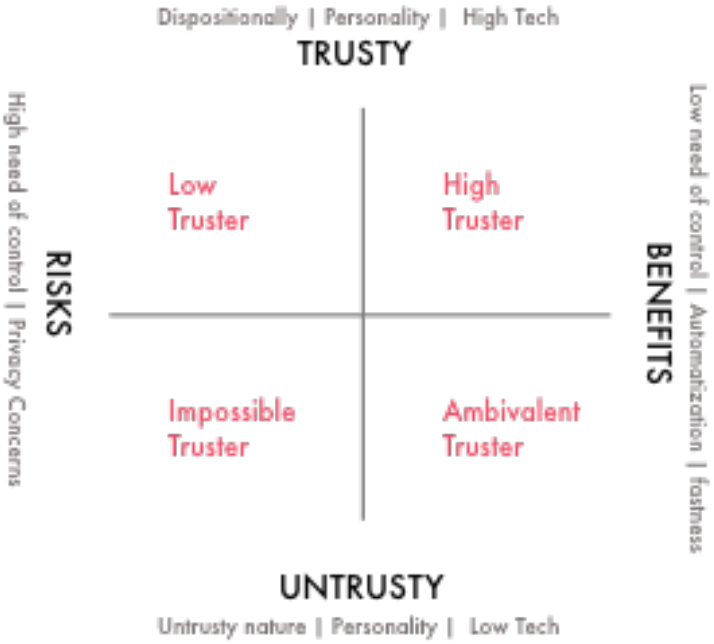
G1. UNDERSTAND YOUR END USERS

"You Are Not the User: […] users have different backgrounds, different experiences with user interfaces, different mindsets, different mental models, and different goals. They are not us." (Raluca Budiu, 2017)

First thing first, in order to design a Voice user experience, the designers need to understand user's needs and goals in the selected context. It is crucial to understand what are benefits and risks perceived by the user in the context and the functionalities.

Understanding user's perceived benefits and risks is fundamental in order to build trust into the technology and in voice activated assistant. User's trade-off (benefits – risks) corresponds to their level of trust. The correspondent level of reported trust is possible to identify and create user trust profiles that combines the user's personality traits, disposition to trust and attitude to technology.

When designing for trust, the design solutions can be different for each of these profiles: High Truster, Low Truster, Ambivalent Truster and Impossible Trust. And it is possible that some of them will never trust it, such as the Impossible truster.



Once done, identify the use case and functionalities you want to create for your use cases' matrix and address how your customer would interact with it.

	Impossible truster	Low truster	Ambivalent truster	High truster
Always				
Most of the time				
Some of the time				
Almost never				

G3. ATTENTION TO THE EMBODIMENT/CHARACTERIZATION OF THE OBJECT OF TRUST:

A Voice Experience is not complete without attaching a persona to the VUI because it would happen implicitly: as Nass (2016) says *"When people hear any voice they automatically and unconsciously assign a personality to it"*.

In order to build trust, it is crucial to focus on AI Agent/Voice Activated Assistant characteristics.

Also, to design a trustworthy voice assistant, the designer needs to make decisions about how: name, gender, role and personality and at the same time it should reflect brand's values and characteristics. All of these have an impact on the perception of the assistant, and the designer should be conscious about the positive/negative consequences of them. It is crucial to create a consistent personality; users are able to trust AI agents that can predict their behavior.

An important decision is whether to use anthropomorphic cues or not. In general, their use in customer care often creates trust and engagement. However, it is important to find a balance with how much those assistants are human-like due to the fact that user's expectations increase with a more representative embodiment. Higher their fidelity of their embodiment to human-like characteristics, higher the chance to be perceived as uncanny.

G4. REDUCE SOCIAL BIASES

Often AI assistants are females, due to stereotypical biases related to the kind of administrative/secretary role they embody. Designers need to consider the kind of stereotypes that can arise from simple design decisions such as names, gender and social roles.

G5. SET THE RIGHT ONBOARDING AND DISCOVERY

It was clear from the free interaction part in first time users the need to be guided especially for those that are not familiar with voice activated assistant. The user needs to be guided about the functions and the possibilities of the devices. Designers need to provide clear examples on the possible ways to interact with the assistant by shaping the right user's mental model and expectations.

G6. SET THE RIGHT EXPECTATION

Mainstream Voice activated assistants (Google Assistant, Alexa, Cortana, Siri) are human-like despite the physical embodiment is reduced to minimal and abstract shape. The abstract embodiment is crucial to set the right expectation and to avoid overpromising functionality capacity.

- Developers/designers should create human-like assistants, however higher the realism higher the expectations would be. Maybe a good compromise could be not to give a human-like embodiment but at the same time design a well-rounded personality.

- All elements as Voice, name, gender, age, embodiment, role and personality can affect the perception, it is important to evaluate their effect on user.

- Also, the designer/developer needs to be clear about the functionalities of the assistant in order not to be over promising and make errors.

- Designers/Developers need to be clear on how well these functionalities can be executed

G7. BUILD CONVERSATION-LIKE

Building trust in AI assistants requires time and fundamental aspects of human conversation: Greetings/Welcoming as a start of the conversation, active listening, echoing and also error management. An assistant that introduces itself and follows social norms can establish a first-time relationship with a user.

G8. REPAIR ERRORS

Technology is improving and error rates are becoming almost the same in a human-like conversation. As found in the Interviews, people really get frustrated when the systems commit mistakes or do not understand correctly their command. It is fundamental to review common errors and fix them with new solutions and be proactive in giving different replies when the end user is not satisfied.

G9. MAKE VISIBLE THE INTERACTION

VUI are challenging to be designed due to the reduced visibility of the system status. In fact, it should always provide conversational markers and states:

- Speaking: visual/sound indicators that human/system is speaking

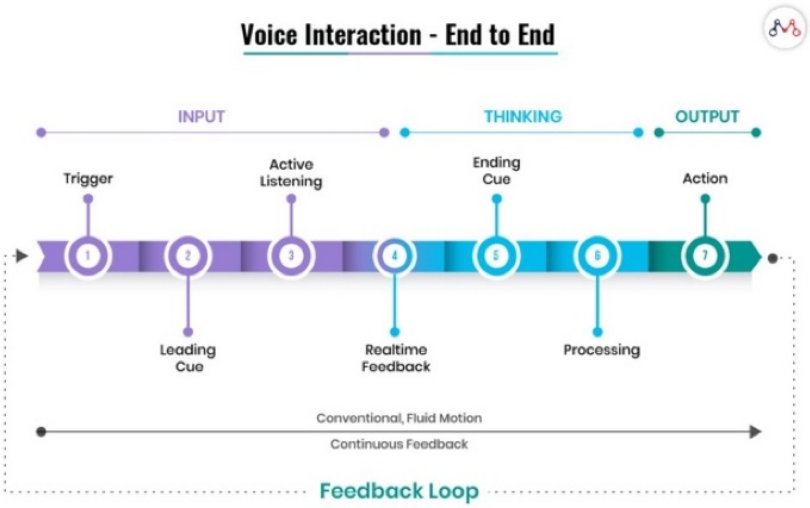
- Listening: visual/sound indicators that the device is listening to user's voice or processing a request.

- Pause: When the conversation is paused

- **Processing**/thinking: visual/sound indicators that human/system is processing the information before giving an answer

- **Ended**: visual/sound indicators that human/system completed tasks

Also, the transition between the states is crucial: to started, processed or finished. Different Feedbacks are implied to convey the states and are present continuously in the voice flow.



Visual cues that show the assistant is listening or proceeding the information increase the perception of human-likeliness and make the interaction faster.

Feedback shows additional information about the system and informs the user about its actions and states. In order to build trust in their usage, especially when the device is actively listening (and recording user data), it is fundamental for the user to know.

Building trust in the interaction requires a specific feedback that informs the user when the camera is switched off or the microphone is off.

G10. GIVE CONTEXTUAL INFORMATION

Provide meaningful information to the user: the user gets frustrated easily when the information received is not relevant to what they asked. In this case, displaying information relevant to user's task or for the specific context could help.

G11. DECIDE THE RIGHT MODALITY OF INTERACTION

It is a priority to understand the physical context in which the interaction takes place when the user engages with a VUI system, in order to design what is the right interaction for user in terms of space and time, also to give the user enough privacy. People in fact in public are less likely to engage through voice,

or in other in which the noise is loud, the system should be able to be effective anyway. After evaluating the use cases, it is necessary to understand how users would interact with it, and ask yourself what are the user's available senses (sight/touch/voice/hearing)?

Man should figure out the best interaction modalities in the use cases and build trust in those interactions.

Visual is best for output, that is able to present a multitude of information and complex outputs. Visual makes information more learnable by reducing cognitive costs (memory and attention) and increasing discoverability.

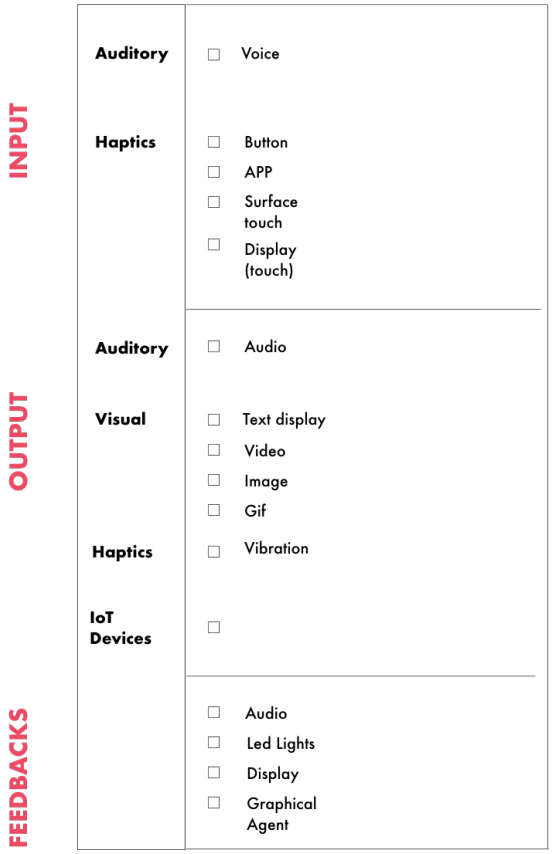
- Relevant Visual cues increase the tendency to trust.
- Include branding to make more evident the credibility and competence. It is also important to show visual elements to help people remind

While Voice is especially strong as input modality due to the fastness, simplicity and hands-free, however, is not the greatest in terms of output due to the intangibility of the modality.

- Relevant Voice cues increase empathy towards the system.
- Voice cues convey implicit information about emotions and personality

However, the combination of visual and auditory modalities can maximize the experience.

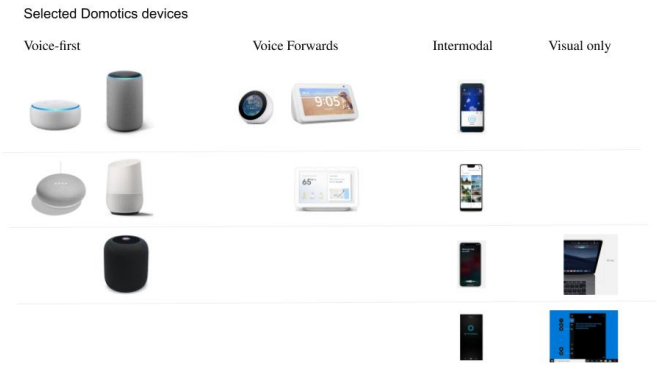
Currently, devices offer the following interaction modalities:



G.12 CHOOSE THE RIGHT DEVICE FOR THE EXPERIENCE

Once the use case is chosen, deciding the device is a matter of deciding what Input/output modalities are the best in the selected context.

The experience needs to design first according to the context and afterwards the device. For first time users, as it was understood in the usability testing, it can be helpful having a screen-based device since the screen can stream the functionality and the visibility of the system is greater.



G13. ADDITIONAL GRAPHICAL USER INTERAFACE

For certain tasks is useful to have an additional graphical interface (mobile/tablet/computer), simply allowing voice through an existing application. The voice can be as an alternative to search and discovery something however the end of the journey should end in the application: as seen in Task 4 shopping through voice was not possible directly there but it was possible to add to the kart /see and discover items but finalize the intention on the app (payment options).

G14. GIVE USER CONTROL

While designing a VUI is important to understand the users' need of control, some users are less likely to trust an assistant when they do not have control over it, and when it is too intrusive or proactive.

- Make clear the decision making behind agent's action/recommendation
- Make sure the users are able to set their preferences
- Provide alternative and do not force the user to use a certain modality
- Provide privacy controls

G15. IMPROVE USABILITY

As it emerged by the interviews, and also by the literature analysis, Ease of use, user's perceived usefulness of the design functionalities can increase user's satisfaction and their overall trust in the technology.

G16. MAKE IT FUNCTIONAL AND MEANINGFUL

As it emerged by the interviews, and also by the literature analysis, if the user perceives the assistant as functional/helpful and meaningful he/she is more likely to adopt or use such technology.

G17. CHOOSE CREDIBLE INFORMATION SOURCES

Users are more likely to trust a voice assistant when the source of information is legit. As mentioned, some people have a higher need of control and might want to change, it should be given to them the possibility.

G18. PRIVACY OPTIONS

It was clear that users were concerned about their personal data and the possibility of device to be always switched on. Few things can be made to decrease user's perceived risks:

- Delete after a certain number of hours the data collected automatically
- The data collected could be sent to third parties only when the user requested a service.
- Make aware the user about the possibility to switch off microphone and screen (the use of the assistant should be voluntary).
- Make clear the indicator of the interaction (wake-up word and cues indicator or listening/responding).

Personal reflection

My thesis journey was challenging in many ways in trying to combine two different worlds: Psychology and Design. Despite the potential synergies and convergence within UX design, these disciplines use different languages and approaches and represent the blend of my personal background. Between research and practice, they can eventually integrate each other, and this thesis is between HCI and Psychology, focusing on the practical project issues in a design discipline experience.

Trust is a crucial factor that predicts user's usage and adoption of a technology, it is one of the easiest things to lose and one of the most critical to have back, that's the reason why trust research became of interest for most companies. Trust in AI agents/voice activated assistants is challenging for two main reasons: personal data/privacy and preference for more traditional ways of interaction. Therefore, I started this thesis with the following question: how do I approach the concept of Trust and apply it to an Artificial Intelligent Agent?

My personal background in Cognitive Psychology influenced the way I saw the problem, and many more questions arise: What exactly can be defined as trust? Is trust between human beings and technology similar or different? Are there any psychological models that can build trust in AI agents? Is their adoption dependent on dispositional factors or rather from agents' factors? Can trust relationships be manipulated and shaped according to agents' different characteristics? Which elements have a stronger influence? Is it the appearance influencing more than its gender or its personality? Does trust change can change along time?

This discussion is not meant to provide a scientifically proven answer to all of these questions but rather to arise awareness about the trust-building relationships between human-beings and AI assistants, and also possible design decision consequences that would shape the user's perception of the assistant.

Designing interacting devices increases the level of complexity in the interaction between AI and Human beings. Can psychology shorter the gap?

Psychology, commonly defined as *"the science of mind and behavior"*, was born as the scientific study of experience. The first experimental psychological Lab created by Wilhelm Wundt in 1897 in Germany was researching *"direct and immediate experience".* Wundt's works focused on understanding (objectively measure and control) the psychological processes by which human beings experience the physical world.

In order to design meaningful experiences, the added value of a psychological approach in the context of human-agent interaction is due to:

1. Psychology providing a scientific understanding of "experience" giving a deeper understanding of the situation, people involved and dynamics, finding aspects that in other ways would be under looked. In fact, the understanding that human behavior is moved by thoughts, mental model, schemas and unconscious emotions can explain why and how people engage and interact with a certain product/system.

2. Psychology providing scientific methods that can be from conducting interviews, doing user research or usability testing, having the analytical eye to formulate valid conclusions about the analyzed experience.

3. Psychology with both qualitative and quantitative methods and tools can help verifying a thesis about the product system involved or rejected making the final solutions as more effective and convicting.

4. The knowledge about cognitive functions can help designing solutions that take into consideration correctly human attention's capability, perception, learning, memory, language, emotions and intelligence in order not to overwhelm the users with too many stimuli. The application of this knowledge in human-agent interaction can provide a strategical relationship to create more useful, usable and accessible devices (agents).

All of these, I believe, are important factors to consider while designing conversational interfaces and voice activated assistants: apparently it seems easy to design for voice, it is immediate, natural and fully accessible to those with less technological literacy. However, there are many factors to be taken into consideration while designing an agent, especially those ones that must be perceived as trustworthy. This naturalness cannot be taken for granted and needs to be designed accordingly.

I approached this doing a literary review of trust and agent's factors that are likely to influence trust behavior. The case studies analysis of mainstream agents was followed by a qualitative research of one of those analyzed that aimed to understand user behavioral trust in a given scenario. The model of trust human-Agent-interaction that I created (see appendix 1) was the result of combining previous research models on the topic, including: personal factors (dispositional/personality related/prior knowledge/tech adoption/need of control), Agent's factors (Aesthetics, perceived anthropomorphism, perceived Functionality/ease of use, helpfulness and predictability) and company factors (Reputations). The results have been interpreted qualitatively, through transcription and questionnaire analysis combining with UX Design tools such as Affinity Diagram, Personas and User Journeys. The analysis ended

understanding that all of these mentioned factors contribute to user's trust behaviors. After a deeper analysis I can say that "trust" can be shaped according to agent's characteristics, however it is not true for all possible end users: there are people that might not be interested in that technology, or they perceived too risky to use it, or dispositionally they are not prone to trust. In fact, a set of guidelines and tools were given to provide possible design directions to build trustworthy agents.

In this sense, I can say that Psychology supports Design in creating natural experiences that match the way the brain evaluates them. In this term, I believe that Psychology and Cognitive Science can provide an additional framework of understanding and methods that can support multidisciplinary teams in developing interactive devices and eventually decrease the digital divide by creating technological solutions that are intuitive and natural, making technology adoption higher.

Future work

Follow-up study

The present analysis is an interpretation of data collected about a post evaluation of a usability test, to be sure about external validity of the proposed approach a following experiment needs to take place, with a wider sample, a control group and a more quantitative analysis.

Usability study

The present Artefact was not tested, it is needed to test whether the proposed solutions are functional or not. The short project-based suggestions of an onboarding was not tested, however to be sure that is valid it should be tested.

Implications for different contexts and scenarios

The interviews took part in a domestic environment, while the user was aware to be part of a "test", in a more realistic situation the results could have been different. It is a known phenomenon in psychology (Rosenthal Effect) that participants in an experiment are more or less consciously motivated to please the experimenter: that could explain why in certain tasks people tried to achieve the task for 8 times, while in a more realistic environment the user might lose the interest more quickly. It would be interesting to see whether the same results would appear in a different context.

Implications for different industry

The approach could be potentially used in different sectors, in which the trust relationship is not established yet.
For example, Voice-shopping need to be analyzed more deeply to make conclusions: In the interviews people were found to overall trust Amazon, and with no particular concern they would even buy with the assistant, probably due to the fact that they could return the objects back.

References

Roger C. Mayer, James H. Davis and F. David Schoorman (1995) An Integrative Model of Organizational Trust

S. Marsh, "Formalizing Trust as a Computational Concept," doctoral thesis, Univ. of Stirling, Scotland, 1994

Supavic Pengnate, Rathindra Sarathy February 2017 An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors Computers in Human Behavior 67:49-60, DOI: 10.1016/j.chb.2016.10.018

Pearl Pu, Li Chen (2006) Trust building with explanation interfaces. IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces January 2006 https://doi.org/10.1145/1111449.1111475

Koslow, S. (2000). Can the truth hurt? How honest and persuasive advertising can unintentionally lead to increased consumer skepticism. The Journal of Consumer Affairs, 4, 245–268.

Luhmann,N. (1979) Trust and Power, introduzione di Gianfranco Poggi. Wiley (trad. it. La fiducia, Il Mulino, Bologna 2002).

Botsman, R. (2017) Di chi possiamo fidarci? trad.it Hoepli.

Kim, D.J., Ferrin, D.L., and Rao, H.R., A Study of the Effect of Consumer Trust in Consumer Expectations and Satisfaction: the Korean Experience, In Proceedings of the Fifth International Conference on Electronic Commerce (ICEC'2003), Pittsburgh, USA, October 1 - 3, 2003

Rempel, J. K., Holmes, J. G. and Zanna, M. P. (1985) 'Trust in Close Relationships', Journal of Personality and Social Psychology.

McCrae, R. R. and Costa, T. J. (2003) Personality in Adulthood: a Five-factor Theory Perspective. New York: Guilford

F. D. Schoorman, R. C. Mayer, and J. H. Davis (2007) An integrative model of organizational trust: Past, present, and future. Academy of Management Review.

Li, X., Hess, T.J. & Valacich, J.S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. Accessed via https://www.sciencedirect.com/science/article/abs/pii/S0963868708000036

McKnight DH, Carter M, Thacher JB, Clay PF(2011) Trust in a specific technology: an investigation of its components and measures. ACM tran Manage Inf Syst 2.

Dinesen, Peter T. (2011) 'Where You Come From or Where You Live? Examining the Cultural and Institutional Explanation of Generalized Trust Using Migration as a Natural Experiment.' European Sociological Review.

Uslaner, E.(2008) 'Where we Stand Depends Upon Where Your Grandparents Sat: The Inheritability of Generalized Trust.' Public Opinion Quarterl.

Leech, P. Psychology for Designers: How to apply psychology to web design and design design process, MPress, 2 Edition.

Poppo L, Zhou KZ, Ryu S.(2008) Alternative origins to interorganizational trust: an interdependence perspective on the shadow of the past and the shadow of the future. Organization Science 19.

Chua RYJ, Ingram P, Morris MW. (2008) From the head and the heart: locating cognition-and affect-based trust in managers' professional networks. Academy of Management Journal.

Doney PM, Cannon JP. (1997) An examination of the nature of trust in buyer-seller relationships. Journal of Marketing.

Gulati R, Sytch M. (2008) Does familiarity breed trust? Revisiting the antecedents of trust. Managerial and Decision Economics.

Biros, D.P., Daly, M., Gunsch, G (2004).: The influence of task load and automation trust on deception detection. Group Decis. Negot.

Lewicki, R.J. and Bunker, B.B. (1996) Developing and Maintaining Trust in Work Relationships.

De Kruijff, R.W. (2018). Technology Acceptance of the Smart Speaker Exploring factors affecting the Use Intention of an emerging technology. Access via http://www.scriptiesonline.uba.uva.nl/document/668183

K. Kelton, K. R. Fleischmann, and W. A. Wallace. Trust in digital information. J. Am. Soc. Inf. Sci. Technol., 59(3):363–374, Feb. 2008. ISSN 1532-2882.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

McKnight, D. Harrison, Larry L. Cummings, and Norman L. Chervany(1998) "Initial Trust Formation in New Organizational Relationships."Academy of Management Review, Vol. 23, No. 3

McKnight, D.H., Chervany, N.L., 2001–2002. What trust means in e-commerce customer relationships: an interdisciplinary conceptual typology. International Journal of Electronic Commerce.

Castaldo, S., Premazzi, K. & Zerbini, F. (2010). The Meaning(s) of Trust. A Content Analysis on the Diverse Conceptualizations of Trust in Scholarly Research on Business Relationships. J Bus Ethics 96, 657–668. https://doi.org/10.1007/s10551-010-0491-4.

Barber, B., 1983. The Logic and Limits of Trust, New Rutgers University Press, Brunswick.

Shapiro, S.P., 1987. The social control of impersonal trust. American Journal of Sociology 93 (3), 623–658

Golembiewski, R., & McConkie, M. (1975). The centrality of interpersonal trust in group process. In C. Cooper (Ed.). New York: Wiley.

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.),
Nebraska Symposium on Motivation, 1962

McKnight DH, Cummings LL, Chervany NL. (1998) Initialtrust formation in new organizational relationships.Academy of Management Review

Lippert, S.K., (2007) Investigating postadoption utilization: an examination into the role of interorganizational and technology trust. IEEE Transactions on Engineering Management. J.

Lee, J. Kim, and J.Y. Moon (2000). "What Makes Internet Users Visit Cyber Stores Again? Key Design Factors for Customer Loyalty," Proc. Human Factors in Computing Systems (SIGCHI 00), ACM Press, New York.

Botsman, Rachel. Di chi possiamo fidarci? (Italian Edition). Hoepli. Kindle Edition. Mark Stephen Meadows, interview dicembre 2016.

Muir, B.M., Moray, N., (1996) Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 39 (3), 429–460.

ZeJda, D. (2010). From Subjective Trust to Objective Trustworthiness in Online Social Networks: Overview and Challenges. Journal of Systems Integration.

Hall, S., McQuay,W. Review of trust research from an interdisciplinary perspective - psychology, sociology, economics, and cyberspace. Proceedings of the IEEE 2010 National Aerospace & Electronics Conference.[1]

Chua, Yong Joo Roy, Michael Morris, and Paul Ingram. "The Social Structure of Affect- and Cognition-based Trust in Chinese and American Managerial Networks." Columbia Business School, 2005.

Zaroukian, E., Bakdash, J. Z., Preece, A. D. and Webberley, W. (2017). Automation bias with a conversational interface: user confirmation of misparsed information. Presented at: IEEE Conference on Cognitive and Computational Aspects of Situation Management, Savannah, GA, USA, 27-31 March 2017. Cognitive and Computational Aspects of Situation Management (CogSIMA), 2017 IEEE Conference on. IEEE, 10.1109/COGSIMA.2017.7929605

Seeger, A.; Pfeiffer, J.; and Heinzl, A. (2017) "When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents". SIGHCI 2017 Proceedings. 15.http://aisel.aisnet.org/sighci2017/15

Cassell,J.; and Bickmore,T. (2000) "External manifestations of trustworthiness in the interface," Communications of the ACM, vol. 43, pp. 50-56, 2000

Cummings, M. (2004). "Automation Bias in Intelligent Time Critical Decision Support Systems" (PDF). AIAA 1st Intelligent Systems Technical Conference (PDF). doi:10.2514/6.2004-6313.

Heerink, M.; Kröse, B; Wielinga, B. Evers, V. (2009) Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. In: Proceedings HCI 2009, Cambridge

Thatcher, J.B., McKnight, D.H., Baker, E.W., Arsal, R.E., Roberts, N.H., (2011). The role of trust in postadoption IT exploration: an empirical examination of knowledge management systems. IEEE Transactions on Engineering Management 58 (1), 56–70.

[1]

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277–301. https://doi.org/10.1080/14639220500337708

Chowdhury, S. (2005). The role of affect- and cognition-based trust in complex knowledge sharing. Journal of Managerial Issues, 17(3), 310-326.

David J.L; Weigert, A. (1985) Social Forces, Volume 63, Issue 4, June 1985, Pages 967–985, https://doi.org/10.1093/sf/63.4.967.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal. https://doi.org/10.2307/256727

Rempel, J.K. ; Holmes, J.G, ; Zanna, M.P.(1985). Trust in close relationships. J Pers Soc Psychol.

Butler,J.K.(1991) Toward understanding and measuring conditions of trust: evolution of a conditions of trust inventory. J of Management

Rousseau, M. D.; Sitkin, S.B.; Burt, R. S. and Camerer, C. (1998). Not So Different After All: A Cross-Discipline View of Trust. Academy of Management Review, 23 (3). pp. 393-404.

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000). The measurement of perceived relationship quality components: A confirmatory factor analytic study. Personality and Social Psychology Bulletin, 26, 340- 354.

Fukuyama F. (1995) Trust: the social virtues and the creation of prosperity. New York: The Free Press.

Larzelere, R. E., & Huston, T. L. (1980). The Dyadic Trust Scale: Toward understanding interpersonal trust in close relationships. Journal of Marriage and the Family, 42(3), 595–604. https://doi.org/10.2307/351903

Luhmann N.(1979) Trust and power. Chichester, UK: Wiley.

Keng, S. ; Wang.W. (2018) Living with Artificial Intelligence -Developing a Theory on Trust in Health Chatbots Conference: Pre-ICIS Workshop on HCI Research in MIS At: San Francisco, CA.

Tatai G., Csordás A., Kiss Á., Szaló A., Laufer L. (2003) Happy Chatbot, Happy User. In: Rist T., Aylett R.S., Ballin D., Rickel J. (eds) Intelligent Virtual Agents. IVA 2003.

McCarthy. J. What is artificial intelligence. Technical report, Stanford University, http://www-formal. stanford.edu/jmc/whatisai.html, 2007.

Bretan, I. and J. Karlgren. Transparent Natural Language Interaction through Multimodality. In ERCIM Workshop on Multimodal HCI. 1993.

Oviatt, S. Breaking the Robustness Barrier: Recent progress on the design of robust multimodal systems. Advances in Computers, 2002.

Oviatt, S. Multimodal Interfaces. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, J. Jacko and A. Sears, Editors. 2003

Oviatt, S., et al. Designing the User Interface for Multimodal Speech and Pen-based Gesture Applications: State-of-the-art systems and future research directions. Human Computer Interaction, 2000.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1-39.

Turing, A. M. (1950). Computing machinery and intelligence.published in Mind

Oliver Budzinski, O. , Noskova, V. & Zhang, X. (2018) The Brave New World of Digital Personal Assistants: Benefits and Challenges from an Economic Perspective. Ilmenau Economics Discussion Papers, Vol. 24, No. 118

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). Pragmatics of Human Communication. A Study of Interactional Patterns, Pathologies and Paradoxes. New York: Norton.

Kääriä, A. (2017) Technology acceptance of voice assistants: anthropomorphism as factor Jyväskylä: University of Jyväskylä, 61 p. Information Systems, Master's Thesis

Chelsea Myers, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. (2018).

Linden, G.; Smith, B.; York, J.; , "Amazon.com recommendations: item-to-item collaborative filtering," Internet Computing, IEEE , vol.7, no.1, pp. 76- 80, Jan/Feb 2003 doi: 10.1109/MIC.2003.1167344 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1167344&isnumber=26323

Suhm, B.; Myers,B. Multimodal Error Correction for Speech User Interfaces. ACM Transactions on Computer-Human Interaction, Vol. 8, No. 1, March 2001, Pages 60–98

Giselle Abramovich Adobe Digital Insights (2019) Study Finds Consumers Are Embracing Voice Services. Here's How.

McTear, Michael, Callejas, Zoraida, Griol Barres, David (2016) The Conversational Interface: Talking to Smart Devices. Springer International Publishing. 10.1007/978-3-319-32967-3

Xiaoxi He(2019) Voice, Speech, Conversation-Based User Interfaces 2019-2029: Technologies, Players, Markets from https://www.idtechex.com/en/research-report/voice-speech-conversation-based-user-interfaces-2019-2029-technologies-players-markets/637

Creative Strategies Inc. Voice Assistant 2.0 Research Study 2017

NN group, By Raluca Budiu and Page Laubheimer on July 22, 2018

Voice Assistant Consumer Adoption Report 2018

J. Jiang, A. H. Awadallah, X. Shi, and R. W. White. Understanding and predicting graded search satisfaction. In WSDM, 2015

Filip Radlinski, Nick Craswell. Interleaving is an online evaluation technique for comparing the for comparing the relative quality of information retrieval CIKM '17; 2017.

Global voice search usage 2019, by region and device. Published by J. Clement. 2019.

Sternberg, R. J. , editor. Handbook of Intelligence. Cambridge University Press, 2000.

Poole. D. 1988. David Poole. A logical framework for default reasoning. Artif. Intell.

Russel S., Norvig. P., Artificial Intelligence: a modern Approach, 3rd edition. Pearson. 2010

Shardlow,N. Action and agency in cognitive science. Master's thesis. Dept. of Psychology, University of Manchester, UK, 1990

Ryan, Nick S., Pascoe, Jason and Morse, David R. (1998): Enhanced Reality Fieldwork: the Context-aware Archaeological Assistant. In: Gaffney, V., Leusen, M. van and Exxon, S. (eds.). "Computer Applications in Archaeology - British Archaeological Reports". Oxford: Tempus Reparatum

Dey, K. (2000) Providing Architectural Support for Building Context-Aware Applications. PhD thesis, College of Computing, Georgia Institute of Technology.

Purington, A; Taft, J.F. (2017) "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo the 2017 CHI Conference Extended Abstracts DOI: 10.1145/3027063.3053246

Ferdig,R.E.; Mishra, P. (2004)Emotional Responses to Computers: Experiences in Unfairness, Anger, and Spite Journal of Educational Multimedia and Hypermedia Volume 13, Number 2, April 2004 ISSN 1055-8896 Publisher: Association for the Advancement of Computing in Education (AACE), Waynesville, NC USA

Guzman, A. L. (Ed.). (2018) Human-Machine Communication: Rethinking communication, technology, & ourselves. In S. Jones (Series Ed.), Digital Formations Series. PeterLang.

Pearl, Cathy. Designing Voice User Interfaces . O'Reilly Media. Kindle Edition. George Lawton 2019 Google Empathy Lab uses 'design feeling' in search of a more human.

Cassell J, Sullivan J, Prevost S, Churchill E (2000) Embodied conversational agents. MIT Press, Cambridge

Bar, M. ; Neta, M. (2006), Humans Prefer Curved Visual Objects. Psychological Science, 17: 645–648. doi: 10.1111/j.1467-9280.2006.01759.x

Cathy Pearl. Designing Voice User Interfaces: Principles of Conversational Experiences. O'Reilly Media 2016

Yee,N.; Bailenson, J.N. ; Rickertsen, K. (2007) A Meta-Analysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces CHI 2007 Proceedings

Andrea Ormaza Cuadra and Ysabelle Rase. The use of personal voice assistants in the context of retail digitalization: An empirical study of consumers' attitude and privacy concerns. Master This at Louvain School of Management. (2018)

Ann Thyme-Gobbel(2018) Learn from the Experts about Voice Technology: Ann Thyme-Gobbel Published O'Reilly Media, Inc.

Clifford Ivar Nass and Scott Brave. 2005. Wired for speech: How voice activates and advances the human-computer relationship. MIT press Cambridge, MA.

Masahiro Mori. 1970. The uncanny valley. Energy 7, 4: 33–35.

Jung, C.G. The Collected Works, Vol. 10: Civilization in Transition; Vol. 11: Mysterium Coniunctionis: An Inquiry into the Separation and Synthesis of Psychic Opposites in Alchemy; Vol. 16: The Practice of Psychotherapy: Essays on the Psychology of the Transference and Other Subjects. Translated by R. F. C. Hull. Pp. xxxvii, 1,699. New York: Pantheon Books, for the Bollingen Foundation, 1964

Carolin Strassmann, Nicole Krämer A Categorization of Virtual Agent Appearances and a Qualitative Study on Age-Related User Preferences. Intelligent Virtual Agents: 17th International Conference, IVA 2017

Meerbeek, B., Saerbeck,M; Bartneck. C.(2009) Iterative design process for robots with personality AISB2009 Symposium on New Frontiers in Human-Robot Interaction Conference, Edingburgh.

De Visser., E. (2016) Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. Journal of Experimental Psychology Applied 22(3) · August 2016 DOI: 10.1037/xap0000092

Sträfling, N., Fleischer, I., Polzer, C., Leutner, D., & Krämer, N. C. (2010). Teaching learning strategies with a pedagogical agent: The effects of a virtual tutor and its appearance on learning and motivation. Journal of Media Psychology: Theories, Methods, and Applications, 22(2), 73–83. https://doi.org/10.1027/1864-1105/a000010

Gulz, A., and Haake, M. Design of animated pedagogical agents – a look at their look. 2006 International Journal of Human-Computer Studies 64(4):322-339 DOI: 10.1016/j.ijhcs.2005.08.006

Caporael, L. R., & Culbertson, G. H. (1986). Verbal response modes of baby talk and other speech at institutions for the aged. Language & Communication, 6(1-2), 99–112. https://doi.org/10.1016/0271-5309(86)90009

Kevin K. Bowden, Tommy Nilsson, Christine P. Spencer, Kubra Cengiz, Alexandru Ghitulescu & Jelte B. van Waterschoot. "I Probe, Therefore I Am: Designing a Virtual Journalist with Human Emotions" (PDF). Proceedings of eNTERFACE'16, Enschede, July 18th – August 12th, 2016

Salem, M.; Rohlfing, K.J; Kopp, S. (2011) A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction Conference: RO-MAN, 2011 IEEE DOI: 10.1109/ROMAN.2011.6005285

Fiske, S.T; Taylor, S.E. (1991). Social Cognition. New York: McGraw-Hill, Inc.
Lester J.C; Conversez, S.A, The Persona Effect: Affective Impact of Animated Pedagogical Agents. CHI '97: Proceedings of the ACM SIGCHI Conference on

Human factors in computing systemsMarch 1997 Pages 359–366https://doi.org/10.1145/258549.258797

B. R. Duffy, "Anthropomorphism and the social robot," Robotics and autonomous systems, vol. 42, pp. 177-190, 2003.

T. Koda and P. Maes. 1996. Agents with faces: The efects of personifcation of agents. In Proc. HCI' 96. 98–103

A. Cafaro et al. 2012. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in frst encounters. In Proceedings of the International Conference on Intelligent Virtual Agents. Springer, Berlin.

Corti, K., Gillespie, A.: Co-Constructing Intersubjectivity with Artificial Conversational Agents: People are More Likely to Initiate Repairs of Misunderstandings with Agents Represented as Human. Comput. Hum. Behav. 58, 431–442 (2016)

Cuicui Cao Ling Zhao* Yingying Hu. Anthropomorphism of Intelligent Personal Assistants (IPAs): Antecedents and Consequences.Twenty-Third Pacific Asia Conference on Information Systems, China 2019

Sproull, L., Subramani, M., Kiesler, S. Walker, J. H., Waters, K.: When the interface is a face. Human Computer Interaction, 11 (1996) 97-124

Hubona, G. S., & Blanton, J. E. (1996). Evaluating system design features. International Journal of Human–computer Studies, 44, 93–118

Cuicui Cao, Ling Zhao, Yingying Hu (2019) Anthropomorphism of Intelligent Personal Assistants (IPAs): Antecedents and Consequences. Twenty-Third Pacific Asia Conference on Information Systems.

Hwang, J.; Park, T. ()The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot Applied Ergonomics Volume 44, Issue 3, May 2013, Pages 459-471 https://doi.org/10.1016/j.apergo.2012.10.010

Scopelliti, M., Giuliani, M. V., D'Amico, A. M., & Fornara, F. (2004). If I had a robot at home: Peoples' representation of domestic robots. In Keates, S., Clarkson, J., Langdon, P., Robinson, P. (Eds.), Designing a More Inclusive World (pp. 257–266). Cambridge: Springer

Breazeal, C.L., Kidd, C.D., Thomaz, A.L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. Paper presented at the International Conference on Intelligent Robots and Systems (IROS 2005).

Lee, K.M., Jung, Y., Kim, J., & Kim, S.R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. International Journal of Human-Computer Studies.

Tapus, A., Tapus, C., & Mataric, M.J. (2007). Emulating empathy in socially assistive robotics. Paper presented at the AAAI Spring Symposium.

Epley, N., Waytz, A., & Cacioppo, J.T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864-886

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J.T. (2008). When We Need A Human: Motivational Determinants of Anthropomorphism.

Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human–robot interaction. Hum Factors 53(5):517–527. doi:10. 1177/0018720811417254

Young, K. S., Parsons, C. E., LeBeau, R. T., Tabak, B. A., Sewart, A. R., et al. (2017). Sensing emotion in voices: negativity bias and gender differences in a validation study of the Oxford Vocal ("OxVoc") Sounds Database. Psychol. Assess. 29, 967–977. doi: 10.1037/pas0000382

Lee, S., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. Human-Computer Interaction, 11(2), 97-124.
Adriana Tapus, Cristian Tapus, and Maja J. Matarić. "User-Robot Personality Matching and Assistive Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy". In Intelligent Service Robotics: Multidisciplinary Collaboration for Socially Assistive Robotics, 1:169-183, Apr 2008.

Kwan Min Lee , Wei Peng , Seung-A Jin & Chang Yan. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction. Journal of Communication ISSN 0021-9916. (2006)

Qvarfordt,P.; Jönsson, A.; Dahlbäck, N.; (2003) The Role of Spoken Feedback in Experiencing Multimodal Interfaces as Human-like. Proceedings of the 5th international conference on Multimodal interfaces.

Voice user interfaces. In: Kortum P (ed) HCI beyond the GUI: design for haptic, speech, olfactory, and other non-traditiona

Caporeal, L.R. Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine. Computers in Human Behavior 2 (1986) 215-234

Perez, M., Saffon, S. and Donis, H. (2018) Everybody is talking about Voice Activated Virtual Assistants, but how are people really adopting and using them? Lessons from a multicountry study. Under review for the 32nd Human Computer Interaction Conference, Belfast, Northern Ireland, 2-6 July 2018.

Pruitt,J.; Adlin, T. (2006) The Personas Lifecycle DOI: 10.1016/B978-0-12-566251-2.X5000-X

Meerbeek, B., Saerbeck, M., & Bartneck, C. (2009b). Iterative design process for robots with personality. AISB2009 Symposium on New Frontiers in Human-Robot Interaction (pp. 94–101). Edingburgh.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574-583.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality 1exicons. In L. Wheeler (Ed.), Review of Personality and social psychology, vol.2. CA, Sege, Beverly Hills.

Fiske, S.T. Stereotyping, prejudice and discrimination. in Gilbert., D.T., Fiske, S.T. and Lindzey, G. eds. The handbook of social psychology, McGraw-Hill, New York, 1998, 357-414.

Piper, Allison M., "Stereotyping Femininity in Disembodied Virtual Assistants" (2016). Graduate Theses and Dissertations. 15792. https://lib.dr.iastate.edu/etd/15792

Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. Frontiers in Psychology, 6, Article 883.

Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. Computers in Human Behavior, 58, 431–442.
Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), Handbook of personality: Theory and research (pp. 244-276). New York: Guilford.

Clifford Ivar Nass and Scott Brave. 2005. Wired for speech: How voice activates and advances the human-computer relationship. MIT press Cambridge, MA.
28. Masahiro Mori. 1970. The uncanny valley. Energy 7, 4: 33–35.

Isbister,K.; Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics Int. J. Human-Computer Studies (2000) 53, 251}267 doi:10.1006/ijhcs.2000.0368

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. Journal of Applied Social Psychology, 29(5), 1093-1110.

Cohen, Giangola, and Balogh, (2004) Voice User Interface

Sitography:

NN group article by Kathryn Whitenton (January 31, 2016) from https://www.nngroup.com/articles/voice-interaction-ux/
NNgroup article by Page Laubheimer (August 23, 2015)https://www.nngroup.com/articles/slips/
https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en
https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616
https://www.statista.com/statistics/1036727/global-voice-search-region-device/
https://quoracreative.com/article/voice-search-statistics-trends
https://voicify.com/learn/statistics-about-voice/

https://voicebot.ai/wp-content/uploads/2018/09/voice-ux-best-practices-ebook-voicebot-applause.pdf
Sara Kleinberg(January 2018) Experience & Design from https://www.thinkwithgoogle.com/consumer-insights/voice-assistance-consumer-
experience/?utm_campaign=VUI_SC&utm_source=hs_automation&utm_medium=email&utm_content=68553672
https://doc.rero.ch/record/9219/files/these.pdf
http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/631051/IPOL_BRI(2019)631051_EN.pd
https://www.accenture.com/us-en/_acnmedia/pdf-39/accenture-pov-dynamic-consumers.pdf
https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020
https://www.gartner.com/en/newsroom/press-releases/2019-29-08-gartner-identifies-five-emerging-technology-trends-with-transformational-impact
https://chatbotsmagazine.com/the-report-2018-voicebots-ahead-269123c40165)
https://www.nngroup.com/articles/intelligent-assistant-usability/

https://www.perficientdigital.com/insights/our-research/digital-personal-assistants-study
https://advertiseonbing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019voicereport.pdf
https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/
https://www.nngroup.com/articles/intelligent-assistant-usability/
https://georgianpartners.com/investment-thesis-areas/overview-conversational-ai/
https://uxdesign.cc/and-how-do-they-work-f692205956ed
Marti Gold Published on Sep 24, 2018 on linkedin: https://www.slideshare.net/martigold/new-frontier-of-multimodal-interfaces-are-you-ready
https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00043900/Diskussionspapier_Nr_118.pdf
https://www.accenture.com/us-en/_acnmedia/pdf-39/accenture-pov-dynamic-consumers.pdf
https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020
https://www.zionmarketresearch.com/report/voice-assistant-market
https://voicebot.ai/wp-content/uploads/2018/09/voice-ux-best-practices-ebook-voicebot-applause.pdf)

https://www.youtube.com/watch?v=auAfeIXtd1g&ab_channel=SmartlyAI

Google I/O 2018 https://www.youtube.com/watch?v=JDakZMlXpQo&ab_channel=GoogleDevelopers

https://chatbotsmagazine.com/the-report-2018-voicebots-ahead-269123c40165
https://vux.world/difference-between-voice-assistant-and-voice-interface/
https://developer.amazon.com/it-IT/alexa/alexa-skills-kit/vui
https://www.technavio.com/report/global-voice-recognition-biometrics-market-2015-2019?utm_campaign=VUI_SC&utm_source=hs_automation&utm_medium=email&utm_content=68553672
https://www.smashingmagazine.com/2018/12/mixing-tangible-intangible-multimodal-interfaces-adobe-xd/
https://www.slideshare.net/martigold/new-frontier-of-multimodal-interfaces-are-you-ready
https://en.wikipedia.org/wiki/Multimodal_interaction
https://careerfoundry.com/en/blog/ux-design/voice-ui-design-and-cognitive-load/

https://thenextweb.com/contributors/2018/12/20/6-things-to-keep-in-mind-when-designing-voice-ui/
https://www.nngroup.com/articles/intelligent-assistants-poor-usability-high-adoption/
https://medium.muz.li/voice-user-interfaces-vui-the-ultimate-designers-guide-8756cb2578a1
https://www.mantralabsglobal.com/blog/voice-user-interface-is-the-next-ux/
https://www.dailymail.co.uk/sciencetech/article-3628362/Are-smart-personal-assistants-doomed-Researchers-say-people-embarrassed-talk-machine-especially-public.html
http://giusepperizzo.github.io/publications/Mensio_Rizzo-ReBM20018.pdf
Google LLC (https://patents.google.com/patent/US20140280296A1/en?q=emotion&q=detection&after=priority:20090101)
UI from https://searchcio.techtarget.com/news/252459180/Google-Empathy-Lab-uses-design-feeling-in-search-of-a-more-human-UI

Bell,K. (2017) from https://mashable.com/2017/01/12/how-alexa-siri-got-names/?europe=true
David Limp interview (2016)from https://fortune.com/2016/07/14/amazon-alexa-david-limp-transcript/
https://www.networkworld.com/article/2221246/steve-jobs-wasn-t-a-fan-of-the-siri-name.html
Kyle Wiggers,2017 from https://www.digitaltrends.com/mobile/microsoft-emotional-assistant-patent/
https://www.businessinsider.com/why-doesnt-googles-smart-assistant-doesnt-have-a-name-like-siri-cortana-alexa-2016-5?IR=T
https://mashable.com/2017/01/12/how-alexa-siri-got-names/?europe=true
https://mashable.com/2017/01/12/how-alexa-siri-got-names/?europe=true
Amazon developer guidelines Alexa https://m.media-amazon.com/images/G/01/mobile-apps/dex/alexa/alexa-skills-kit/marketing-guidelines/Echo_Alexa_Guidelines_External.pdf
UNESCO report (2019) I'd blush if I could: closing gender divides in digital skills through education
Eric Hal Schwartz, H. E. 2020 february 21 article published https://voicebot.ai/2020/02/21/voice-assistants-very-prone-to-accidentally-waking-up-and-recording-long-audio-clips-study/
Alex Cook, Jonathan Jarvis, and Jonathan Lee 2019 in https://design.google/library/evolving-google-identity/
Deborah Harrison Interview by Malini Goyal Mar 09, 2019 “We wanted Cortana to be funny but not shame people: Microsoft” from https://economictimes.indiatimes.com/tech/software/we-wanted-cortana-to-be-funny-but-not-shame-people-microsoft/articleshow/68335713.cms
Patently Apple, 2018 article from: https://www.patentlyapple.com/patently-apple/2018/05/siris-reconstructed-personality-to-arrive-at-wwdc-2018.html

Deborah Harrison Interview by Malini Goyal Mar 09, 2019 "We wanted Cortana to be funny but not shame people: Microsoft" from https://economictimes.indiatimes.com/tech/software/we-wanted-cortana-to-be-funny-but-not-shame-people-microsoft/articleshow/68335713.cms
Aliza Vigderman (May 20, 2019) published on https://securitybaron.com/news/amazon-alexa-voice-detection/
Catherine Gao Nov 26, 2019 "Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience" from https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2019/11/new-alexa-emotions-and-speaking-styles?linkCode=w61&imprToken=wIJHgmYGN9bLYhAVSmRiFA&slotNum=2&tag=theverge02-20
Fussel, S. (12 October 2018) "Alexa Wants to Know How You're Feeling Today" from https://www.theatlantic.com/technology/archive/2018/10/alexa-emotion-detection-ai-surveillance/572884/
Marcus Ash(2015) interview from https://blogs.windows.com/windowsexperience/2015/02/10/how-cortana-comes-to-life-in-windows-10/
https://www.expertreviews.co.uk/software/operating-systems/1402658/how-to-tell-if-cortanas-feeling-satisfied-or-embarrassed
Arianna stefanoni interview from https://medium.com/ideas-at-igenius/humanizing-ai-is-a-job-for-a-woman-8913a3d5fff2
https://clearbridgemobile.com/7-key-predictions-for-the-future-of-voice-assistants-and-ai/
https://www.youtube.com/watch?v=auAfeIXtd1g&ab_channel=SmartlyAI
Jordan Novet (2019) from https://www.cnbc.com/2019/06/04/why-siri-is-not-as-smart-as-alexa-google-assistant.html