Vaasan yliopisto
UNIVERSITY OF VAASA

Amanda Sahlstedt

Building trust in AI Systems

School of Technology and Innovation Management
Master's thesis in Information Systems
Technical Communication

Vaasa 2021

UNIVERSITY OF VAASA
School of Technology and Innovation Management
Author: Amanda Sahlstedt
Title of the Thesis: Building trust in AI Systems
Degree: Master of Science in Economics and Business Admistration
Programme: Technical Communication, Information Systems
Supervisor: Tero Vartiainen
Year: 2021 **Pages:** 84

ABSTRACT:
Artificial Intelligence has integrated as a part of humans' daily life while at the same time the AI-enabled services and applications are widely considered distrustful. Because the majority of the users are not expert of Machine Learning, not to mention Deep learning, it is important to create trustworthy AI services that understand humans but also explains themselves in an easily understandable way. This type of approach to Artificial Intelligence is called Explainable Human-Centered thinking and it has been discovered as a solution for the distrust problem between human-AI interaction. This research is a qualitative study of User-Experience of different AI-based applications and services that are used in daily life activities such as navigation or checking grammar mistakes. The goal is to find UX elements that affect to user's trust-perception of the service or application and create a united list of these elements based on previous literature. This list can be used for designing better, explainable, and human-centered AI, but it also fulfills its purpose by gathering together and validating research of the field. The results showed that even in the most strongly trusted services and applications, users can notice problems such as privacy issues or missing explainability. However, many of the commonly used services provide added value for its user and they are relatively better than the other similar services. Based on the results, this study discusses also critically whether implementing HAI is only a UX-design problem but rather a part of sharing knowledge of trustworthy AI and not accepting non-transparent functions and data usage.

KEYWORDS: Artificial Intelligence, UX-Design, Trust, Human-Centered AI, Explainable AI

VAASAN YLIOPISTO
Tekniikan ja Innovaatiojohtamisen yksikkö

Tekijä:	Amanda Sahlstedt
Tutkielman nimi:	Building trust in AI Systems
Tutkinto:	Kauppatieteen maisteri
Oppiaine:	Tekninen viestintä, Tietojärjestelmätiede
Työn ohjaaja:	Tero Vartiainen
Vuosi:	2021 **Sivumäärä:** 84

TIIVISTELMÄ:

Tekoäly on integroitunut osaksi ihmisten jokapäiväistää elämää, mutta yleisesti tekoälyperustaisia palveluja ja sovelluksia ei pidetä luotettavina. Tämän lisäksi välillä palveluja tai sovelluksia käyttäessä on mahdotonta todentaa, onko käyttäjä kosketuksissa ihmisen vai koneen kanssa ja mihinkä käyttäjän saama informaatio, kuten ohjeet tai ehdotukset, perustuvat. Koska tämän tyyppinen käyttäjäkokemus lisää epäluottamusta ihmisen ja tietokoneen kanssakäymisessä ja koska suurin osa käyttäjistä ei ole koneoppimisen asiantuntijoita, on tärkeää luoda luotettavia tekoälypalveluja, jotka ymmärtävät ihmisiä ja selittävät omaa toimintaansa helposti ymmärrettävällä tavalla. Tämän tyyppistä lähestymistapaa tekoälyyn kutsutaan selittäväksi ihmiskeskiseksi (explainable human-centered) ajatteluksi ja sitä on pidetty ratkaisuna nimenomaiseen ihmisen ja tekoälyn välisen epäluottamuksen ongelmaan.

Tämä kvalitatiivinen tutkimus tarkastelee käyttäjäkokemusta erilaisissa tekoälypohjaisista sovelluksista ja palveluista, joita käytetään jokapäiväisessä elämässä, kuten navigoinnissa tai esimerkiksi kieliasun tai kielioppivirheiden tarkastuksessa. Tavoitteena on löytää UX-elementit, jotka vaikuttavat käyttäjän kokemukseen luottamuksesta käyttäessään palvelua tai sovellusta, ja luoda yhtenäinen luettelo näistä elementeistä aiemman kirjallisuuden perusteella. Tätä luetteloa voidaan käyttää apuna ihmiskeskeisessä tekoälysuunnittelussa, mutta se täyttää tarkoituksensa myös kokoamalla yhteen ja validoimalla alan aiempaa tutkimusta nimenomaan tekoälyperusteisista sovelluksiin liittyen.

Kirjallisuuskatsaus esittelee tutkimuksen keskeiset käsitteet, kuten tekoälyn, luottamuksen ja käyttäjäkokemuksen. Lisäksi tässä osiossa kerätään yhteen tärkeimmät edellisissä tutkimuksissa jo identifioidut UX-elementit, jotka vaikuttavat käyttäjän kokemaan luottamukseen muun muassa web-suunnittelussa. Itse tutkimus jakaantuu kolmeen vaiheeseen, jossa ensimmäisenä tekoälyperustaiset sovellukset listataan perustuen alan kirjallisuuden tyyppimääritelmiin sekä käyttäjämäärä arvioiden mukaan. Toisessa vaiheessa, valitut sovellukset ja palvelut listattiin luotetuimmasta epäluotettavimpaan perustuen lyhyeen kyselytutkimukseen. Viimeiseksi syvähaastattelu, perustuen kriittisten tapahtumien tekniikkaan, suoritettiin kyselyyn vastanneille. Avoimilla kysymyksillä kartoitettiin tietoja tapahtumasta, jossa käyttäjä tunsi luottamusta tai epäluottamusta käyttäessään valittua tekoälyperusteistasovellusta tai palvelua.

Tulokset analysoitiin teemoittamalla havaitut UX elementit, jotka lisäävät luottamusta tai vähentävät epäluottamusta ja vertaamalla niitä listaan alan edellisistä havainnoista luottamukseen liittyen. Tuloksena saatiin tutkimuksen tavoitteen mukainen lista, jossa on validoitu kirjallisuuden havaintoja, että lisätty uusia havaintoja luottamukseen vaikuttavista UX-elementeistä perustuen tehtyihin käyttäjähaastatteluihin.

Kaiken kaikkiaan tämän tutkimuksen tärkeimmät havainnot vahvistivat luettelon tärkeistä UX-elementeistä, jotka on otettava huomioon luotaessa käyttäjien ja tekoälyjärjestelmien välistä luottamusta, mutta samalla vain luotettavien palvelujen suunnittelu ei riitä. Yksi tutkimuksen johtopäätös onkin, että kyselyn osallistujat käyttivät näitä palveluja, vaikka monet olivat huolissaan esimerkiksi omasta yksityisyydestään tai järjestelmän epämääräisestä datakäytöstä.

Näin ollen nämä tulokset osoittavat, että käyttäjät hyväksyivät nämä käytännöt, koska sovelluksen tai palvelun käyttäminen toi suhteellista etua muihin palveluihin verrattuna tai merkittävää lisäarvoa käyttäjän jokapäiväiseen elämään. Näiden tulosten perusteella, tässä tutkimuksessa keskustellaan myös kriittisesti siitä, onko HAI:n (Human Centered Artificial intelligence) eli ihmiskeskeisen tekoälyn käyttöönotto vain UX-suunnittelun ongelma, vaan pikemminkin osa koulutusta ja tiedon jakamista luotettavasta tekoälystä jolloin käyttäjät eivät hyväksy läpinäkymättömiä toimintoja tai tietojen väärinkäyttöä, vaan vaativat luotettavia ja avoimia käytäntöjä, jotka selitetään heille erilaisten käyttöliittymäelementtien kautta.

AVAINSANAT: Tekoäly, UX-suunnittelu, Luottamus, Ihmiskeskeinen tekoäly, Selittävä tekoäly

Content

Figures

Tables

Abbreviations

AI	Artificial Intelligence
CPU	Central Processing Unit
DL	Deep Learning
DNN	Deep Neural Network
HCAI	Human-Centered Artificial Intelligence
HCI	Human-Computer Interaction

HCXAI	Human-Centered Explainable Artificial Intelligence
IPA	Intelligence Personal Assistant
IS	Information Science
ML	Machine Learning
NLA	Natural Language Analysing
NLP	Natural Language Processing
NLUI	Natural Language User Interface
NS	Neural-Symbolic AI
UX	User Experience
UXD	User Experience Design
XAI	Explainable Artificial Intelligence

1 Introduction

1.1 Background

Artificial Intelligence (AI) can be defined, for example as, an application or service that is based on machine learning and deep learning methods. Humans are increasingly interacting with systems that are based on AI in their daily activities such as using navigation maps on phones or listening to customized playlists according to user's tastes (FCAI, 2020). Interestingly, simultaneously over 40 percent of users did not have any trust in AI-based services in the United States (Davenport, 2019). Sometimes, it is impossible to say if the user is interacting with a machine or human being or where the information, such as guidelines or recommendations, is based on. These types of features in User Experience (UX) create distrust in human and computer interaction (Ribeira & Lapedriza, 2019).

Human-Centered Artificial Intelligence (HCAI) has been discovered as a solution for the distrust problem in human-AI interaction (Ribeira & Lapedriza, 2019; Rield, 2019). HCAI based systems give tools to solve problems and support humans but are based on complete transparency and justice. These systems are built to understand humans from a socio-cultural perspective, but they are also helping humans to understand them. Explainability is increasingly important since more often the user is non-expert on AI systems and the explanations must be produced in a format that the non-expert user can understand the information (Rield, 2019). It is important to sift towards Human-Centered Explainable Artificial Intelligence (HCXAI) to make AI trustworthy while maintaining the benefits of AI systems. HCXAI can be achieved by concentrating on two aspects of the high level of automatization and also the preservation of human control in AI-design decisions (Shneiderman, 2020b).

The purpose of this paper is to identify UX-design elements that enhance trust or prevent distrust in AI-based services and applications for the development of Explainable Human-Centered AI. However, the content of distrustful and trustful user experiences to identify the elements that enhance trust or cause distrust is already researched by Seckler, Heinz, Forde, Tuch, and Opwis (2014), However, in the paper, these elements are not identified in AI-enabled services, and applications but in the context of websites.

In turn, studies by Ribeiro, Singh, and Guestrin (2016) and Arriata, Díaz-Rodríguezb, Del Sera, Bennetot, Tabikg, Barbadoh, Garciag, Gil-Lopeza, Molinag, Benjaminsh, Chatila,f and Herrerag (2020) argue that explainability of individual predictions of any model was found to be important and different explainability models found to be crucial in building trust. Even though explainability of functions of AI-system itself is perceived as a part of user experience (Amershi, Weld, Vorvoreanu, Fourney, Nushi, Collisson, Suh, Iqbal, Bennett, Inkpen, Teevan, Kikin-Gil and Horvitz, 2019), explainability in this research only applies elements of User Experience and explainability models in algorithmic level are not further researched.

All in all, research in this topical field tends to stay more at the theoretical level (e.g., Shneiderman, 2020a & b; Davenport, 2019) and there are no studies about what are the elements that cause trust or distrust towards AI-based services and applications.

1.2 Research area, goal, and question

This research is an empirical study of User-Experience of different AI-based Systems that are used in daily life activities such as navigation or checking your text grammar and that are limited to Shneiderman’s (2020a) definition of consumer and professional applications. Furthermore, this research paper explores: what are the UX elements of AI-based systems that contribute to service or application to be perceived as trustful by users?

The goal of this research is to gather opinions of users of the services or application to provide a list of elements that affect the perception of trust in multiple AI systems in order to develop trustworthy HCXAI since trust plays a crucial sub-part of trustworthiness (Cho, Xu, Hurley, Mackay, Benjamin and Beaumont, 2019).

As said, the research is done from the interpretive perspective. The purpose is to gather opinions of users, and previous literature findings to make interpretations of them to understand the current social reality of trust in human-computer interaction.

1.3 Structure of the thesis

The first section of this paper is a literature review that explores the theory behind this empirical study. Firstly, the concepts of AI systems focusing on Machine and Deep Learning and trust with trustworthiness and Human-Centered Artificial Intelligence are presented. Secondly, User Experience is defined, including AI-UX key elements and guidelines for UX for trust.

The second section of this paper presents the two-part methodology of the study and data analysis. In the results, commonly used AI-enabled services and applications are ranked based on trust perceived by users, and eventually, UX elements that increase trust or cause distrust in services and applications are defined. The list of these UX elements is compared with the existent research on UX elements that affect trust.

Lastly, in the conclusion and discussion part, the key findings of the research are presented and how the new results can be used in order to develop an explainable human-centered approach in AI is discussed. In addition, limitations of the study and further research questions are considered in this part of the paper.

2 Literature review

In this section, the main concepts of the research are presented based on previous scientific literature. The articles and books referred are selected from tritonia.finna.fi database research. Databases used are ABI Inform Complete, ACM Digital Library, IEEEXplore - IEEE/IET Electronic Library, Webofknowledge, SAGE Research Methods, ScienceDirect, and Taylor & Francis Online Journal Library. Exceptions are the book of Rouhiainen (2019), the book of Lew and Schumacher (2020), the book of Knight (2019), the book of Hartson and Pyla (2018), and material of the Finnish Center for Artificial Intelligence (2020). All of the refereed papers and books are published between 2010 - February 2021 and this literature review is done by prioritizing the most recent research.

Since the research is part of multidisciplinary natured Human-Computer Interaction (HCI), the field of the search is not limited to one discipline but covers Information Science, design, and psychology. HCI is an area of research that tries to understand how people interact with computers and apply psychological principles to the design of computer systems. It is an area that is especially important now when the key factor for the success of AI technology is the user experience (Lew & Schumacher, 2020).

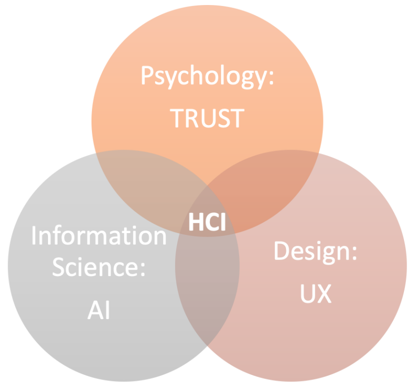

Figure 1. Multidisciplinary field of HCI in relation to topics of trust, AI, and UX.

2.1 AI systems

A system can be part of Artificial Intelligence if the system is autonomous and adaptive. The system must be able to deliver tasks at least partly independently in a complex environment and learn and develop its performance continuously (FCAI, 2020). However, the definition of AI is controversial and for example Campesato (2020) suggest using definition of ability of machines to do things that would require intelligence as if task would have been done by human, originally written by Raphael already in 1976.

According to Rouhiainen (2019), Artificial intelligence can be divided into three different levels of intelligence, that are Weak AI, Strong AI, and Super AI. Weak AI can deliver singular tasks while Strong AI can deliver multiple tasks simultaneously in a similar way as humans do. The most important difference between strong and weak AI is that the strong AI is capable of independent thinking while weak AI can only imitate human intelligence. However, weak AI is only currently used level of AI, but strong AI might be available shortly. Furthermore, Super AI represents the level of Intelligence that delivers tasks better than humans bypassing human intelligence and capability. It is improbable if this type of AI will ever be achieved.

Overall, Artificial Intelligence is a field of Information Science (IS) that is widely used in a variety of applications, such as natural language processing, automatic programming, machine vision, automatic content recommendation, and automatic image processing. Machine Learning (ML) is a sub-concept of Artificial Intelligence and refers to systems that improve their performance by learning from previous experiences or data. Deep learning (DL), in turn, is a sub-concept of Machine Learning based on more complex and deep mathematical models. The parallel concept of artificial intelligence is data science, which refers to algorithms and data management. Even though Data Science is partially tangential to all the subfields of AI (FCAI, 2020), that is not the focus of this paper. This research refers to AI systems as systems that are based on Artificial Intelligence or, more precisely, Machine or Deep Learning.

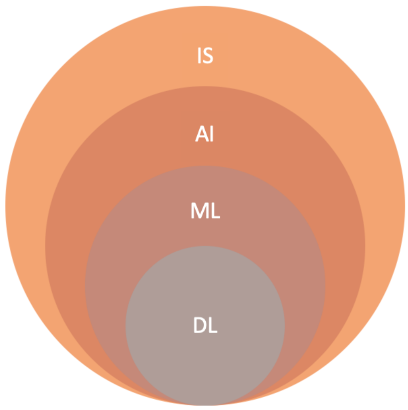

Figure 2. Artificial Intelligence as a sub-part of Information Science (FCAI, 2020).

2.1.1 Machine Learning

Campesato (2020) define Machine Learning as a subset of AI that can proceed tasks that are infeasible for more conventional programming languages. ML includes multiple different algorithms, but the selection of data is the most crucial part of ML, whether it is labelled or unlabelled. Hence, Machine Learning can be divided into three categories of supervised, unsupervised, and reinforcement learning based on the used data (FCAI, 2020; Shalev-Shwartz & Ben-David, 2014).

Supervised learning refers to simple classification. For example, recognizing whether the mail is *spam* or *non-spam* is a simple binary classification problem where the email *is* or *is not* belonging to a particular category. In this type of spam detection task, the learner has received training, and based on that the learner will figure out a rule for labeling (FCAI, 2020; Shalev-Shwartz & Ben-David, 2014).

Unsupervised learning, in turn, refers to a situation where there are no premade classes and data itself is unlabeled. Input data is being explored and the same type of inputs are clustered or classified. Besides, data can also be presented by a few key variables or dimensions. Moreover, data visualization and generating new data based on the input can also be counted as unsupervised learning. Unsupervised spam detection task would differ from supervised way of learning, by providing a large bundle of emails without any training or relabeling the messages. It is the learner's task to make up the rule and detect what are unusual messages and which are not (Campesato, 2020; FCAI, 2020; Shalev-Shwartz & Ben-David, 2014).

Reinforcement learning refers to situations where there is more information in training examples than the test examples. Hence the learner has to predict information for the test examples. For example, a self-driving car has to operate in a complex environment where feedback from decisions may be delayed. Reinforcement learning is also used in games where the outcome is only decided at the end of the game, such as chess (FCAI, 2020; Shalev-Shwartz & Ben-David, 2014).

These three categories of ML are not clearly defined, and many ML problems and methodologies are not possible to categorize. For example, semi-supervised learning refers to problems and methods in between supervised and unsupervised learning and where some datapoints are labeled and some unlabeled (Campesato, 2020; FCAI, 2020).

2.1.2 Deep Learning

Deep Learning refers to a network of simple ML processing unit layers and it is a subset of Machine learning (Campesato, 2020; FCAI, 2020). The information, inputted into the system, passes through the layers. The structure is inspired by the complex functioning of the human brain and consists of operating components called neurons. The nature of the depth and the layers of neural networks allows the learning of complex structures and rules without large amounts of data (FCAI, 2020).

According to Erdal, Kächele, and Schewenker (2016), neural networks have enabled learning abstract high-level projection from raw data. Thus, Neural Networks have enabled major development in areas of ML such as natural language processing by recognizing emotions in speech (FCAI, 2020; Erdal & alt, 2016).

Neural networks have few special features that differ from traditional computer systems that make the development of ML possible. Firstly, neural networks are capable to process data in each neuron independently, in order to process a vast amount of data simultaneously. In contrast, in traditional computer architecture, the computing and data processing happens in the central processing unit (CPU) that can only receive a small amount of data from the memory unit and process it and store it back to memory before receiving more data packages (FCAI, 2020).

As described earlier, in traditional systems data storage and processing are divided into two components but in neural networks, the components, neurons, can store and process data. The neurons have short-term storage while long-term memory is based on the connection between neurons. The information is stored as connections (FCAI, 2020).

The idea that deep learning is the best approach to develop AI is based on the fact when we simulate sub-symbolic data processing by neurons and neural networks the result is intelligence. This argument is based on logic where developing human-level intelligence requires simulating human abstract thinking and our symbolic reasoning that is based on certain concepts through logical reasoning (FCAI, 2020). One interesting aspect of deep learning is Neuro-Symbolic AI (NS), which will be further explained in chapter 2.3.2, which combines neural networks techniques with symbolic reasoning and could be an answer for truly explainable AI (Doran, Schulz and Besold, 2018; Confalonieri, Coba, Wagner & Besold, 2020).

2.2 Trustworthiness and trust

Trustworthiness is agreed to be one of the most important aims of explainable human-centered Artificial Intelligence (Arrieta & alt, 2020; Ribeiro, Singh, Guestrin, 2016; Shneiderman, 2020a). However, Arrieta and others (2020) remind that trustworthy AI is not directly equal to an explainable one, but trustworthiness is a basic requirement for AI systems and their models to be considered as explainable.

According to Cho, Xu, Hurley, Mackay, Benjamin, and Beaumont (2019) concept of trustworthiness is referring to a good state of quality in information, systems, or entities in Information Science. Furthermore, Arrieta and others (2020) explain trustworthiness in AI as the confidence of whether a model will act as intended. Often, trustworthiness is mixed with trust, but these terms should be distinguished from each other. Trustworthiness refers to an objective view of trust that is based on evidence. Trust, in turn, refers to a subjective view of humans (Cho, 2015).

However, trust is one of the four attributes of trustworthiness thus subjective view is in some terms reviewed in trustworthiness. Other attributes are security, resilience, and agility (Cho & alt, 2019). According to Pfleeger, Pfleeger, and Margulies (2015), security in computing systems can be perceived as a danger or threat-free.

Cho and others (2019) define resilience as a system’s ability to withstand disruption and recover from it within acceptable delay and cost. Besides withstanding, resilience refers to the capability to reduce the duration and magnitude of these disruptions. In turn, agility can be defined by the speed and cost of adaptability when facing a sudden change or unexpected circumstances. (Cho & alt, 2019). However, this research's focus is on this subjective view of trust, and it will be presented in more detail next.

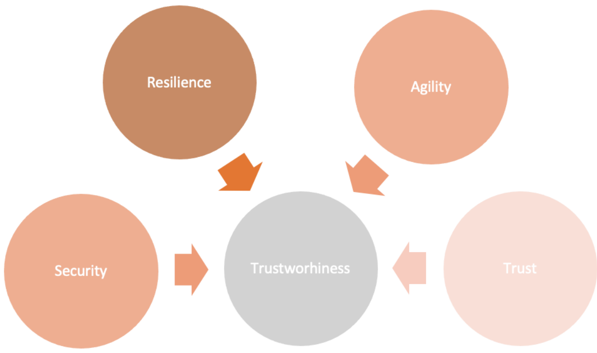

Figure 3. Trust as a sub-part of trustworthiness (Cho & alt, 2019).

2.2.1 Trust

Cho, Swami and Chen (2011) define trust as a subjective belief of trustor behaving as expected when trustee takes a risk in an uncertain situation. Based on the definition trust in Human-Computer Interaction (HCI) can be concluded as an interaction where the trustee is the human and where the trustor is the computer or the system. Thus, trust is a relationship between two actors, trustee and trustor, and it is usually based on past interactions between these actors (Cho & alt, 2011). However, Pengnate and Sarathy (2017) have studied trust in the context of unfamiliar websites where there are no prior experiences. Trust in this kind of context is called initial trust that is based on cognitive cues and rapid first impressions.

Seckler, Heinz, Forde, Tuch, and Opwis (2014) explain that trust always involves vulnerability. They argue that trust is mainly needed in an uncertain and risky environment. This type of environment can be for example that user visits even though they are uncertain about risks and full consequences involved. In this research, these types of environments are considered in AI-enabled services and applications.

Cho and all (2019) define trust at the systems level on sub-attributes that are predictability, reliability, and safety. Predictability is specified as high certainty that the trustee behaves or decides as predicted or otherwise successfully performs a task. The level of this certainty is affected by lack, vagueness, or ambiguity of information. In turn, safety is the absence of risks or threats. Safety includes aspects of cybersecurity, environmental security, physical security, and logical security. Lastly, reliability intends to provide the same, uniform results according to the need whenever the user uses the system (Cho & all, 2019).

2.2.2 Distrust

Seckler, Heinz, Forde, Tuch, and Opwis (2014) argue that the antonym of trust is not distrust but these perceptions can actually exist simultaneously. Even when the user has a lack of trust for a particular product or service, it does not necessarily mean that they have perceived the product or service distrustful. According to Benamati, Serva, and Fuller (2010) distrust refers to the unwillingness to become vulnerable to the trustee. Distrust includes a belief that the trustee behaves in a harmful, neglectful, and incompetent way.

2.3 Explainable Human-Centered Artificial Intelligence

2.3.1 Human-Centered Artificial Intelligence

People are increasingly in contact with Artificial Intelligence-based systems even without considering that their daily decisions are based on AI-enabled algorithmics. The purpose of the Human-Centered Artificial Intelligence approach is to create AI systems designed for a system composed of different people, such as users and customers. The role of artificial intelligence is not to replace people but rather to give people new tools for solving everyday problems based on transparency and fairness (Rield, 2019; Lew & Schumacher, 2020). HCAI focuses on enchanting human performance in ways that the system becomes reliable, safe, and trustworthy. HCAI brings humans to the center of the design and emphasizes the importance of User Experience Design (Shneiderman, 2020a) that will be presented later in chapter 2.4. HCAI design can be divided into two parts. Firstly, to AI systems that help people understand them, and secondly to systems that understand people from a socio-cultural perspective (Rield, 2019).

When the user does not know which algorithm the AI system uses or if it is almost impossible to check it due to the complexity of the system, one can talk about a BlackBox situation that is opposite for transparency (Rield, 2019; Arrieta, Díaz-Rodríguezb, Sera, Bennetot, Tabikg, Barbadoh, Garciag, Gil-Lopeza, Molinag, Benjaminsh, Chatilaf & Herrerag, 2020). In a BlackBox situation, the user may make decisions without understanding what the instructions, recommendations, or other outputs provided by the system are based on. To prevent such a situation from arising, AI systems should be designed in such a way that users can better understand their decisions. It is important to understand what people need to know about an AI system thus they can base their decisions on it and whether it is possible to implement the system in a way that delivers this information to the user in an understandable way (Rield, 2019). Explainability (see chapter 2.3.2) and causability (see chapter 2.3.3) are important components of making the system more transparent and understandable for humans.

However, even with AI systems, it may be impossible to understand a human being because the system does not have the same inputs as a human. In many situations, people behave on the basis of general knowledge derived from socio-cultural beliefs or norms. Intelligent systems must have some input on general human knowledge on how people behave and why they behave that particular way. In this way, AI systems would be able to make better predictions about human activity and avoid mistakes in things that humans take for granted. Understanding socio-cultural norms and beliefs make systems more reliable and secure, even for people who do not have an expert level understanding of Information Science or more specifically in Artificial Intelligence.

In the future, it may be possible to create systems that monitor their own behaviour in accordance with ethical and fair values (Rield, 2019).

All in all, developing an AI system to a) help humans to better understand them and b) understand humans from a socio-cultural perspective, requires focusing on measuring performance and satisfaction of humans, valuing their needs, and also ensuring sufficient human control where automatization of the system enables high-level of human performance (Shneiderman, 2020a & b).

Figure 4. HCAI systems (Rield, 2019).

2.3.2 Explainable Artificial Intelligence

Explainable AI (XAI) and furthermore Human-Centered Explainable AI (HCXAI) refers to the system's explanations of outputs in a format that all humans, including the non-expert user can understand the information (Rield, 2019; Ding, 2018). Arrieta and others (2020) explain XAI as a system that produces details and reasons for its functions for its audience to easily understand it better. Moreover, explainable AI enables humans to understand, trust, and effectively manage AI systems. Doran, Schulz, and Besold (2018) have listed traits of explainable systems: confidence, trust, safety, ethics, and fairness. These traits are part of explainable systems, but they are not only dependent on the system's explainability but rather on the interaction between the user and the system. For example, trust is a user's own perception of the system.

To achieve the trustworthiness of the AI system, which is one of the goals of XAI, detailed explanations of algorithmic outputs are necessary. These explanations should give an insight into AI processes, data, and models that are used to draw a conclusion (Doran, Schulz, and Besold, 2018; Arrieta & alt, 2020). However, previously referred Black Box situations are common, and users frequently accept the outcome of AI without critically analyzing it. This can be caused by the necessity of using these AI-based services or applications or by the choice of accepting the situation (Doran, Schulz, and Besold, 2018; Rield, 2019).

Doran, Schulz, and Besold (2018) characterized three notions of explainable AI currently existing and define one truly explainable notion. The first one is called opaques systems and it refers to systems that offer no perception of the algorithmic mechanisms. Interpretable systems, in turn, provide access for its user to mathematically analyze the algorithmic mechanisms. Comprehensive systems provide visual elements and symbols such as words, to explain how the conclusions are reached. Interpretable and comprehensive systems are near to explainable systems, but truly explainable systems should automatically reason their outputs without a human generative process. This last notion uses human-centered featured to input data and reason explanations of occurred outputs. That also support Ding's (2018) interpretation of explainability where the systems must build trust by providing explanation behind AI system's conclusions and reasonings. Doran, Schulz and, Besold (2018) argue that leaving the explanation generation to humans be dangerous since human interpret is dependent on their background and expertise. Truly explainable systems differ from interpretable and comprehensive systems by producing the explanations themselves rather than only enabling the explaining of the occurring events.

One interesting way to develop systems to be truly explainable is within neural-symbolic integration and Neuro-Symbolic AI which combines neural network techniques with symbolic reasoning (Doran & alt, 2018; Confalonieri, Coba, Wagner & Besold, 2020). These symbolic systems operate at a symbolic level where reasoning is operated at an abstract level. Symbolic process and reasoning resemble humanlike explanations and that is why the process levels up the explainability. Comparing to current deep learning systems, NS systems have multiple benefits. NS is more data-efficient and requires less sample complexity by abstract learning logic. In addition, NS can generalize its sample distribution rather well and thus it can transfer its learnings. Most importantly, NS systems can be trained to process in a humanlike way that can be understood easily by humans and communicated transparently. Bennetot, Laurent, Chatila, and Díaz-Rodríguez (2019) propose NS, where the system provides a fair explanation of its reasoning and correct its biases thoroughly automatically, making the system as well explainable as human-centered. NS is especially important for explainability for Deep Neural Networks that tend to be very complex because of their nature of layers (Arrieta & alt, 2020).

However, as important technical aspect of developing XAI and HCXAI is, Ehsan, and Rield (2020) reminds that explainability is especially an HCI problem since understanding the explanations depends on who is on the receiving end. Breaking out of black-boxed AI systems is highly dependent on the question of who is the human in the loop? Understanding the who is crucial since it names the explanation requirements. This comes even more important since AI systems are increasingly implemented in socially situated applications for example in health care and in communications. All in all, Ehsan and Rield (2020) advocate for a reflective sociotechnical approach that includes both social such as interaction design and technical elements such as explainable models for developing AI systems towards HCXAI systems. In this research, the explanation interface is in focus in the perspective of UX and explainability models, which are for example studied by Arriata and others (2019), are left out of the scope.

2.3.3 Causability

According to Holzinger, Langs, Denk, Zatloukal, and Müller (2019) while explainability highlights relevant parts of the algorithmic decision making, causability extent to the level of causal understanding that human receives with effectiveness, efficiency, and satisfaction in the specific content. Causability refers to the human-understandable model. Shin (2021) explains that while explanations generate users’ trust, causability afford users emotional confidence. Causability justifies what and how should be explained, thus it has an indirect effect on the perceived trust by the user. That is why both, explainability and causability are necessary for reducing the opacity of Black-Boxed AI systems and build up trust within (Holzinger & alt, 2019). Shin (2021) highlights that including causability and explainability within AI systems, increases trust and supports access to high-quality explanations. Especially, trust-based feedback loops, that connect causability with explainability, are considered useful designing AI interfaces and UX.

2.4 User Experience

According to Lew and Schumacher (2020), User Experience or UX refers to a design approach to view new technologies as experiences, not as products. User Experience Design, UXD, refers to the whole process of creating elements that are relevant for the user and enables them to interact with the world around them effectively and efficiently, for example with Intelligent Virtual Assistant (Unger & Chandler, 2012; Lew & Schumacher, 2020).

User experience takes into consideration for example, information architecture, visual design, Interaction design and Industrial design (Knight, 2019). Thus, User Experience must be considered as a holistic approach rather than a specific design field. Knight (2019) considers UX honeycomb model, originally created by Morville (2004), as a great representation of complex and multidisciplinary character of UX.

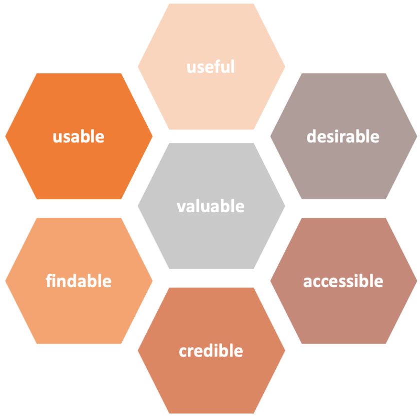

Figure 5. The UC Honeycomb (Morville, 2004).

Morville’s (2004) model consist of seven facets of UX. Usable refers to the ease of use and learnability. Useful refers to the fact that a product or a service must fulfil its users’ needs or solve a specific problem. Desirable, in turn, refers to the emotional aspect of UX. The user should feel connection to the brand, an identity, or to the product. Accessible refers to the fact that services and products should be available for everyone and when designing them people with different disabilities should be kept in focus. In addition, user must be able to find everything that they need, this refers to term findable. Credible refers to the trustworthiness of the product and whether the user trust and believes what they are told to. Lastly, the product must bring value to its users and other stakeholders (valuable) (Knight, 2019; Morville, 2004).

As well as a holistic approach, UXD can be viewed as an integrated and iterative part of entire process of generating and integrating a product (Knight, 2019). Even though UX processes very depending on the project, Hartson and Pyla (2018) have identified four main elements of design processes: understanding needs, prototype candidates, and evaluate UX. Where the first step is to understand the user and their needs. The second step refers to creating a concept of design and interaction within the experience. The third step refers to prototyping and understanding the design alternatives. Lastly, evaluating UX refers to measuring whether and in which ways the design matches the user needs and how the design can be refined. All the four components are evaluated for example by testing or analysing preferably with the user. This model is called as The Wheel.

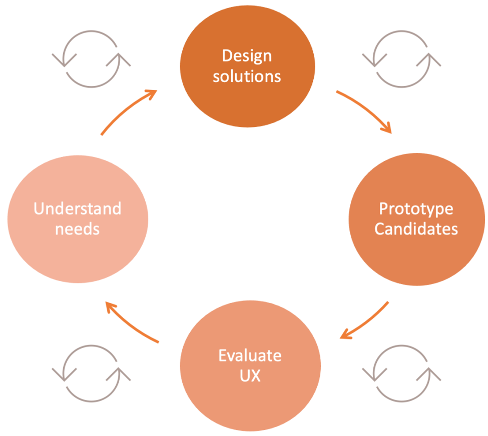

Figure 6. UX Wheel (Hartson & Pyla, 2018).

2.4.1 UX of AI

When it comes to the design of AI systems, they have been thought of mainly by their functionality. While this has been very important for the rapid development of AI, it is equally important to think about the experience of using AI-based applications and services (Lew & Schumacher, 2020). After all, there is already plenty of AI functions that have been developed but they are not accessed or used by users, and even more on, they are not understood by the user.

Lew and Schumacher (2020) list three principles for AI-UX: context, interaction, and trust. Context is all outside information that an AI system uses to perform a certain task and it can include information about the user and why are they requesting a certain task. Context can also include information about the external world. For UXD, it is important to understand the context of systems outputs to gain knowledge on the meaning of the output and its expectations (Lew and Schumacher, 2020).

Interaction is important since AI systems can potentially make significant actions on their user's behalf and this needs to be communicated even before performing the task. Interaction can be designed in various ways, for example by engaging the user with a message or push notification (Lew & Schumacher, 2020).

Lastly, trust refers to the feeling of the user that a particular AI system successfully performs a task demanded by a user without unexpected outcomes. Lew and Schumacher (2020) define unexpected outcomes as unnecessary additional tasks that the user did not demand, or these tasks are breaching the user's privacy. Trust is an important principle of the User experience of AI especially in the user adaption since trust is easily lost and gaining it back is difficult. Vice versa, it is likely that if the user once trusts the service or application, they will continue to keep trusting it. For example, when Siri, Apple's Intelligence personal assistant, IPA, was introduced, users tend to express negative emotions and frustration. Despite transformational technical capabilities at the time (Vimalkumar, Sharma, Singh & Dwivedi, 2021), many things such as too easily and many times accidentally triggered Siri and its speech recognition was not robust and did not match customer's expectations (Lew and Schumacher, 2020). In a very short time, people stopped using it and this distrust prevented people to try any similar products. For example, Microsoft's Cortana IPA did suffer from distrust in the product markets. This era of distrust towards IPAs changed only when Amazon introduces its personal assistant Alexa where its user experience was well received by the users. As the example showed product's perception by its users is a sum of previously experienced that the user has with the product. If the product provides the value that users expected, they will be more likely to trust the product (Lew and Schumacher, 2020).

2.4.1 Key elements of UX of AI

Lew and Schumaher (2020) define key elements of UX of AI as utility, usability, aesthetics, and scale of weirdness. Additionally, Shneiderman (2020b) reminds that feeling of the position of control and fully autonomous are important factors for UX of AI and especially regarding trust in AI systems. The balance between automatizing tasks with AI and control of human users is one of the key elements to take into consideration.

Utility or functionality refers to the fit of the application to the purpose that it is designed for. The application should have the features and functions expected and needed by the user. Designing frictionless and effortless interactions for the user to experience utility and functionality are base of a successful product. In turn, usability refers to the efficiency, effectiveness, and safety of users to perform functions within an AI system (Lew and Schumaher, 2020). In comparison to the Morville's (2004) UX honeycomb, utility and usability are similar to the facets of usability and usefulness (see chapter 2.4).

Utility and usability are directly affecting the trust in interaction. It is common that users cannot trust what the AI system delivers as an output if they are off point. People have a low tolerance for mistakes made by machines or their insufficient utility or poor usability. The same can be applied with AI-enabled products. For example, previously mentioned long-lasting distrust for IPA-products was repeal by Amazon's Alexa. This newly built trust in IPAs was based on the utility and usability of Alexa in specific areas, such as playing the news or telling the weather. However, there have been many useful skills implemented but because of their low usability, they have been discarded from the system (Lew and Schumaher, 2020). Furthermore, according to Vimalkumar and others (2021) users who believe that the system offers great utility, are more likely to adopt the technology regardless of other trust issues such as issues related to privacy or usage of personal data.

The aesthetic usability effect, in turn, describes that user will define the product as more usable when it is also visually pleasing. While utility and usability construct the base for a good user experience, aesthetics is an attribute that differentiates products that have an equal level of utility and usability (Lew and Schumaher, 2020). Interestingly, a study by Pengnate and Sarathy (2017) revealed that aesthetics can sometimes affect user's evaluation of trust in greater impact than usability. Aesthetics can communicate feelings and emotions to the user and emotions have a profound effect on the user's perception of the product (Lew and Schumaher, 2020). That is why also the trustworthiness of the AI-enabled service or application can be communicated by aesthetics. In comparison to the UX honeycomb (Morville, 2004) desirable this element is rather similar to the desirability (see chapter 2.4).

The weirdness of AI interactions refers to uncomfortable, awkward, and unusual situations for humans. Since AI systems have continuously more data about its user, it is important to scale their outputs and actions by how appropriate or inappropriate they are. This is called the weirdness scale (Lew and Schumaher, 2020). This scaling of suggestions based on personal patterns of behavior and habits requires a user-centric and moreover human-centric approach to understand humans from their socio-cultural perspective. Lew and Schumaher (2020) describe the line between acceptable and unacceptable triggered actions as UX questions and weirdness of the actions can be measured with the help of the user. For example, the AI engine can capture the times when the user clicked on a recommendation and when the user reported the recommendation disruptive or when the recommendation was left unnoticed. Scaling triggered actions and looking at AI from a UX perspective can help overcome fears for continuous monitoring and observation of users and build trustworthy and responsible interaction for humans and AI systems (Lew and Schumaher, 2020). For example, digital assistants, such as earlier mentioned Alexa and Siri have raised concerns on privacy and security since they collect sensitive data such as user's location history and browsing history and on top of that users believe them to be passive listeners of them (Vimalkumar & alt, 2021).

2.4.2 UX for trust

According to Yan, Kantola, and Zhang (2011) trust in HCI can be enhanced by proper UI design, by information and notifications visualization, and by trust management for computer systems and computer communications. These are all part of the User Experience and affect the perceived trust of the system. In addition, Yan and others (2011) define three main roots for trust in user experience. The first of them is interaction intension that refers to the degree of willingness to interact with the system. The second root cause is (computer) System Trust which means the perceived trustworthiness of the system. The last root cause was identified to be Communication Trust that refers to trustworthiness in communications via the system.

Seckler, Heinz, Forde, Tuch, and Opwis showed in their study (2014) that trustful and distrustful user experiences on the web, differ in terms of perceived honesty, competence, and benevolence. To prevent distrustful experiences resources should be directed to enhance honesty and benevolence. In turn, to enhance web trust resources should be directed to focus on competence and honesty.

In table 1, Seckler and others (2014) and Yan and others (2011) research results are combined. UX elements causing trust and preventing distrust are listed and grouped with corresponding trust root construct. This table can be compared with the results of this study to see whether the UX elements causing trust and preventing distrust will differ in AI-based systems.

Table 1. UX-elements that affect trust in the context of the web (Yan & alt, 2011; Seckler & alt, 2014).

Trust root construct (Yan & alt, 2011)		**Identified UX elements causing trust and preventing distrust** (Seckler & alt, 2014)
Interaction intension	**Social factors**: Internalization of user's reference group culture, and interpersonal agreements with others, in specific social situations	**Friends' social proof:** Recommendations by friends, family members, or colleagues
	Perceived Usefulness: How much user believes that using the system would enhance their benefits or profits	

	Personal Motivation: The perception of the user that interaction with the system delivers added value, such as increased performance, other benefits, or profits	
	Reputation: Public opinion on the system and its influence on the intention of HCI	**User's social proof:** user ratings and reviews
		Image or brand of the system operator
	Perceived Ease of Use: The perception of the level of easiness of use and freedom of effort	
	Personality: user's individual combination of emotional, attitudinal, and behavioral response	
	Relative Advantage: the perception that using the particular system is better than other similar ones	
	Recommendation: suggestions for user behavior	
System Trust	**System Quality:** The degree of how well the system can perform	**Plausible promises:** promises that the system kept on the perception of the user
		Prior experience
	Information Quality: the accuracy, correctness, and the reliability of the information	**Content:** the credibility of information, accurate, up to date information
		Expertise: Competence and quality of knowledge
	UI Quality: the feeling and ease of use	**Pop-ups or ads**
		Visual design: colors, layout complexity, and photos
		Usability: effectivity and efficacy of the UI elements such as navigation, links, and task flow

Communica-tion Trust	**Perceived privacy:** belief of user that the system is free of concerning private information disclosure	**Privacy: secondary use:** user's belief that their information is not being used for another reason that it is collected for
		Privacy: collection: user's belief that data is not collected by the system operator
		Demands: demands to share a link, download a piece of software, or create an account to get access to a website or a service
	Perceived identity: Belief of user the level that the system recognizes entities in communications	**Customer service:** availability of customer service agents, provision of service to customers after purchase
		Real-world link: For example, location or life experience
	Communication context: All the information that characterize the situation of communications	**(Web) address**
		Policy: Policy, general terms, and conditions

3 Research process and methods

Since the research is done from interpretive perspective where subjective opinions of users are gathered to understand the concept of trust in AI systems, the qualitative methods were chosen to be used. This research process is divided into three parts that consist of literature review, survey, and one-on-one interview methods.

Firstly, different AI-based services and applications are studied based on literature, and the most commonly used services are chosen by classifying them by the number of users globally and by everyday use type of use of these applications and services. This literature review method is chosen to investigate how AI-based services or applications can be categorized and which services and applications are most commonly used.

Secondly, different services are ranked via a short survey (see appendix 1) of AI-based applications and services, to classify them by perceived trust or distrust from most trustful service or application to least trustful or even distrustful one. The survey was conducted using Typeform (https://www.typeform.com) and the subjects of the study are ten adults between 20 to 30 years old. The responders were regular users of some of the applications or services are chosen or they use similar services or applications in daily use. Participants were informed about data use and the anonymous character of the survey.

The survey questionnaires are based on interview questions of Seckler, Heinz, Forde, Tuch, and Opwis's (2014) research in trustful and distrustful user experiences. The survey is divided into three parts to measure honesty, benevolence, and competence. In addition, the disposition trusting stance of participants is measured.

In this part of the research, survey was chosen to be used to map out which services the participants have used, in order to assign them one of these services or applications to review in the interview part of this study. This method also enabled to review the interview result and the chosen services or applications in more depth when analyzing the results. It revealed which services were the most trusted and which ones were not. In addition, the measurement of disposition trusting stance enabled to see whether the participants were usually likely to trust for example services or applications than distrust.

Lastly, an interview is conducted (see Appendix 2), in which the survey responders are randomly divided into small groups of one to three persons and in which the experimental groups perform an interview with open-ended questions examining which elements affect application or service to be perceived trustful. Other test groups, the control groups, do a similar interview for less trustful or distrustful ones. All the interviews are done individually via Zoom call or phone call.

The one-on-one interview is based on a technique called critical incidents technique (CIT) by Münscher and Kühlmann (2015) were asking an open-ended question from users to gather facts about an incident when the user felt trust or distrust towards the system open-ended critical incident question in this research was “Please describe whether you felt trust or distrust when using the service (or application) and describe the situation and environment in detail”.

Using CIT method within the interview was chosen since the it is more likely that the user recalls important experience issues related to that hence the data collected reflect the most important experience that affect trust. In order to help user to recall these memories, possible additional questions were asked to help the interviewee identify and describe the incidence.

The data of interviews were gathered by transcribing them during the interviews (see appendix 2). These short transcripts include a short description of the critical incidences experienced. This text data was further analyzed by identifying UX elements that were mentioned or experienced by the users when describing the incidence.

Finally, these elements were listed and analyzed by comparing and thematizing UX elements that enhance trust. Furthermore, themes are compared to previous research of

UX elements that affect trust in the general web or online context (see table 1). In addition, the results will be compared to previously identified key elements of UX of AI (see chapter 2.4.1). Results will reveal what is the list of elements that are specific to AI-based services or applications.

4 Results

4.1 Commonly used AI services and applications

The list of commonly chosen services is based on Elliot's (2019) and FCAI's (2020) examples of everyday services and applications of AI (Table 2) that are divided into three main categories of traveling and transportation applications and Intelligence Personal Assistants, and communication applications. Different categories are included in this study to have more wide racing understanding of trust in AI-based services since viewing only one category might reflect trust at the specific field or function rather than AI-based services overall. However, in this research commonly used AI services and applications are limited to Shneiderman's (2020a) definition of consumer and professional applications that are not consequential or life-critical for human life.

Application and services were chosen base on the commonness of their use globally. Estimation of users can be found in table 2. To be recognized in the category of commonly used, services or applications must have over 10 million users a month. However, many of the chosen applications and services far exceed this estimation.

4.1.1 Traveling and transportation applications

Google Maps and Uber were chosen to use as an example of traveling and transport category. Google Maps is a web mapping service that is based on deep learning algorithmic such as *Graph Neural Networks* that fetch data from various sources. For example, location, historic and current traffic, governance, and feedback data from users are used to develop the mapping system and predict traffic for navigation services (Lau, 2020).

Furthermore, Uber is an AI-powered platform for transportation. The platform is based on machine learning to predict accurate travel times, best possible routes to the destination, and to connect drivers and users nearby each other. Uber also uses deep learning to improve the service's understanding of cities and traffic (Uber, 2021).

4.1.2 Intelligent Personal Assistants

AI-based assistants services are usually based on the use of Natural Language User Interfaces, NLUIs, to enable communication between humans and computers. Intelligent Personal Assistants, IPAs, analyze user's speech or gestures to provide services or support functions for the user (De Barcelos Silva, Comes, Da Costa, Da Rosa Righi, Barbosa, Pessin, De Doncker, & Federizzi, 2020).

For this study, Apple's Siri and Amazon's Alexa were chosen to explore. Both of them are personal assistants that are based on speech synthesis and deep neural networks, DNNs. They provide hands-free access to different functions such as playing music, setting timers or lists, calls, or for example controlling lights of the room. Where Siri's NLUIs are embedded on Apple's devices such as computers and smartphones, Alexa is requiring a smart speaker and it is possible to access smart home functions. Both of the AI's are based on companies' clouds (Apple, 2017; Kim, 2018 & Amazon, 2020)

4.1.3 Communication applications

For communication services, Grammarly was selected. Grammarly is an AI-based writing assistant that analyzes natural language to spot mistakes, spelling errors, or the tone of the text. It uses ML, DL, and Natural Language Processing NLP. The advantages of Grammarly are that it can be used in most communications functions and on the browser (Grammarly, 2019).

Table 2. List of commonly used AI services and applications.

AI-based service or application	**Category**	**Users**
Google Maps	Travelling and transportation applications	one billion/month (Google, 2020)
Uber	Travelling and transportation applications	78 million/month (Statista, 2020)
Siri	IPAs	374 million / month (Maggio, 2018)
Alexa	IPAs	unknown

		Alexa has more than 60,000 supported devices (Statista, 2019).
Grammarly	Communication applications	30 million/day (Grammarly, 2020)

4.2 Trusted AI services and applications

To compare perceived trust of listed AI services, a short survey was concluded using Type form-platform. The survey evaluated perceived trust by measuring perceived honesty, benevolence, and competence of services and applications of all the categories listed. Survey questions are attached as appendix 1.

According to the survey Uber was perceived to be the most trustful of compared services and applications. It was ranked to the highest level of honesty, benevolence, and competence of its services while selected IPAs, Alexa and Siri, were ranked to be the least trustful of all the services and applications. The results can be found in table 3, in which levels of honesty, benevolence, and competence and moreover calculated overall perceived trust are based on the average of all submitted responses of the concluded survey.

Table 3. Perceived trust of commonly used AI services and applications.

Ranking from most trusted to least trusted	AI-based service or application	Category	Honesty (where 100 % strongly perceived honesty)	Benevolence (where 100 % strongly perceived benevolence)	Competence (where 100 % strong perceived compe-	Overall trust level (where 100 % strong perceived trust)

					tence)	
1.	Uber	Travelling and transportation applications	73 %	90 %	97 %	87 %
2.	Grammarly	Communica-tion applica-tions	64 %	73 %	89 %	76 %
3.	Google Maps	Travelling and transportation applications	50%	47 %	82 %	60 %
4.	Siri	IPAs	49 %	44 %	60 %	51 %
4.	Alexa	IPAs	20 %	33 %	100 %	51 %

What can be noticed of all the services and applications, is that they all have very high competence levels while perceived honesty and benevolence vary. In addition, results show that honesty does only reach the level of 20% to 73 % of 100% (strongly perceived honesty) whereas benevolence differs between 33% to 90% of 100% (strongly perceived benevolence). Honesty is always perceived lower, except in the case of Google Maps where honesty and benevolence are the rather same level.

All in all, it can be noted that the honesty and benevolence levels are quite low comparing how popular are these services or applications. In addition, from all of the services, only Uber and Grammarly reach over 50% level in all three components. This aligns with Davenport’s (2019) study where 40 percent of users did not have any trust in AI-based services in the United States.

4.3 Analysis of the services and application

4.3.1 Participants

Almost all of the participants had used some of the listed services or applications and only one of the participants did not use any of the services. In the survey, the trusting stand of participants was measured by the question “I usually trust people until they give me a reason not to trust them”. Most of the participants agreed or strongly agreed to this while few answered neutral positions. The Median was 4 (agree). Measured trusting stand reveals that the study’s participant sample was more likely to trust service or application than distrust. The high trust stand might affect the results of the study.

All of the participants who had used the services or applications were randomly assigned one service or application that they use or have previously used in order to describe the critical incidence. There were between one to three interviewees for service and application.

4.3.2 Google Maps

Google Maps was one of the most used services of all of the listed commonly used services in this study and according to Google’s (2020) user statistics. The average perceived honesty of Google Maps was 50%, and the median rating of biases, notifying of changes and transparent sources of Google maps, was 2 out of 5. However, the results show that there is more variation of ratings in the question of regarding biased information of Google Maps.

Perceived benevolence was under 50%. However, the average ratings of question: “I believe that only the necessary information for Google Maps to function is being collected” was only 1,8 of 5 (36%) and median 1, which refers that most of the participants believe that Google Maps collects information of the users that is not required for the service to function.

Surprisingly, perceived competence was 82% that communicates that usage of the service might lie on its liability, accuracy, and overall functionality. The median of all the questions was 4.

Table 4. Google Maps: perceived honesty, benevolence, and competence.

Google Maps (n=9)	**honesty**	**benevolence**	**competence**
	50%	47 %	82 %

Furthermore, Critical Incidence interviews support the results of the survey. In this part, two types of situations were identified. On the other hand, the interviewees said that they trust Google Maps based on its accuracy and its logical successions. Thus, information quality refers to the accuracy, correctness, and reliability of information and systems quality (Yan & alt, 2011), and content refers to the accurate and up-to-date information where the suggestions are logical to the user (Seckler & alt, 2014) are validated. However, both of the interviews noticed that the accuracy and functionality of the Google Maps service depend on the environment where it is used. For example, outside the city areas or in certain countries, such as South Korea, the Google Maps service is not reliable, and the user used Korean apps instead of Google Maps. This validates the relative advantage concept of Yan and others (2011).

Furthermore, a situation where the location was not familiar beforehand for the user, deepened the trust for the Google Maps service. This support in an unfamiliar situation can be used in UX design by identifying situations that are not familiar to the user and support in those situations. This is also related to the human-centered approach since it is crucial to empathize with the user and understand them from the socio-cultural situation.

On the other hand, an interviewee, who felt distrust towards the service, described that the distrust is based on the Google company image and how much data they generally collect from their users. This validates the UX elements identified by Seckler and others (2014) and Yan and others (2011) of reputation, image, or brand of the system causing trust and preventing distrust. The participant also referred to the privacy concern of collection of data (Seckler & alt, 2014) as an element that affects their perception of trust.

In summary, the UX components identified that enhance trust or diminish distrust are listed in table 5.

Table 5. Google Maps: identified UX elements that enhance trust or prevent distrust.

Google Maps				
New element	Element	Description	Origin	Interview quotes
	Information quality	The accuracy, correctness, and reliability of the information	Yan & alt (2011)	*" if the route was logical for you, then I have felt trust towards the service."*
	System quality	The degree of how well the system can perform	Yan & alt (2011)	*"The map service works well, in my point of view, and it gives very clear guidance to find the place I want to go to."*
	Content	The credibility of information, accurate, up to date information	Seckler & alt (2014)	*"-- sometimes Google Maps is not reliable, for example outside the city areas, and it does not know all the roads or best routes there. "*
	Relative advantage	The perception that using the particular system is better than other similar ones	Yan & alt, (2011)	*"When I was an exchange student in Seoul, I would use Korean apps to get more accurate info. I would use Google Maps alongside these apps, but I*

				trusted the Korean map apps over Google Maps."
Support in an unfamiliar situation		Support of users in situations that they perceive as new or difficult		*"I truly trust Google Maps, when I am going for place of which location, I am not sure."* *"Trust to Google Maps deepens in situations when I am in unfamiliar place, and I don't know the routes or roads there. In these situations, I feel that I blindly trust the service since I have no other choice."*
	Reputation	Public opinion on the system and its influence on the intention of HCI	Yan & alt (2011)	*"I distrust Google as a corporation in general as they collect so much information"*
	Image or brand	Image or brand of the system operator	Seckler & alt (2014)	*"I distrust Google as a corporation in general—"*
	Privacy: collection	User's belief that data is not collected by the system operator	Seckler & alt (2014)	*"--as they collect so much information—"*

4.3.3 Uber

The perceived trust for Uber was the highest based on the three components: honesty, benevolence, and competence. However, Uber was not used as much as Google Maps by the participant of the survey. Perceived honesty of Uber service was 73% while benevolence rating was 90%. Competence of the service was perceived 97% that refers the great accuracy, information quality, and personalization.

Table 6. Uber: perceived honesty, benevolence, and competence.

Uber (n=4)	**honesty**	**benevolence**	**competence**
	73 %	90 %	97 %

The interviews support the results of the survey and all of the interviewees felt trust towards the service when describing the Critical incidence.

In the interviews, it came clear that recommendations by friends, family members, or colleagues, so-called friends' social proof (Seckler & alt, 2014), is an extremely important element when building trust. Also, the effect of social factors (Yan & alt, 2011), such as that the users felt belonging to the exchange student group who all used uber, was validated.

Another factor that came up in the interviews and that confirms Seckler & others' (2014) findings that the real-world link enhances trust. Real-world links can be things such as real location or life experience that the user experience or brick-and-mortar shop in case of online shop. In the case of Uber, it has many real-life links, for example, drivers, cars, and location to list a few. Furthermore, the availability of great customer service by the drivers was found to deepen the trust (Seckler & alt, 2014).

Another user experience element that came up also with Google Maps (see chapter 4.3.2) is that unfamiliar location, enhances the trust in the services. For example, one of the interviewees described a situation where they were abroad in an area that was often the scene of theft and, in major accidents, threats with weapons. But using Uber, made them trust uber even more in this area since they were always safe during these rides.

> *"--Us exchange students mostly used Uber, especially when it came to getting home from a party in the middle of the night. I still remember a situation where I*

took an Uber alone and I fell asleep in the passenger seat of the Uber and only woke up again when we were at my house. I was a bit perplexed about that and it was a bit uncomfortable for me to have fallen asleep next to a stranger. However, I had no worries that I was not safe with him - or her –"

Another interviewee identified that their trust in the service also relies on the fact that alternative services (taxis) are not that reliable or not available (car). That also confirms that Yan and others' (2011) finding that relative advantage**,** that the perception that using the particular system is better than other similar ones, fits also for AI-based systems.

Lastly, the excellent competence level was discussed by the interviewee. They explained that they trust that Uber has accurate and fare prices and routes. The interviewee further explained that the trust by accuracy is based on the technology and the fact that the route is defined automatically, without input from the driver. This validates the importance of information quality and system quality (Yan & alt, 2011). In addition, Seckler and others (2014) identified elements of credibility of information, and accurate and up-to-date information (content), and that the system keeps its promises on the perception of the user (plausible promises).

An element that was also found to enhance trust was the autonomy of the user, even though the route was defined automatically, the user can change it or choose another one when needed. This refers to the Shneiderman's (2020b) article of importance of the feeling of the position of control and fully autonomous are important factors for UX of AI and especially regarding trust in AI-systems. This balance between automatizing tasks with AI and control of human users has been designed well in the case of Uber, and that also has a high rate of trust in the service.

Table 7. Uber: identified UX elements that enhance trust or prevent distrust.

Uber

New element	Element	Description	Origin	Interview quotes
	Friends' social proof	Recommendations by friends, family members, or colleagues	Seckler & alt (2014)	*"Us exchange students mostly used Uber"*
	Social factors	Internalization of user's reference group culture, and interpersonal agreements with others, in specific social situations	Yan & alt (2011)	*"Us exchange students mostly used Uber"*
	Real-world link	For example, real-life location	Seckler & alt (2014)	*" I lived in the city of Valparaíso. -- Us exchange students mostly used Uber"*
	Customer service	Availability of customer service before, during, and after the experience	Seckler & alt, (2014)	*"Uber drivers were usually very chatty, open and helpful. They gave you tips and always made you feel safe."*
Support in an unfamiliar situation		Support of users in situations that they perceive as new or difficult		*" When I had my semester abroad in Chile" (being abroad)*
	Relative advantage	The perception that using the particular system is better than other similar ones	Yan & alt, (2011)	*"I need to get to a destination by car, and I do not have a car of my own"* *" This contrasts with taxis, whom are known to not rarely misguide the users into longer routes so that they can milk a loftier fee for the service."*
Autonomy of the user		User's feeling of being in control	Shneiderman	*"--maps technology to define a route automatically, without in-*

			(2020b)	*put from the driver, unless requested by the user. "*
	Information quality	The accuracy, correctness, and reliability of the information	Yan & alt (2011)	"*Uber will take me to the destination in the cheapest price and in the fastest time available*"
	System quality	The degree of how well the system can perform	Yan & alt (2011)	*"My trust is based on their use of maps technology to define a route automatically, without input from the driver –"*
	Content	The credibility of information, accurate, up to date information	Seckler & alt (2014)	"*--take me to the destination in the cheapest price and in the fastest time available."*
	Plausible promises	Promises that the system kept on the perception of user	Seckler & alt (2014)	"*I trust that Uber will take me to the destination –"* *"--I fell asleep in the passenger seat of the Uber and only woke up again when we were at my house."*

4.3.4 Siri

Apple's IPA, Siri, was rated 49 % of honesty, where the median was 2 out of 5. This refers that the responders disagree that Siri is honest in its functions. Also, benevolence was rated as 44% where interestingly participants strongly disagreed with the sentence "I believe that only the necessary information for Siri to function is being collected". The competence of Siri was rated the lowest out of all the services or applications (60%).

Table 8. Siri: perceived honesty, benevolence, and competence.

Siri (n=3)	**honesty**	**benevolence**	**competence**
	49 %	44 %	60 %

In the Critical Incidence part, the interviewee expressed distrust towards the service that also supports the fact that Siri was ranked as one of the least trustful applications in this study.

In Siris's case, the low competence rating was explained by for example in previous experiences with using the application. The interviewee mentioned that they felt that using Siri in their own native language did not work accurately and fluently hence they stopped using it. This validates that Seckler and others (2014) defined prior experience, and usability, and Yan and others (2011) defined perceived identity as UX elements causing trust and preventing distrust also in the case of AI-based systems.

Furthermore, the critical incidence reveals the explosive development of recommendation systems from Yan and other's study from 2011. While recommendations and personalization are still important User Experience elements, the AI development has made personalization almost too smart in many cases:

> *"I stopped using it, but with my new phone I haven't remember to switch it off and it seems to be very accurate sometimes, it scares me a bit when I do a safari [browser] search on my phone Siri already recommends something that I am thinking, or I could be interested in."*

Even though recommendations are still a strong factor in building trust, the definition must be rephrased, and human-centric factors have to be reconsidered when defining it. This also refers to Shneiderman's (2020b) finding of the importance of the feeling of the position of control when interacting with AI-based systems and Lew and Schumaher's (2020) scale of weirdness (see chapter 2.4.1). In addition to that participant reported of the feeling of that Siri listens to them all the time, what also refers to the Shneiderman's (2020b) autonomy concept.

The interviewee also mentioned that they do not perceive the privacy to be at a high level: *"I don't feel like my personal data is safe with it since I don't understand how and what parts of the data is used."* This validates the perceived privacy or belief of the user that the system is free of concerning private information disclosure (Yan & alt, 2011) is concerning also AI-based systems.

Another factor that was found to cause distrust was the feeling that the user does not understand how the system works. Hence explainability and explainable UI can be seen as a factor that prevents distrust and possibly enhances trust.

Lastly, Yan and other's (2011) personal motivation plays a role in the distrustful user experience. The interviewee quoted as "-- *I don't feel like I really need the service." which refers that the* user does not perceive any added value by interacting with the AI-based system.

Table 9. Siri: UX elements that enhance trust or prevent distrust.

Siri				
New element	Element	Description	Origin	Interview quotes
	Prior experience		Seckler & alt (2014)	"*It always said that "I can't understand." I stopped using it –"*
	Perceived identity	The belief of the user the level that the system recognizes entities in communications	Yan & alt (2011)	"*--It did not work at all in my native language [Finnish]. It always said that "I can't understand."*
Autonomy of the user		User's feeling of being in control	Shneiderman (2020b)	"*It scares me a bit when I do a safari [browser] search on my phone Siri already recommends something that I am thinking, or I could be interested in.* "

Recommendations with scaling of the weirdness		Suggestions for user behavior by letting the user report whether the results are propriate or appropriate.	Yan & alt (2011); Lew and Schumaher (2020)	*"I haven't remember to switch it off and it seems to be very accurate sometimes, it scares me a bit when I do a safari [browser] search on my phone Siri already recommends something that I am thinking, or I could be interested in."*
	Usability	Effectivity and efficacy of the UI elements such as navigation, links, and task flow	Seckler & alt (2014)	*"I have used it in the past when it first came with iPhone and it did not work at all in my native language."*
	Perceived privacy	The belief of user that the system is free of concerning private information disclosure		*"--it feels like it listens to me all the time. I don't feel like my personal data is safe with it--"*
Explainable data usage		Explaining how and what data the systems use		"I *don't feel like my personal data is safe with it since I don't understand how and what parts of the data is used."*
	Personal motivation	The perception of the user that interaction with the system delivers added value, such as increased performance, other benefits, or profits	Yan & alt (2011)	"--I don't feel like I really need the service. "

4.3.5 Alexa

Amazon's IPA, Alexa, was perceived as the lowest honesty and benevolence levels of all services and applications compared in this study. Honesty level was only 20% and ratings for all of the questions measuring honesty were rated only as 1 (strongly disagree). The same goes with benevolence level. However, "I think that Alexa is concerned with the present and future interests of its users" were only rated as 3 (not strongly disagree nor agree). Surprisingly, Alexa's competence was rated as 100%, and the median was 5.

Table 10. Alexa: perceived honesty, benevolence, and competence.

Alexa (n=2)	**honesty**	**benevolence**	**competence**
	20%	33%	100%

The interviewee reported that the Alexa is extremely easy to use, and they found it adding benefits to their life, such as playing music just by voice command. This validates once again the effect of personal motivation (Yan & alt, 2011) and perceived ease of use and quality of the system.

While there is a clear mitch match between competence and other elements of the trust, in the critical incidence interview, it came clear that questions regarding personalization and recommendations might be causing that. The interviewee said that while Alexa works smoothly and gives helpful recommendations, it sometimes conquers invading their personal space.

> *"-- sometimes even so well that I feel that the system knows me too well. Sometimes its recommendations almost forecast our wants and needs, and in those situations, I have begun to feel distrust towards Alexa. I feel that I am constantly listened –"*

However, this situation refers to Lew's and Schumaher's (2020) identifications of the weirdness of AI interactions that is presented in chapter 2.4.1. While personalization enhances the user experience until a particular point and is a part of great competence and functionality, it can also increase distrust if the recommendations feel inappropriate, for example, they are based on very personal information. Recommendations or suggestions should be done from a human-centric perspective and systems should be designed to understand humans from their socio-cultural perspective. This can be done for example measuring triggering actions by allowing the user to report recommendations that do not feel appropriate.

Also, Shneiderman's (2020b) concept of autonomy and loss of control was experienced by the participant, since they felt listened to even when they switched the system off.

Table 11. Alexa: identified UX elements that enhance trust or prevent distrust.

Alexa				
New element	Element	Description	Origin	Interview quotes
	Personal Motivation	The perception of the user that interaction with the system delivers added value, such as increased performance, other benefits, or profits	Yan & alt (2011)	*"we love to listen to music via Alexa speaker"*
	Perceived Ease of Use	The perception of the level of easiness of use and freedom of effort	Yan & alt (2011)	*"-- change songs by saying "Hey Alexa, next song please" and so on, it is just so easy."*
	System quality	The degree of how well the system can perform	Yan & alt (2011)	*"--it has been working unbelievable well—"*
Recommendations with scal-		Suggestions for user behavior by letting	Yan & alt (2011);	*"Sometimes even so well that I feel that the system knows me*

ing of the weirdness		the user report whether the results are propriate or appropriate.	Lew and Schumaher (2020)	*too well. Sometimes its recommendations almost forecast our wants and needs, and in those situations, I have begun to feel distrust towards Alexa—"*
Autonomy of the user		User's feeling of being in control	Shneiderman (2020b)	*"I feel that I am constantly listened, even when I know that the system is switch off."*

4.3.6 Grammarly

Grammarly's honesty and benevolence were perceived at a good level while competence in almost at an excellent level that seems to be the pattern according to the results of this study.

Table 12. Grammarly: perceived honesty, benevolence, and competence.

Grammarly (n=3)	**honesty**	**benevolence**	**competence**
	64%	73%	89%

Again, the interview results support the data of the initial survey, even though another responder described the feeling of distrust towards the service.

The interviewee who felt trust towards the service described the experience with Grammarly as very smooth and effortless to use. However small mistakes were found in the usage but not so major that the user would not continue to use the service or make them distrust it. This validates the Yan's and others (2011) defined perceived Ease of Use and UI quality as UX elements that enhance trust.

The interviewee mentioned that element that would make them stop using the service would be if they noticed ads or other marketing based on their topics in the texts that they have corrected or analyzed by Grammarly. This is concerning the element of perceived privacy (Yan & alt, 2011) thus it an element that enchases trust to continue using the service and becoming vulnerable to it.

However, another interviewee's distrusts towards Grammarly were mainly based on four elements: personality, the fact that missing explainability of the functions of the system, the system's access to the personal data, and lack of control.

Firstly, according to Yan and others (2011) personality can be defined as a user's individual combination of emotional, attitudinal, and behavioral responses. This definition matches well with the interviewee's emotional and passionate attitude towards data security, that they have been gaining a lot of knowledge in the past year, hence the interviewee felt skeptical of Grammarly's policies, privacy, and data usage.

> *"Since this year I have been taking data security more seriously, after learning a lot about data misusage of companies, I feel very vulnerable, since I don't know how my data have been used by Grammarly."*

Secondly, the interviewee also mentioned that not knowing how their data is used, increased the distrust. This validates the importance of the explainability of the system. The third cause of distrust or feeling of less trust was caused simply by that the system has access to the user's personal data. This element of access differs from other privacy concerns since the system might believe to use the data but simply allowing them to have access to the personal data of the user, might cause the feeling of less trust or distrust. Also, a concern of data collection (Seckler & alt, 2014) was expressed in the interviews.

Fourthly, this feeling of less trust or even distrust was partly caused by a lack of control (Shneiderman, 2020b). The user felt that it was what felt like the easiness of use at first became impossible to control:

> *"Sometimes I also feel that the service is almost too easy to use, for example today, I noticed that I have accidentally accepted to use Grammarly in my online presentation tool [Canva]. I almost feel like that I am not anymore in control of*

this program, even though at first, I felt that it is great and very easily usable tool for example checking email grammar or important text before I sent it to client.”

This balance between the user’s control and ease of use must be balanced. This is part of Lew and Schumaher’s (2020) scaling of weirdness concept but furthermore, it is systems behavior of rechecking if the user wants to continue with accepted policies and weather, they feel still accurate and appropriate. This relates to Seckler and other’s (2014) concept of policy in the web context but when concerning a continuously changing and learning AI-based system, these policies should be rechecked and reagreed with the user. This rechecking should be done in an explainable human-centric way.

Even though the user felt distrust, they continued using the service that refers to Yan’s and others' (2011) perceived usefulness and relative advantage. The user believes that using the service brings some added value or the system is relatively better than other similar services. Since the user continued to use the service, we can argue that they felt instead of distrust, they felt less trust towards Grammarly since according to Benamati, Serva, and Fuller (2010) distrust refers to the unwillingness to become vulnerable to the service or application. In this case, the user believes that Grammarly is less trustworthy but does not believe that the system necessarily behaves in a harmful, neglectful, and incompetent way.

Table 13. Grammarly: identified UX elements that enhance trust or prevent distrust.

Grammarly				
New element	Element	Description	Origin	Interview quotes
	Perceived ease of Use	The perception of the level of easiness of use and freedom of effort	Yan & alt (2011)	*”it has been very smooth, and effortless to use. I have no major complains, just small things that are bugging me—”* *“I also feel that the service is*

				almost too easy to use—"
	UI quality	The feeling and ease of use	Yan & alt (2011)	*"it has been very smooth, and effortless to use."* *"I felt that it is great and very easily usable tool for example checking email grammar or important text before I sent it to client."*
	Perceived privacy	The belief of user that the system is free of concerning private information disclosure	Yan & alt (2011)	*"I may stop using Grammarly if I notice that ads are targeted according to the topic that Grammarly has corrected or analyzed—"*
	Personality	User's individual combination of emotional, attitudinal, and behavioral response	Yan & alt (2011)	*"Since this year I have been taking data security more seriously, after learning a lot about data misusage of companies, I feel very vulnerable, since I don't know how my data have been used by Grammarly."*
Explainable data usage		Explaining how and what data the systems use.		*"I feel very vulnerable, since I don't know how my data have been used by Grammarly."*
Access to user's personal data		Notifying when and what kind of data is used by the system.		*" Another thing that worries me, is that I have accepted Grammarly for my web browser and other tools that I use regularly, so they have access almost all of my data and all my work [in digital environment]."*
	Privacy: collection	User's belief that data is not collected by the sys-	Seckler & alt (2014)	*"Now I am worried that the company will collect data of my confidential business mes-*

		tem operator		*sages.”*
Continuously rechecked policy		Rechecked and reagreed policy, general terms, and conditions		*”I noticed that I have accidentally accepted to use Grammarly in my online presentation tool [Canva]. I almost feel like that I am not anymore in control of this program”*
	Perceived usefulness	How much user believes that using the system would enhance their benefits or profits	Yan & alt (2011)	
	Relative advantage	The perception that using the particular system is better than other similar ones	Yan & alt (2011)	*“I occasionally use it for correcting my grammar since I haven’t found a better alternative”*

4.4 List of UX elements that build trust in AI-systems

Finally, to answer the research question, the UX elements of AI-based systems that contribute to service or application to be perceived as trustful by users are listed in table 14.

These elements are based on previous research by Yan, Kantola, and Zhang (2011), Seckler, Heinz, Forde, Tuch, and Opwis (2014), and also finding by Lew and Schumaher (2020) and Shneiderman (2020b). While these elements are validated in the context of AI-based services through this study, the list also includes new elements that have been identified through Critical Incidence interviews.

The list summaries the elements and it has been categorized into three main categories by Yan and others (2011): Interaction intension, system trust, and communication trust.

Table 14. UX-elements that enhance trust or prevent distrust in AI-based applications and services.

Category	Element	Description	Origin	Validated by the user of
Interaction intension	**Friends' social proof and other social factors**	Internalization of user's reference group culture, interpersonal agreements with others, and recommendations by friends, family members, or colleagues	Seckler & alt (2014) ; Yan & alt (2011)	Uber
	Reputation, image, or brand	Public opinion, image, or brand on the system and its operator and its influence on the intention of HCI	Yan & alt (2011) ; Seckler & alt (2014)	Google Maps
	Personal motivation	The perception of the user that interaction with the system delivers added value, such as increased performance, other benefits, or profits	Yan & alt (2011)	Siri, Alexa
	Perceived usefulness	How much user believes that using the system would enhance their benefits or profits	Yan & alt (2011)	Grammarly
	Perceived Ease of Use	The perception of the level of easiness of use and freedom of effort	Yan & alt (2011)	Alexa, Grammarly
	Relative advantage	The perception that using the particular system is better than other similar ones or even being the only one available	Yan & alt, (2011)	Google Maps, Uber, Grammarly
	Recommendations	Suggestions for user be-	Yan & alt	Siri, Alexa

	with scaling of the weirdness	havior by letting the user report whether the results are propriate or appropriate.	(2011); Lew and Schumaher (2020)	
	Personality	User's individual combination of emotional, attitudinal, and behavioral response	Yan & alt (2011)	Grammarly
System Trust	**Information quality**	The accuracy, correctness, and reliability of the information	Yan & alt (2011)	Google Maps, Uber
	System quality	The degree of how well the system can perform	Yan & alt (2011)	Google Maps, Uber, Alexa
	UI quality	The feeling and ease of use	Yan & alt (2011)	Grammarly
	Prior experience		Seckler & alt (2014)	Siri
	Plausible promises	Promises that the system kept on the perception of user	Seckler & alt (2014)	Uber
	Content	The credibility of information, accurate, up to date information	Seckler & alt (2014)	Google Maps, Uber
	Usability	Effectivity, efficacy, and safety of users to perform functions within an AI system	Seckler & alt (2014) ; Lew and Schumaher (2020)	Siri
	Support in an unfamiliar situation	Support of user in situations that they perceive as new or difficult		Google Maps, Uber
	Customer service	Availability of customer service before, during, and after the experience	Seckler & alt, (2014)	Uber
	Autonomy of the	User's feeling of being in	Shneiderman	Uber, Siri,

	user	control	(2020b)	Alexa
Communication Trust	**Perceived privacy**	The belief of user that the system is free of concerning private information disclosure	Yan & alt (2011)	Siri, Grammarly
	Privacy: collection	User's belief that data is not collected by the system operator	Seckler & alt (2014)	Google Maps, Grammarly
	Customer service	Availability of customer service before, during, and after the experience	Seckler & alt, (2014)	Uber
	Real-world link	For example, real-life location	Seckler & alt (2014)	Uber
	Perceived identity	The belief of user the level that the system recognizes entities in communications	Yan & alt (2011)	Siri
	Explainable data usage	Explaining how and what data the systems use		Siri, Grammarly
	Privacy: Access	Notifying when and what kind of data is used by the system.		Grammarly
	Continuously rechecked policy	Rechecked and reagreed policy, general terms, and conditions		Grammarly

4.5 Comparison to web trust elements

4.5.1 Interaction intension

While the list of UX-elements that enhance trust or prevent distrust in AI-based applications and services lists almost all the elements listed in table 14 from the original interaction intension category (chapter 2.4.2, table 1), element of user's social proof that is based on user ratings and reviews were not validated. This element might still play a role in enhancing trust, especially since for example Uber, which had the highest perceived trust level in this study, has an intergraded review system of the drivers and the passengers (Uber, 2021).

Interestingly in the interaction intension category, the three most strongly trustful perceived services and applications, in the case of Uber, Grammarly, and Google Maps, users identified relative advantage user experience element that enhances trust. In these cases, the system has been found to be better than any similar one, or even only one in the market. This might be also one of the reasons why these services and applications are extremely popular despite concerns regarding privacy or other elements. Users simply accept the fact that they do not know anything about the AI system, or they were not given any explanation, and that they were not in control of the situation, to use this service that has a relative advantage over others.

Where else, the new element of recommendations with scaling of the weirdness, can be defined as suggestions for user behavior by letting the user report whether the results are propriate or appropriate. This element is inspired by Yan's and others (2011) identified element of recommendations but updated with the scale of weirdness concept by Lew and Schumaher (2020). This element was validated by users of both of the IPAs of this study. In both of the cases, the user felt that the system knows them too well and gives too accurate recommendations based on personal information. Especially in the case of IPAs, which were perceived to be the least trustful in this study, letting the user affect the suggestions would perhaps prevent the distrust enhance the trust.

4.5.2 System Trust

In the system trust category, Seckler's and other's (2014) expertise, pop-up or ads, and visual design were not mentioned in the critical incidence interviews. Expertise, however, might be a factor partly under Information quality identified by Yan and others (2011). Pop-ups or ads were not mentioned since the services selected does not include advertisement or other pop-ups. However, visual design, such as colors, layout complexity, and photos, was perhaps not mentioned by the users since the overall visual look of the selected services and applications are already developed to be excellent, hence they there are already commonly used and scaled. Visual elements might be a more significant element in the perception of trust in services and applications that are newly founded or otherwise less developed.

Furthermore, there are also new elements in the category of system trust. The first one of them is support in an unfamiliar situation where the user can get support in situations that they perceive as new or particularly difficult to them. This element was validated in the traveling and transportation applications.

The second new element is the autonomy of the user that refers to the user's feeling of being in control when interacting with the AI system. The autonomy of users is originally identified by Shneiderman (2020b) but it is validated in this study by the users of multiple different services: Uber, Siri, and Alexa.

4.5.3 Communication Trust

In the category of communication trust, up to four elements of the list in chapter 2.4.2 (table 1) were not validated by concluded study. Yan and others (2011) identified communication context, that is the information that characterizes the situation of communications, is dropped out of the list of UX elements that build trust in AI systems even though it might still be baseline elements to perceived trust.

In addition, an element identified by Seckler and others (2014) that is a privacy concern of users that the information is not being used for another reason that it is collected for, was dropped out of the list. Even though this element is partly handled in Perceived privacy-element, it was not directly mentioned during the interviews of this study. However, the secondary use of data is closely related to the elements of a collection of the data and explainable data usage hence it can be considered to be an important element in building trust in AI.

Also, demands to share a link, download a piece of software, or create an account to get access to a website or a service (Seckler & alt, 2014), was not found through analyzing the critical incidence interviews.

However, there were also new elements identified in the Communication trust category. Firstly, an element called Explainable data usage that refers to a system explaining how and what data the system uses. This was identified as an element that users of Siri and users of Grammarly would have needed to feel more trust towards the service. Explainable data usage is also a great way to developed systems towards explainable human-centered AI.

Secondly, privacy concern towards access to the user's personal data was identified. The system should always notify when and what kind of data is used by the system and on what it has access to, in order for the user to feel secure when giving access to their digital environment or data.

Lastly, the continuously rechecked policy was found to be an element that enhances trust and indirectly supports another element, the autonomy of the user.

4.6 Comparison to AI-UX key elements

According to Lew and Schumaher (2020) and Shneiderman (2020b) key elements of UX of AI are utility, usability, aesthetics, the scale of weirdness, and autonomy of the user. While the scale of weirdness, autonomy of the user, and usability has been found as an element that enhances trust or prevents distrust, utility and aesthetics have not been directly identified during the critical incidence interviews.

However, Lew and Schumaher (2020) define the utility as functionality that refers to the fit of the application to the purpose that it is designed for. The system should function in an expected frictionless and effortless manner. As seen, the utility can be divided into listed elements (table 14) of System quality, Perceived Ease of Use (Yan & alt, 2011), and Plausible promises (Seckler & alt, 2014).

Aesthetics of the services and applications were not referred by the interviewees but that does not necessarily mean that this does not affect the trust. In reality, the situation might be the opposite since the visual elements of any online service can be considered as a hygiene element, and service without a great visuality and aesthetics are not succeeding in the first place. Selected AI-based services are widely used, hence the results of measuring less popular or rather new services might have provided different results.

5. Discussion & conclusions

The goal of this interpretive research, to gather the opinions of users of the services or application to provide a list of elements that affect the perception of trust in multiple AI systems in order to develop trustworthy HCXAI was achieved. The validated list, based on previous literature of the field as well as to new findings of the study, is provided in chapter 4.4.

The List of UX elements that build trust in AI systems is meant to help to design a system that is explainable, trustworthy, and empathize human while still maintaining all the benefits of AI. For example, automatization is an important part of AI services, but they should still allow humans to be in control. This refers to the element of autonomy of the user. The list can work as a checklist when designing AI-based services and applications, but it also fulfills its purpose in gathering field’s research together by validating them in the specific area of commonly used services and applications.

5.1 Reflection of the results with existing literature

Table 15. Reflection of the results with existing literature.

Result	**Reflection with literature**	**Origin**
Overall level of trust was quite low when measuring it though components of honesty, benevolence, and competence. Especially it can note that the honesty and benevolence levels are quite low comparing how popular are these services or applications.	Result supports findings of previous literature. For example, according to Davenport’s study in 2019, 40 percent of users did not have any trust in AI-based services in the United States.	Davenport (2019)
Category of Intelligence Personal Assistants are perceived the least trustful.	While according to Vimalkumar & others (2021) the trust towards IPAs is very low, Lew and	Lew & Schumaher (2020), and Vimalkumar, Sharma, Singh & Dwivedi (2021)

	Schumaher (2020) present that release of Amazon's Alexa would have repealed the distrust towards IPAs. However, they do not particularly argue that the trust would have been build towards this type of services or applications.	
Four other new elements were identified: autonomy of the user, explainable data usage, privacy concern of access to the user's data, and continuously rechecked policy.	Autonomy of the user is based on based on theory of Shneiderman (2020b). Explainable data usage is based on theory of explainability that for example Doran, Schulz, and Besold (2018), Rield (2019), and Arrieta and others (2020) advocate for. Privacy concern of access to users' data is a completely new finding of this research while continuously rechecked policy is modified from Seckler's and others (2014) defined policy.	Shneiderman (2020b), Doran, Schulz & Besold (2018), Rield (2019), Arrieta, Díaz-Rodríguezb, Del Sera, Bennetot, Tabikg, Barbadoh, Garciag, Gil-Lopeza, Molinag, Benjaminsh, Chatilaf & Herrerag (2020), and Seckler, Heinz, Forde, Tuch & Opwis (2014)
UX-element, visual design, were left out of the final list.	While visual design and aesthetics are considered as one of the most impactful elements of creating trust, in this study participants did not mention any elements considering visuality or aesthetics when describing their experiences considering trust.	Lew & Schumaher (2020), and Pengnate & Sarathy (2017)
Users accepted their status as under the control of the system regardless of issues with privacy etc., because usage of the ser-	The result support findings of Vimalkumar and others (2021) that when services and applications offer great utility and	Vimalkumar, Sharma, Singh & Dwivedi (2021)

vice or application brings relative advantage.	need, users adopt technology regardless of appeared trust issues.	
Transition to HAI requires users to demand trustworthy and transparent practices that are explained for them via UI elements.	Doran, Schulz, and Besold (2018), and Rield (2019) have noted that simple acceptation of untransparent functions and trust misusage enhances the creation of untransparent and unfair AI.	Doran, Schulz, and Besold (2018), and Rield (2019).

This study showed that overall level of trust was quite low when measuring it though components of honesty, benevolence, and competence. Especially it can note that the honesty and benevolence levels are quite low comparing how popular are these services or applications. In addition, from all the services, only Uber and Grammarly reach over 50% (from 100%) level in all three components. Result supports findings of previous literature. For example, according to Davenport's study in 2019, 40 percent of users did not have any trust in AI-based services in the United States.

Comparing the least trustful and most strongly trusted services and applications via the survey showed that Intelligence personal assistants are perceived the least trustful. Users described these services many times to be too personal and invading with their recommendations and suggestions, hence the element of recommendations was revised to be recommendations with scaling of the weirdness. This brings control and the possibility to report appropriate recommendations for the user.

When reflecting results to the previous literature, the results low trust level is not surpirising. According to Vimalkumar & others (2021) the trust towards IPAs is very low, while Lew and Schumaher (2020) present that release of Amazon's Alexa would have repealed the distrust towards IPAs. However, they do not particularly argue that the trust would have been build towards this type of services or applications.

In addition, four other new elements were identified: autonomy of the user, explainable data usage, privacy concern of access to the user’s data, and continuously rechecked policy. Most of the newly identified elements are loosely based on previous literature. Autonomy of the user is based on based on theory of Shneiderman (2020b) of importance that user feels that they are in position of control. Explainable data usage is not previously presented as UX-element that enhance trust, but it is based on theory of explainability that for example Doran and others (2018) Rield (2019), and Arrieta and others (2020) advocate for. Privacy concern of access to users’ data is a completely new finding of this research while continuously rechecked policy is modified from Seckler’s and others (2014) defined UX-element of policy.

While most of the previous literature findings were validated through this study some elements such as aesthetics and visual design (Lew & alt, 2020, Pengnate & alt, 2017) were left out of the final list. These aspects remain important elements of UX of AI, but they might be considered as the basis of today’s applications and services, especially in the case of commonly used services. This might be the reason why the participants of this study did not mention any elements considering visuality or aesthetics when describing their experiences considering trust.

Lastly, participants of the survey used these services, many of them were concerned for example privacy, data usage of the system, and functionality of the system, and they were left out without any explanation by the system. Hence interestingly these results show that users accepted their status as under the control of the system because it has a relative advantage compared to other services or the system brings some benefits or value to the user’s daily life. This support the findings of Vimalkumar and others (2021) that users who believe that the system offers great utility, are more likely to adopt the technology regardless of other trust issues such as issues related to privacy or usage of personal data.

It is equally important to increase the level of critical thinking, action, and education of fair and transparent services and applications, in addition to creating trustworthy AI systems that understand stand humans but also explain themselves. Simple acceptation of untransparent functions and data misuse enhances the creation of untransparent and unfair AI that is also earlier noticed by Doran, Schulz, and Besold (2018) and Rield (2019).

All in all, the major findings in this research confirmed the list of crucial UX elements to consider when building trust between users and AI-based systems but at the same time simply designing trustworthy services is not enough. Transition to HAI also need users to demand trustworthy and transparent practices that are explained for them via UI elements.

5.2 Recommendations

5.2.1 Recommendations to researchers

Table 16. Recommendations to researchers.

Recommendations to researchers	**Origin**
Using the list as a summary of the research of the field.	
Cognitive phycological approach: e.g., measuring cognitive gestures and expressions	Lazar, Feng & Hochheiser (2017)
Prototyping by by manipulating the UX features.	Pengnate & Sarathy (2017)
Repeating this research with life-critical or consequential applications and services.	Shneiderman (2020a)
Researching ways to educate and make users aware of human-centered AI.	

For future research, the result of this study: the list of UX-elements that enhance trust or prevent distrust in AI-based applications and services (table 14) can be used as a summarizing report of previous literature finding that are validated in context of AI and with real users.

In addition, including the cognitive phycological approach is suggested, for example by measuring cognitive gestures and expressions since they might reveal deeper emotions related to feelings of trust. Methods such as eye-tracking tools, galvanic skin response, heart rate or brain activity measurements can be used to receive physical and emotional responses (Lazar, Feng & **Hochheiser, 2017).**

To find more detailed information on the components of UX that made the experience feeling trustful, creating, and testing prototypes of the same service but with different UX features could give more exact data on the features that affect the experience of trust. For example, a study by Pengnate and Sarathy (2017) has given great results in investigating trust in website context by manipulating the UX features. This study could be repeated in the context of AI services and applications.

I suggest repeating this research with life-critical or consequential applications and services as they are defined by Shneiderman (2020a) to see whether UX elements affecting the perceived trust will differ from commonly used AI services and applications.

Lastly, it is important to research ways to educate and make users aware of human-centered AI and how to spot unfair or untransparent AI-systems in order to them to demand better practices.

5.2.2 Recommendations to practitioners

Table 17. Recommendations to practitioners.

Recommendations to practitioners	**Origin**
Using list (table 14) as a checklist when designing AI-based services and applications.	
Encouragement for users to demand trustworthy and transparent practices that are explained for them via UI elements.	Doran, Schulz, and Besold (2018) and Rield (2019).

To practitioners, the result of this research, the list of UX-elements that enhance trust or prevent distrust in AI-based applications and services, can be used to help to design AI-systems in a Human-centered way, since the UX-elements are validated by the actual users. Designers can check that as many as possible of these elements are considered in the design of AI-based system or they can spot flaws of already existing systems that might cause distrust or use the list to enhance the trust of the users within the system.

In addition, it is equally important to encourage the users of any AI-based system to demand trustworthy and transparent practices that are explained for them via UI elements, since accepting untransparent functions and data misuse further enhances the creation of untransparent and unfair AI (Doran, Schulz & Besold, 2018; and Rield, 2019). To transit to truly human-centered AI, designers need an input from the users when they feel that the system is not trustworthy or transparent. I suggest that designers of these systems, especially UX-professionals, and users work increasingly together to spot these flaws that increase distrust. In addition to that, designers must also provide opportunities for users to spot and report obscure or otherwise untrustworthy situations when using the system.

5.3 Limitations and evaluation of the research

The validity of the results from external point of view includes the problem of small sample groups that is present in qualitative research. In this research the survey and interviews are based on the perspective of people in the age group 20 to 30 and repeating the study with different age groups might give different results. Also, participants of this study also were evaluated to have a high trusting stand and repeating the study with a sample group of a lower level of trusting stand, might give different results. Also comparing studies done with different trusting stand-level sample groups might lead to interesting results. The findings of this research can be generalized only within the age group of the study and the results might be limited since the participants happen to have rather high trust level. In addition, the results can be generalized only when considering AI-systems that can be used in similar everyday services as in the study.

From the internal point of view, the observation of this research matches with the theoretical ideas that is the benefit of using qualitative methods. In this research, the experts of the topic are the user and direct quotations of them are used to draw the results.

From the reliability point of view, the study can be repeated using by using the same set of questions of the survey and the interview (appendices 1&2) and by observing the same AI-based applications and services hence the chosen instruments are reliable. However, using a CIT method the observations might vary a lot when repeating the study since they are very personal and based on the experiences of individual people. On the other had the experience with commonly use AI-based services and application might change greatly because the development of these services is fast, and the emphasis of different UX-elements might change when the study is repeated later hence the results are not completely comparable.

On the internal reliability point of view the study is however limited since there was only one observer in the study and the results are based on the interpretation of on researcher. A research team would have a benefit of comparing their analysis of the results and perhaps even more UX-elements would have been noticed by several people. Also, member of the team could compare their opinions and observations and produce results that have been more critically evaluated and hence more edited. On the other hand, the interviews of this study were recorded and transcribed hence the answers of the participants can be viewed also in the future (see appendix 2).

The main problematic issues, in this study, is the principle of multiple interpretations, which refers to the sensitivity of the differences of interpretations among the participants (Klein and Myers, 1999). There were only few interviewees per one applications and service. For example, in the case of Siri there were only one participant of the interviews hence the results are based on their subjective view. Overall, there were multiple different services and applications selected in the study and in combination of subjective experiences of user, forms a whole picture of the trust in AI systems. Furthermore, the results are considered as a contribution for the development of Human-centered AI. This strengthen usage of the fundamental principle of the hermeneutic circle, that refers to the understanding of interdependent parts and that they are forming the complex whole (Klein & Myers, 1999). Also, the principle of abstraction and generalization, that means relating the details of interpretative data to general concepts that describe human understanding and social action (Klein & Myers, 1999), is important principle in this study. The theories of previous literature of UX-elements and trust were modified and validated on basis of the user study. This resulted list of UX-elements that affect trust in AI-systems which can be fruitful for researcher and practitioners in the similar area.

All in all, the research instruments chosen to measure the personal experience of the trust of the users of chosen services and applications, but the results cannot be generalized to all people, but rather to young adults. In addition, the results of this study can be repeated in some extend since the survey and interview questions are available in the appendices (1 & 2). However, because of the interpretive character of this research, the context, such as people’s views and the technology itself, is not static but changes rapidly. In addition, this research matches the principles of Klein and Myers (1999) for conducting and evaluating interpretive field study.

References

Amazon (2020). What is Alexa? Retrieved 2021-1-5 from https://developer.amazon.-com/en-US/alexa

Apple (2017). Deep Learning for Siri's Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis. Retrieved 2021-1-5 from https://machine-learning.apple.com/research/siri-voices

Arrieta, A. B., Díaz-Rodríguezb, N., Del Sera, J., Bennetot, A., Tabikg, S., Barbadoh, A., Garciag, S., Gil-Lopeza, S., Molinag, S., Benjaminsh, R., Chatilaf, R. & Herrerag, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58: 82-115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012

Benamati, J., Serva, M. A. & Fuller, M. A. (2010). The Productive Tension of Trust and Distrust: The Coexistence and Relative Role of Trust and Distrust in Online Banking. Journal of Organizational Computing and Electronic Commerce 20(4). DOI: https://doi-org/10.1080/10919392.2010.516632

Bennetot, A.; Laurent, J-L.; Chatila, R.; Díaz-Rodríguez, N. (2019). Towards Explainable Neural-Symbolic Visual Reasoning. 14th International Workshop on Neural-Symbolic Learning and reasoning.

Campesato, O. (2020). Artificial Intelligence, Machine learning and Deep Learning. Mercuri Learning and Information. ISBN: 9781683924678

Cho, J-H. (2015). Tradeoffs between trust and survivability for mission effectiveness in tactical networks. IEEE Transactions on Cybernetics 45(4). DOI:10.1109/TCYB.2014.2335744

Cho, J-H.; Xu, S.; Hurley, P. M.; Mackay, M.; Benjamin, T. & Beaumont, M. (2019). STRAM: Measuring the Trustworthiness of Computer-Based Systems. ACM Computing Surveys 51(6). DOI:10.1145/3277666

Cho, J. H.; Swami, A. & Chen, I. R. (2011). A survey of trust management in mobile ad hoc networks. IEEE Communications Surveys and Tutorials 1(4): 562–583. DOI: 10.1109/SURV.2011.092110.00088

Confalonieri, R., Coba, L., Wagner, B. & Besold, T.R. (2020). A historical perspective of explainable Artificial Intelligence. WIREs Data Mining and Knowledge Discovery 11(1). URL: https://doi-org.proxy.uwasa.fi/10.1002/widm.1391

Davenport, T. H. (2019). Can we solve AI's 'trust problem'? *MIT Sloan Management Review*, *60*(2). Retrieved 2020-10-12 from https://www-proquest-com.proxy.uwasa.fi/scholarly-journals/can-we-solve-ais-trust-problem/docview/2161593896/se-2?accountid=14797

De Barcelos Silva, A., Comes, M. M., Da Costa, C. A., Da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., De Doncker, G. & Federizzi, G. (2020). Intelligent personal assistants: A systematic literature review. Expert Systems with Applications 147. DOI: https://doi.org/10.1016/j.eswa.2020.113193

Ding, L. (2018). Human Knowledge in Constructing AI Systems – Neural Logic Networks Approach towards an Explainable AI Networks Approach towards an Explainable AI. Procedia Computer Science 126 (2018) 1561–1570. DOI: https://doi.org/10.1016/j.procs.2018.08.129

Doran, D., Schulz, S. & Besold, T.R. (2018). What does explainable ai really mean? A conceptualization of perspectives. CEUR Workshop Proceedings 2071. International Workshop on Comprehensibility and Explanation in AI and ML, CEX 2017; Bari, Italy.

Ehsan, U. & Rield M. O. (2020). Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach. HCII 2020. Lecture Notes in Computer Science 12424. DOI: https://doi.org/10.1007/978-3-030-60117-1_33

Elliot, A. (2019). *The Culture of AI: Everyday Life and the Digital Revolution* (1st Edition). Routledge. https://doi-org.proxy.uwasa.fi/10.4324/9781315387185

Erdal M., Kächele M. & Schwenker F. (2016) Emotion Recognition in Speech with Deep Learning Architectures. ANNPR 2016. Lecture Notes in Computer Science 9896. Springer, Cham. DOI: https://doi-org.proxy.uwasa.fi/10.1007/978-3-319-46182-3_25

Finnish Center for Artificial Intelligence, FCAI (2020). Elements of AI. Retrieved 2020-10-13 https://www.elementsofai.com/fi?_ga=2.142613110.2132112837.1595414941-704589728.1595414941

Google (2020). Maps Metrics and Infographics. Retrieved 2021-1-6: https://sites.-google.com/a/pressatgoogle.com/google-maps-for-iphone/google-maps-metrics

Grammarly (2019). How We Use AI to Enhance Your Writing. Retrieved 2021-1-6 from https://www.grammarly.com/blog/how-grammarly-uses-ai/

Grammarly (2021). About Us. Retrieved 2021-1-6 from https://www.grammarly.com/about

Hartson, R. & Pyla, P. S. (2018). The UX Book: *Agile UX Design for a Quality User Experience.* Elsevier Science. ISBN: 9780128010624

Holzinger, A., Langs, G., Denk, H., Zatloukal, K. & Müller, H. (2019). WIREs Data Mining and Knowledge Discovery 9(4). DOI: https://doi.org/10.1002/widm.1312

Kim, Y-B. (2018), The Scalable Neural Architecture behind Alexa’s Ability to Select Skills. Amazon Science. Retrieved 2021-1-6 from https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills

Klein, H. K. & Myers, M. D. (1999). A set of Principles for conducting and evaluating Interpretive Field Studies in Information Systems. MIS Quartely 23(1) 67-93. DOI: https://doi-org.proxy.uwasa.fi/10.2307/249410

Knight, W. (2019). UX for developers: how to integrate user-centered design principles into your day-to-day development work. Apress. ISBN: 9781484242278

Lau, J (2020). Google Maps 101: How AI helps predict traffic and determine routes. Retrieved 2021-1-6 from https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/

Lazar, J., Feng, J. H. & Hochheiser, H. (2017). Research Methods in Human-Computer Interaction (2nd digital edition). Morgan Kaufmann. ISBN: 9780128093436

Lew, G. & Schumacher, R. M. (2020). AI and UX. *Why Artificial Intelligence Needs User Experience*. Apress. ISBN: 978-1-4842-5775-3. DOI: 10.1007/978-1-4842-5775-3

Maggio, E. (2018). Apple says that 500 million customers use Siri. Business Insider. Retrieved 2021-2-4 https://www.businessinsider.com/apple-says-siri-has-500-million-users-2018-1?r=US&IR=T

Morville, P. (2004, June 21). User Experience Design. Semantic Studios. http://semanticstudios.com/user_experience_design/

Münscher, R., & Kühlmann, T. M. (2015). Using critical incident technique in trust research. In G. Lyon, G. Möllring, M. Saunders, & J. Hatzakis, Handbook of research methods on trust (2nd edition): 161-172 Cheltenham, UK: Edward Elgar. ISBN: 978-178254741-9;978-178254740-2

Pengnate, S. & Sarathy, R. (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. https://doi.org/10.1016/j.chb.2016.10.018

Pfleeger, C. P., Pfleeger, S. L. & Margulies (2015) Security in Computing (5th Edition). Prentice Hall. ISBN 978-0-13-408504-3

Raphael, B. (1976). The thinking computer: mind inside matter. Freeman. ISBN: 9780716707226.

Ribeira, M. & Lapedriza, A. (2019). Can we do better explanations? A proposal of User-Centered Explainable AI. CEUR Workshop Preceedings 2327. Explainable Smart Systems 2019 conference, Los Angeles.

Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier, Publication: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 1135-1144. DOI: https://doi-org /10.1145/2939672.2939778

Rield, M. O. (2019). Human-centered artificial intelligence and machine learning. Human behavior and Emerging Technologies 1(1).

Rouhiainen, L. (2019). Artificial Intelligence: 101 Things You Must Know Today About Our Future: 1-5, 30-31. ISBN: 9781982048808.

Seckler, M., Heinz, S., Forde, S., Tuch, A. N. & Opwis, K. (2014). Trust and distrust on the web: User experiences and website characteristics. Computers in Human Behavior 45: 39–50. DOI: http://dx.doi.org/10.1016/j.chb.2014.11.064

Shalev-Shwartz, S. & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107298019

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies 146. DOI: https://doi.org/10.1016/j.ijhcs.2020.102551

Shneiderman, B. (2020a). Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI. AMC Transactions on Interactive Intelligent Systems (26). DOI: https://doi.org/10.1145/3419764

Shneiderman, B. (2020b). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International journal of human-computer interaction 36(6): 495-504. DOI: https://doi-org/10.1080/10447318.2020.1741118

Statista (2019). Total smart home devices supported by various voice assistants in 2018 and 2019. Retrieved 2021-1-18 from https://www.statista.com/statistics/933551/worldwide-voice-assistant-supported-smart-home-devices/

Statista (2020). Monthly number of Uber's active users worldwide from 2017 to 2020, quarter (in million). Retrieved 2021-2-2 from https://www.statista.com/statistics/833743/us-users-ride-sharing-services/

Uber (2021). About us. Retrieved 2021-1-6 from https://www.uber.com/fi/en/about/

Unger, R. & Chandler, C. (2012). A Project Guide to UX Design. For User Experience Designers in the Field or in the Making (2nd edition). New Riders. ISBN 13: 978-0-321-81538-5

Vimalkumar, M., Sharma, S. K., Singh, J. B. & Dwivedi, Y. K. (2021). 'Okay google, what about my privacy?': *User's privacy perceptions and acceptance of voice based digital assistants*. Computers in Human Behavior 120. DOI: https://doi.org/10.1016/j.chb.2021.106763

Yan, Z., Kantola, R. & Zhang, P. (2011). A research Model for Human-Computer Trust Interaction. International Joint Conference of IEEE TrustCom 1: 274-281.

Appendices

Appendix 1. Questions of the survey

part 1: Personal Background and online experience

a. I usually trust people until they give me a reason not to trust them (strongly disagree (1) to strongly agree (5))
b. Do you use any of these services regularly? (yes/no)

Google Maps, Uber, Siri, Alexa, Grammarly

part 2: for all of the selected services (in part 1):

Honesty (strongly disagree (1) to strongly agree (5))

a. I think that the information offered by this service is not biased
b. I believe that the service will notify me if important changes that affect my decisions are made
c. I believe that the service's information sources are transparent

Benevolence (strongly disagree (1) to strongly agree (5))

a. I believe that only the necessary information for the service to function is being collected
b. I think that this service is concerned with the present and future interests of its users
c. I think that this service takes into account the repercussions that their actions could have on the user

Competence (strongly disagree (1) to strongly agree (5))

a. I think that service works as intended
b. I think that this service provides all the information I need
c. I think that this service knows its users well enough to provide recommendations adapted to their needs

Appendix 2. Critical incidence interview

a. Question: Please identify a situation when you felt trust or distrust in Google Maps, and if possible, explain the reasons that led you to trust it or not to trust it. Describe the situation and environment.

Answer 1:

“I truly trust Google Maps, when I am going for place of which location, I am not sure. The map service works well, in my point of view, and it gives very clear guidance to find the place I want to go to. I also use Google Maps to save good restaurants and cafe in the map.

However, sometimes Google Maps is not reliable, for example outside the city areas, and it does not know all the roads or best routes there. “

Answer 2:

“I distrust Google as a corporation in general as they collect so much information, but regarding the functionality of the Google Maps app, I usually do not rely solely on the app alone for information about routes, restaurants, etc.

An example of reasons for distrust: Google Maps clearly functions better in certain countries. When I was an exchange student in Seoul, I would use Korean apps to get more accurate info. I would use Google Maps alongside these apps, but I trusted the Korean map apps over Google Maps.”

Answer 3:

“A situation when I have not trusted the Google Maps service have been situations where the route suggestion has not been logical, if you know better that there is better route to take (shorter, better for selected vehicle etc) and vice versa, if the route was logical for you, then I have felt trust towards the service. Trust to Google Maps deepens in situations when I am in unfamiliar place, and I don’t know the routes or roads there. In these situations, I feel that I blindly trust

the service since I have no other choice. However, if I would not trust the service at all I would have a paper map with me. "

b. Question: Please identify a situation when you felt trust or distrust in Uber, and if possible, explain the reasons that led you to trust it or not to trust it. Describe the situation and environment.

Answer 1:

" I trust Uber whenever I need to get to a destination by car, and I do not have a car of my own. I trust that Uber will take me to the destination in the cheapest price and in the fastest time available. My trust is based on their use of maps technology to define a route automatically, without input from the driver, unless requested by the user. This contrasts with taxis, whom are known to not rarely misguide the users into longer routes so that they can milk a loftier fee for the service."

Answer 2:

" When I had my semester abroad in Chile, I lived in the city of Valparaíso. Even though Chile is generally considered quite safe, Valparaíso is often the scene of theft and, in major accidents, threats with weapons. Us exchange students mostly used Uber, especially when it came to getting home from a party in the middle of the night. I still remember a situation where I took an Uber alone and I fell asleep in the passenger seat of the Uber and only woke up again when we were at my house. I was a bit perplexed about that and it was a bit uncomfortable for me to have fallen asleep next to a stranger. However, I had no worries that I was not safe with him - or her (I don't even remember if it was a man or a woman). My trust in Uber was generally really high as Uber drivers were usually very chatty, open and helpful. They gave you tips and always made you feel safe."

c. Question: Please identify a situation when you felt trust or distrust in Grammarly, and if possible, explain the reasons that led you to trust it or not to trust it. Describe the situation and environment.

Answer 1:

“I started using it to quickly check my grammar in the online environment and it has been very smooth, and effortless to use. I have no major complains, just small things that are bugging me such as Grammarly wants to add “a” before certain words where it absolutely should not be one and wrong prepositions at times. Hence, I trust to continue to use it. If you look from the other side, I may stop using Grammarly if I notice that ads are targeted according to the topic that Grammarly has corrected or analyzed, especially since I am using it as add-on In Chrome.”

Answer 2:

“I actually feel distrust towards Grammarly services, even though I occasionally use it for correcting my grammar since I haven’t found a better alternative. Since this year I have been taking data security more seriously, after learning a lot about data misusage of companies, I feel very vulnerable, since I don’t know how my data have been used by Grammarly. Another thing that worries me, is that I have accepted Grammarly for my web browser and other tools that I use regularly, so they have access almost all of my data and all my work [in digital environment]. Sometimes I also feel that the service is almost too easy to use, for example today, I noticed that I have accidentally accepted to use Grammarly in my online presentation tool [Canva]. I almost feel like that I am not anymore in control of this program, even though at first, I felt that it is great and very easily usable tool for example checking email grammar or important text before I sent it to client. Now I am worried that the company will collect data of my confidential business messages.”

d. Question: Please identify a situation when you felt trust or distrust in Alexa, and if possible, explain the reasons that led you to trust it or not to trust it. Describe the situation and environment.

   Answer 1:

   "I bought Alexa speaker with my boyfriend last year, it has been working unbelievable well, sometimes even so well that I feel that the system knows me too well. Sometimes its recommendations almost forecast our wants and needs, and in those situations, I have begun to feel distrust towards Alexa. I feel that I am constantly listened, even when I know that the system is switch off. However, we still continue using Alexa, since we love to listen to music via Alexa speaker and change songs by saying "Hey Alexa, next song please" and so on, it is just so easy."

e. Question: Please identify a situation when you felt trust or distrust in Siri, and if possible, explain the reasons that led you to trust it or not to trust it. Describe the situation and environment.

   Answer 1:

   "I don't like Siri at all. I have used it in the past when it first came with iPhone and it did not work at all in my native language [Finnish]. It always said that "I can't understand." I stopped using it, but with my new phone I haven't remember to switch it off and it seems to be very accurate sometimes, it scares me a bit when I do a safari [browser] search on my phone Siri already recommends something that I am thinking, or I could be interested in.

   [I feel] distrust towards Siri since it I don't understand how it works and it feels like it listens to me all the time. I don't feel like my personal data is safe with it since I don't understand how and what parts of the data is used. In addition, I don't feel like I really need the service. "