Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Development of a Concept for Privacy Explanations and its Prototypical Evaluation

Masterarbeit

im Studiengang Informatik

von

Jakob Richard Christian Droste

Prüfer: Prof. Dr. rer. nat. Kurt Schneider
Zweitprüferin: Dr. rer. nat. Jil Ann-Christin Klünder
Betreuer: M. Sc. Wasja Brunotte

Hannover, 19.04.2022

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 19.04.2022

Jakob Richard Christian Droste

Abstract

Online privacy as well as privacy in software systems is a sensible topic that concerns everyone who regularly connects to the Internet. Technologies such as smartphones and Internet of Things (IoT)-devices are entering both private and professional spaces. As the daily contact with software becomes continuously more inevitable, users also share increasing amounts of their personal data with others. In an effort to educate and protect end-users, the European Union (EU) introduced the General Data Protection Regulation (GDPR), which secures a "Right to Explanation" for all EU citizens. So far, this right has manifested itself in documents such as privacy policies, cookie banners and app privacy notices.

Contemporary research on these documents has shown that they are not fit for the explanation of privacy aspects to end-users. This is caused by their structure and contents, or lack thereof. If users do not understand how their personal data is being processed, they are not able to give genuine informed consent to those data processing practices. Consequently, there is a need for dedicated explanations of end-user privacy, which are easy to understand and satisfying for those users. The topic of explainability is already deeply rooted in computer science, specifically in artificial intelligence, and lends itself to being adopted for privacy in software systems.

This thesis provides a view on privacy explanations through the lens of explainability. To this end, relevant literature from the fields of privacy and explainability is identified and reviewed. Different types of established forms of explanation are used and adapted to develop a novel concept for privacy explanations. This concept is realized within an interactive software prototype. By conducting an extensive exploratory user study, the prototype and its underlying concept are thoroughly examined and evaluated by 61 participants. Through this, we gain first insights on what privacy explanations should look like and what they need to contain. The results of the study show that the developed approach was well received by the participants. Furthermore, they indicate that privacy explanations can increase end-users' privacy awareness.

Zusammenfassung

Privatsphäre online, sowie im Zusammenhang mit Software-Systemen, ist ein sensibles Thema, welches all diejenigen betrifft, die sich regelmäßig mit dem Internet verbinden. Technologien, wie Smartphones und Internet of Things (IoT)-Geräte, dringen immer weiter in private Räume und Arbeitsplätze vor. Da der tägliche Kontakt mit Software sich immer weniger vermeiden lässt, teilen Nutzer auch mehr persönliche Daten mit anderen. Um diese Nutzer aufzuklären und zu schützen, hat die Europäische Union (EU) die Datenschutz-Grundverordnung (DSGVO) verabschiedet, welche allen EU-Bürgern ein "Recht auf Erklärung" einräumt. Zurzeit wird dieses Recht vor allem in Form von Dokumenten wie Datenschutz-Erklärungen, Cookie-Bannern und Privatsphäre-Benachrichtigungen umgesetzt.

Wissenschaftliche Arbeiten zu derartigen Dokumenten haben gezeigt, dass diese sich nicht eignen, um Nutzern Privatsphäre-Aspekte verständlich zu erklären. Das liegt zum einen an ihrer Struktur, aber auch an ihren Inhalten. Wenn Nutzer nicht verstehen, wie ihre persönlichen Daten verarbeitet werden, ist es ihnen auch nicht möglich, ein echtes informiertes Einverständnis zu den beschriebenen Datenverarbeitungs-Praktiken zu geben. Daraus resultiert, dass explizite Privatsphäreerklärungen benötigt werden, welche verständlich und zufriedenstellend für Nutzer sind. In der Informatik, speziell im Fachgebiet der künstlicher Intelligenz, hat die Erklärbarkeit als Thema bereits tiefe Wurzeln geschlagen. Sie lässt sich aber auch auf das Fachgebiet der Privatsphäre in Software anwenden.

Diese Masterarbeit betrachtet Privatsphäreerklärungen aus Sicht der Erklärbarkeit. Dazu wird zunächst fachliche Literatur aus den Forschungsfeldern der Erklärbarkeit und der Privatsphäre herangezogen. Dann werden etablierte Erklärungsformen identifiziert und angepasst, um ein neuartiges Konzept für Privatsphäreerklärungen zu entwickeln. Dieses Konzept wird außerdem software-prototypisch umgesetzt. Anschließend wird es im Rahmen einer explorativen Nutzer-Studie mit 61 Teilnehmern erforscht und bewertet. Dadurch ist es möglich, erste Einblicke darin zu erhalten, wie Privatsphäreerklärungen aussehen müssen und welche Inhalte sie benötigen. Die Ergebnisse zeigen, dass die Teilnehmer den entwickelten Ansatz gut aufgenommen haben. Zusätzlich deuten sie darauf hin, dass Privatsphäreerklärungen Nutzer bezüglich ihrer Privatsphäre sensibilisieren können.

Contents

Chapter 1

Introduction

1.1 Motivation

Privacy of personal data is an ever-present topic that concerns everyone who regularly connects to the Internet. In the age of smartphones and IoT-devices almost no one is an exception to this rule. Every day, people use mobile software applications in their private lives and at their workplace [33, 59]. As they interact with them in their personal and professional spaces, users share all kinds of personal data with the software, such as their location data or their hobbies and interests. This data is either actively shared by the users or passively collected by the software. The companies, that provide the software, receive and process the data, and often even share it with third parties [9, 46].

According to recent data protection laws, such as the European GDPR [23], every user has a right to be informed about their privacy. Among other things, this includes how their data is collected, processed and shared. Privacy statements that include this information are often extensive and complicated, and usually only available in the form of privacy policies or similar legal documents [66]. Users are confronted with these documents while browsing the Internet or installing software. They are supposed to give their informed consent to the described data processing practices [66]. For them to be able to give their informed consent, it is of utmost importance that users can easily read and understand the documents.

1.2 Problem Statement

Despite how widely privacy policies are used, contemporary research has deemed them not suitable for explaining privacy aspects to end-users, due to their language and structure [12, 21, 40, 66]. Other privacy declarations, like website cookie banners, were even found to contain dark design patterns, which aim to mislead their users [8, 44]. These designs result in a severe

negative impact on end-users privacy and the software's trustworthiness.

While they do not serve as a replacement for the legal document that is the privacy policy, dedicated privacy explanations could help to increase end-users' privacy awareness. This includes them understanding whether their personal information is shared with others, which information is shared and how the information is processed [52]. However, privacy explanations can only succeed as long as they are designed in a way that is understandable and satisfying for end-users. Such a design encompasses an interactable user interface, which allows users to set their privacy preferences while they read the corresponding explanations [21]. Furthermore all necessary information concerning end-user privacy needs to be provided in a way that is actually readable and understandable for those users. At this point, there is a lack of research on how such privacy explanations need to be structured and what information they need to contain [11]. Privacy explanations should efficiently educate end-users, raise their understanding of privacy and increase their privacy awareness. The goal of this thesis is to develop a concept for privacy explanations that satisfy those needs, built upon literature from computer science, law and social sciences.

1.3 Solution Approach

This thesis examines privacy explanations in the context of explainability. It starts with an extensive analysis of existing scientific literature on explainability and its impact on privacy and trustworthiness. Closely related topics, such as dark design patterns in privacy explanations and the privacy paradox, are also explored. In addition to the traditional literature, data from a recent online survey on the effectiveness of privacy explanations was provided by this thesis' supervisor.

Taking the gathered knowledge into account, this thesis develops a complete and novel concept for privacy explanations. Based on different types of explanations found in literature and by adapting the ideas of layered and personalized explanations in artificial intelligence, different types of privacy explanations are offered in a compartmentalized manner, which serves as a simplified way of personalizing them. This concept is also implemented as a software prototype and tested within an exploratory user study with 61 participants. Thereby, we are able to gain insights on the viability of the concept of compartmentalized privacy explanations and the types of privacy explanations that end-users prefer.

1.4 Thesis Contribution

Although there has been a lot of research concerning explainability as a non-functional requirement, so far, it has mostly been in the context of artificial

intelligence. Furthermore, there is still insufficient research on the systematic engineering of explainability [10]. Explainable privacy is a fairly explored field, but it has mostly been focused on the usefulness and readability of privacy policies and privacy notices. This thesis explicitly connects the non-functional requirements privacy and explainability, and is the first work that develops a complete concept for privacy explanations with an actual focus on explainability.

1.5 Thesis Structure

This chapter serves as an introduction to the work contained in this thesis. Chapter 2 is focused on the background knowledge that is needed in order to design a concept for privacy explanations. To this end, related work is examined and discussed. Chapter 3 describes the process of acquiring the primary literature and additional data for this work. As the core section, chapter 4 defines this thesis' research questions, develops the concept for privacy explanations and describes its prototypical implementation. Chapter 5 goes over the conducted user study and its evaluation, while chapter 6 discusses the results and their implications. Finally, chapter 7 concludes this thesis and suggests possible future work.

Chapter 2

Background and Related Work

This chapter explores the topics of explainability and privacy, which form the background foundation for this thesis. To this end, related work from those fields was identified, reviewed and put into context. A detailed description of how the primary literature was acquired is provided in chapter 3.

2.1 Explainability as a non-functional Requirement

Simply said, explainability is the ability of a software system to explain itself. Within the context of requirements engineering, it is seen as a non-functional requirement of software systems. Chazette, Brunotte and Speith [15] offer a refined definition of explainable systems, taking different aspects of the system as well as context into account:

Definition 2.1.1. *A system **S** is explainable with respect to an aspect **X** of **S** relative to an addressee **A** in context **C** if and only if there is an entity **E** (the explainer) who, by giving a corpus of information **I** (the explanation of **X**), enables **A** to understand **X** of **S** in **C**. [15]*

A simplified visualization of this definition can be found in figure 2.1. Generally said, if there is a software system of interest (S), explainability is not concerned with the system as a whole, but rather with specific aspects of the system (X) for a given context (C). The explanation is provided by an explainer (E) to an addressee (A) and includes all relevant information (I), so that the addressee can understand the desired system aspects within the given context.

Explainability does not stand alone, but interacts with other non-functional requirements of software systems [15]. Following definition 2.1.1, explanations are provided with the goal of the addressee understanding how the software works. Thus, explainability is strongly linked to understandability. Providing information about a software system also interacts with its transparency, as insights on the inner workings of the software might be required to be able to provide an understandable explanation.

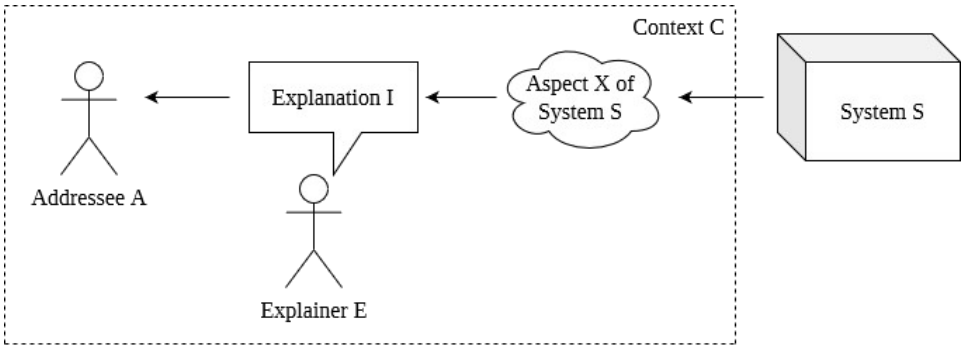

Figure 2.1: Aspects of explainable systems

2.1.1 Impact on related Requirements

As indicated in the previous section, explainability is related to a plethora of other non-functional requirements. These relationships can be bidirectional and result in positive and negative impacts between the requirements. Chazette, Brunotte and Speith [15] provide an extensive overview of these impacts. In the following section, some of those related non-functional requirements and their relation to explainability will be examined.

Transparency and Understandability

Following the definition of explainable systems 2.1.1, it becomes clear that transparency and understandability are closely related to explainability as a non-functional requirement. Chazette, Brunotte and Speith [15] also describe these as the two foundational qualities of their model of explainability. Leite and Capelli [37] provide definitions of transparency and understandability:

Definition 2.1.2. *A software system is transparent if information on its inner workings is* ***accessible***, ***usable***, ***informative***, ***understandable*** *and* ***auditable****. (according to [37])*

Definition 2.1.3. *Understandability is the quality of comprehensible language or thought. (according to [37])*

Adida and Berrada [1] conducted a literature review on the research field of Explainable Artificial Intelligence (XAI). They identified explainability as a promising approach to solve concerns regarding transparency in intelligent software systems. This is underlined by Chazette and Schneider [17], who state that "transparency is becoming increasingly necessary as a non-functional requirement" and point towards explainability as a possible solution to the problem. Doran et al. [22] as well as Rosenfeld and Richardson [56] define transparency as a key attribute of interpretable

systems. While interpretable systems are not necessarily explainable by themselves [22], interpretation is required when forming an understandable explanation [56, 28]. Thus, an explainable system must be interpretable, which in turn means that it needs to be transparent as well.

Regarding the possible downsides of explanations, Chazette and Schneider [17] identified a *Double-Edged Sword* effect of explainability. They found that explanations can interfere with the understandabilty of a system, if they are not presented in a way that is appropriate for their target audience. Furthermore, in case of too much transparency, the user is flooded with unnecessary information, which might lower their understanding of the system as well. Similarly, Kizilcec [32] as well as Springer and Whittaker [62] found too much transparency to have a negative influence on user trust and understandability.

Summing up these findings, transparency plays an important role when providing explanations to end-users. However, careful consideration is necessary so that the degree of system transparency is appropriate for the addressee of the explanation. This way, it is possible to provide end-users with understandable explanations that make the software more understandable.

Trust and Trustworthiness

Nunes and Jannach [45] identify user trust as one of the main goals of explanations within the realm of XAI. They define trust as the notion of a system being perceived as trustworthy by its user. It appears that, when evaluating the impact of explainability on trust and trustworthiness, the terms must first be distinguished from one another. In the following, trust is defined according to Nunes and Jannach [45]:

Definition 2.1.4. *A system S is trusted if it is perceived as a trustworthy system. (according to [45])*

Kästner et al. [36] similarly defined trust and differentiated it from trustworthiness. Following them, trust stems from users' subjective perception of the software and is not a controllable property of the system. This coincides with the definition of trust 2.1.4 according to Nunes and Jannach [45]. Trustworthiness, on the other hand, is defined as a controllable property of the system. Kästner et al. [36] offer the following definition for trustworthiness:

Definition 2.1.5. *A system S is trustworthy to a stakeholder H in a context C if and only if*
(a) S works properly in C, and
(b) H would be justified to believe that (a) if H came to believe that (a). [36]

According to this definition, a software system is not simply trustworthy because users are trusting it. For a system to be trustworthy, the trust must be warranted, meaning that the user is correct in their beliefs about how the system works. Following Kästner et al. [36], users trusting a system even though it might not be trustworthy is called "trust without trustworthiness, or *unwarranted trust*". On the other hand, users can fail to trust a system, even though it is trustworthy. Kästner et al. [36] call this "trustworthiness without actual trust, or *failed trust*".

Contemporary research has already found evidence of the impact of explainability on user trust. Notably, this impact can be both positive and negative. Pu and Chen [51] investigated the impact of explanation interfaces for recommender systems on user trust. They found that those explanations could build user trust, especially if the interface enables users to compare different options available to them. Similarly, Glass et al. [25] found system transparency to be a deciding factor in building user trust. However, transparency itself is not sufficient, as the raw information that was made available still needs to be explained to the addressee [22].

Pieters [49] claims that explanations for trust need to be detailed. According to them, explaining why the software does what it does is not enough. How it is done also needs to be explained. They state that explaining how a system works is not only explaining for trust, but also increases system transparency. However, they note that too much detail would not help trust, as the user would no longer be able to understand the information presented to them. This is supported by Chazette and Schneider [17], who found that explanations can have both a positive or a negative effect on user trust.

Similarly, Bussone et al. [13] found that explaining why the software does something is important for building user trust. At the same time, they found that those explanations could also help users mitigate unwarranted trust. If an aspect of a software system, that is not trustworthy, is explained in an appropriate manner, users are able to understand that their trust might be unwarranted and decide to not use the software. This is underlined by Ribeiro et al. [54], who found that explanations can help users assess their trust in a software system. Kästner et al. [36] propose that an influence on trust must not strictly be positive. Among other things, they discuss the works of Hoffman et al. [30] and Cai et al. [14], both of which propose that explainability could lead to an "appropriate" level of user trust.

Kästner et al. [36] argue that developers should engineer for trustworthiness instead of trust. While user trust appears as a desirable goal when developing software, they state that a system needs to be both trusted and trustworthy, so that the trust is actually warranted. As trustworthiness is more easily controlled than user trust, it should be prioritized over trust in system design. Furthermore, they state that an untrustworthy system that is wrongfully trusted can lead to more devastating consequences than a trustworthy system that users fail to trust. However, it appears that the

trustworthiness of a system is rather hard to gauge. To this end, Kästner et al. [36] declare a need for ways to empirically assess trustworthiness.

Judging from these findings, it becomes clear that, if used appropriately, explanations can positively influence user trust where it is warranted. However, for that to be possible, the underlying software system needs to be trustworthy in the first place. If developers want to design software systems that their users can rightfully trust, it is of utmost importance that the software not only works properly, but is also explained in a manner that is appropriate for those users.

Privacy and Privacy Awareness

Compared to the other non-functional requirements discussed so far, end-user privacy and privacy awareness have not been as much of a focus in the research field of explainability. Examining the effects of explainability on end-user privacy will be a major part of the later chapters of this thesis (chapter 5 and 6). However, some relevant findings were made by past research and they will be discussed in the following. Pötzsch [52] defines privacy and privacy awareness as follows:

Definition 2.1.6. *Privacy is the right to select what personal information about me is known to what people. (according to [52])*

Definition 2.1.7. *Privacy awareness is achieved if an individual knows **whether** their information is shared with others, **which** information is shared and **how** the information is processed. (according to [52])*

Notably, Pötzsch [52] differentiates between privacy of personal sphere and privacy of personal data. The definitions of privacy 2.1.6 and privacy awareness 2.1.7, used within this thesis, refer only to privacy of personal data, as privacy of personal sphere is not a part of this work's scope.

Schneider and Handali [58] found that addressee (they call it "explainee") privacy must be taken into account when providing personalized explanations. As personalized explanations require background information on the addressee they are being personalized for, the addressee needs to have some of their personal data collected and processed. As is common practice, this processing must respect all relevant laws and regulations. For EU citizens this necessarily includes compliance with the GDPR. This is underlined by Ras et al. [53], who voiced concerns about the privacy of those, whose data is ultimately used to provide explanations in intelligent systems. Arrieta et al. [2] support this notion, stating that explanations of intelligent systems "may also compromise the differential privacy of the data origin".

While not strictly from the realm of explainability, privacy priming has been researched as a way of explaining privacy aspects to end-users. Kulyk et al. [35] designed a flyer to prime end-users on privacy related

information and tested it within a user study. Their findings suggest "that the flyer can improve the ability of participants to make more informed decisions to protect their privacy", implying that they were able to raise their participants' privacy awareness. This is underlined by Chong et al. [18], who were successful in promoting safer app selection behavior of end-users through privacy priming.

It becomes clear that paying attention to privacy when providing explanations to users is not only a legal concern. Privacy explanations hold the potential to positively influence privacy awareness, leading to safer and more educated user behavior.

2.1.2 Special Types of Explanations

Examining possible challenges that can arise for explanations, it appears that different users have different needs when it comes to explanations [25, 50, 56, 58, 61, 64, 67]. To tackle this issue, numerous research works have proposed the idea of personalized [25, 31, 58, 61, 64] or layered explanations [50]. By personalizing or layering the explanation, the contained information can be tailored specifically to the needs of each individual end-user. However, when designing personalized explanations, one cannot only customize the contents of the explanation, but also the way in which the information is presented to the addressee. In the following, two special types of explanations, that deviate from the regular textual explanation, will be examined.

Contrastive Explanation

In XAI, so-called contrastive explanations are used as a contrast to regular explanations [1, 38, 42, 51, 56]. While a regular explanation simply states why something happened, a contrastive explanation instead states why something happened instead of something else. For example, a regular explanation could explain why the evening sky is red. A corresponding contrastive explanation would explain why the evening sky is red instead of being blue. Figure 2.2 shows a visualization of this process.

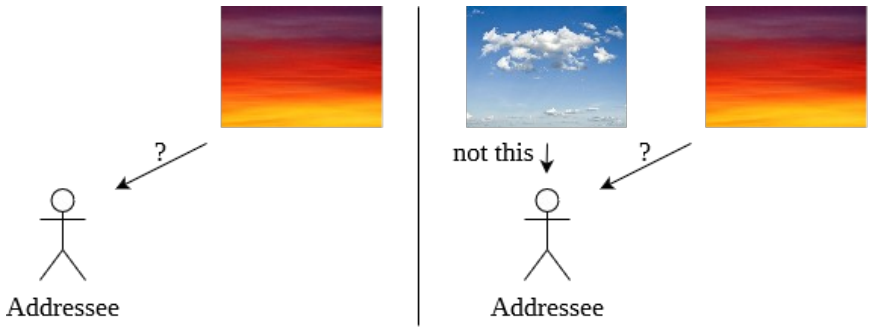

Figure 2.2: Regular explanation (left) and contrastive explanation (right)

Miller [42] indicates the importance of contrastive explanation when making a connection between XAI and the social sciences. They state that, due to the complexity of intelligent software systems, complete regular explanations can put too much of a cognitive burden on the addressee. By offering the addressee a point of comparison, this cognitive burden can be lowered. Following Miller [42], the challenge of providing a viable contrastive explanation lies in finding that point of comparison. Preferably, it should be something the addressee wrongfully expects. In the previous example, the evening sky not being blue is used, as blue is the usual color of the sky at day. If possible, the addressee can provide their wrongful expectation to the explainer. In all other cases, the explainer needs to infer it themselves.

Example-based Explanation

Another form of conveying information to an addressee would be to provide example-based explanations. Within the context of XAI, such example-based explanations have proven to be a viable alternative to regular textual explanations [1, 14, 28].

Example-based explanations can come in different forms, depending on what is being explained. Adadi and Berrada [1] identify two kinds of example-based explanations. The first type is "Prototypes and criticism". Prototypes are examples from a data set that show what the whole set of data could look like. In contrast, criticisms are examples from the same data set, which are not well represented by the prototypes chosen before. Together, prototypes and their corresponding criticisms can provide a fair impression of a data set without overgeneralizing it [1].

The second type of example-based explanations identified by Adadi and Berrada [1] are "Counterfactual explanations". In automated decision systems, counterfactual explanations describe what needs to be done to sway the decision of the system into another direction. Adadi and Berrada [1] mention a bank loan being approved as an example. If a bank customer was denied a bank loan, they might not only want to know why they were denied, but also what they can do about it. A counterfactual explanation would provide them with exactly that information, offering an example of what the customer needs to do in order to become eligible for a loan. Wachter et al. [65] explain that there can be multiple counterfactual explanations at the same time and that their specific relevance differs from case to case.

Adadi and Berrada [1] declare that "Amongst agnostic methods, visualization is the most human-centered technique". Consequently, when designing explanations for end-users, example-based explanations, especially visual representations, can be an effective tool. However, visual examples must also be chosen with the user in mind. They need to meet an appropriate degree of complexity, so that they are informative, but still understandable to their target audience [1].

2.2 Privacy Explanations

Privacy explanations are verbatim explanations concerning the topic of privacy. As the terms are very close to each other, privacy explanations can easily be confused with privacy policies. This section defines privacy explanation as a term and differentiates it from privacy policies. Furthermore, it examines prevalent documents that regularly contain privacy explanations.

2.2.1 Definition and Differentiation

Using the definition of explainable systems provided by Chazette, Brunotte and Speith [15], privacy explanations can be formally defined. The different aspects of privacy explanations are described accordingly:

Aspects of privacy explanations:

System S - The software system to be used or installed
Privacy Aspect PX - The privacy aspect(s) in question
Addressee A - The recipient of the privacy explanation
Context C - The context of interaction with the software system
Privacy Explanation PE - The privacy explanation itself
Information I - The information concerning the addressee's privacy

Taking definition 2.1.1 and the above listed aspects into account, privacy explanations are defined as follows:

Definition 2.2.1. *A privacy explanation **PE** explains one or more privacy aspects **PX** of system **S** to addressee **A** in context **C**, by providing information **I** that is relevant in **C** to **A**, so that **A** understand **PX** of **S** in **C**. (using the definition of explainable systems by [15])*

For end-users to be thoroughly enlightened about their privacy, a notable number of privacy aspects need to be covered. According to the GDPR's article 15 [23], this includes the purposes of processing, what kinds of data are being processed, who receives it, for how long it is stored, where it is not being collected, if it is used for automated decision-making and the different rights available to end-users. This corroborates the definition of privacy awareness 2.1.7 according to Pötzsch [52], which covers whether an individual's information is shared, which of their information is shared and how it is processed. In other words, users understanding the various aspects of their privacy coincides with them achieving privacy awareness.

Within in the definition 2.2.1 of privacy explanations, the privacy explanation acts as the explainer that conveys the privacy-related information to the end-user. Such an explanation could come in textual form, but it

could also be presented visually or via audio. In any case, it either needs an user interface or a document within which it can be provided. Widespread examples of such documents would be privacy policies, app privacy notices and cookie banners. Notably, they are not privacy explanations themselves, but are documents that usually contain various paragraphs of textual information, some of which can include privacy explanations. This differentiation may be unclear to end-users, especially in the German language, and will be subject to further discussion in chapter 6.3.

2.2.2 Documents containing Privacy Explanations

Documents containing privacy explanations are regularly encountered by users when they browse the Internet or install software on their devices. Their practical application is for end-users to read them, understand them and decide if they give their informed consent to the therein described data processing practices. According to the GDPR's article 12 [23], the processing of end-users' personal data must be explained to them "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". At present, these documents are not fit to fulfill these purposes, as they are neither readable nor understandable for end-users [12, 21, 66].

Privacy Policies

Waldman [66] discusses privacy policies as a failed attempt of implementing the so-called "notice-and-choice" concept. The idea of the notice-and-choice concept is that end-users are supposed to be notified about data processing practices (for example through a privacy policy). They are then supposedly free to choose if they want to consent to these practices. However, according to Waldman [66], privacy policies miss their purpose. They find that "no one reads privacy policies in part because they are long and difficult to understand". Furthermore, they state that "Even privacy experts find them misleading". This failure is attributed to privacy policies not being designed with the correct target audience in mind. Instead of being written as educational documents for end-users, they are legal documents aiming to assure the developers' compliance with laws and regulations. Finally, Waldman [66] states that it is not sufficient to just present privacy information in an appropriate language. Their findings suggest that the visual and structural design of privacy policies plays an important role in increasing system transparency to the benefit of end-users.

Cranor [21] came to a similar conclusion, stating that privacy policies miss their purpose of providing notice-and-choice, as they "are poor mechanisms for communicating with individuals about privacy". Similar to Waldman [66], Cranor [21] attributes this to lackluster contents and structure of privacy policies, stating that "These policies are long, complicated, full of

jargon, and change frequently".

In the context of notice-and-choice, Martin [40] examined how end-users privacy expectations relate to the actual contents of privacy notices. They found that even in scenarios where privacy notices are technically correct, users may feel like their expectations were violated. This is ascribed to users projecting their own privacy expectation onto the privacy notices, rather than understanding the actual contents of the document as they are. Consequently, Martin [40] states that "privacy notices are insufficient to meet privacy expectations". Furthermore, they argue that the misunderstanding is rooted in the designed obscurity of privacy notices. Their results suggest that "if firms were actually clear about privacy practices, consumers would share less information". If the opacity of the documents leads to users trusting the systems, developers might feel no need to provide end-users with in-depth privacy information that they do not want to divulge.

Brandtzaeg et al. [9] found apps that do not comply with their own Terms of Service (ToS) and privacy policies, misinforming their users and processing their data in undisclosed ways. On top of that, they found that "some apps tracked users when the app was not in use, violating the app terms of use and privacy policies". Similar concerns were voiced by Okoyomon et al. [46], who found that privacy policies use ambiguous and confusing statements, which mislead end-users. In particular, they highlight examples of privacy policies being unclear or dishonest about children's privacy, third party data sharing and secure data transmission.

In summation, privacy policies are clearly not fit to explain privacy to end-users. Their length, structure and language is problematic, both in terms of readability and understandability. The danger posed by this state of the art is twofold. On the one hand, most users cannot feasibly read current privacy policies and stay completely unaware of how their data is being processed. On the other hand, users might become increasingly reluctant towards looking at privacy explanations in the first place, as they do not expect them to be readable by default.

App Privacy Notices

When installing a mobile app from places like the *Apple App Store* or the *Google Play Store*, users are presented with a privacy policy and an app privacy notice. These privacy notices specify the system permissions that the app in question asks from the user. This is often connected to a user's personal data, such as their contacts and their location data. Examples of parts of privacy notices from both, the *Apple App Store* and the *Google Play Store*, can be seen in figure 2.3. Individual screenshots of and links to these examples can be found in the appendix C of this thesis.

As can be seen in the examples, these privacy notices are a stark contrast to privacy policies. They are shorter, use bullet points and contain only

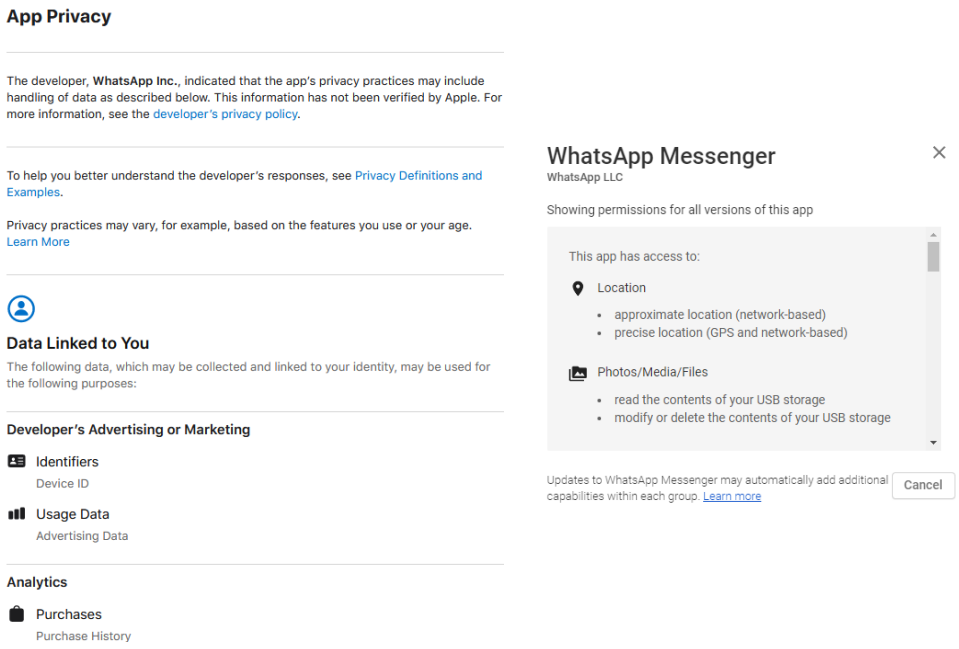

Figure 2.3: App privacy notices by *Apple* (left) and *Google* (right)

minimal amounts of information in comparison. The notice from the *Google Play Store* is the less detailed one, naming only the system functionalities, which are accessed from the user's device. Data processing is not directly explained and must be inferred by the user. The only alternative offered by the *Google Play Store* is a link to the app developer's privacy policy.

The privacy notice from the *Apple App Store* is a bit more specific. It is categorized by the different processing purposes and lists all data types that apply to each purpose. In addition to the developer's privacy policy, a link to a textual list of examples for the different types of data[1] is available.

The shortness of these privacy notices makes reading them more feasible for end-users. However, they leave out important privacy information, such as how long the data is stored, who receives it and what laws and rights apply. The list of data examples given in the *Apple App Store* is better than nothing, but is rather general and stands in no specific context to the app in question. As such, end-users will not be able to get a thorough understanding of their privacy, just by reading them, and the only other available option is the developer's privacy policy. Consequently, end-users who want to install apps like these are not educated in a way that is both readable and understandable for them, while offering all relevant privacy information at the same time.

[1] https://apps.apple.com/story/id1539235847

2.3 Dark Patterns in Privacy Design

On top of identified problems with readability and understandability, end-users are confronted with additional hurdles when they encounter privacy explanations. When interacting with software on the Internet or on their personal devices, they can encounter so-called *Dark Patterns*, both knowingly and unknowingly. These malicious design patterns are aiming to manipulate and exploit end-users. Chromik et al. [19] examined dark patterns in the context of explainability. They found that these patterns are usually used when explanations "are not primarily designed with the users' benefits in mind, but purposely deceive users for the benefit of other parties".

Within privacy explanations, dark patterns might be used to lead users into sharing their personal data, even though they might not want that. This section explores existing dark patterns, used within documents that contain privacy explanations, and the effects that they have on end-users.

2.3.1 Cookie Banners

Bongard-Blanchy et al. [8] investigated dark patterns in online spaces and offer a classification of those patterns based on their findings. Among these classes of dark patterns, many apply to the so-called cookie banners. Cookie banners are a form of Consent Management Platforms (CMPs), used to let end-users choose their cookie preferences when they visit a website. Bongard-Blanchy et al. [8] cite Nouwens et al. [44], who examined CMPs from the United Kingdom (UK)'s top 10.000 websites. Among the tested CMPs, only 11.8% did not employ dark patterns. Following Bongard-Blanchy et al. [8] and Nouwens et al. [44], dark patterns that appear in cookie banners can be classified as follows:

- Loss-gain framing → ambiguously claiming that the site will not "work properly" without cookies or framing cookies as necessary for security
- Pre-selection → setting options for data processing to "yes" by default, especially for "legitimate interests"
- False hierarchy → showing buttons to accept data processing more prominently than buttons to deny it (using color, size and position)
- Hidden information → hiding "legitimate interests" on secondary page and making them not easy to find
- Bundled consent → prominently offering buttons to consent to all data processing at once
- Forced consent → claiming that the website cannot be used at all, unless data processing is accepted, or inferring implicit consent from users simply visiting or navigating the website

Gray at al. [27] conducted a survey on the effects of such dark patterns on end-users. They found that, while end-users might not be able to precisely describe dark patterns, they are indeed generally able to identify issues that make them feel like they are being manipulated. Bongard-Blanchy et al. [8] support this notion, stating that "individuals are aware of manipulative designs' potential influence on their behaviour and rather capable of recognising such designs". They also state that just increasing end-users' awareness of dark patterns is not enough to protect them, but that, in order to adequately protect end-users, "design, technical, educational, and regulatory measures" need to be taken.

2.3.2 Third Party Sharing

North [43] examined the ToS agreements of the 50 most popular apps from the *Apple App Store*, from each the free and the paid section. They found that 80% of the examined ToS agreements admitted to collecting data from end-users and that 71% shared such user data with third party companies. Brandtzaeg et al. [9] investigated personal data flows of mobile apps and found that data from Europe is sometimes shared with third parties abroad. They state that this endangers end-user privacy, as EU data protection laws might not apply on other continents.

Okoyomon et al. [46] checked the privacy policies of mobile apps with regards to third party data sharing. They found that, among the apps they investigated, "over 75% of mobile apps make use of third-party services, only around 22% actually disclose the names of those services, while 10% do not mention any information about their affiliates at all". Furthermore, their findings show evidence of developers ridding themselves of accountability by stating that they are not responsible for the privacy policies and practices of third parties.

Balebako et al. [6][2] researched the understandability of app privacy notices and found that end-users struggle with understanding third party entities, unless they recognize the names of the respective companies. This is critical, as privacy explanations should be provided with the purpose of educating end-users. If users are only provided with a third party's name or general category (such as data resellers or government agencies), they might not be able to infer how their personal data is being processed.

Third party data sharing has become a ubiquitous practice on websites and mobile apps. As it stands, users are often provided with nothing more than names or general categories when they encounter third parties in documents that contain privacy explanations. Consequently, users are not sufficiently educated about their privacy, as they cannot understand how and why their data is used.

[2]This technical report is not peer reviewed, but is co-authored by Lorrie Cranor, who is trusted and respected in the privacy research community

2.4 Privacy Paradox

When studying the privacy behavior of end-users in online spaces, contemporary research has found a phenomenon called the *Privacy Paradox*. Barnes [7] coined the term when investigating teenagers' privacy behavior on social networks in the United States of America (USA). They found that adults are generally very concerned about their privacy, suspecting government agencies and big tech companies to amass and process citizens' personal data. At the same time, teenagers who are active on social media share the same kinds of personal data voluntarily with the public. This discrepancy in privacy behavior is what Barnes describes as the privacy paradox. They also note that, while solving this issue is not an easy task, "Awareness is key to solving the solution" [7]. This underlines the importance of explainability's impact on privacy awareness when designing privacy explanations. In order for end-users to actually want to protect themselves, they need to be educated about possible privacy issues first.

Pentina et al. [48] introduce a wider definition of the privacy paradox and link it to the concept of *Privacy Calculus*. Throughout their research, they found that millennial mobile end-users use apps that access their personal information, even though they might have privacy concerns about those same apps. Consequently, the privacy paradox does not only exist across generations, but also within the users themselves. According to Pentina et al. [48], end-users who are aware of privacy issues of mobile apps might still want to use them, if the social or economic benefits obtained by using those apps outweigh the privacy risks that come with them. They address this calculated trade-off between privacy and service as the the privacy calculus theory. These findings are concerning, as end-users might engage in unsafe behavior if they wrongfully feel like the trade-off is worth it. Dark patterns such as *loss-gain framing* (introduced in section 2.3) can give end-users false impressions about the risks and benefits, and push them towards accepting privacy trade-offs, even though they are not really beneficial to them.

Hargittai and Marwick [29] researched the connection between the privacy paradox and the online apathy of end-users. They found that "the privacy paradox cannot be attributed solely to either a lack of understanding of or a lack of interest in privacy". Instead, they attribute the lack of secure privacy behavior to the apathy end-users developed towards privacy in online spaces. According to their findings, this apathy is rooted in both the opaque design of contemporary privacy protection regulations and in the belief that end-users cannot do anything about the violation of their privacy anyways. This is mirrored by Waldman [66], who states that documents such privacy policies are not designed for users, but are rather "written by lawyers and for lawyers". Research by Sunyaev et al. [63] reinforces the privacy paradox, finding that "apps are being highly rated and successfully sold although privacy policies are either absent, opaque, or irrelevant".

End-users of software systems should feel like their privacy and time is respected by the developers. Documents such as privacy policies present them with endless pages of legal language, while cookie banners and app privacy notices use questionable design patterns and miss out on important privacy information. Neither solution is acceptable, especially if the goal is to get the users' genuine informed consent. The fatal result of these practices is that users stop reading privacy explanations, even though they might care about their privacy. Consequently, the documents miss their purpose of educating users about their privacy and fail as designs.

The existence of the privacy paradox poses an important challenge to designers of privacy explanations. Providing privacy explanations that are readable and understandable for end-users could be an important step towards solving this privacy paradox and raising privacy awareness. Cranor [21] claims that standardized privacy notice mechanisms "have failed users and they will continue to fail users unless they are accompanied by usable mechanisms for exercising meaningful choice and appropriate means of enforcement". Preferably, such operability would be directly implemented within privacy explanations, so that users can set their privacy preferences while they are being informed. This can reinforce end-users' right to decide over their own personal data and could expectably provide them with a feeling of control and safety.

Chapter 3

Data and Literature Analysis

This chapter describes the ways in which the primary literature and data used for this work was chosen and analyzed. The process consists of a manual inspection, starting with the review of a baseline paper that was provided by this thesis' supervisor. Using this as a starting point, a snowballing process under the use of grounded theory is conducted. To further supplement and diversify the acquired literature, a database search with three different search strings was planned and conducted on the web search engine *Google Scholar*[1]. Finally, the data set resulting from a survey on privacy explanations, which was also provided by this thesis' supervisor, is both described and analyzed.

3.1 Literature Selection Criteria

Works selected for the literature analysis of this thesis are subject to a set of strictly defined inclusion and exclusion criteria. They must fulfill all of the inclusion criteria and none of the exclusion criteria in order to be chosen. The criteria are defined as follows:

Inclusion Criteria:

(i) The work must be concerned with either one of the following topics:

  (a) Explainability or related non-functional requirements (e.g. trustworthiness, privacy, ...)

  (b) Privacy explanations

  (c) Documents that contain privacy explanations

(ii) The work must be peer reviewed.

(iii) The work must be freely available for reading.

[1] https://scholar.google.com/

Exclusion Criteria:

(i) The work must not be written in a language other than English or German.

(ii) Works from after 2019 should not only be focused on the ongoing *Covid-19* pandemic.

3.2 Manual Inspection

The first part of this thesis' literature analysis was a manual inspection, akin to a systematic literature analysis, but smaller in scope. A visualization of this approach can be seen in figure 3.1. This section of the thesis describes each step of the process in detail and provides overviews of the corresponding results. Detailed listings of the results of each step can be found in the appendix A of this thesis.

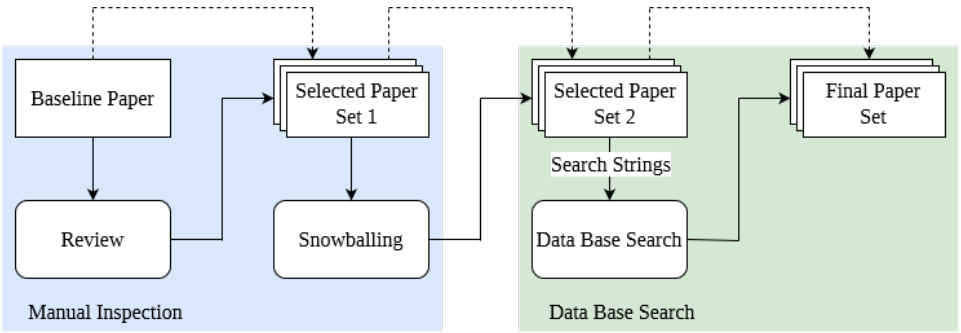

Figure 3.1: Process of literature analysis

Review of the Baseline Paper

The baseline for the literature analysis of this work is a rather general research paper on explainability by Chazette, Brunotte and Speith [15]. It serves as a comprehensive introduction into the world of explainability and offers definitions for all necessary terms. Furthermore, it develops a model as well as an extensive knowledge catalog for explainability and its impact on related non-functional requirements. Understanding the baseline paper and reviewing its references was critical in developing a thorough understanding of explainability as a non-functional requirement.

After the paper by Chazette, Brunotte and Speith [15] was read in full, all papers cited by the baseline paper were identified. Additionally, using the *cited by*-function of *Google Scholar*, all paper citing the baseline paper at the time were identified. If a paper could not be immediately included or excluded via its title, its abstract and conclusion were scanned for relevant

information. At the point in time when the review was conducted, this yielded a total of 88 cited papers and 2 citing papers respectively. Judging by the abstracts of the papers and a quick in-document search, they are classified by whether they deal with explainability, trustworthiness or privacy. Related terms such as transparency or trust were also identified and noted. Table 3.1 shows the total number of reviewed papers, the number of included papers that meet the criteria defined in section 3.1 and the number of included papers associated with explainability, trustworthiness (short: trustw.) and privacy. Notably, some of the included works were concerned with none or more than one of these non-functional requirements.

Table 3.1: Results of baseline paper review

Type	Reviewed	Included	Explainability	Trustw.	Privacy
Cited	88	42	34	10	5
Citing	2	2	2	1	0
Total	90	44	36	11	5

A total of 44 papers were included and considered for further reading. Together with the initial baseline paper, they form *Selected Paper Set 1* (see figure 3.1). Notably, there is a strong focus on explainability, while trustworthiness and privacy only appear in a smaller amount of works.

The Process of Snowballing

By reviewing the initial baseline paper, the first set of selected papers was obtained. From there, a snowballing process was conducted. Wohlin provides a formal approach to snowballing in their "Guidelines for Snowballing in Systematic Literature Studies [...]" [68]. The approach chosen in this work is derived from the therein described process.

Snowballing was conducted in both backward and forward direction. Backward snowballing examines all of the cited papers of the work in question, while forward snowballing examines all of the citing papers. Normally, the snowballing process is iterated until no new papers can be found. For this work, the snowballing process was not complete, but ended when a "theoretical saturation" was achieved, meaning that no new concepts appeared. Wolfswinkel et al. [69] defined this as the goal of rigorously reviewing literature, using grounded theory. The approach was also followed by Chazette, Brunotte and Speith [15].

According to Wohlin [68], the snowballing procedure starts with a tentative start set of papers. The papers from the starting set are evaluated for relevance and either included or excluded from the first step of snowballing. In this work, the snowballing started from the first set of selected papers, which was obtained in the previous step. Four papers from

the first set, that were particularly promising, were chosen as starting points. A paper by Kästner et al. [36] was selected from the initial citing papers. Kästner et al. define the difference between trust and trustworthiness, and explain the difficulties with evaluating the latter.

Among the many initially cited works, a paper from Chazette and Schneider and a paper from Kizilcec, as well as the supplementary material of the baseline paper [16] were chosen for snowballing. In their work, Chazette and Schneider [17] identified a double-edged sword effect of explainability. Similarly, Kizilcec [32] found that too much transparency in software systems can erode user trust. Both are critical findings when engineering explainability for privacy. Finally, the supplementary material of the baseline paper [15] was chosen for snowballing as it provides an extensive overview and classification of literature concerned with the impact of explainability on related non-functional requirements. The results of snowballing can be seen in table 3.2.

Table 3.2: Results of snowballing

Direction	Reviewed	Included	Explainability	Trustw.	Privacy
Backward	263	39	17	4	1
Forward	251	4	0	0	0
Total	514	43	17	4	1

Summary of Manual Inspection Findings

Together with the papers from the first set, the results of the snowballing form the new set of selected papers. Summing up the findings from both steps of the manual inspection, we arrive at a grand total of 87 included papers. Table 3.3 provides an overview of the acquired literature.

Table 3.3: Combined results of manual inspection

Step	Included	Explainability	Trustw.	Privacy	XAI
Review	44	36	11	5	39
Snowballing	43	17	4	1	21
Overall	87	53	15	6	60

Notably, the acquired literature leans heavily in the direction of explainability as a non-functional requirement, specifically in the context of XAI. However, to successfully develop a concept for privacy explanations, there is also a need for literature that is concerned with those explanations and the documents in which they appear. To this end, a database search was conducted.

3.3 Database Search

As shown in the summary of findings, the literature found through the manual inspection in section 3.2 was heavily focused on explainability as a non-functional requirement and the role that it plays within the context of XAI. To further supplement the available literature in the direction of privacy, a database search was planned and conducted on *Google Scholar*. Preliminary searches for works regarding privacy returned millions of results. Clearly, working through that many papers was not feasible. Furthermore, the search could be subdivided into three broader categories. Therefore, it was performed over three iterations with individual search strings.

The definition of the search strings is a direct result from the forgoing findings of the manual inspection. Considering the focus of the previous results, the goal was to find works that are concerned with the relationship between explainability, privacy and trustworthiness. Furthermore, there was still a need for literature that is concerned with privacy explanations and the documents which include them. To this end, two new exclusion criteria were added to the previously ones:

Exclusion Criteria:

(iii) The work must not be concerned with only explainability.

(iv) The work must not be focused on XAI.

In order for a paper to be included for further reading, it needed to fulfill all of the criteria defined in section 3.1 as well as the newly added exclusion criteria. If a paper could not be immediately excluded by its title, it became subject to closer inspection. Figure 3.2 shows the process by which papers were included or excluded from further reading. Detailed listings of the results of the database search can be found in the appendix A of this thesis.

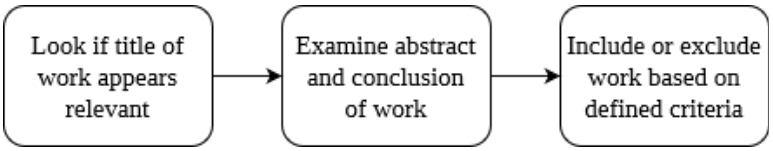

Figure 3.2: Process for inclusion from database search

The First Search

The goal of the first search was to find literature on the relationship between explainability and privacy. Additionally, it should be concerned with either privacy explanations or explainable privacy in specific. Alternatively, it could include the related non-functional requirements usability or

trustworthiness. The first search string was defined as follows:

(explainability **AND** privacy) **AND** (usability **OR** trustworthiness **OR** "privacy explanation" **OR** "explainable privacy") **AND** (-ai -xai -machine -neural -recommender)

The first part of the search string, '(explainability AND privacy)', ensures that both the terms explainability and privacy are included within the examined work. The second part of the search string, '(usability OR trustworthiness OR "privacy explanation" OR "explainable privacy")', is OR-connected, which means that either one of those four terms must be included within the work. The last part of the string, '(-ai -xai -machine -neural -recommender)', excludes prevalent terms from the field of XAI.

At the time, this query yielded a total of 83 papers, 12 of which were subject to closer examination. Among them, only four fit the constraints for inclusion and were considered for further reading. Table 3.4 shows the results for the first search string.

Table 3.4: Results of the first database search

Total	Examined	Included	Explainability	Trustw.	Privacy
83	12	4	1	0	4

The Second Search

For the second search, the goal was to gather insights on the state of privacy explanations and the documents that can contain them. Dedicated privacy explanations are not regularly encountered by themselves, but are usually included within documents such as privacy statements or privacy notices. The search string therefore includes these terms. As the goal of these documents is to get end-users' informed consent, consent is also included for the search. As a result, the second search string was defined as follows:

allintitle: (privacy) **AND** (explanation **OR** statement **OR** notice **OR** constent) **AND** (-ai -xai -machine -neural -recommender)

When using the search string without 'allintitle:', *Google Scholar* returned a total of more than 3 million works. Hence, the query was reduced to works with the search terms in the title. '(privacy) AND (explanation OR statement OR notice OR constent)' ensures that the term privacy as well as either one of the other terms must appear. Similar to the first search string, '(-ai -xai -machine -neural -recommender)' is used to remove works focused on XAI. Table 3.5 shows the results for the second search string.

Table 3.5: Results of the second database search

Total	Examined	Included	Explainability	Trustw.	Privacy
173	14	13	0	0	13

The Third Search

To expand the literature findings from the second search into the realm of mobile applications, which play an especially important role in today's society [59], a third and final database search was conducted. The third search string looked as follows:

allintitle: (app **AND** privacy) **AND** (mobile **OR** smartphone) **AND** (-covid)

Similar to the second search string, this query made use of 'allintitle:' to limit the number of search results to a more reasonable amount. The results had to be concerned with both apps and privacy, either on smartphones or on mobile devices in general. As results of a preliminary search were overwhelmingly focused on app privacy in the context of the ongoing *Covid-19* pandemic, '(-covid)' was used to further filter the query results. An overview of those results can be seen in table 3.6.

Table 3.6: Results of the third database search

Total	Examined	Included	Explainability	Trustw.	Privacy
123	32	24	0	0	24

Summary of Database Findings

Overall, the database study achieved its goal of expanding and diversifying the included literature in the direction of privacy. The combined results of all three searches can be found in table 3.7.

Table 3.7: Combined results of the database search

Search	Included	Explainability	Trustw.	Privacy
first	4	1	0	4
second	13	0	0	13
third	24	0	0	24
overall	41	1	0	41

3.4 Privacy Explanation Study

In addition to the regular literature, data from an online questionnaire regarding privacy explanations was provided by this thesis' supervisor. As literature on privacy explanations, especially on a conceptual level, is in short supply, the provided study grants important insights on what end-users expect from those explanations. This section will give a brief introduction to the study, analyze a part of the obtained data and provide a short discussion of the results.

3.4.1 Data Set Analysis

The study is a short online survey on the topic of privacy explanations, gauging if there is an interest in those explanations and what they should look like. 155 participants completed the survey in either German, English or Portuguese language. At the start, participants were presented with a hypothetical scenario that contained examples of privacy explanations. Within the context of an app for sightseeing tours, it was stated that both their location data as well as their date of birth would be processed. Participants were then presented with explanations for why each processing is necessary. After reading the scenario, they answered questions concerning the following topics:

1. How useful are the privacy explanations within the given scenario?
2. Are participants generally interested in privacy explanations?
3. Did participants feel more comfortable after reading the privacy explanations?
4. Can privacy explanations increase the trust in a software system?
5. What should privacy explanations contain?
6. What are the benefits of privacy explanations?
7. When should privacy explanations be presented?

There were both closed- and open-ended questions. The answers for the closed-ended questions were evaluated by using descriptive statistics. For the open-ended questions, many participants gave extensive and detailed answers. By conducting a simplified cycle of *in vivo* coding [57], these answers were summed up and categorized. This allows for a short and superficial analysis of the participants sentiments.

3.4.2 Discussion

Judging from the participants diverse answers, a number of general observations on privacy explanations can be made. In the following, we go over the different subjects of the questions asked within the survey, and take a closer look at the related answers.

Interest in and Usefulness of Privacy Explanations

Participants interest in and their perceived usefulness of privacy explanations was gauged on a 7-point Likert scale. More than 80% of participants answered that they are at least "moderately interested" in those explanations, more than two thirds of which were at least "very interested".

Concerning the usefulness of the privacy explanations presented in the context of the proposed scenario, answers deviated between the two example data types. While more than 50% participants said that the explanation for the date of birth was at least "slightly useful", almost 80% said the same for the processing of the location data. Judging from this stark contrast, it appears that privacy explanation's usefulness depends on types of data which is processed and the context in which the data is processed. On the face of it, location data feels more natural than date of birth in the context of a sightseeing app, which is likely why it was more accepted by participants. This coincides with Ortloff et al. [47], who found that users prefer privacy policies when they are presented in a contextual manner.

Summing up this section of the discussion, it becomes clear that participants appear to be generally interested in privacy explanations. Furthermore, those explanations can be useful to users, so long as they make sense within the context of the explained software system.

Effects on Trust and Comfort

When evaluating the effects of privacy explanations on the trust in software systems, participants answered again answered on a 7-point Likert scale. More than 80% of participants said that they at least "somewhat agree" that privacy explanations can increase the trust in a software system, more than two thirds of which said that they at least "agree".

Concerning the comfort given by privacy explanations, participants were asked if they felt more comfortable or not. If not, they were given the possibility to reason their answer in an open-ended manner. More than 50% of participants answered that they felt more comfortable after reading the explanations in the proposed scenario. Among the ones that did not think so, more than 66% gave a reason for their answer. Most of the given reasons were related to quality issues of the given privacy explanations. A number of participants stated the explanations were either not credible or trustworthy, were too superficial or did not make sense.

Other participants stated that they did not want to disclose the data, that the processing was not appropriate for the app's described features or that the features were not wanted in the first place. These are likely not problems with the explanation, but rather participants not agreeing to the trade-off between privacy and service described within the scenario.

Generally speaking, it appears that privacy explanations can have a positive effect on both the trust and the comfort of end-users. However, participants answers made clear that those explanations need to fulfill high quality standards, especially concerning their credibility, completeness and understandability.

Elements and Timing of Privacy Explanations

Participants were asked what they think a privacy explanation should contain, answering in an open-ended manner. The responses can be divided into 3 categories: contents, form and additional requests.

Concerning the actual contents of the explanations, participants said they wanted to know what data is being processed and the reasoning behind it. They wanted to know how the data is used and stored, who can access it and when it is deleted. Participants also stated that they want to be educated on the benefits of sharing their data and that they want to know about possible third party sharing. When it came to the form of privacy explanations, participants stated that they need to be exact and detailed, not leaving anything out. At the same time, the explanations should be short and precise, use simple language and be well-structured. Visual help, for example in the form of icons, was also suggested.

In addition to contents and form, some participants stated that they want privacy explanations to give them a sense of security. Furthermore, they demanded their right to control their own data. Therefore should be operable, allowing users to control, deny and restrict data disclosure. Another question of the survey was concerned with the timing of privacy explanations. When asked if they prefer to see the explanation, more than 66% answered that they want to be automatically notified when something regarding their privacy changes. Another fourth of the participants wanted to see an explanation every time they request it.

The answers discussed in this section grant important insights on what privacy explanations need to contain and how they need to be structured, in order to be satisfying for end-users. Notably, this is not only limited to the explanations themselves, but also includes the functionalities of such explanations. Judging from the participants' preferences, operability and the ability for users to exercise control over their own data may be key to developing a successful concept for privacy explanations.

Benefits of Privacy Explanations

When asked about the benefits of privacy explanations, participants offered a variety of ideas and wishes concerning privacy explanations, not all of which were necessary answering the question itself. Concerning the topic of data disclosure, they mentioned that privacy explanations offer transparency, delivery information on the disclosure of user data and the usage by third parties. Additionally, those explanations could contain reasonings and justifications and ensure the legal compliance of the company behind the software.

Participants stated that privacy explanations can be used to educate end-users, increase their privacy awareness and ensure the security of users and their data. If operable, they could also help users control their own data and guarantee their privacy. Another effect of the explanations is that they can help the user to build trust in the software and the company, and support them in comparing and choosing apps and services. At the same time, it was mentioned that companies could use the explanations to better communicate with their users, explain the software to them and also protect themselves legally. Privacy explanations would also offer room for them to explain how they want to serve advertisements or possibly sell user data to affiliated third parties.

While not specifically part of the question, participants used their answer to express their worries about privacy explanations. Some stated that those explanations are not for end-users, but only of use for the companies that provide them. As they were a formality, privacy explanations would not be trustworthy and there would not be a way to check the truthfulness of their contents. Another problem was that participants perceived current privacy explanations as not understandable and badly designed. Besides that, them not being actionable and there seeming to be no alternatives was mentioned as problematic.

What we can gather from this last section of answers is that privacy explanations could prove to be a viable opportunity to educate and support end-users. However, this can only work if those explanations are designed in a user-centric manner, addressing the worries and fears of their users, and respecting their right to control their own data.

Chapter 4

Concept Development

This chapter, as the core of this work, develops a novel concept for privacy explanations. To this end, four research questions are defined. The concept for privacy explanations is then developed with those questions in mind. Lastly, the concept is implemented within a prototypical software system, so that it can be tested within a practical user study.

4.1 Research Questions

As was thoroughly explored in section 2.2, little is known on how to actually design and present privacy explanations, so that they are understandable and satisfying for end-users. Hence, the first two research questions are concerned with how to provide desirable structure and contents of privacy explanations. Furthermore, different types of privacy explanations will be tested and compared. These explanation types are then going to be examined through the third research question. Finally, the possible impacts of privacy explanations on end-users' privacy awareness will be examined within the context of the forth and final research question.

RQ1: Privacy Explanation Structure

Throughout section 2.2, it became clear that contemporary documents, which contain privacy explanations, employ structures that make them difficult to read and understand. In order for privacy explanations to be useful sources of privacy information, they need to be presented within a structure that allows users to properly read and understand them. Research question 1 tackles this problem, investigating the structural elements necessary for understandable privacy explanations:

> **RQ1:** How must privacy explanations be structured, so that they are understandable for end-users?

RQ2: Privacy Explanation Contents

Similar to the question of structure, there is a need to examine the required contents of privacy explanation. Privacy explanations should not just comply with law and regulations, but put the end-user and their needs into the focus. Users need to feel sufficiently educated on their privacy matters, in order to give genuine informed consent to the processing of their personal data. Research question 2 is concerned with the privacy explanation contents that are necessary to satisfy end-users' want for privacy information:

> **RQ2:** What must privacy explanations contain, so that they are satisfying for end-users?

RQ3: Privacy Explanation Types

Following the findings of section 2.1.2, it becomes clear that privacy explanations can not only be presented in a direct and textual manner. Consequently, there is a need to investigate the viability of different privacy explanation types, such as contrastive or example-based privacy explanations. These explanation types have proven to be effective within the realm of XAI and it is reasonable to assume that they can be viable when used within privacy explanations as well. Research question 3 examines if end-users prefer different kinds of privacy explanations and the reasons for their preferences:

> **RQ3:** Do end-users prefer different kinds of privacy explanations?

RQ4: Privacy Explanations' Impact on Privacy Awareness

As discussed in section 2.4, increasing end-users' privacy awareness can be critical in helping them to employ safer privacy behavior. Findings from section 2.1.1 imply the potential of privacy explanations to increase privacy awareness in end-users. However, at the current point in time, there is no significant research to support or deny this notice. With that in mind, research question 4 seeks insights on whether privacy explanations are actually able to impact the privacy awareness of end-users:

> **RQ4:** Can privacy explanations influence end-users' privacy awareness?

4.2 Influential Factors for Privacy Explanations

With the research questions in mind, this chapter will go on to develop a concept for viable privacy explanations, that are both understandable and satisfying for end-users. To that end, it is necessary to identify the different factors that influence privacy explanations. Figure 4.1 shows a comprehensive overview of all of these factors.

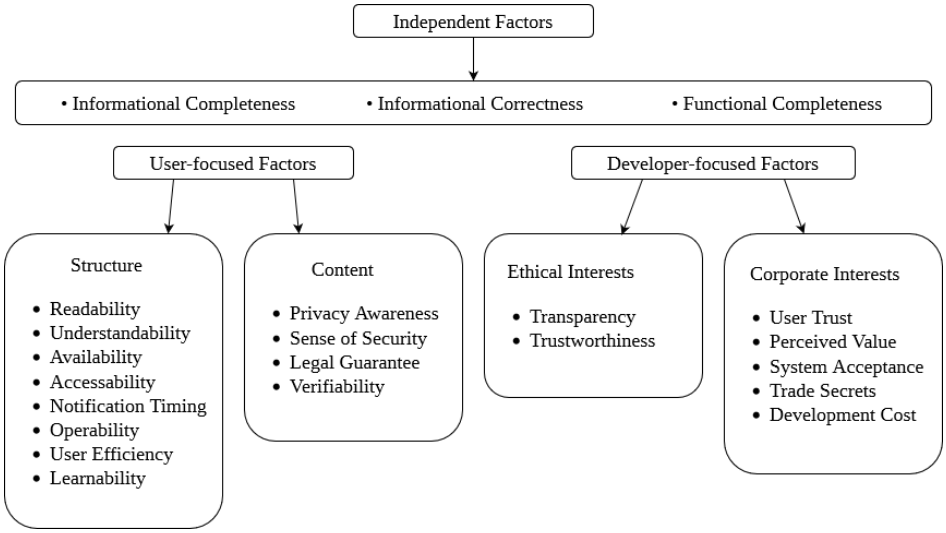

Figure 4.1: Influential factors for privacy explanations

The factors are classified by who they are targeted at. User-focused factors have the needs and goals of end-users in mind. Developer-focused factors describe the ways in which privacy explanations can be useful to the developers of the software which the privacy explanations are related to. The independent factors include qualities that are not focused on users or developers, but are generally needed from both sides.

User-focused Factors

In accordance with RQ1 and RQ2, the user-focused factors are subdivided into qualities that either concern the structure or the contents of privacy explanations. First and foremost among the structural factors stand readability and understandability. Reading and understanding privacy explanations are necessary steps before users can give genuine informed consent within a notice-and-choice framework [66]. Following definition 2.2.1, privacy explanations are only useful as long as users are able to understand them.

In order to be able to read the explanations in the first place, they need to be available and accessible to end-users. Research by Sunyaev et al. [63]

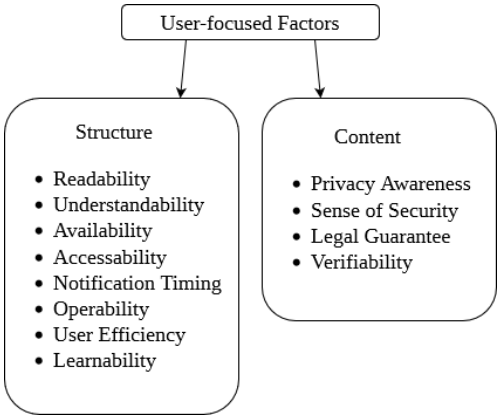

Figure 4.2: User-focused factors for privacy explanations

found that mobile apps often do not provide privacy policies. Taking this as well as the findings of section 2.3 into account, it appears that users of those apps currently suffer from a poor availability of relevant privacy information. This is critical, as it denies users their right to be informed about their privacy, which is granted by the GDPR.

The timing of the notification is also an important factor when providing privacy explanations to end-users. Results from section 3.4 showed that end-users want access to privacy explanations whenever something happens concerning their privacy. Yet, Balebako et al. [5] found that privacy explanations should already be present before the software is installed, so that users are able to make a more well-founded decision about whether or not they want to use the software in question. They also state that those explanations should be operable by the users, so that they can make their privacy preference choices while reading the explanations. If privacy explanations are not operable themselves, users might forget some of the privacy information before reaching the privacy settings within the application. This is also mitigated by enabling users to make their choices immediately, before they start the application for the first time.

Research by McDonald and Cranor [41] reveals that if the average end-user were to just skim over every privacy policy they encounter, they would spend an average of 154 hours by year doing so. It would take even longer to fully read and understand them. This is clearly not acceptable and highlights the need for a more efficient solution for explaining privacy. Results by Kreuter et al. [34], who found that privacy explanations are often not carefully read by users, support this notion. They attribute this unsafe behavior to the "considerable increased cognitive burden" that is placed on the readers of privacy explanations. It follows that privacy explanations need to be designed in a way that promotes user efficiency. By shortening

the time needed to navigate and read the explanation, it should be possible to decrease the cognitive burden on end-users. Related to this are findings from Cranor [21], which suggest that standardizing privacy explanations can improve their learnability. This way, end-users get used to the structure and navigation of the privacy explanations, increasing the efficiency of use over time.

Regarding the contents of privacy explanations, several goals that users might have when reading privacy explanations come to mind. Primarily, privacy explanations can be used to educate end-users about their privacy and thereby increase their privacy awareness. As shown in sections 2.1.1 and 2.1.7, privacy awareness is critical in promoting safe privacy behavior and addressing the privacy paradox. Furthermore, as indicated by the results of section 3.4, privacy explanations can give their users a sense of security, as they have a better understanding about what happens to their personal data. This is supported by Lin et al. [39], who found that "properly informing users of the purpose of resource access can ease users' privacy concerns to some extent".

Results of section 3.4 imply that some users see privacy explanations as a legal guarantee concerning their privacy. Privacy explanations need to apply to laws and regulations and are expected to be faithful and honest. The problem arising from this, is that while a privacy explanation might be technically correct and honest in its statements, it is not possible for end-users to actually check the validity of those statements. Following Kästner et al. [36], users would have no way to know if their trust in the explanation is warranted. Riegelsberger et al. [55] found that trust-supporting systems must be traceable and accountable. In the context of privacy explanations, this clearly implies a need for verifiability. That means that all statements made within privacy explanations need to be able to be checked for their correctness by end-users. If this is possible, end-users' would know that their trust in the privacy explanations is warranted.

Developer-focused Factors

The developer-focused factors are comprised of ethical and corporate interests. As was thoroughly discussed in section 2.1.1, privacy explanations can increase system transparency and, by extent, trustworthiness. Of course, this is heavily reliant on developers being honest and correct when providing privacy explanations. Both are ethical interests that might be important to developers.

Privacy explanations can also have an influence on developers' corporate interests. According to section 2.1.1, explanations can increase user trust. Findings by Zanker [71] also reveal a positive influence of explanations on a system's perceived value. Cramer et al. [20] showed that transparency, and

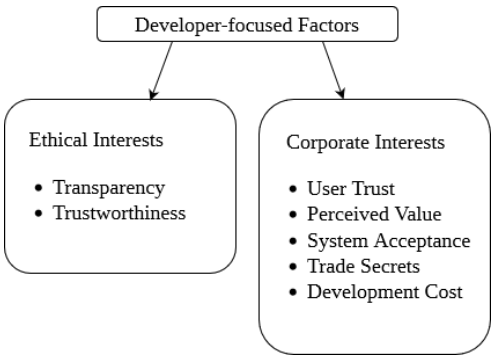

Figure 4.3: Developer-focused factors for privacy explanations

by extend explanations, positively influence system acceptance. Ultimately, these factors can lead to a higher customer retention, which would be in the corporate interest of the developers.

Providing appropriate privacy explanations also entails factors that might be undesirable to developers. As mentioned before, they can increase system transparency, which is not always to the benefit of the developers. If not handled carefully, they could reveal trade-secrets, endangering the security of sensitive corporate information [19, 26]. Furthermore, privacy explanations come with their own development cost, as they are provided in addition to privacy policies, not instead of them.

Independent Factors

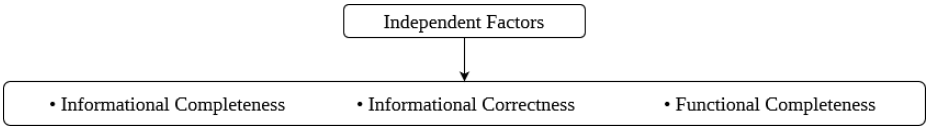

Figure 4.4: Independent factors for privacy explanations

Finally, the independent factors describe general qualities of privacy explanations that are important to both users and developers. To comply with all relevant law and legislature, privacy explanations require both informational completeness and informational correctness. This way they offer a legal guarantee to end-users, while serving as a proof of legal compliance for the developers at the same time.

In order to provide the operability of privacy explanations, described in the user-focused factors, and to provide general usability to the system, the interface that provides those privacy explanations requires functional completeness. This is needed by both users and developers, as an unusable system would be useless to both sides.

4.3 Types of Privacy Explanations

Section 2.1.2 highlighted the effectiveness of personalized explanations in XAI. Explanations can be personalized in different ways, using varying degrees of complexity. Several research works mentioned in section 2.1.1 raise concerns about the privacy of the addressees of personalized explanations [2, 53, 58]. To provide personalized explanations, the explainer must first collect and process information from the addressee, in order to learn about their preferences. In the context of privacy explanations, this creates a never-ending loop, as addressees of personalized privacy explanations would yet again require a personalized privacy explanation about the preference data collected from them. This goes on infinitely, until either no privacy explanation is offered to the users or no further personalization is being done. Figure 4.5 provides a visualization of this problem.

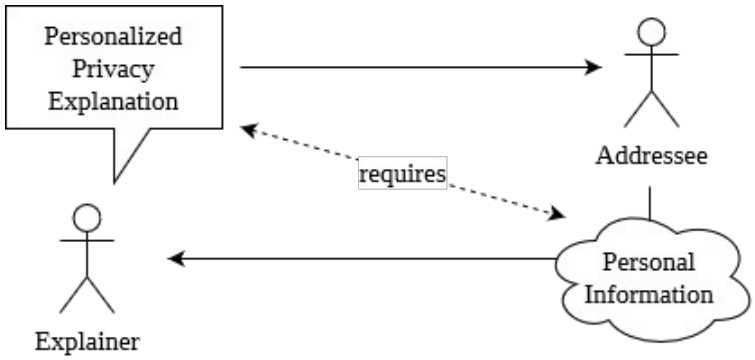

Figure 4.5: Loop of personalized privacy explanations

Sokol and Flach [61] argue that personalization should be left to the end-user. That could be realized by allowing those users to directly interact with the explanations themselves. Following Sokol and Flach [61], this approach does not only remove the need for complex computational processes, but would also make "the whole process feel more natural, engaging and less frustrating". Taking this into account, the chosen approach of this work is compartmentalization rather than complex personalization. In such an approach, the entirety of privacy information is divided into several compartments. These compartments house different types of privacy explanations, that explain the different aspects of privacy, and are presented to the users on separate pages. This way, users are able to view and explore all the different aspects of privacy information, allowing them to navigate between navigation explanation types on their own accord.

The idea of compartmentalization is also supported by Cranor [21]. They state that "A layered approach to privacy notices would make very simple notices readily available with links to more detailed notices". If privacy

explanations are provided within in a compartmentalized structure, the different pages can work akin to layers of complexity, displaying the different aspects of privacy. Furthermore, if appropriate, privacy explanations can offer links to external documents, such as relevant legislature, offering more details for interested users. On another note, compartmentalization solves the issue of the never-ending privacy explanation loop (see figure 4.5) that arises from personalizing privacy explanations, as no personal information needs to be collected from the addressees of the explanations.

In the following, five different types of privacy explanations are developed. Each explanation type provides only part of all necessary information on end-user privacy. Together, they contain all information necessary for users to fully understand the processing of their personal data, according to the GDPR [23], and achieve privacy awareness as described in definition 2.1.7.

Base-level Explanation

The first compartment contains the base-level explanation. This explanation type is fairly similar to the app privacy notices from the *Apple App Store*, discussed in section 2.2.2. It states the types of data that are going to be processed and the ways in which they will be used. Providing this information is not only necessary according to the GDPR's article 15 [23], but also plays a critical role in end-users understanding their privacy. Users will encounter the base-level explanation first, hence it needs to be understandable on its own, without any further context.

Data processing can be subdivided into two types of categories. The first category are the main functionalities of the software in question. For example, a travel guide app might want to access their users' location data in order to suggest close-by places of interest. The other category are secondary functionalities of the software, which can take advantage of data processing. This entails services like providing targeted advertising, creating profiles of the users or improving the software.

As some users might not want to read beyond the first explanation, it is important to mention system functions which will not work if the user decides to object to data processing. This way, they are able to make more informed decisions, as they get a better understanding about whether software can provide the desired functions without processing personal data. In summation, the base-level explanation needs to contain the following:

- What type of data is being processed
- How the data is primarily being processed
- Possible secondary uses of the data
- The consequences of denying data processing

Contrastive Explanation

As shown in section 2.1.2, contrastive explanations have proven to be an effective tool within the realm of XAI. Unlike explanations in XAI, privacy explanations do not explain the outcome of an algorithm. Hence, contrastive explanations cannot be directly applied. However, Waldman [66] used a surprisingly similar approach when they successfully tested tabular designs for privacy policies. One of the table columns contained "Info we DON'T collect". Unfortunately, it does not seem like Waldman [66] evaluated this particular part of their design.

Following Miller's [42] work on contrastive explanations in XAI, comparing reality to wrongful expectations improves understandability. This could be expected to hold true for privacy explanations as well. By telling end-users how their data is not being processed, we can help them understand how it is actually being processed, offering a point of comparison. Furthermore, dealing with users' wrongful expectations about data processing might ease their pre-existing doubts and fears towards the software. Since the contrastive explanation serves as a direct contrast to the contents of the base-level explanation, it should be placed within the second compartment. The contrastive explanation contains the following:

- How the data is not going to be processed

Example-based Explanation

Section 2.1.2 discussed the importance of example-based explanations in XAI. Perhaps most critically, Adadi and Berrada [1] stated that "Amongst agnostic methods, visualization is the most human-centered technique". Mirroring this in the context of end-user privacy, Brandtzaeg et al. [9] suggest that data flows of mobile apps should be visualized to offer more transparency. Given that privacy explanations should be designed with users in mind, visual examples may serve as a prime tool to explain privacy to end-users. More specifically, they could be used to show users what certain data types, which are collected from them, look like.

For example, the privacy explanation of a software system could state that the software processes its users' browser history to provide its services. However, there is no guarantee that every end-user knows what a browser history is or how it works. Throughout their research, Won et al. [70] "found that many people were not even aware that history exists". This is critical, as users might fail to understand what personal data is being processed, even though the privacy explanation tells them about it. In such a case, proper privacy awareness cannot be achieved. If provided with visual examples, users might get a better understanding of the data that is asked from them and how it will be used.

In accordance with Ortloff et al. [47], privacy explanations are context-sensitive. This especially applies to example-based privacy explanations. By taking context into account, privacy explanations are an opportunity to provide examples that are related to the actual functionalities of the software in question. For example, the privacy explanation of a travel guide app could display an example map of a city with indicators that show users the confines of the location data that the app wants to collect from them. Corroborating this example, Fu and Lindqvist [24] found that while end-users generally have a good understanding of what precise location data is, the same does not necessarily apply to approximate location data. Following Fu and Lindqvist [24], misunderstanding the confines of data processing may lead users into making decisions that do not match their actual privacy preferences. Contextual examples might help the issue, giving users a better idea of what data is being processed precisely. As the example-based explanation is closely tied to the contents of the base-level explanation, it is placed within the third compartment. It contains the following:

- A visualization of the data that is being processed

Explanation of further Details

The base-level explanation on its own does not contain all information that is legally required by the GDPR's article 15 [23]. It explains what data is being processed and why, but it is missing out on further details, such as how long the data is kept, who can access it and what rights are available to the user. Notably, how long the data is kept is directly linked to the point at which it is deleted or anonymized. This should also be specified within the explanation. Furthermore, users' rights depend on the responsible jurisdiction.

A separate explanation of further details offers an opportunity to convey this information to those users that are interested in it. At the same time, putting these details on another page keeps the base-level explanation from being cluttered with information. This could expectably help users that feel overwhelmed with long text documents such as privacy policies. In summation, the explanation of further details has to contain the following:

- Where the data is stored (important to determine jurisdiction)
- Who can access the data
- For how long the data is stored
- At what point is the data deleted or anonymized
- What rights are granted to the user

Third Party Explanation

According to the GDPR [23], end-users have a right to know who their data is being shared with. Section 2.3 highlights the importance of educating users about third party data sharing and the problems related to that. Third parties and their privacy-related practices have a history of being poorly communicated to end-users and privacy explanations offer an opportunity to improve this state of the art. In theory, it would be optimal for a privacy explanation to provide additional complete privacy explanations for each third party. However, this is not feasible, as it would blow the original privacy explanation out of proportion, rendering it unreadable.

Taking the findings of Balebako et al. [6] into account, there is a low understandability of third parties if users do not recognize the names of the respective companies. To counteract this, a third party explanation should list the intended data processing purposes of each third party. This is similar in design to the base-level explanation and allows end-users to infer how their data is being used. Of course, this only covers what types of data are being used, by whom they are used and how they use them. Additional information, such as the points covered by the before-described explanation of further details, would not be included.

By offering external links, users that are interested in further details are able to visit the respective third parties' privacy policies. Considering the findings of section 2.2.2, this is in no way an optimal solution, as privacy policies are not fit to communicate privacy information to end-users. Hopefully, in the future, if third party companies were to provide comprehensive privacy explanations, those could be linked in place of the privacy policies. Consequently, the third party explanation contains the following information:

- With which third parties the data is being shared
- Links to each third parties privacy policy (subject to change)
- How each third party uses the data

4.4 Prototype Development

In order to be practically evaluated, the concept of the different privacy explanation types was implemented within an interactive software prototype. The following section describes the prototyping software used to build the prototype - *Axure RP*. Furthermore, it explains the prototype's architecture and prominent design elements used within it. Lastly, the topic of immediate language, which also played a role in the prototypes design, will be discussed. Screenshots of each privacy explanation type found within the prototype are provided in the appendix B of this thesis.

4.4.1 Axure Prototyping

Axure RP is a competent prototyping tool developed by *Axure Software Solutions*. Access to the tool was granted via a student license in the context of the *Axure RP Academic Perks Program*[1].

As the focus of this thesis lies on the readability and understandability of privacy explanations, rather than on the interface's usability and optics, the prototype's design is kept rather simple. It is only built from *Axure RP*'s internal elements and no further frameworks were used. For the purpose of being presented in this thesis, the prototype was first developed in the English language. Afterwards, for the purposes of the user study, it was also faithfully translated to German. The translated prototype has the exact same contents as the original version.

4.4.2 Prototype Architecture

The prototype serves as an interactive platform for the different privacy explanation types. As such, the concept of compartmentalization is a key element of its architecture. It was also designed with the use on mobile systems in mind, somewhat akin to an app privacy notice. Figure 4.6 shows the prototype's first page and highlights its navigation elements.

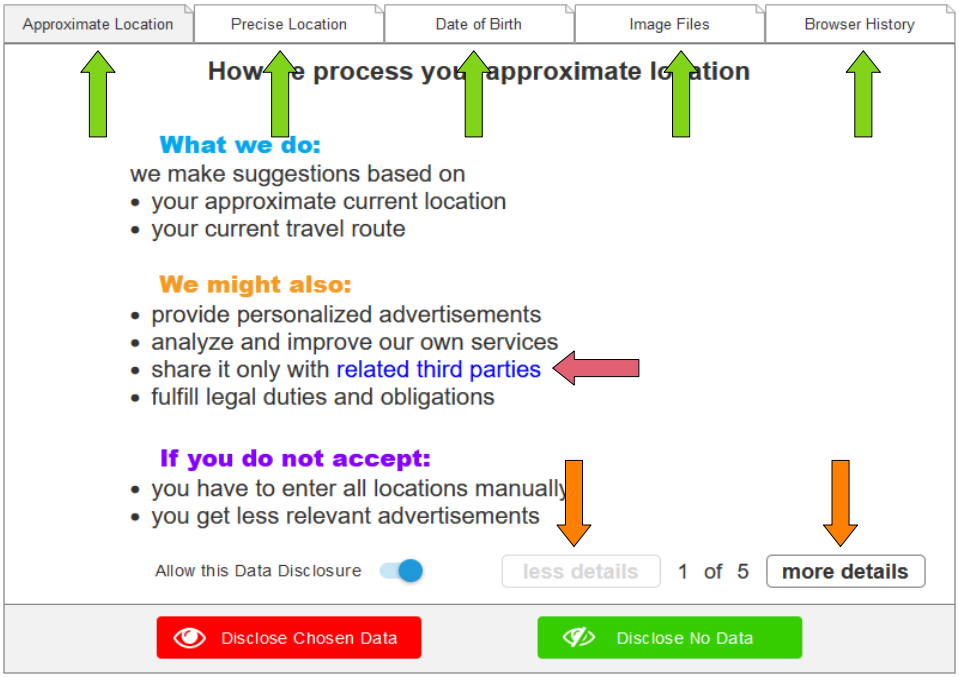

Figure 4.6: The software prototype's navigation elements

[1] https://www.axure.com/edu

The prototype offers privacy explanations for 5 different example data types, each of which comes with all the explanation types developed in section 4.3. The only exception is the date of birth data type, which does not include an example-based explanation. Users can navigate between the data types using the tabs at the top of the application (green arrows). This way, they can immediately see what kind of data is collected from them at one glance. Furthermore, this enables users to freely navigate between the data types in their own preferred order.

The buttons in the right bottom of the prototype (orange arrows) are used to navigate between the different explanation types, which are divided into the compartment pages described before. Users can switch to either the next or the previous compartment by pressing the buttons or by swiping over the page. Additionally, the first page, which contains base-level explanation, contains a direct link (pink arrow) to the last page, that houses the third party explanation. If the last page was accessed via this link, it will offer a button that leads back to the first page. A schematic visualization of the possible navigation within the prototype can be seen in figure 4.7

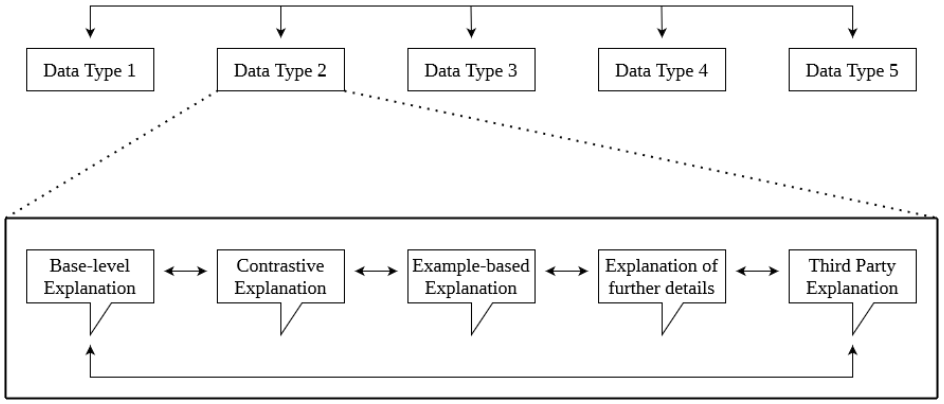

Figure 4.7: Navigation routes within the software prototype

4.4.3 Prominent Design Elements

The prototype includes a number of prominent design elements. These elements are not necessarily optimal in the context of usability, but are expected to invoke reactions from the participants of the user study. They are either based on insights from the literature discussed in chapter 2 or on the comments of the participants of the survey presented in section 3.4. Figure 4.8 highlights these design elements, using colored arrows.

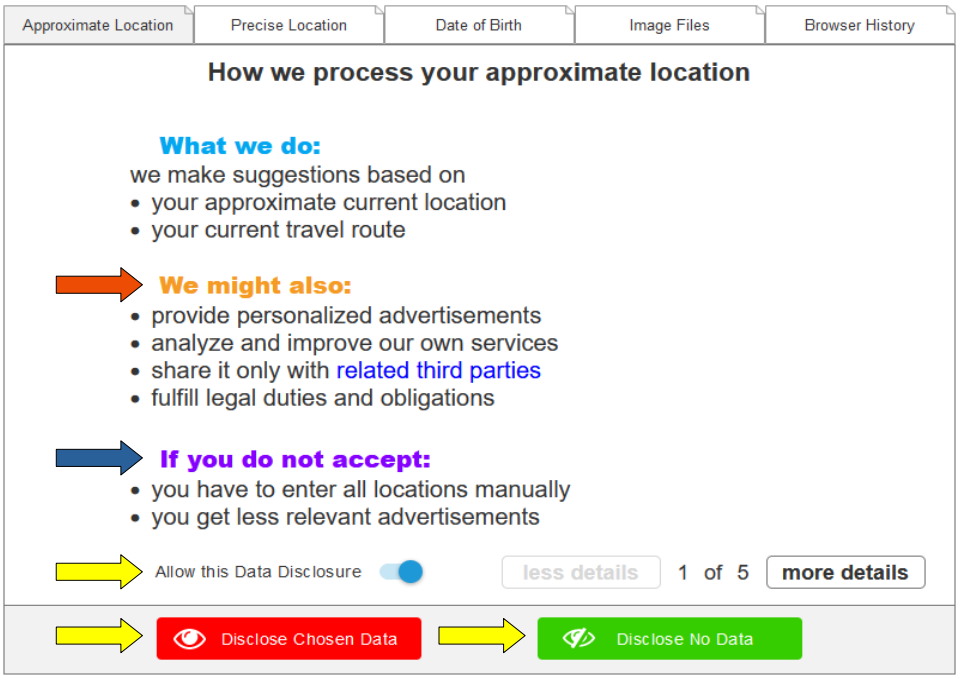

Figure 4.8: The software prototype's prominent design elements

Operability

As discussed in sections 2.4 and 3.4, operability is a feature that is wanted by many end-users, but currently missing from documents that contain privacy explanations. Cranor [21] declares that standardized privacy notice mechanisms "have failed users and they will continue to fail users unless they are accompanied by usable mechanisms for exercising meaningful choice and appropriate means of enforcement". Balebako et al. [5] found that users rarely remember the privacy information that is presented to them when they install the software. Adding operability to privacy explanations might mitigate this problem, as users are able to make informed privacy decisions while they are being educated.

The prototype contains these elements of operability (yellow arrows). For each single data type, there is a permission switch on the bottom left that controls whether or not the data type is allowed to be collected from the user. Once users have made their selection, they can conclude the process by pressing the bottom left button to disclose the chosen data to the software. In addition, the button on the bottom right enables users to disclose no data to the software at all. Using this button requires no interaction with data types' individual permission switches.

Existing notice-and-choice mechanisms notify users about the entirety of their data processing practices and then have them make their choice. In the

case of app privacy notices, users are not able to make individual choices, but can choose if they want to agree to all practices or to none. If they choose to share no data, they are unable to install and use the app. In the prototype, the individual permission switches allow users to make their choice while they are being notified about a specific type of data processing. Furthermore, they are able to make their choices regarding each data type individually. This allows users to exercise actual control over their personal data, rather than forcing them to decide between privacy and service [48].

Reverse Dark Patterns

Section 2.3 discussed the concerning existence of dark patterns in contemporary privacy related documents, such as cookie banners. The prototype developed for this work takes these dark patterns into account, either avoiding them or using reverse patterns, which could expectably invoke reactions from the study participants.

- Loss-gain framing is avoided by stating the disadvantages of not disclosing data, without judging them (blue arrow). The base-level explanation simply specifies which functions will not be available, enabling users to make an educated decision without being manipulated.
- The individual permission switches are pre-selected to *yes*. However, users have the option to opt out of all data processing in one click (yellow arrows, green button). This way, users can demand a bundled decline, which stands in stark contrast to the bundled consent buttons, which are prevalent in cookie banners. Only users that want to disclose data will be required to interact with the permission switches, unless they want to share everything.
- Within the prototype, the button to share no data is colored in green, while the button used to share data is red, which clearly indicates a hierarchy. However, this is a reverse pattern of false hierarchy, a dark pattern usually encountered in form of prominently highlighted bundled consent buttons. Employing this reverse pattern will expectably invoke reaction from users, who are aware of false hierarchy designs.
- There is no hidden information, as secondary data processing practices, that might be unwanted by the user, are clearly stated within the base-level explanation (brown arrow).
- Instead of a bundled consent button, a bundled decline button is available as a reverse pattern (yellow arrows, green button).
- Forced consent is not present within the prototype, as users have the option to opt out of all data processing.

Icons and Text Colors

The use of compartmentalization divides the entirety of privacy information into the five different privacy explanation types. However, the information provided within the individual explanations can be broken down further, into subcategories. For instance, as can be seen in figure 4.9, the explanation of further details is subdivided into topics like the location and duration of user data storage. As discussed in section 2.2.2, long text forms of explaining privacy have failed as a concept and in practice. To tackle this issue, the prototype provides explanations in subcategories and via bullet points. This way, the information is broken down as much as possible, expectably increasing its readability for end-users.

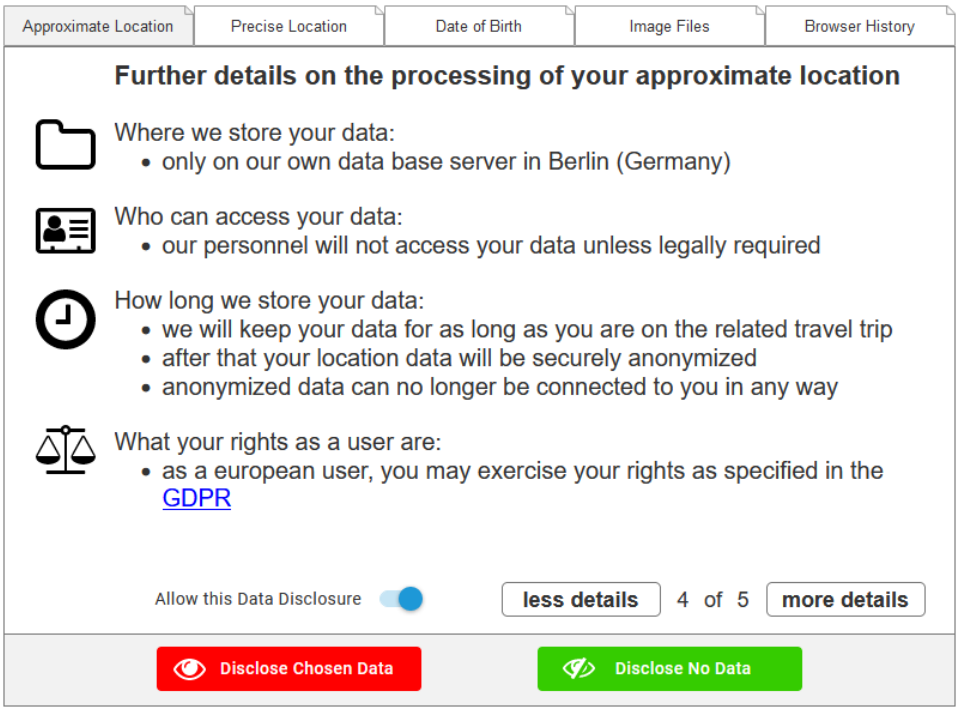

Figure 4.9: The explanation of further details and its icons

To highlight the subdivision of the explanations and provide easier orientation for users, the prototype uses both icons and text colors to distinguish the subcategories. Icons are used within the contrastive explanation and the explanation of further details, as well as on the red and green buttons at the very bottom of the prototype. They are rather simple in design and are expected to improve readability and understandability for the users. Bal [4] researched the effects of privacy indicator icons within the context of app privacy notices. They found that those icons improve the understandability of the privacy notices and promote safer privacy behavior.

Figure 4.9 shows the explanation of further details and its icons. Screenshots of the contrastive explanation and its icons are included in the appendix B of this thesis. Notably, all icons included in the prototype are part of *Axure RP*'s original icon library.

Instead of icons, the base-level explanation uses a variation of text colors for its three sub-headers: *What we do*, *We might also* and *If you do not accept*. This can be seen in figure 4.10. Currently, there is no sufficient research on the effects of text colors within privacy explanations. The use of different text colors could possibly help users with reading and understanding the explanations, but might also be unwanted or distracting to some. In order to understand the effect of text colors on the end-users of privacy explanations, they were examined in the context of the user study of this thesis.

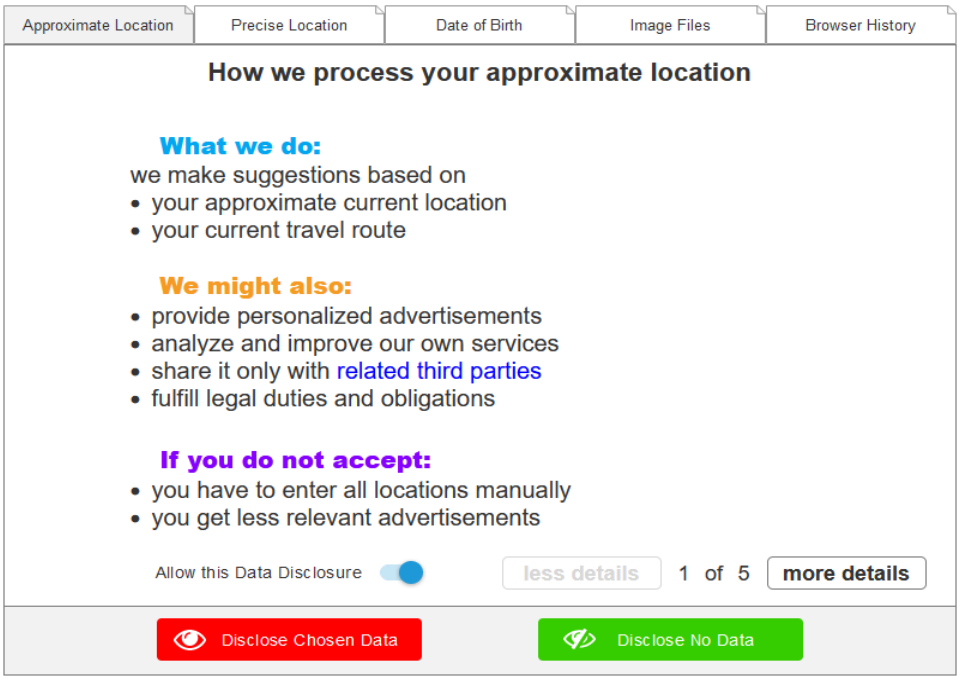

Figure 4.10: The base-level explanation and its text colors

Immediate Language

This work introduces the concept of immediate language to the topic of privacy explanations. The prototype was developed in two versions, using different kinds of formulations when addressing the users. Notably, the first version addresses users directly, using personal pronouns like *your data* or *your rights*. In contrast, the second version is formulated in a more indirect manner, using terms like *user data* or *users' rights*.

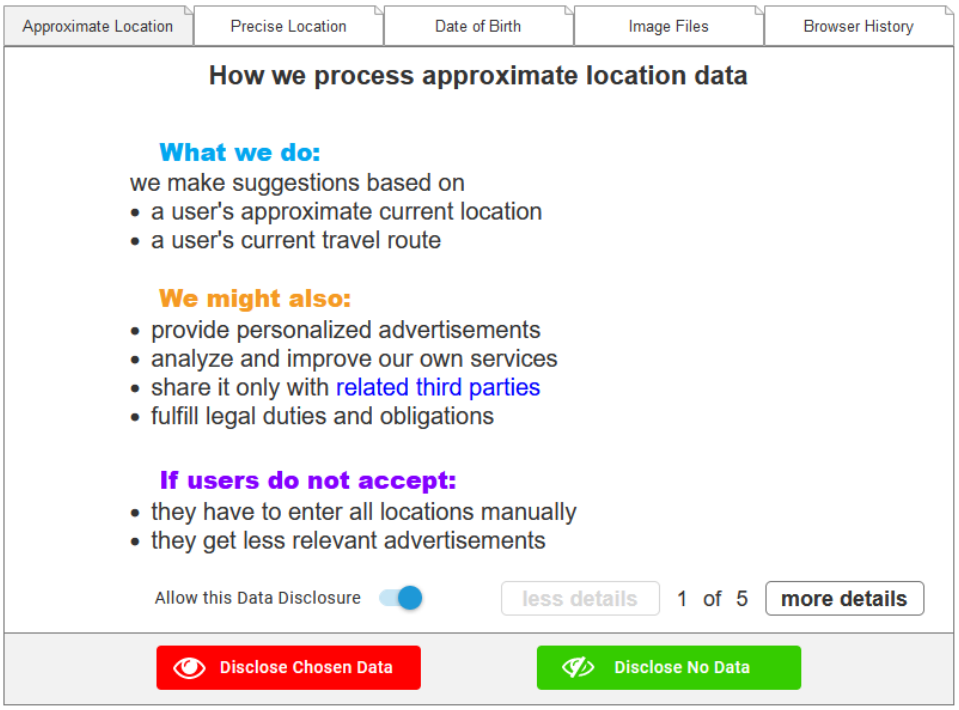

Figure 4.11: Indirect speech used within the prototype

For comparison, figure 4.10 shows the prototype addressing its user directly, while figure 4.11 shows the same page and explanation, using indirect speech instead. The goal of this design is to examine if users react differently when addressed directly. As follows from definition 2.1.1, explanations are interactions between explainers and addressees. At this point, privacy explanations are designed for humans by humans, so it is reasonable to assume that the language used by the explainer has an effect on the explanation's addressees. Baker [3], who researched the effects of immediate communication on students in online learning environments, supports this notion. They found that using immediate speech, such as addressing students by name or using their personal pronouns, has a positive impact on the cognition of those students. It stands to reason, that when addressing the end-users of privacy explanations, using immediate language, such as personal pronouns, might have a positive effect on their cognition as well.

Chapter 5

Evaluation

This chapter describes the user study conducted within the context of this thesis. First, the study goals are described, taking the research questions defined in section 4.1 into account. Then, the methodology of the study is explained. This includes the used technology and the overall design of the study and its structure. Following is an analysis of the participants' demography as well as their privacy awareness at the beginning of the study, before they interacted with the software prototype. Finally, the results of the study are presented and analyzed.

5.1 Study Goals

Chapter 4 describes the development of a concept for privacy explanations and its implementation within a software prototype. The goal of the user study was to evaluate this concept and its prototype, as well as testing their impact on end-users' privacy awareness. Doing this aligns with the research questions defined in section 4.1. All exercises and questions within the study's questionnaire were designed to answer those research questions. That includes an examination of the following topics:

Before using the prototype:

1. How privacy aware are the participants?

After using the prototype:

2. Did participants enjoy the privacy explanations' structure?
3. Did participants understand the privacy explanations' contents?
4. Did partaking in the study influence participants' privacy awareness?

5.2 Methodology

For the study of this thesis, semi-structured interviews were conducted either at the interviewer's workspace or via remote control. For the most part, the study followed a predefined set of questions. In some cases, participants were asked additional follow-up questions, to get a better understanding of their reasoning. The complete questionnaire is provided as described within appendix D of this thesis.

5.2.1 Study Design

This study consisted of three main sections. The first questions were concerned with the participants' demography and their privacy awareness at that point in time. Then, participants navigated the prototype and examined its explanation types. The hypothetical scenario, in which the prototype was used, is described in detail in the appendix B of this thesis. Afterwards, they answered questions concerning the prototype, as well as questions concerning the study's influence on their privacy awareness. Notably, the user study was first pilot tested with two members of the research staff of the *Software Engineering* group at *Leibniz University Hannover*. After the pilot testing, minor adjustments to the questionnaire and prototype were made.

Within the study, participants were asked to answer the questionnaire in a think-aloud manner. This includes, that they were supposed to explain their way of thinking and give reasons for their answers. The interviewer was present throughout the entirety of all study sessions. If participants did not give reasons for some of their answers, they were verbally reminded to do so. Furthermore, if they made unclear or unexpected remarks, they were subject to follow-up questions by the interviewer. All interviews were recorded with consent of the participants.

This resulted in about 50 hours worth of video material, which was then fully reviewed and coded. Coding consisted of 2 cycles, the first of which was *in vivo* coding, following the method of Saldaña [57]. Participants' comments and remarks were scanned for contents related to the research questions of this thesis. Notably, they had to be translated from German to English. In the second coding cycle, the results of the *in vivo* coding were broken down further, using pattern coding as described by Saldaña [57]. The pattern strictly follows the research questions, meaning that the codes can be divided into the following four categories:

1. Privacy Explanation Structure
2. Privacy Explanation Contents
3. Privacy Explanation Types
4. Privacy Awareness

The results of these two coding steps are provided as described within the appendix D of this thesis.

The focus of this study is the evaluation of the developed concept for privacy explanations. Its purpose is not to evaluate the usability of the prototypical implementation, which is only used as a platform to present the explanations. Therefore, usability tests will not be part of this study. Instead, the participants' subjective experiences and perceptions are the metrics used to answer the research questions.

5.2.2 Technology

Due to the ongoing *Covid-19* pandemic, not all of the interviews could be done in person. As the interaction with the software prototype required participants to have access to the interviewer's computer system, remote control software and voice chat applications had to be used.

Remote Control Software

AnyDesk[1] is a remote control software, that allows end-users to control another's device. Within the study, remote participants used *AnyDesk* to fill out the questionnaire and navigate the software prototype.

Voice Chat Applications

Remote participants were allowed to use a variety of voice chat applications, depending on their personal preferences. This was needed to allow the interviewer to communicate with the participants, and for them to be able to answer the questionnaire in a think-aloud manner. The utilized voice chat applications were *BigBlueButton*[2], *Skype*[3], *Discord*[4] and *Facetime*[5].

Recording Software

OBS Studio[6] was the recording software of choice for this user study. It was used to record the voices of the participants and the interviewer, as well as the computer screen on which the study's questionnaire and prototype interaction took place.

[1] https://anydesk.com
[2] https://bigbluebutton.org
[3] https://skype.com
[4] https://discord.com
[5] https://apps.apple.com/de/app/facetime/id1110145091
[6] https://obsproject.com

5.3 Participants' Demography

In the following section, the study participants' demography will be described and analyzed. On one hand, this includes general demographic factors, such as age or education. On the other hand, participants' privacy awareness at the beginning of the study, before they interacted with the software prototype, will be subject to examination.

General Demography

Participants were recruited from personal acquaintances, their friends and from among the research and education staff of the *Software Engineering* group at *Leibniz University Hannover*. A total of 61 participants partook in the study, 38% of which identified as female and 62% of which identified as male. None of them identified themselves with non-binary gender. All participants answered the questionnaire in German.

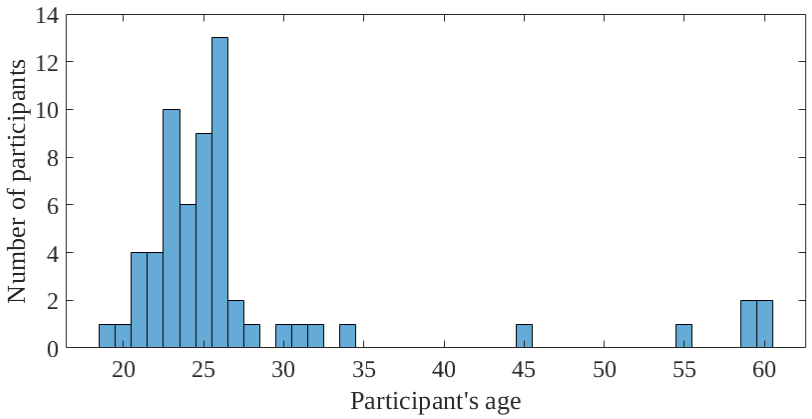

Figure 5.1: Participants' age distribution

All of the participants were at least 18 years old, with the median age being 25. Following the definition used by Chazette and Schneider [17], 90% of the participants were so-called digital natives, as they were born after 1980. The other 10% were digital immigrants, being born before that point in time. Figure 5.1 shows the participants' age distribution.

As for their highest formal education, 28% of participants had a high school degree and 15% had completed an apprenticeship or something similar. The remaining participants all had academic degrees. 36% had a bachelor's degree and 20% had a master's degree. Only one participant had a doctor's degree.

As for their frequency of Internet use, 46% of participants stated that they spend more than 5 hours of their free time on the Internet each day. 26% each said that they spend between three and five hours or between one

and three hours online. Only one participant claimed that they spend less than an hour a day of their free time on the Internet.

Every participant stated that they use software within their workspace. For those who were students, the university was considered to be their workspace. 89% of participants use software within their workspace on a daily basis. 8% said they use software at work every week and only 3% claimed to do so less frequently.

Many participants were also active on social networks. Only 13% claimed to never use social networks. 51% said that they were active on one or two, while 28% used between three and five. 8% even stated to use more than five social networks. Moreover, 59% of all participants said that they only share personal data with their friends on social networks. However, 24% stated that they at least sometimes share data with strangers. The remaining 17% do not use social networks at all or only use them anonymously.

In summation, the participants' demography is rather young on average, the vast majority of them being digital natives. A large number of them had an academic background. Furthermore, most of the participants use software in their workspaces and spend a considerable amount of their free time online. They are also active on social networks. From these facts, it is reasonable to assume that the majority of participants is probably well educated and technologically literate.

Participants' Privacy Awareness

Within the first half of the questionnaire, all 61 participants answered a couple of questions concerning their privacy awareness at the beginning of the study. In the following, the results of these questions will be supplemented with results from the coding of the participants' verbal comments.

First, they were asked how important the protection of their personal data is to them in general. The following questions examined how important their awareness of data processing practices is to them, depending on the place of data processing. Figure 5.2 shows the results of these questions.

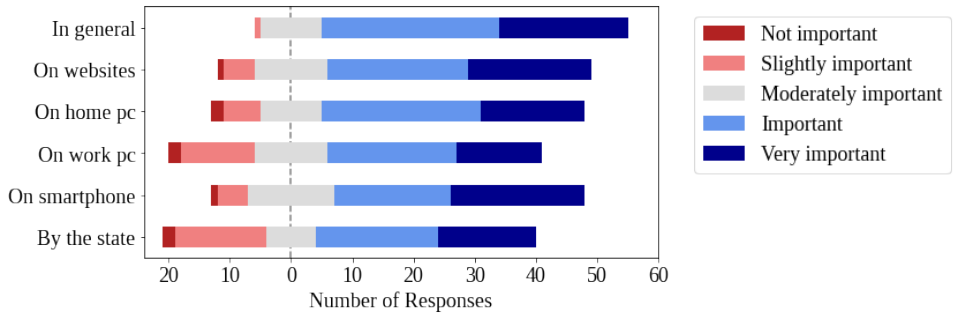

Figure 5.2: Privacy of personal data in different places of data processing

In general, an overwhelming number of participants stated that the protection of their personal data is important or very important to them. However, they were able to differentiate between different places of data processing. In particular, data processing on work place computers, as well as data processing by state institutions appeared to be less of a concern, in comparison. Notably, 23% of participants explicitly stated that privacy in their personal sphere is more important than privacy at their work place. Furthermore, 21% stated that they trust state institutions with their personal data.

Participants did not only differentiate between the places of data processing, but also between different types of personal data. They were asked about how important their awareness of data processing practices is to them, depending on the type of personal data. The results of these questions can be seen in figure 5.3.

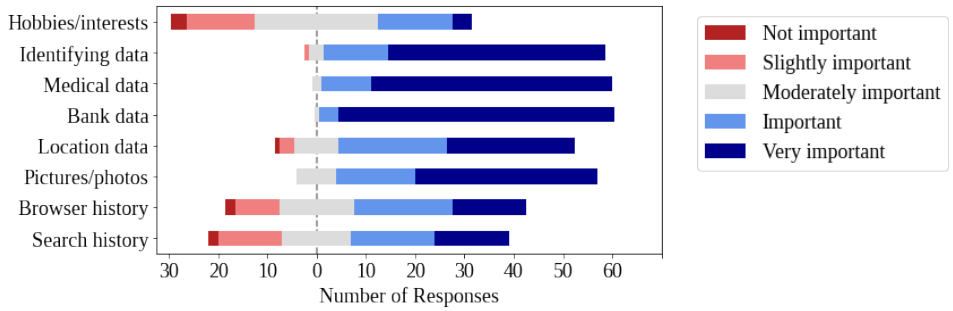

Figure 5.3: Privacy of personal data depending on the data type

For these findings, it becomes clear that the participants had a distinguished view on the importance their personal data, depending on the data type. Things such as identifying data, medical data, bank data and photos were of high importance. These types of data were seen as particularly vulnerable and five participants specifically mentioned their fear of criminal activity as a factor for their answers. The other data types, namely interests, location data, browser and search history were not as critical for the participants. Especially in the context of these low priority data types, 16% of participants stated that they have nothing to hide in the first place. On another note, 38% mentioned that they consciously allow the processing of those data types depending on if the trade-off between privacy and service is worth it. This is corroborated by the findings of Pentina et al. [48], who presented the concept of privacy calculus (see section 2.4), which supports this kind of transactional end-user behavior.

Underlining these observations, is the participants' stance towards the sharing of their personal data with third parties. While they were not fond of third party data sharing, the vast majority of participants said that they were fine with it, as long as it is necessary in order to use the software. This

can been seen in figure 5.4.

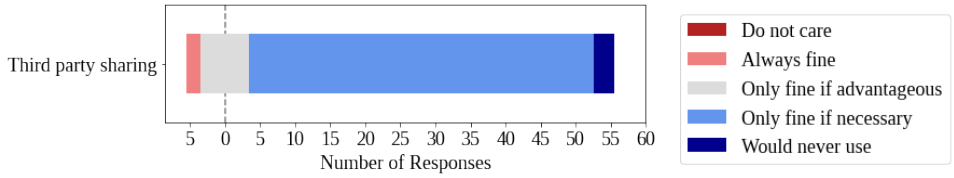

Figure 5.4: Acceptance of third party data sharing

Notably, none of the 61 participants were completely unaware of the ubiquity of third party data sharing. 11% stated that they did not explicitly know about these practices, but that they expected them to be happening. The remaining 89% said that they were well aware of their data being shared with third parties.

Despite all participants being EU citizens, not all of them were aware of the GDPR. Only 5% had read major parts of the regulation. 46% knew at least some of its contents and 36% stated to know what the GDPR is, but none of its specific contents. 11% had at least heard the name before, but the remaining one participant did not know of it at all.

Taking these results into account, it appears that the participants are generally aware of their privacy. They want their personal data to be safe and they want to be notified if their data is being processed. This is especially true for sensible data such as identifying information or photos. However, many participants also consciously use their personal data as a resource and trade it for services that they desire.

5.4 Findings

After the demographics questions, but before using the prototype, there was another set of questions. Participants were asked about their opinion on two documents that contain privacy explanations, namely privacy policies and end-user agreements. In the following, both of these documents will be addressed as *privacy agreements*. For the first question of this section, participants stated what they think the purpose of these privacy documents is. They were able to give multiple answers. None of the 61 participants stated that the documents serve no purpose at all. With 97% agreement, almost every participant said that privacy agreements serve as a legal insurance for companies, but only 49% thought that they did the same for end-users.

Concerning the informative purposes of the documents, 8% of participants claimed that privacy agreements use obscure designs to confuse end-users on purpose, only pretending to protect them. However, 69% stated that privacy agreements can be used to educate end-users about their privacy.

Despite that, when asked about their privacy agreement reading habits, 54% of participants admitted that they do not read them at all. 18% only read the parts relevant for them and 26% said that it depended on the software in question. Only one participant stated that they tend to read the documents in full.

Judging from these findings, it appears that a majority of participants often does not read privacy agreements, even though they find them to have an educational purpose. This coincides with the privacy paradox phenomenon, which was discussed in section 2.4. In the following questions, participants were asked to give reasons for why they do not read privacy agreements in full or at all. Yet again, they were able to give multiple reasons. Considering the findings of section 2.2.2, it is not surprising that 95% of participants found the documents to be too long to be read. Moreover, 26% said that they had problems with the language used and 28% did not understand the explanations within the documents. Surprisingly, 46% of participants said that they had no interest in privacy agreements and 13% said that they do not trust them.

Participants were also asked to gauge the readability of privacy agreements, and express how often they have trouble reading them. Figure 5.5 shows the results of this exercise.

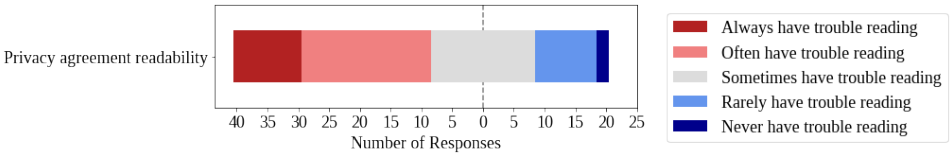

Figure 5.5: Perceived readability of privacy agreements

These results look rather grim, with the majority of participants either always or often having problems when reading privacy agreements. Considering the demography of the study participants, they look even worse, as the majority of them are digital natives and most likely technologically literate. Throughout the study, 20% of participants mentioned a feeling of privacy apathy. They stopped interacting with privacy explanations on their own accord, even though they are generally interested in their privacy. 15% of participants mentioned that they feel like they do not really have a choice and are forced into sharing their data. These sentiments are likely linked to the bad readability of privacy agreements. End-users expect these documents to be hardly readable in the first place, so they have stopped interacting with them.

5.4.1 Privacy Explanation Structure

After interacting with the privacy explanation prototype, participants answered questions regarding the privacy explanations' structure and its contents. Furthermore, they made several remarks regarding privacy explanations while and after interacting with the prototype. As described in section 5.2.1, these remarks were coded and evaluated with regards to the research questions.

Section 4.3 developed the concept for privacy explanations in a compartmentalized manner. Additionally, instead of a long text form, the explanations were broken up into subcategories and presented in the form of bullet points. This was done under the hypothesis that breaking up and highlighting the information could help users understand that information better. After all, long text documents like privacy policies have proven that they are not fit for this purpose (see section 2.2.2).

Count and Order of Privacy Explanations

The prototype included five example data types, each of which housed five pages that included the different privacy explanation types. Only the date of birth data type had four pages, as it had no example-based explanation. This results in a total of 24 pages with privacy explanations. The order of these pages was described in 4.3. Within the questionnaire, participants answered questions concerning the count and order of the explanations. In both cases, they answered on a 5 point Likert scale (5 - Very good to 1 - Very poor). Figure 5.6 shows how participants answered regarding privacy explanation count and order.

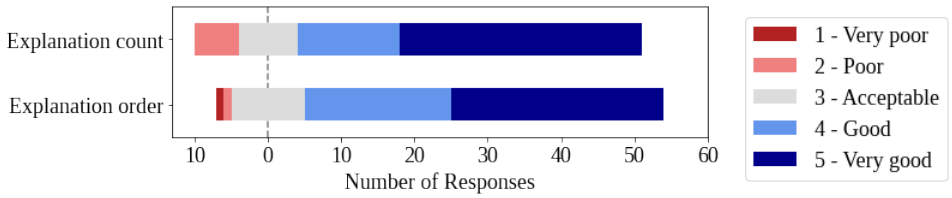

Figure 5.6: Evaluation of count and order of the privacy explanations

The descriptive statistics used for the qualitative analysis of data are mean M and standard deviation SD. Unless stated otherwise, this will be the norm within this work. The findings for privacy explanation count ($M = 4.21$, $SD = 1.01003$) indicate that the participants did not feel overwhelmed with the number of explanations. Similarly, the order the privacy explanations ($M = 4.23$, $SD = 0.89431$) was perceived as overall good. Unfortunately, at this point, there is no point of comparison for these findings, which is why they cannot be tested for statistical significance and have to stand on their own.

Icons and Text Colors

Section 4.4.3 described the prominent design elements used within the prototype. As part of the questionnaire, participants were specifically asked if they noticed the included icons and text colors, and what they thought about them. Notably, 13% of participants did not notice the icons and 11% did not notice the text colors. Figure 5.7 shows the opinions of the participants that noticed the icons and text colors.

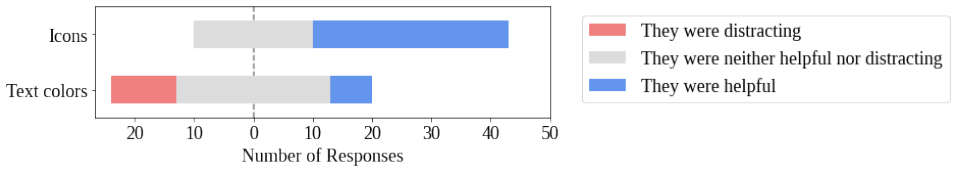

Figure 5.7: Evaluation of icons and text colors

The icons were well received by the participants, with the majority finding them helpful. None of the participants said that they found the icons distracting. However, the same is not true for the text colors. While the relative majority had no strong opinion on them, there were more participants that found them distracting, rather than helpful.

Length of Privacy Explanations

Participants also gave their opinion on the lengths of the different privacy explanations. In particular, they rated the length of each different explanation type on a bidirectional 5 point scale (Far too much to Far too little). The results of this are visualized in figure 5.8.

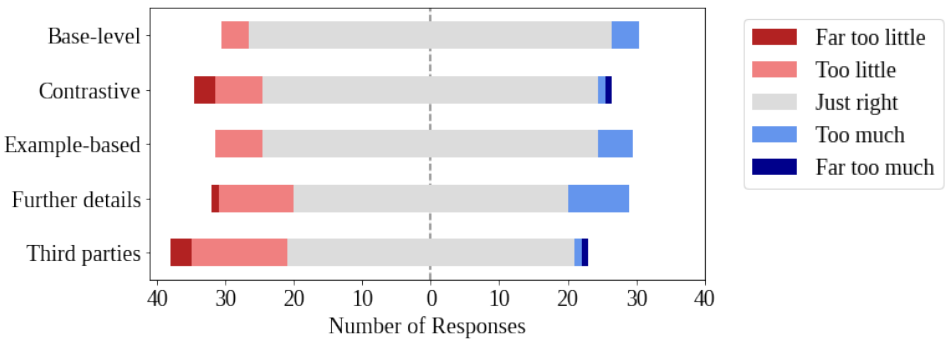

Figure 5.8: Perceived length of the privacy explanations

Overall, it seems like the length of the privacy explanations was well received by the participants. Especially the base-level explanation, the example-based explanation and the explanation of further details appear

to be well-balanced. However, the results for the contrastive explanation and the third party information show a tendency of being too short.

Participants' Remarks concerning Privacy Explanation Structure

By examining the coding of participants' remarks and comments throughout the study, it is possible to gain further insights on their preferences regarding privacy explanation structure. 26% of participants highlighted the importance of keeping privacy explanations short. Underlining this, 11% stated that repetitive information needs to be avoided. In contrast, four of the 61 participants stated that privacy explanations should not be too short. Otherwise, it could seem like the explanations are hiding something from the users.

39% of participants stated that they enjoyed the compartmentalized approach of the privacy explanations. Indeed, they explicitly described them as well-structured. These remarks suggest that breaking up the privacy information was successful in increasing participants' understanding of the explanations. In the same context, 33% of participants mentioned that the navigation of privacy explanation interfaces needs to be intuitive. 23% said that they appreciate the availability of useful links (internal and external) within privacy explanations. 11% wanted a professional look for privacy explanations. This included qualities such as the privacy explanation interface using modern design frameworks and it not being too colorful.

41% of participants mentioned the necessity of operability within privacy explanations. 31% stated that consent management within privacy explanations needs to be intuitive and easy to use. In contrast, 16% demanded the operability to be more precise than just choosing whether or not a certain type of personal data may be processed. They wanted the consent management to be more complex and precise, allowing them to manage data processing with regards to individual system functionalities.

Another point of contention were the reverse dark patterns used within the prototype. 26% of participants welcomed the availability of a prominent bundled decline button. However, 15% thought that dark patterns need to be avoided completely, instead of being reversed. In particular, they perceived the coloring of the consent management buttons as a false hierarchy.

5.4.2 Privacy Explanation Contents

The contents of the privacy explanations were tested between-groups. As described in section 4.4.3, the software prototype existed both with immediate and with more indirect phrasing. Participants were randomly assigned to either one of the two prototypes. Out of the 61 participants, 29 saw the prototype with the immediate phrasing. The other 32 saw the prototype that used the more indirect speech.

Liking of Content

Participants were asked to gauge their general liking of the privacy explanations. The two-tailed null hypothesis regarding the liking of the privacy explanations is constructed as follows:

> H_0: There is no significant difference between immediate and indirect speech, when it comes to the liking of privacy explanations.

Concerning their general liking of the privacy explanations, participants answered on a 5 point Likert scale (5 - Very good to 1 - Very poor). Statistical significance was tested at a 5% significance level, using the Mann-Whitney U test. As this test is non-parametric, it does not require the populations to be normally distributed [60]. The test resulted in a Z-ratio of 0.09387 and a p-value of 0.92828. Therefore, the null hypothesis is not rejected at the 5% significance level. This indicates that the type of phrasing did not significantly influence participants' liking of the privacy explanations. The following qualitative analysis is visualized through figure 5.9. To enable an easier comparison between the test groups and the sum of all participants, figure 5.9 uses percentages instead of absolute numbers.

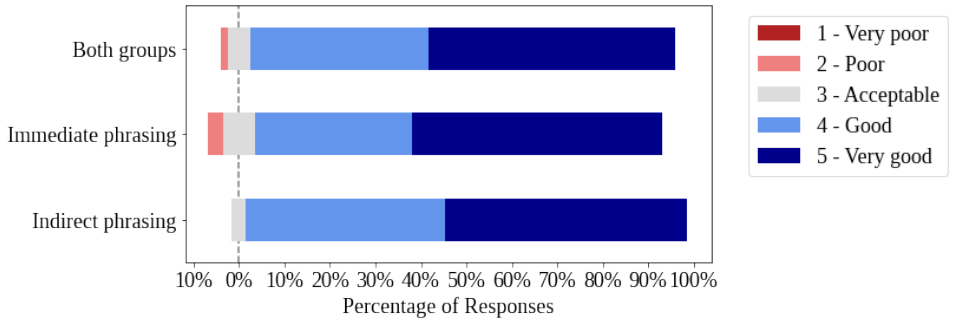

Figure 5.9: General liking of the privacy explanations

Judging from the results, participants generally liked the contents of the privacy explanations ($M = 4.46$, $SD = 0.66711$). Notably, there was no significant difference between the group that saw the immediate phrasing ($M = 4.41$, $SD = 0.76642$) and the group that saw the indirect phrasing ($M = 4.5$, $SD = 0.55902$).

Credibility of Content

Participants also evaluated the credibility of the privacy explanations. The corresponding two-tailed null hypothesis is built as follows:

H_0: There is no significant difference between immediate and indirect speech, when it comes to the perceived credibility of privacy explanations.

Similar to their liking of privacy explanations, participants gauged the credibility on a 5 point Likert scale (5 - Very good to 1 - Very poor). Statistical significance was tested at a 5% significance level, using the Mann-Whitney U test. For the perceived credibility, the Z-ratio was -0.38993 and the p-value was 0.69654. Consequently, the test failed at rejecting the null hypothesis at the 5% significance level. Apparently, participants' perception of the privacy explanations' credibility was not significantly influenced by the type of phrasing. The results of the qualitative analysis are contained in figure 5.10. Similar to figure 5.9, figure 5.10 uses percentages for easier comparison.

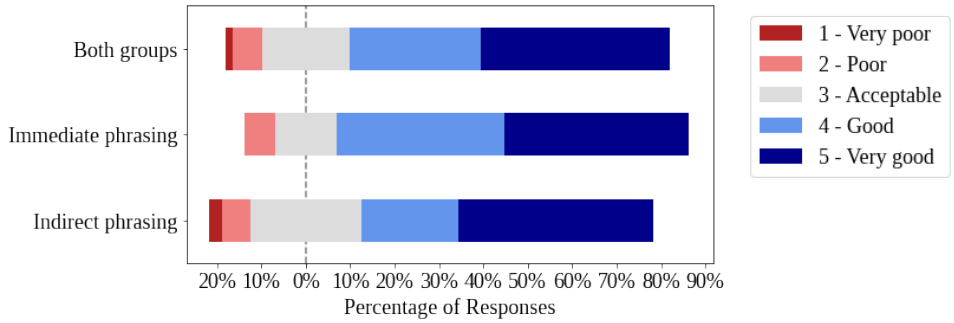

Figure 5.10: Perceived credibility of the privacy explanations

From the looks of it, participants found the credibility of the privacy explanations to be good ($M = 4.05$, $SD = 1.01507$). As with the liking of the explanation, the results of the group that saw the immediate phrasing ($M = 4.14$, $SD = 0.8992$) and the group that saw the indirect phrasing ($M = 3.97$, $SD = 1.10353$) did not differ significantly.

Participants' Remarks concerning Privacy Explanation Contents

In addition to answering these questions, participants gave their own thoughts on their desired contents of privacy explanations. In particular, 48% mentioned that unclear and imprecise wording needs to be avoided. Furthermore, 23% stated that misunderstandings and perceived contradictions can be a problem in privacy explanations.

13% of participants said that privacy explanations build user trust and 39% expressed that they help users understand their privacy. Concerning transparency, five of the 61 participants specifically mentioned it as a

requirement to privacy explanations. In a similar vein, 38% of participants highlighted the importance of honesty from the side of the developers, especially when it came to things that might have a negative impact on their privacy. However, 13% did not this was sufficient, and they demanded the contents of privacy explanations to be verifiable by the user.

5.4.3 Privacy Explanation Types

The privacy explanation types were tested within-groups. This means that every participant had access to all five privacy explanation types and was asked to evaluate them in relation to each other.

Relevancy of Privacy Explanation Types

First, participants were asked to evaluate the relevancy of the different explanation types. The base-level explanation covers the most basic privacy information (namely data type and processing purposes) and its contents are comparable to those of an app privacy notice. Hence, it is used as the point of comparison for the other explanation types. The two-tailed null hypothesis concerning explanation type relevancy was formulated as follows:

> H_0: There is no significant difference between the base-level explanation and the other privacy explanation types concerning their relevancy.

Consequently, 4 different cases were examined and evaluated:

- $R_{contrastive}$: the compared relevancy of the contrastive explanation
- $R_{example}$: the compared relevancy of the example-based explanation
- $R_{details}$: the compared relevancy of the explanation of further details
- $R_{thirdparties}$: the compared relevancy of the third party explanation

Participants answered on a 5 point Likert scale (5 - Very important to 1 - Not important). All statistical tests were done with a 5% significance level, using the Wilcoxon Signed-Rank test. Similar to the Mann-Whitney U test, this test is also non-parametric and requires no normal distribution of the populations [60]. Table 5.1 includes the test cases and their corresponding z-values and p-values.

The results of every test succeeded in rejecting the null hypothesis. Participants perceived every other privacy explanation type to not be as relevant as the base-level explanation. In all four cases, the p-values indicate statistical significance under the 5% significance level. Figure 5.11 shows the qualitative analysis of the participants' answers.

Table 5.1: z-values and p-values for the relevancy of privacy information

Case	z-value	p-value
$R_{contrastive}$	-3.5797	.00034
$R_{example}$	-6.3931	p < .00001
$R_{details}$	-3.9719	.00008
$R_{thirdparties}$	-3.8097	.00014

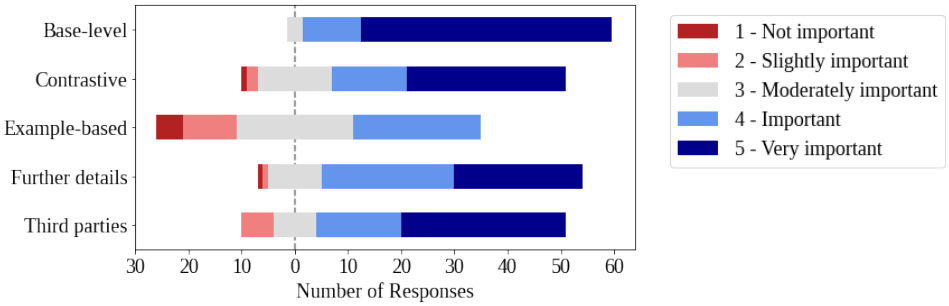

Figure 5.11: Perceived relevancy of privacy explanation types

Participants clearly thought that the base-level explanation was the most important to them ($M = 4.72$, $SD = 0.54716$). All other explanation types ranked significantly lower. With almost equal results for their means, the contrastive explanation ($M = 4.15$, $SD = 0.98122$), the explanation of further details ($M = 4.15$, $SD = 0.86529$) and the third party explanation ($M = 4.18$, $SD = 1.00013$) came in second. The example-based explanation was clearly perceived to be the least important ($M = 3.07$, $SD = 0.93859$).

Understandability of Privacy Explanation Types

Participants were also asked to gauge their understandability of the different explanation types. Yet again, the base-level explanation was used as the point of comparison. The resulting two-tailed null hypothesis concerning explanation type understandability looks as follows:

> H_0: There is no significant difference between the base-level explanation and the other privacy explanation types concerning their understandability.

Consequently, 4 different cases were examined and evaluated:

- $U_{contrastive}$: the compared understandability of the contrastive explanation

- $U_{example}$: the compared understandability of the example-based explanation
- $U_{details}$: the compared understandability of the explanation of further details
- $U_{thirdparties}$: the compared understandability of the third party explanation

Similar to the relevancy of the privacy explanation types, participants answered on a 5 point Likert scale (5 - No problems understanding to 1 - Did not understand at all). All statistical tests were done with a 5% significance level, using the Wilcoxon Signed-Rank test. The resulting z-values and p-values are shown in table 5.2.

Table 5.2: z-values and p-values for the understandability of privacy information

Case	z-value	p-value
$U_{contrastive}$	-0.3823	.70394
$U_{example}$	-2.334	.0198
$U_{details}$	-0.0284	.97606
$U_{thirdparties}$	-2.5732	.01016

Only the results of $U_{example}$ and $U_{thirdparties}$ successfully reject the null hypothesis at the 5% significance level. This indicates that participants found the understandability of the example-based explanation and the third party explanation to be significantly different than that of the base-level explanation. $U_{contrastive}$ and $U_{details}$ did not succeed in rejecting the null hypothesis. Hence their perceived understandability did not significantly differ from that of the base-level explanation. The results of the qualitative analysis can be seen in figure 5.12.

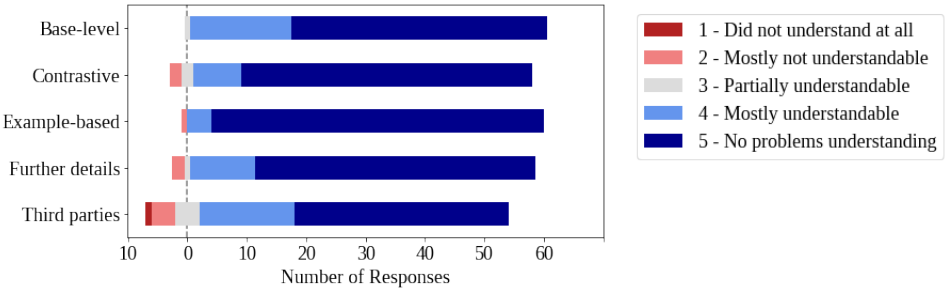

Figure 5.12: Perceived understandability of privacy explanation types

The base-level explanation seemed to pose no severe problems to the participants ($M = 4.69$, $SD = 0.49724$). The vast majority of them

claimed to have understood at least most of the information contained within it. From among the statistically significant results, the example-based explanation achieved an even better understandability with the participants ($M = 4.89$, $SD = 0.44715$). However, the third party explanation was perceived as significantly less understandable than the base-level explanation ($M = 4.34$, $SD = 0.97317$). The understandability of the contrastive explanation ($M = 4.7$, $SD = 0.68579$) and the explanation of further details ($M = 4.69$, $SD = 0.66631$) did not differ significantly from that of the base-level explanation.

Participants' Remarks concerning the Privacy Explanation Types

Throughout the interview, participants made various remarks concerning the privacy explanation types. For instance, even though the example-based explanation ranked last when it came to perceived significance, 43% of the participants explicitly highlighted the importance of providing helpful examples. In contrast, 15% mentioned that example do not necessarily need to be provided.

Concerning the contrastive explanation, 30% of participants made remarks that underlined the usefulness of addressing pre-existing privacy worries of end-users. However, 18% stated that providing contrast is not necessarily needed, especially if users are able to infer it themselves.

Six of the 61 participants, commented on the base-level explanation, stating that they are interested in knowing what exact functions they are missing when they do not allow the processing of a specific data type. Finally, 25% of participants said that they demand complete privacy explanations, missing none of the privacy information present in the prototype.

5.4.4 Privacy Awareness

Part of this study was also to examine whether or not privacy explanations can influence end-users' privacy awareness. This is not a trivial task, as there are no metrics to easily measure privacy awareness. Furthermore, an increased privacy awareness is really only useful if the effects are long-term. However, the questionnaire contains a number of questions that can possibly be linked to privacy awareness. Moreover, participants made some remarks and comments concerning their privacy awareness, which will also be subject to closer examination.

Effects of the Privacy Explanations

For starters, after navigating the prototype, participants were asked to evaluate how they felt after using the prototype. More specifically they were asked if the privacy explanations had an impact on the following factors F_i:

- $F_{security}$: The perceived security of their data
- $F_{privacy}$: Their perceived privacy
- $F_{trustworthiness}$: The perceived trustworthiness of the software that is described in the privacy explanations
- $F_{readiness}$: Their readiness to use the software

The effects on these factors were tested between-groups. Again, the participants were split by whether they saw the prototype with the immediate phrasing or that with the more indirect phrasing. The two-tailed null hypothesis is formulated as follows:

> H_0: There is no significant difference between immediate and indirect speech, when it comes to the perceived factors F_i.

Participants answered whether there was an effect, and if it was positive or negative. As the data is nominal, a chi-square test was used at the 5% significance level. Table 5.3 shows the results of the statistical tests.

Table 5.3: chi-square statistics and p-values for the effects of the privacy explanations

Case	chi-square statistic	p-value
$F_{security}$	1.0603	.588521
$F_{privacy}$	2.3247	.312743
$F_{trustworthiness}$	1.6183	.445241
$F_{readiness}$	2.4161	.298785

In all cases, the tests fail to reject the null hypothesis at the 5% significance level. It appears that the type of phrasing used in the prototype had no significant impact on the factors F_i. Figure 5.13 provides an overview of the privacy explanations' effects on the total participants population. Individual figures, that show the effects for the individual groups, can be found in the appendix C of this thesis.

Even though there was no significant differences between the test groups, the overall results suggest that a notable majority of participants perceived the privacy explanations to have positive effects on all of the examined factors. Furthermore, many of the participants, that found the privacy explanations to have a negative effect, stated that they did so because they did not like the specific contents of the explanation. For example, a number of participants did not like the scope of personal data that the app wanted to access.

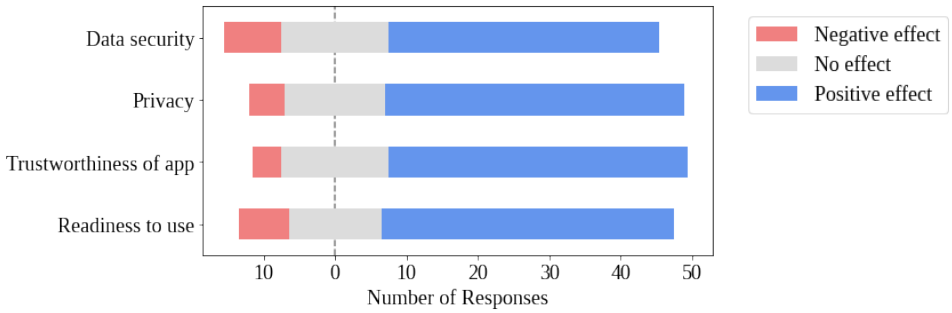

Figure 5.13: Overview of the effects of the privacy explanations

Privacy Priming through the User Study

At the end of the user study, participants answered a couple of follow-up questions regarding possible effects of privacy priming through the user study. In particular, they gave their opinion on the following topics:

- If they are more worried about their privacy
- If they want to learn more about data protection
- If they learned something about data protection
- If they want their data to be protected better
- If they want to actively work for better data protection
- If they feel like their privacy awareness increased

Participants stated if the study had an effect or not, or if they were unsure. Figure 5.14 shows the results of this exercise.

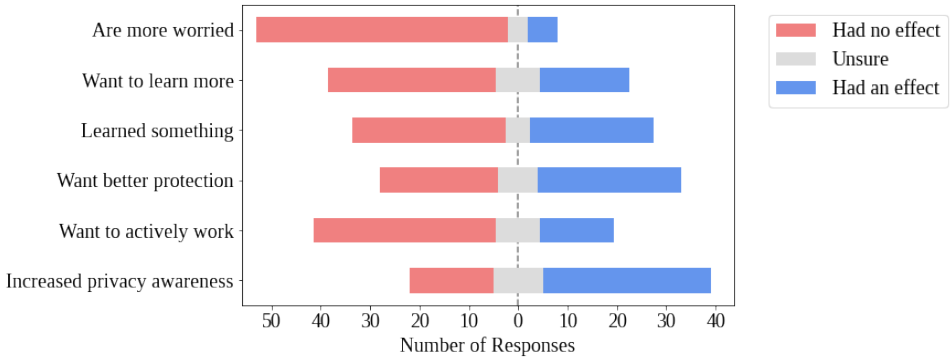

Figure 5.14: Effects of privacy priming through the user study

Concerning their worry about their privacy, the vast majority of participants said that it was not increased through their participation in the study.

Similarly, most participants stated that they did not want to learn more about data protection after their participation. In the same vein, only few participants said that they want to actively work for more data protection. Curiously, both of these are active processes.

Participants were more open towards more passive processes. For instance, a notable part of the participants said they learned something through the course of the study. Furthermore, a relative majority of participants demanded better protection of their data, now that they had participated. Notably, when asked directly if they feel like their privacy awareness was increased through the study, the majority of participants thought that this was the case.

Participants' Remarks concerning their Privacy Awareness

Finally, the participants made a variety of remarks concerning privacy awareness throughout the study. Before seeing the prototype, 44% of participants did not know what privacy explanations are or could not differentiate them from privacy policies. 39% had a vague idea about the term, but were not sure about it. Only three out of all 61 participants knew what a privacy explanation is and what it entails.

Concerning the trustworthiness of privacy explanations, 11% of participants stated that they do not trust them by principle. 20% of participants would not read an entire privacy explanation, regardless of if it is well-structured.

15% of participants explicitly stated that the privacy explanations increased their privacy awareness. Interestingly, 13% of participants mentioned that their privacy awareness did not increase, but that they attributed this to them already having a high privacy awareness. Concluding the findings are 5 of the 61 participants, who said that these effects of privacy priming can only work short-term.

Chapter 6

Discussion

This chapter discusses the implications of the findings presented in this thesis. First, the results presented in section 5.4 are interpreted and set into context. Then, the limitations of this work and the challenges that came with conducting the research are examined.

6.1 Interpretation of Results

Privacy Explanation Structure

> **RQ1:** How must privacy explanations be structured, so that they are understandable for end-users?

Throughout the literature analysis, which was conducted before the user study, it already became clear that long text documents, such as privacy policies, are not fit to provide understandable privacy explanations. The results of the study underline this. A majority of participants had problems reading these documents, even though their demographic leaned towards a young and technologically literate audience. By far, most of those participants attributed this issue to the length of the documents.

The most obvious solution to this problem was to shorten the explanations, while still providing the entirety of all necessary privacy information. This was done by breaking up the information and compartmentalizing the explanation. The number of privacy explanations and their order was generally well received by the participants. However, this might change if more different data types are up for processing. In such a case, the number of privacy explanations could easily become too large for users to reasonably read. For this problem, there is no easy answer. If there is a lot of data to be processed, there is also a lot to be explained. However, by using compartmentalized privacy explanations, users could simply read those explanations that are the most interesting to them personally.

Concerning the length of the individual explanation compartments, the majority of participants thought it was just the right amount. Indeed, many highlighted the advantages of breaking up the text and compartmentalizing privacy information. Most participants found the explanations to be understandable for them, even though they were much shorter than what they usually encounter. Only the third party explanation deviated from this trend. However, this was rooted in the contents of the explanation, rather than its structure. Overall, the findings of the study indicate that compartmentalizing privacy explanations can make them more readable and understandable for end-users.

Regarding the prominent design elements of the prototype, participants found the operability to be the most important. This underlines the findings of the privacy explanation survey, which was discussed in section 3.4. Clearly, privacy explanations need to contain these elements of operability. Specifically, these findings suggest that users need to be able to manage their privacy settings while reading the privacy explanations, and they need to be able to do it in an intuitive manner.

The icons used within the prototype were unanimously well received by the participants. They succeeded in supporting those who needed them, and could simply be ignored by others. However, the same was not true for the use of different text colors within the explanations. Participants found these to be more distracting than helpful. Some even explicitly stated that a privacy explanation being too colorful makes it look unprofessional.

Finally, the reverse dark patterns in the prototype received mixed feedback. While a notable number of participants enjoyed the prominent bundled decline option, some criticized that this also forced hierarchy upon them. Consequently, while an option of bundled decline should be present in operable privacy explanations, it should not be highlighted or otherwise be forced upon the user.

Privacy Explanation Contents

> **RQ2:** What must privacy explanations contain, so that they are satisfying for end-users?

Among the prominent design elements of the prototype, there was also immediate language, which was tested between-groups. As it turned out, the type of phrasing seemed to have no significant effect on participants' liking of the privacy explanation contents. On another note, there was a slight trend towards immediate language, when it came to the credibility of the explanation. However, this trend was not statistically significant and would need to be investigated further, before any reasonable deduction can be made.

Overall, the contents of the privacy explanations were well received and rated as credible by the participants. Indeed, many participants did not only want to know the purposes of data processing, but were also interested in a contrastive view, examples, further details and a third party explanation. Results from the coding of participants comments and remarks highlight the importance of clear and understandable language. In order to achieve this, it is of utmost importance that developers of privacy explanations are not only honest about their practices, but also convey their explanations in a language that is easily understandable for the average end-user.

Concerning the specific contents, the privacy explanations were successful in providing a suitable amount of information to the participants. As mentioned before, the majority of participants enjoyed the length of the privacy information. Yet, there were problems with the understandability of the third party explanation for some users. When asked about the roots of their misunderstandings, many participants stated that they did not understand the technical terminology or wanted more in-depth information on the third parties. This poses a conflict to developers, as the third party explanation is very limited in space and cannot possibly include a complete privacy explanation for each third party. A possible remedy to this issue would be industry-wide norms and regulations for privacy explanations. If every third party was required to provide an appropriate privacy explanation, instead of an unreadable privacy policy, interested users could use external links to navigate between the original privacy explanations and those of the third parties.

Privacy Explanation Types

RQ3: Do end-users prefer different kinds of privacy explanations?

Results from the study show that participants clearly had preferences among the different privacy explanation types. The base-level explanation was seen as the most important, even though it provided the most general information. The other explanation types were significantly less important to users, with the example-based explanation being the most far off. Many participants seemed to be able to imagine what the data would look like and did not need a visual example. Yet, a large number of participants stated the usefulness of helpful examples, especially if the data is complex.

Interestingly, participants rated the example explanation as the most easily understandable. It was even significantly ahead of the base-level explanation in this regard. As a consequence, the example-based explanation gains in importance. While the examples might not have been important for the particular users in this study, it seems clear that they can be helpful for understanding what kind of data is processed. This is especially true for lay

users who are not that technologically literate and cannot infer what data looks like by themselves.

Finally, none of the explanations types were seen as overwhelmingly unnecessary and a notable amount of participants stated that they would not want to miss any of the provided information. Even though participants showed their preferences, it should still be advisable to provide the privacy explanation as a whole and to not cut any of the explanation types.

Privacy Explanations' Impact on Privacy Awareness

RQ4: Can privacy explanations influence end-users' privacy awareness?

Judging from participants direct answers, it appears that the study was successful in raising participants privacy awareness. Furthermore, after interacting with the prototype, the majority of participants reported positive effects on the perceived security of their data, their privacy, the trustworthiness of the software and their readiness to use said software.

However, it is unclear if these findings are consequential. At the end of the study, a majority of participants did not want to inform themselves more about data protection. Similarly, most participants said that they did not want to actively work towards improving data protection for themselves and others. This coincides with the feeling of privacy apathy, which was reported by a number of participants, and with the phenomenon that is the privacy paradox (see section 2.4).

At this point, it needs to be stated that these findings concerning privacy awareness are only preliminary. While it might seem like the privacy explanations were able to increase participants' privacy awareness, this is an effect that would need to be examined long-term. That was also pointed out by some of the participants. To this end, further research on privacy explanations' effects on privacy awareness is needed, especially over a long term.

6.2 Limitations and Threats to Validity

This thesis and the user study conducted within it are subject to a couple of limitations and threats to validity. First and foremost, the snowballing process of the manual inspection, for the literature analysis of this thesis, was not complete. Instead, it ended after "theoretical saturation" according to Wolfswinkel et al. [69] was achieved. This means that some relevant literature might have been missed. However, a complete snowballing procedure would not have been feasible within the scope of this work. As an additional means to counteract this, a database search was conducted to supplement the manual inspection.

On another note, the demography of the participants of the user study was fairly limited. While a total number of 61 people participated, the vast majority of them were digital natives (born after 1980) and likely technologically literate. Most of them were well aware of privacy policies and their structure and contents. Only two of the 61 participants requested to look at a privacy policy before answering the questions concerning the topic. The consequences of this are twofold. On the one hand, this means that, in all likelihood, most participants were able to give qualified answers. They knew what terms like privacy policy and third party data sharing entail, and they were therefore able to give educated answers. On the other hand, this implies a bias regarding the readability and understandability of privacy explanations. They are more likely to understand technological language and are more able to infer information from simple statements. People who are less acquainted with online privacy and the digital space might have had more problems understanding and evaluating the privacy explanations.

It is also likely that a number of participants were biased when it came to the effects of privacy explanations on privacy awareness. As digital natives, most of the participants were already aware of how they share their personal data. For example, many explained that they deliberately share their data in order to gain access to certain software services. Some participants even explicitly stated that their privacy awareness did not increase throughout the study, as it was already high from the start. If the demography had been more balanced, it is possible that the privacy explanations could have had a stronger impact on the privacy awareness of the participants.

The privacy explanations in this study were considerably shorter than the typical privacy policy that is encountered on the Internet. That said, when navigating the prototype, participants still had a lot of information to take in. Hence, when they answered the questions regarding the prototype and its privacy explanations, participants were explicitly allowed to look back at the prototype at their own discretion. However, some of the participants did not want to look at the prototype again, as they thought that doing so would falsify their answers. In those cases, the answers might be influenced by forgetfulness on part of the participants.

While this thesis offers a complete concept for privacy explanations, it was not possible to test the entire concept within the scope of the user study. Not all of the influential factors for privacy explanations, defined in section 4.2, could be examined. Informational completeness and informational correctness could not be gauged, as the software described within the prototypical privacy explanations does not exist. In the same vein, verifiability could not be implemented. Notably, the lack of verifiability within the prototype was also criticized by some of the participants. Corporate factors like trade secrets and development cost also could not be tested, as the implementation was only prototypical and not related to a real product. Lastly, the learnability of privacy explanations could not be

evaluated, as the study only consisted of a single session per participants. If the concept was tested within a corporate environment and over a long term, these missing factors could possibly be examined as well.

6.3 Challenges

The design of the concept for privacy explanations, which was developed within this thesis, as well as the user study that accompanied it, came with several challenges. At this point, there has been barely any research on privacy explanations as a concept. On the other side, the field of explainability in artificial intelligence and the contemporary research on end-user privacy offer a vast amount of related literature. However, the insights from these areas of research can not always be directly applied to privacy explanations, but must first be put into context and need to be adapted. While this enables researchers to be creative and test a plethora of designs, it is hard to predict how effective a design will be before it is tested. For a thesis like this, that is strictly limited in time and resources, this means that the results are not final, but rather give a strong indication towards the correct direction. This state of affairs will likely change once privacy explanations become more of a focus in contemporary research.

On another note, the user study conducted within the context of this thesis was designed to be interactive, along the lines of semi-structured interviews, and the interviewer needed to be present throughout the whole study. That said, in light of the ongoing *Covid-19* pandemic, it was not possible to have all sessions in person. Indeed, a variety of software tools had to be used to conduct the study with the vast majority of participants (see section 5.2.2). This limits the coding of their comments and remarks to their verbal feedback, as it was not possible to witness their non-verbal gestures and expressions.

Lastly, there appears to be a profound misunderstanding with the term privacy explanation in the German language. Before seeing the prototype, almost half of the participants had no idea what privacy explanations are or could not differentiate them from privacy policies. Most of the rest only had a vague idea about the term. The German translation for the term privacy policy is *Datenschutzerklärung*. In the German language, the word *Erklärung* translates to both explanation and declaration. Hence, *Datenschutzerklärung* is often understood as data protection explanation, even though it really is a data protection declaration. This can easily be confused with the term privacy explanation, which is *Privatsphäreerklärung* in German. Notably, all study participants were German speaking. It is reasonable to assume that this played a critical role in participants misunderstanding of the term privacy explanation.

Chapter 7

Conclusion and Future Work

This final chapter concludes the thesis. The results and gathered insights found and discussed throughout the previous chapters are set into context with the initial goals of this thesis. Furthermore, possible avenues for future research are proposed and discussed.

7.1 Conclusion

The goal of this thesis was to investigate the structure and contents that are necessary to provide comprehensible and enlightening privacy explanations for end-users. In addition to that, the effects of these privacy explanations on end-users' privacy awareness needed to be examined. Using an extensive data and literature analysis as the foundation, a novel concept for privacy explanations was developed. This concept was then implemented within a software prototype and thoroughly tested and evaluated in the context of an interactive user study.

Within the literature analysis, research works from the fields of explainability and privacy were consulted. At the start, a manual inspection, starting from a baseline paper on explainability, was performed. The objective was to understand the concept of explainability and ways in which explainability influences privacy and trustworthiness. After the initial manual inspection, a database search was conducted to further supplement the acquired literature in the direction of privacy. This was necessary in order to understand the current state of privacy explanations and what is needed in order to improve them. Lastly, the results of a previously conducted online survey on privacy explanations were analyzed.

From the outset of the literature analysis, it became clear that long-form text documents such as privacy policies are not fit to appropriately convey privacy explanations to end-users. Based on the concepts of personalized and layered explanations in XAI, this work develops privacy explanations in a compartmentalized manner. This included the examination and

adaptation of different types of explanations found in XAI and in the social sciences. In particular, the concepts of contrastive and example-based privacy explanations were developed. Together with the base-level explanation, the explanation of further details and the third party explanation, they cover all of the necessary privacy aspects.

After being implemented within a software prototype, the compartmentalized privacy explanation was rigorously evaluated through a user study. This included the statistical analysis of an expansive questionnaire and the coding of participants comments and remarks. The results of the user study corroborate the viability of the developed concept. Most of the participants enjoyed the compartmentalized approach and were able to understand the provided privacy explanations. Moreover, the newly developed contrastive and example-based privacy explanations have proven to be appropriate means to convey privacy information to end-users. Notably, the example-based privacy explanation has shown to be significantly more understandable than the purely textual base-level explanation.

Finally, participants answered questions regarding the effects of the privacy explanations on their privacy awareness. Among other things, they reported a heightened sense of security and an increased readiness to use the software in question. This coincides with the findings of Brunotte et al. [11] and indicates that privacy explanations can, indeed, positively contribute to end-users' privacy awareness.

7.2 Future Work

End-users are becoming increasingly aware of malicious privacy practices and are stating their need for comprehensive privacy information. Consequently, privacy explanations arise as important components of today's competent software systems. While this thesis provides important insights on the needed contents and structure of those explanations, there is still much work to be done. In the following, a couple of possible avenues for the future research of privacy explanations are suggested:

- **Refine interface usability**
  Participants of the user study in this thesis mentioned a couple of usability problems, which they encountered while navigating the prototype. Usability was not in the focus of this thesis, as it is not a feature of the privacy explanations themselves, but rather a quality of the explanation interface. However, usability is an important non-functional requirement of software systems. If privacy explanations were to be ported to real life scenarios and were used for real products, optimizing the usability of their interface would become a primary concern and require further research.

- **Employ actual personalization**
  Personalizing explanations can require complex computations, which are out of scope for the prototypical implementation in this thesis. Furthermore, the privacy problem that arises when using personalized privacy explanations was discussed in section 4.3. If this problem were to be solved, privacy explanations could possibly be personalized according to each user's individual needs. Personalizing the explanations locally on users' end-devices, without sharing data with external parties, could be an avenue to success. Future research should investigate this and other possible solutions of employing personalization without endangering end-user privacy.

- **Cooperate with legal sciences**
  It will also be important to find ways to implement verifiability and legal guarantee within privacy explanations. Participants of the user study in this thesis mentioned these qualities as important ways to build end-user trust. Following Kästner et al. [36], a privacy explanation is only trustworthy as long as the users' trust in it is warranted. By implementing verifiability, developers of privacy explanations can show users that their trust is, indeed, warranted. In the same vein, these explanations need to be legally binding, so that they offer a legal guarantee to their users. To this end, it will be necessary to cooperate with peers from the legal sciences.

- **Conduct long-term studies**
  This thesis discussed and examined the impact of privacy explanations on privacy awareness. The user study was successful in providing first insights on these effects. However, it is unclear if these effects have lasting consequences on end-users' privacy behavior. Therefore, long-term studies are needed to further examine them. Incidentally, such long-term studies would also offer an opportunity to investigate the learnability of privacy explanations.

Appendix A

Literature Analysis Tables

Chapter 3 describes the process through which the primary literature used for this thesis was acquired. This appendix shows the steps of that process and the literature included in each step.

Table A.1: Results of the documented literature analysis

Analysis Step	Acquired Literature
Baseline Paper	[L25]
Baseline Paper Review	[L1], [L2], [L3], [L7], [L6], [L15], [L19], [L20], [L26], [L27], [L34], [L44], [L45], [L50], [L55], [L58], [L66], [L67], [L72], [L73], [L74], [L76], [L77], [L83], [L85], [L90], [L94], [L95], [L96], [L98], [L101], [L104], [L105], [L112], [L113], [L114], [L115], [L116], [L120], [L121], [L122], [L126], [L127], [L128]
Snowballing	[L5], [L13], [L17], [L22], [L24], [L28], [L30], [L31], [L32], [L33], [L36], [L37], [L38], [L39], [L40], [L42], [L43], [L47], [L49], [L51], [L52], [L54], [L57], [L59], [L60], [L63], [L65], [L69], [L75], [L80], [L87], [L93], [L97], [L99], [L102], [L103], [L108], [L109], [L110], [L117], [L119], [L124], [L125]
Database Search	[L4], [L8], [L9], [L10], [L11], [L12], [L14], [L16], [L18], [L21], [L23], [L29], [L35], [L41], [L46], [L48], [L53], [L56], [L61], [L62], [L64], [L68], [L70], [L71], [L78], [L79], [L81], [L82], [L84], [L86], [L88], [L89], [L91], [L92], [L100], [L106], [L107], [L111], [L118], [L123], [L129]

Appendix B

Prototype Scenario and Screenshots

Prototype Scenario

This section of the appendix provides the hypothetical scenario in which the participants of the user study interacted with the software prototype. In particular, these are the German description text that the participants read, as well as a faithful English translation of that text.

German Description:

Stellen Sie sich folgendes Szenario vor: Sie wollen in ihrem Urlaub auf eine Reise ins Ausland gehen (z. B. nach Paris). Sie haben im Vorfeld recherchiert und haben dabei eine Reiseführer-App gefunden, die Sie verwenden wollen. Die App soll Ihnen Vorschläge für Sehenswürdigkeiten oder Ereignisse machen, die sich in Ihrer Nähe befinden. Als Sie die App aus dem Appstore installieren wollen, öffnet sich eine Privatsphäreerklärung (siehe Software-Prototyp). Mithilfe der Tabs können Sie zwischen den Daten-Arten wechseln. Mit "mehr Details" oder "weniger Details" wechseln Sie zwischen den Erklärungen. Navigieren Sie den Prototypen zunächst frei nach Interesse. Danach gibt es einige kleine Aufgaben darin zu erfüllen.

English Translation:

Imagine the following scenario: You want to go on a vacation abroad (e. g. to Paris). Before leaving, you have read up on travel guide apps and found something that you want to use. The app in question is supposed to give you travel suggestions based on what is close to you. As you are moving to install the app from the app store, a privacy explanation opens up (see software prototype). Using the tabs, you can navigate the different types of data. By clicking "more details" or "less details", you can switch between the explanation types. Please navigate the prototype at your own discretion. Afterwards, there might be some exercises for you to complete.

Prototype Screenshots

The following figures show the five privacy explanation types for the precise location data type. Specifically, it shows the English prototype that uses immediate language. All four prototypes (English/German, immediate/indirect phrasing) are provided with the enclosed USB drive (see appendix D).

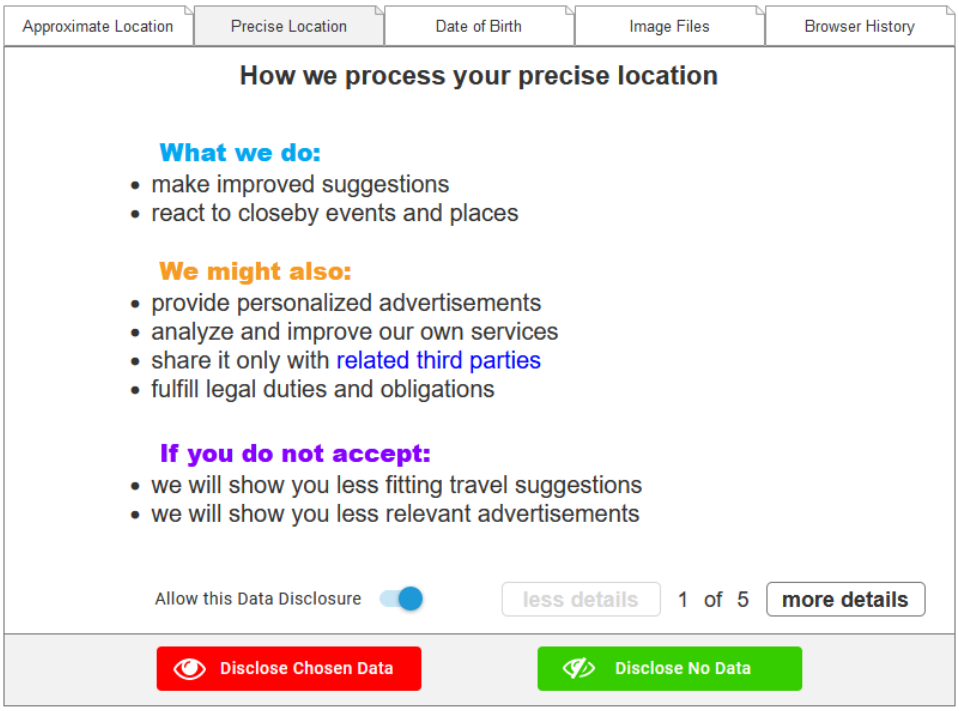

Figure B.1: Base-level explanation for precise location

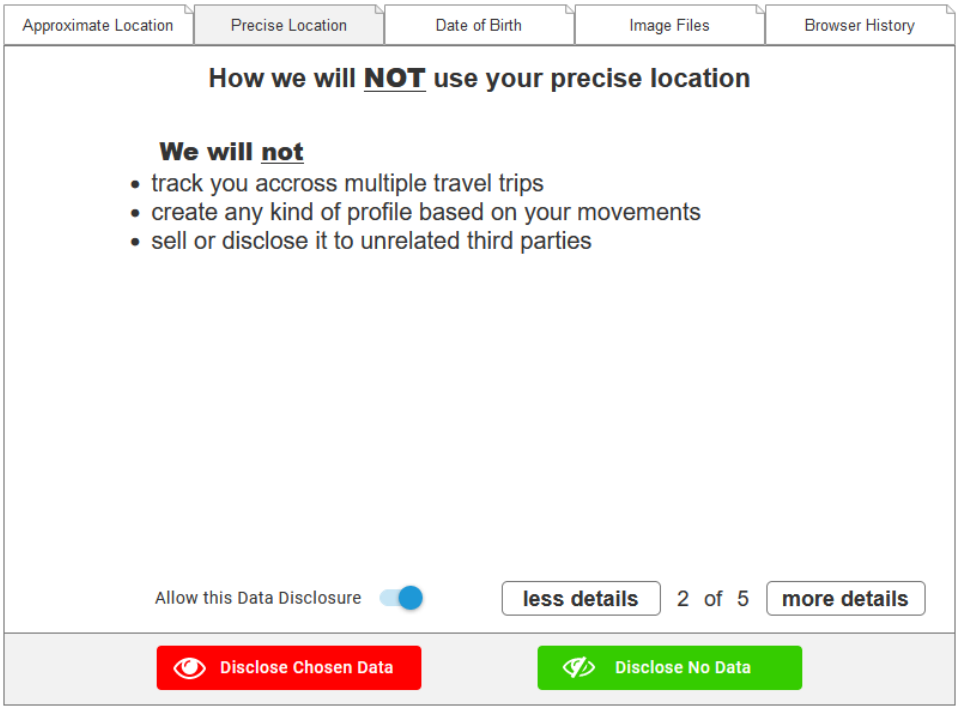

Figure B.2: Contrastive explanation for precise location

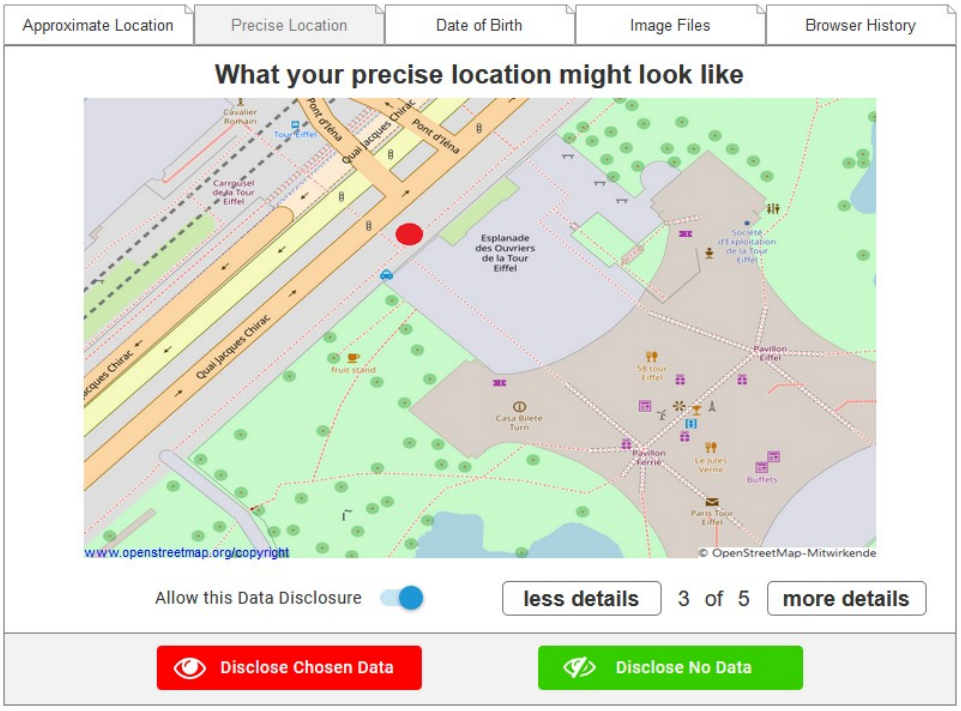

Figure B.3: Example-based explanation for precise location

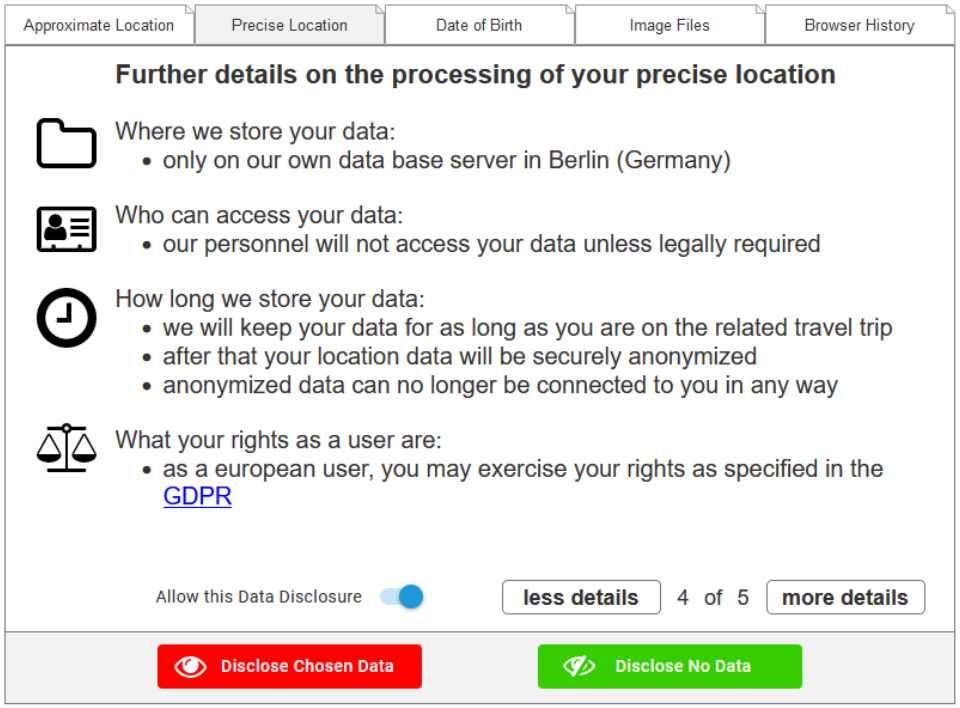

Figure B.4: Explanation of further details for precise location

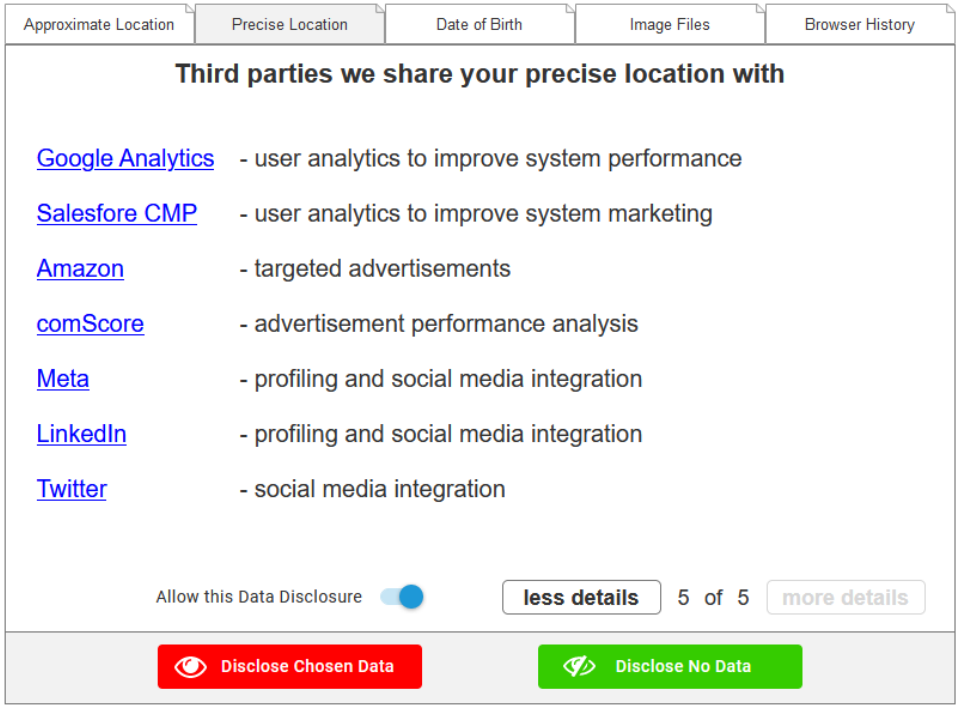

Figure B.5: Third party explanation for precise location

Appendix C

Supplemental Figures

App Privacy Notice Screenshots

These are high-resolution screenshots of and links to the app privacy notices discussed in section 2.2.2.

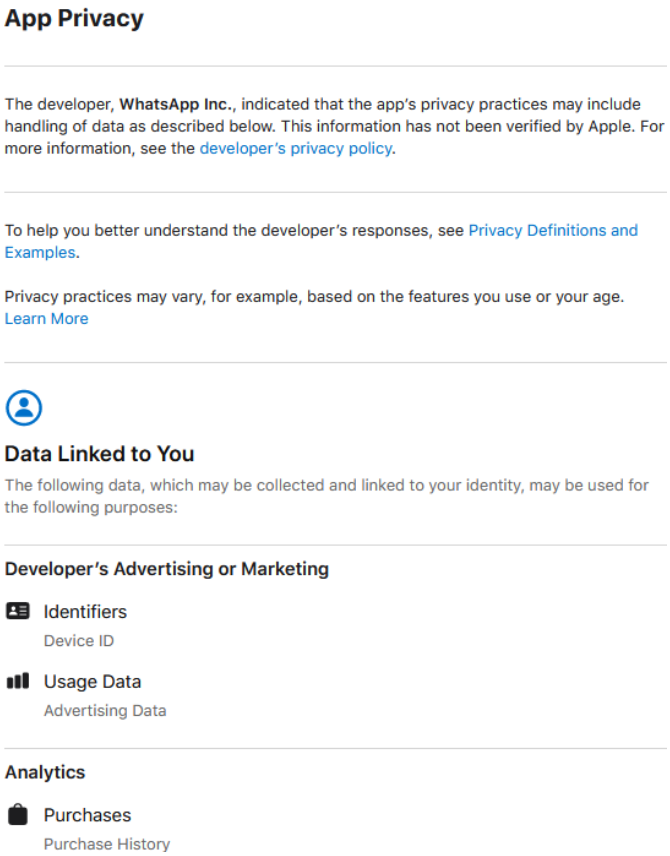

Figure C.1: The *WhatsApp Messenger* in the *Apple App Store* `https://apps.apple.com/us/app/whatsapp-messenger/id310633997`

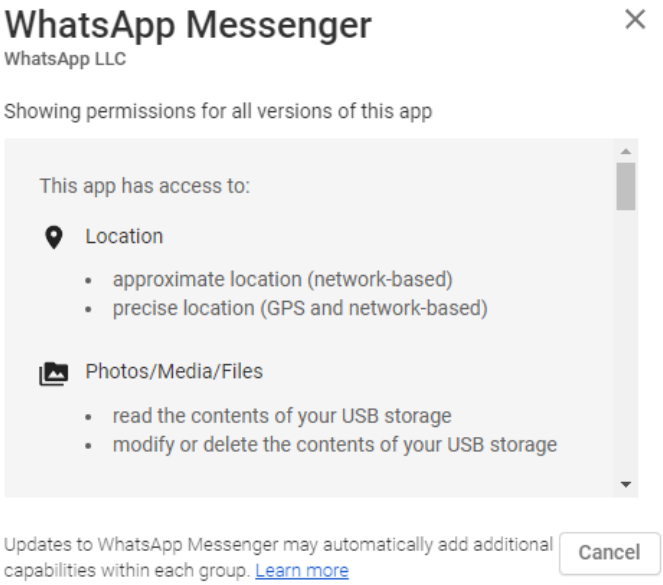

Figure C.2: The *WhatsApp Messenger* in the *Google Play Store* `https://play.google.com/store/apps/details?id=com.whatsapp&gl=US`

Effects of the Privacy Explanations

These are supplemental figures to the findings discussed in 5.4.4. They show how participants from the two groups (immediate vs. indirect phrasing) answered concerning the effects of the privacy explanations.

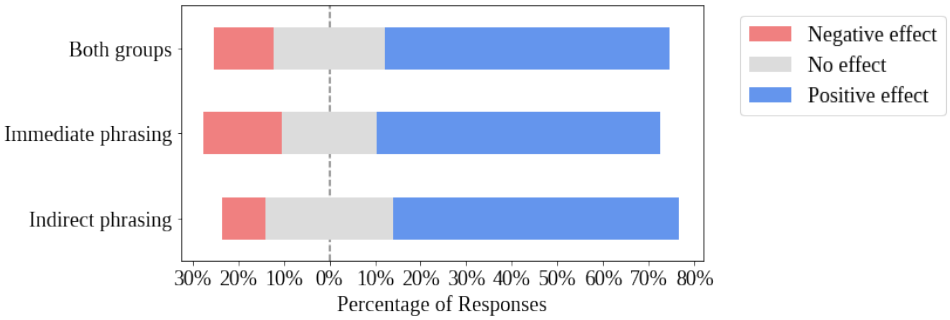

Figure C.3: The effects of the privacy explanations on participant's perceived security of data

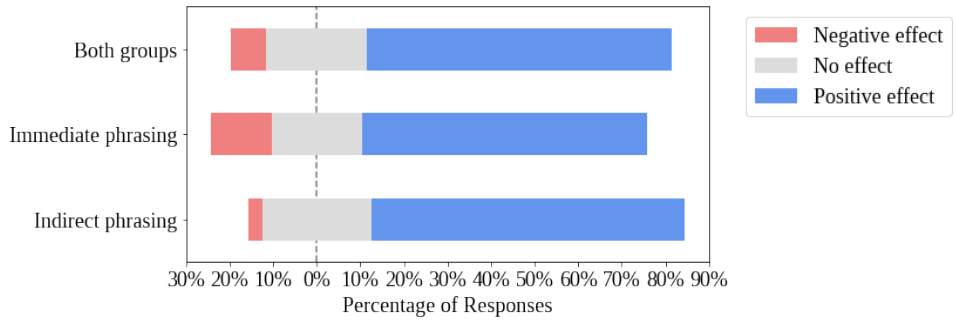

Figure C.4: The effects of the privacy explanations on participant's perceived privacy

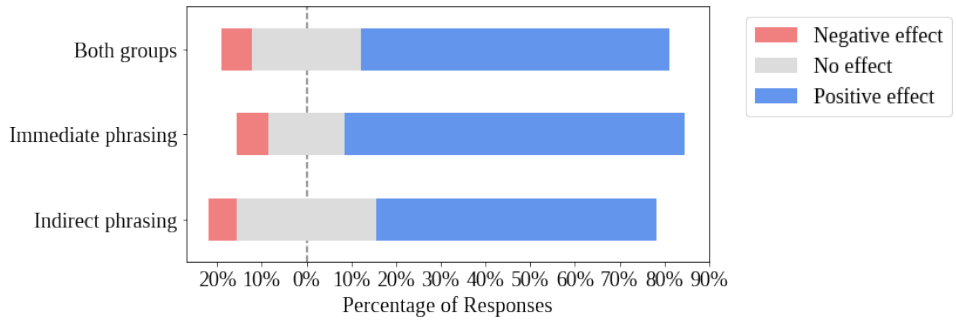

Figure C.5: The effects of the privacy explanations on the perceived trustworthiness of the described software

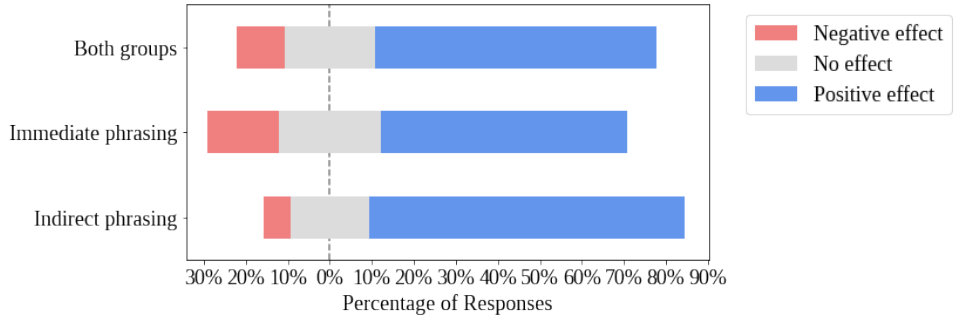

Figure C.6: The effects of the privacy explanations on participant's readiness to use the described software

Appendix D

Contents on the USB Drive

The USB drive provided with this thesis contains the following:

- The *LaTeX* archive used to build this document
- A *PDF* file version of this document
- The *Python* scripts used to build some of the figures in this thesis
- The results of the data and literature analysis
- The prototypes, both as an *Axure* project and in *HTML* format
- A directory called "Survey Data" which contains the following:
  - The questionnaire of the online survey provided by the supervisor of this thesis
  - The results of that survey
  - The *MatLab* workspace used to analyze the results
- A directory called "My Survey" which contains the following:
  - The user study's questionnaire
  - The results of the user study's questionnaire
  - The video files recorded over the course of the user study
  - The protocols of the *in vivo* coding of each individual video
  - The results of the pattern coding (second cycle)
  - The *MatLab* workspace used to analyze the results
- The *Git* repository that was used throughout the work on this thesis

List of Figures

List of Tables

Acronyms

Notation	Description
CMPs	Consent Management Platforms 16
DSGVO	Datenschutz-Grundverordnung vii
EU	European Union v, vii, 9, 17, 57
GDPR	General Data Protection Regulation v, 1, 9, 12, 13, 36, 40, 42, 43, 57
IoT	Internet of Things v, vii, 1
ToS	Terms of Service 14, 17
UK	United Kingdom 16
USA	United States of America 18
XAI	Explainable Artificial Intelligence 6, 7, 10, 11, 24–26, 34, 39, 41, 77, 78

Bibliography

[1] A. Adadi and M. Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160, 2018.

[2] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[3] C. Baker. The impact of instructor immediacy and presence for online student affective learning, cognition, and motivation. *Journal of Educators Online*, 7(1):n1, 2010.

[4] G. Bal. Designing privacy indicators for smartphone app markets: A new perspective on the nature of privacy risks of apps. In *Proceedings of the 20th Americas Conference on Information Systems*, 2014.

[5] R. Balebako, F. Schaub, I. Adjerid, A. Acquisti, and L. Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 63–74, 2015.

[6] R. Balebako, R. Shay, and L. F. Cranor. Is your inseam a biometric? evaluating the understandability of mobile privacy notice categories. *CMU, Tech. Rep. CMU-CyLab-13-011*, 2013.

[7] S. B. Barnes. A privacy paradox: Social networking in the united states. *First Monday*, 11(9), Sep. 2006.

[8] K. Bongard-Blanchy, A. Rossi, S. Rivas, S. Doublet, V. Koenig, and G. Lenzini. "i am definitely manipulated, even when i am aware of it. it's ridiculous!"-dark patterns from the end-user perspective. In *Designing Interactive Systems Conference 2021*, pages 763–776, 2021.

[9] P. B. Brandtzaeg, A. Pultier, and G. M. Moen. Losing control to data-hungry apps: A mixed-methods approach to mobile app privacy. *Social Science Computer Review*, 37(4):466–488, 2019.

[10] W. Brunotte, L. Chazette, V. Klös, and T. Speith. Quo vadis, explainability? – a research roadmap for explainability engineering. In V. Gervasi and A. Vogelsang, editors, *Requirements Engineering: Foundation for Software Quality*, pages 26–32, Cham, 2022. Springer International Publishing.

[11] W. Brunotte, L. Chazette, and K. Korte. Can explanations support privacy awareness? a research roadmap. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 176–180, 2021.

[12] W. Brunotte, L. Chazette, L. Köhler, and K. Schneider. What about my privacy? helping users understand online privacy policies. In *Proceedings of the International Conference on Software and System Processes and International Conference on Global Software Engineering (ICSSP'22)*, ICSSP '22, New York, NY, USA, 2022. Association for Computing Machinery.

[13] A. Bussone, S. Stumpf, and D. O'Sullivan. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*, pages 160–169. IEEE, 2015.

[14] C. J. Cai, J. Jongejan, and J. Holbrook. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th international conference on intelligent user interfaces*, pages 258–262, 2019.

[15] L. Chazette, W. Brunotte, and T. Speith. Exploring explainability: A definition, a model, and a knowledge catalogue. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 197–208. IEEE, 2021.

[16] L. Chazette, W. Brunotte, and T. Speith. *Supplementary Material for Research Paper "Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue"*, July 2021. Available at `https://doi.org/10.5281/zenodo.5114922`, last visited on 2022-04-09.

[17] L. Chazette and K. Schneider. Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering*, 25(4):493–514, 2020.

[18] I. Chong, H. Ge, N. Li, and R. W. Proctor. Influence of privacy priming and security framing on mobile app selection. *Computers & Security*, 78:143–154, 2018.

[19] M. Chromik, M. Eiband, S. T. Völkel, and D. Buschek. Dark patterns of explainability, transparency, and user control for intelligent systems. In *IUI workshops*, volume 2327, 2019.

[20] H. Cramer, V. Evers, S. Ramlal, M. Van Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted interaction*, 18(5):455–496, 2008.

[21] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[22] D. Doran, S. Schulz, and T. R. Besold. What does explainable AI really mean? A new conceptualization of perspectives. *CoRR*, abs/1710.00794, 2017.

[23] European Parliament and Council. General data protection regulation, 2016-05-04. Available at `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504`, last visited on 2022-03-16.

[24] H. Fu and J. Lindqvist. General area or approximate location? how people understand location permissions. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 117–120, 2014.

[25] A. Glass, D. L. McGuinness, and M. Wolverton. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 227–236, 2008.

[26] B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3):50–57, 2017.

[27] C. M. Gray, J. Chen, S. S. Chivukula, and L. Qu. End user accounts of dark patterns as felt manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25, 2021.

[28] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.

[29] E. Hargittai and A. Marwick. "what can i really do?" explaining the privacy paradox with online apathy. *International journal of communication*, 10:21, 2016.

[30] R. R. Hoffman, G. Klein, and S. T. Mueller. Explaining explanation for "explainable ai". In *Proceedings of the Human Factors and Ergonomics*

Society Annual Meeting, volume 62, pages 197–201. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[31] H. Johnson and P. Johnson. Explanation facilities and interactive systems. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 159–166, 1993.

[32] R. F. Kizilcec. How much information? effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2390–2395, 2016.

[33] N. Koehler, O. Vujovic, and C. McMenamin. Healthcare professionals' use of mobile phones and the internet in clinical practice. *Journal of mobile technology in medicine*, 2(1):3–13, 2013.

[34] F. Kreuter, G.-C. Haas, F. Keusch, S. Bähr, and M. Trappmann. Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38(5):533–549, 2020.

[35] O. Kulyk, P. Gerber, K. Marky, C. Beckmann, and M. Volkamer. Does this app respect my privacy? design and evaluation of information materials supporting privacy-related decisions of smartphone users. In *Workshop on usable security (USEC'19). San Diego, CA*, pages 1–10, 2019.

[36] L. Kästner, M. Langer, V. Lazar, A. Schomäcker, T. Speith, and S. Sterz. On the relation of trust and explainability: Why to engineer for trustworthiness. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 169–175. IEEE, 2021.

[37] J. C. S. d. P. Leite and C. Cappelli. Software transparency. *Business & Information Systems Engineering*, 2(3):127–139, 2010.

[38] B. Y. Lim, A. K. Dey, and D. Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2119–2128, 2009.

[39] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 501–510, 2012.

[40] K. Martin. Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, 34(2):210–227, 2015.

[41] A. M. McDonald and L. F. Cranor. The cost of reading privacy policies. *I/S: A Journal of Law and Policy (ISJLP)*, 4:543, 2008.

[42] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[43] M. North. An examination of mobile app privacy policies and third-party data sharing. *Issues in Information Systems*, 14(2), 2013.

[44] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal. Dark patterns after the gdpr: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[45] I. Nunes and D. Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3):393–444, 2017.

[46] E. Okoyomon, N. Samarin, P. Wijesekera, A. Elazari Bar On, N. Vallina-Rodriguez, I. Reyes, Á. Feal, S. Egelman, et al. On the ridiculousness of notice and consent: Contradictions in app privacy policies. In *Workshop on Technology and Consumer Protection (ConPro 2019), in conjunction with the 39th IEEE Symposium on Security and Privacy*, 2019.

[47] A.-M. Ortloff, L. Güntner, M. Windl, D. Feth, and S. Polst. Evaluation kontextueller datenschutzerklärungen. In R. Dachselt and G. Weber, editors, *Mensch und Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V.

[48] I. Pentina, L. Zhang, H. Bata, and Y. Chen. Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison. *Computers in Human Behavior*, 65:409–419, 2016.

[49] W. Pieters. Explanation and trust: what to tell the user in security and ai? *Ethics and information technology*, 13(1):53–64, 2011.

[50] A. Preece, D. Harborne, D. Braines, R. Tomsett, and S. Chakraborty. Stakeholders in explainable ai. *arXiv preprint arXiv:1810.00184*, 2018.

[51] P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542–556, 2007.

[52] S. Pötzsch. Privacy awareness: A means to solve the privacy paradox? In *IFIP Summer School on the Future of Identity in the Information Society*, pages 226–236. Springer, 2008.

[53] G. Ras, M. van Gerven, and P. Haselager. Explanation methods in deep learning: Users, values, concerns and challenges. In *Explainable*

and interpretable models in computer vision and machine learning, pages 19–36. Springer, 2018.

[54] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[55] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, 62(3):381–422, 2005.

[56] A. Rosenfeld and A. Richardson. Explainability in human–agent systems. *Autonomous Agents and Multi-Agent Systems*, 33(6):673–705, 2019.

[57] J. Saldaña. *The Coding Manual for Qualitative Researchers.* SAGE Publications Inc., Thousand Oaks, CA, USA, 2nd edition, 2013.

[58] J. Schneider and J. Handali. Personalized explanation in machine learning: A conceptualization. *arXiv preprint arXiv:1901.00770*, 2019.

[59] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2347–2356, 2014.

[60] S. Siegel. Nonparametric statistics. *The American Statistician*, 11(3):13–19, 1957.

[61] K. Sokol and P. Flach. One explanation does not fit all. *KI-Künstliche Intelligenz*, 34(2):235–250, 2020.

[62] A. Springer and S. Whittaker. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th international conference on intelligent user interfaces*, pages 107–120, 2019.

[63] A. Sunyaev, T. Dehling, P. L. Taylor, and K. D. Mandl. Availability and quality of mobile health app privacy policies. *Journal of the American Medical Informatics Association*, 22(e1):e28–e33, 2015.

[64] N. Tintarev and J. Masthoff. Effective explanations of recommendations: user-centered design. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 153–156, 2007.

[65] S. Wachter, B. Mittelstadt, and C. Russell. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841, 2017.

[66] A. E. Waldman. Privacy, notice, and design. *Stanford Technology Law Review*, 21:74, 2018.

[67] A. Weller. Transparency: motivations and challenges. In *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 23–40. Springer, 2019.

[68] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[69] J. F. Wolfswinkel, E. Furtmueller, and C. P. Wilderom. Using grounded theory as a method for rigorously reviewing literature. *European journal of information systems*, 22(1):45–55, 2013.

[70] S. S. Won, J. Jin, and J. I. Hong. Contextual web history: using visual and contextual cues to improve web browser history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1457–1466, 2009.

[71] M. Zanker. The influence of knowledgeable explanations on users' perception of a recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 269–272, 2012.

Results of the Literature Analysis

[L1] B. Abdollahi and O. Nasraoui. Transparency in fair machine learning: the case of explainable recommender systems. In *Human and machine learning*, pages 21–35. Springer, 2018.

[L2] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–18, 2018.

[L3] A. Adadi and M. Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160, 2018.

[L4] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times! a field study on mobile app privacy nudging. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 787–796, 2015.

[L5] D. Ameller, C. Ayala, J. Cabot, and X. Franch. How do software architects consider non-functional requirements: An exploratory study. In *2012 20th IEEE international requirements engineering conference (RE)*, pages 41–50. IEEE, 2012.

[L6] S. Anjomshoae, A. Najjar, D. Calvaresi, and K. Främling. Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, pages 1078–1088. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[L7] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies,

opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[L8] G. Bal. Designing privacy indicators for smartphone app markets: A new perspective on the nature of privacy risks of apps. In *Proceedings of the 20th Americas Conference on Information Systems*, 2014.

[L9] R. Balebako, A. Marsh, J. Lin, J. I. Hong, and L. F. Cranor. The privacy and security behaviors of smartphone app developers. *USEC '14*, 2014.

[L10] R. Balebako, F. Schaub, I. Adjerid, A. Acquisti, and L. Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 63–74, 2015.

[L11] R. Balebako, R. Shay, and L. F. Cranor. Is your inseam a biometric? evaluating the understandability of mobile privacy notice categories. *CMU, Tech. Rep. CMU-CyLab-13-011*, 2013.

[L12] Z. Belkhamza, M. Niasin, and A. Faris. The effect of privacy concerns on smartphone app purchase in malaysia: Extending the theory of planned behavior. *International Journal of Interactive Mobile Technologies*, 11(5), 2017.

[L13] B. Boehm and H. In. Identifying quality-requirement conflicts. *IEEE software*, 13(2):25–35, 1996.

[L14] K. Bongard-Blanchy, A. Rossi, S. Rivas, S. Doublet, V. Koenig, and G. Lenzini. "i am definitely manipulated, even when i am aware of it. it's ridiculous!"-dark patterns from the end-user perspective. In *Designing Interactive Systems Conference 2021*, pages 763–776, 2021.

[L15] R. Borgo, M. Cashmore, and D. Magazzeni. Towards providing explanations for ai planner decisions. In *IJCAI/ECAI 2018 Workshop on Explainable Artificial Intelligence (XAI)*, pages 11–17, 2018.

[L16] P. B. Brandtzaeg, A. Pultier, and G. M. Moen. Losing control to data-hungry apps: A mixed-methods approach to mobile app privacy. *Social Science Computer Review*, 37(4):466–488, 2019.

[L17] T. BREAUX. Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations. In *Proc. of 14th IEEE International Requirements Engineering Conference, 2006*, 2006.

[L18] C. Buck, C. Horbel, and T. Eymann. Dealing with privacy and security risks: App consumers in mobile ecosystems. *Tagungsband*

Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014), pages 64–74, 2014.

[L19] A. Bussone, S. Stumpf, and D. O'Sullivan. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*, pages 160–169. IEEE, 2015.

[L20] C. J. Cai, J. Jongejan, and J. Holbrook. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th international conference on intelligent user interfaces*, pages 258–262, 2019.

[L21] R. Calo. Against notice skepticism in privacy (and elsewhere). *Notre Dame L. Rev.*, 87:1027, 2011.

[L22] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8):832, 2019.

[L23] C. Castelluccia, S. Guerses, M. Hansen, J. Hoepman, J. van Hoboken, B. Vieira, et al. Privacy and data protection in mobile applications: A study on the app development ecosystem and the technical implementation of gdpr. 2017.

[L24] Centre for International Governance Innovation (CIGI). Cigi-ipsos global survey on internet security and trust 2019. Available at `https://www.cigionline.org/cigi-ipsos-global-survey-internet-security-and-trust/`, last visited on 2022-04-14.

[L25] L. Chazette, W. Brunotte, and T. Speith. Exploring explainability: A definition, a model, and a knowledge catalogue. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 197–208. IEEE, 2021.

[L26] L. Chazette, W. Brunotte, and T. Speith. *Supplementary Material for Research Paper "Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue"*, July 2021. Available at `https://doi.org/10.5281/zenodo.5114922`, last visited on 2022-04-09.

[L27] L. Chazette and K. Schneider. Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering*, 25(4):493–514, 2020.

[L28] H.-F. Cheng, R. Wang, Z. Zhang, F. O'Connell, T. Gray, F. M. Harper, and H. Zhu. Explaining decision-making algorithms through

ui: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–12, 2019.

[L29] I. Chong, H. Ge, N. Li, and R. W. Proctor. Influence of privacy priming and security framing on mobile app selection. *Computers & Security*, 78:143–154, 2018.

[L30] M. Chromik, M. Eiband, S. T. Völkel, and D. Buschek. Dark patterns of explainability, transparency, and user control for intelligent systems. In *IUI workshops*, volume 2327, 2019.

[L31] L. Chung and J. C. S. d. Prado Leite. On non-functional requirements in software engineering. In *Conceptual modeling: Foundations and applications*, pages 363–379. Springer, 2009.

[L32] L. Coles-Kemp, R. B. Jensen, and C. P. Heath. Too much information: questioning security in a post-digital society. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[L33] C. L. Corritore, B. Kracher, and S. Wiedenbeck. On-line trust: concepts, evolving themes, a model. *International journal of human-computer studies*, 58(6):737–758, 2003.

[L34] H. Cramer, V. Evers, S. Ramlal, M. Van Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted interaction*, 18(5):455–496, 2008.

[L35] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[L36] L. F. Cranor and S. Garfinkel. Guest editors' introduction: Secure or usable? *IEEE security & privacy*, 2(5):16–18, 2004.

[L37] L. M. Cysneiros, M. Raffi, and J. C. S. do Prado Leite. Software transparency as a key requirement for self-driving cars. In *2018 IEEE 26th international requirements engineering conference (RE)*, pages 382–387. IEEE, 2018.

[L38] J. L. De La Vara, K. Wnuk, R. Berntsson-Svensson, J. Sánchez, and B. Regnell. An empirical study on the importance of quality requirements in industry. In *SEKE*, pages 438–443, 2011.

[L39] V. Distler, M.-L. Zollinger, C. Lallemand, P. Roenne, P. Ryan, and V. Koenig. Security–visible, yet unseen? how displaying

security mechanisms impacts user experience and perceived security. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI2019)*, 2019.

[L40] D. Doran, S. Schulz, and T. R. Besold. What does explainable AI really mean? A new conceptualization of perspectives. *CoRR*, abs/1710.00794, 2017.

[L41] F. Ebrahimi, M. Tushev, and A. Mahmoud. Mobile app privacy in software engineering research: A systematic mapping study. *Information and Software Technology*, 133:106466, 2021.

[L42] C. Flavián, M. Guinalíu, and R. Gurrea. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & management*, 43(1):1–14, 2006.

[L43] G. Friedrich and M. Zanker. A taxonomy for generating explanations in recommender systems. *AI Magazine*, 32(3):90–98, 2011.

[L44] L. H. Gilpin, C. Testart, N. Fruchter, and J. Adebayo. Explaining explanations to society. *arXiv preprint arXiv:1901.06560*, 2019.

[L45] A. Glass, D. L. McGuinness, and M. Wolverton. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 227–236, 2008.

[L46] N. Good, R. Dhamija, J. Grossklags, D. Thaw, S. Aronowitz, D. Mulligan, and J. Konstan. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 43–52, 2005.

[L47] B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3):50–57, 2017.

[L48] C. M. Gray, J. Chen, S. S. Chivukula, and L. Qu. End user accounts of dark patterns as felt manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25, 2021.

[L49] E. C. Groen, S. Kopczyńska, M. P. Hauer, T. D. Krafft, and J. Doerr. Users—the hidden software product quality experts?: A study on how app users report quality aspects in online reviews. In *2017 IEEE 25th international requirements engineering conference (RE)*, pages 80–89. IEEE, 2017.

[L50] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.

[L51] P. Gutmann and I. Grigg. Security usability. *IEEE security & privacy*, 3(4):56–58, 2005.

[L52] R. Hardin. *Trust and trustworthiness*. Russell Sage Foundation, 2002.

[L53] M. Hatamian. Engineering privacy in smartphone apps: A technical guideline catalog for app developers. *IEEE Access*, 8:35429–35445, 2020.

[L54] K. Hawley. *How to be trustworthy*. Oxford University Press, USA, 2019.

[L55] M. Hind, D. Wei, M. Campbell, N. C. Codella, A. Dhurandhar, A. Mojsilović, K. Natesan Ramamurthy, and K. R. Varshney. Ted: Teaching ai to explain its decisions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 123–129, 2019.

[L56] M. Hintze. In defense of the long privacy statement. *Md. L. Rev.*, 76:1044, 2016.

[L57] R. R. Hoffman, G. Klein, and S. T. Mueller. Explaining explanation for "explainable ai". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 197–201. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[L58] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1312, 2019.

[L59] M. Hosseini, A. Shahri, K. Phalp, and R. Ali. Four reference models for transparency requirements in information systems. *Requirements Engineering*, 23(2):251–275, 2018.

[L60] A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 624–635, 2021.

[L61] M. J. Jafar, A. Abdullat, et al. Exploratory analysis of the readability of information privacy statement of the primary social networks. *Journal of Business & Economics Research (JBER)*, 7(12), 2009.

[L62] S. Joeckel and L. Dogruel. Default effects in app selection: German adolescents' tendency to adhere to privacy or social relatedness features in smartphone apps. *Mobile Media & Communication*, 8(1):22–41, 2020.

[L63] H. Johnson and P. Johnson. Explanation facilities and interactive systems. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 159–166, 1993.

[L64] R. Kesler, M. Kummer, and P. Schulte. Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings*, volume 2020, page 20978. Academy of Management Briarcliff Manor, NY 10510, 2020.

[L65] D. E. Kieras and S. Bovair. The role of a mental model in learning to operate a device. *Cognitive science*, 8(3):255–273, 1984.

[L66] R. F. Kizilcec. How much information? effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2390–2395, 2016.

[L67] M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender. Explainability as a non-functional requirement. In *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pages 363–368. IEEE, 2019.

[L68] F. Kreuter, G.-C. Haas, F. Keusch, S. Bähr, and M. Trappmann. Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38(5):533–549, 2020.

[L69] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W.-K. Wong. Too much, too little, or just right? ways explanations impact end users' mental models. In *2013 IEEE Symposium on visual languages and human centric computing*, pages 3–10. IEEE, 2013.

[L70] O. Kulyk, P. Gerber, M. El Hanafi, B. Reinheimer, K. Renaud, and M. Volkamer. Encouraging privacy-aware smartphone app installation: What would the technically-adept do. In *USEC'16-Usable Security Workshop, 21 February 2016, San Diego*. Internet Society, 2016.

[L71] O. Kulyk, P. Gerber, K. Marky, C. Beckmann, and M. Volkamer. Does this app respect my privacy? design and evaluation of information materials supporting privacy-related decisions of smartphone users. In *Workshop on usable security (USEC'19). San Diego, CA*, pages 1–10, 2019.

[L72] L. Kästner, M. Langer, V. Lazar, A. Schomäcker, T. Speith, and S. Sterz. On the relation of trust and explainability: Why to engineer for trustworthiness. In *2021 IEEE 29th International Requirements*

Engineering Conference Workshops (REW), pages 169–175. IEEE, 2021.

[L73] M. Langer, K. Baum, K. Hartmann, S. Hessel, T. Speith, and J. Wahl. Explainability auditing for intelligent systems: A rationale for multi-disciplinary perspectives. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 164–168. IEEE, 2021.

[L74] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, A. Sesing, and K. Baum. What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. *Artificial Intelligence*, 296:103473, 2021.

[L75] J. C. S. d. P. Leite and C. Cappelli. Software transparency. *Business & Information Systems Engineering*, 2(3):127–139, 2010.

[L76] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4):611–627, 2018.

[L77] Q. V. Liao, D. Gruen, and S. Miller. Questioning the ai: informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[L78] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 501–510, 2012.

[L79] B. Liu, J. Lin, and N. Sadeh. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? In *Proceedings of the 23rd international conference on World wide web*, pages 201–212, 2014.

[L80] A. F. Markus, J. A. Kors, and P. R. Rijnbeek. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, 113:103655, 2021.

[L81] K. Martin. Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, 34(2):210–227, 2015.

[L82] A. M. McDonald and T. Lowenthal. Nano-notice: Privacy disclosure at a mobile scale. *Journal of Information Policy*, 3(1):331–354, 2013.

[L83] J. McInerney, B. Lacker, S. Hansen, K. Higley, H. Bouchard, A. Gruson, and R. Mehrotra. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM conference on recommender systems*, pages 31–39, 2018.

[L84] A. N. Mehdy and H. Mehrpouyan. Modeling of personalized privacy disclosure behavior: A formal method approach. In *The 16th International Conference on Availability, Reliability and Security*, pages 1–13, 2021.

[L85] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[L86] G. R. Milne, M. J. Culnan, and H. Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing*, 25(2):238–249, 2006.

[L87] B. Mittelstadt, C. Russell, and S. Wachter. Explaining explanations in ai. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 279–288, 2019.

[L88] F. Mosca. Value-aligned and explainable agents for collective decision making: Privacy application. In *AAMAS*, pages 2199–2200, 2020.

[L89] M. North. An examination of mobile app privacy policies and third-party data sharing. *Issues in Information Systems*, 14(2), 2013.

[L90] I. Nunes and D. Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3):393–444, 2017.

[L91] E. Okoyomon, N. Samarin, P. Wijesekera, A. Elazari Bar On, N. Vallina-Rodriguez, I. Reyes, Á. Feal, S. Egelman, et al. On the ridiculousness of notice and consent: Contradictions in app privacy policies. In *Workshop on Technology and Consumer Protection (ConPro 2019), in conjunction with the 39th IEEE Symposium on Security and Privacy*, 2019.

[L92] I. Pentina, L. Zhang, H. Bata, and Y. Chen. Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison. *Computers in Human Behavior*, 65:409–419, 2016.

[L93] R. Pierrard, J.-P. Poli, and C. Hudelot. A new approach for explainable multiple organ annotation with few data. In *IJCAI 2019 Workshop on Explainable Artificial Intelligence (XAI)*, 2019.

[L94] W. Pieters. Explanation and trust: what to tell the user in security and ai? *Ethics and information technology*, 13(1):53–64, 2011.

[L95] A. Preece, D. Harborne, D. Braines, R. Tomsett, and S. Chakraborty. Stakeholders in explainable ai. *arXiv preprint arXiv:1810.00184*, 2018.

[L96] P. Pu and L. Chen. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 93–100, 2006.

[L97] P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542–556, 2007.

[L98] V. Putnam and C. Conati. Exploring the need for explainable artificial intelligence (xai) in intelligent tutoring systems (its). In *IUI Workshops*, volume 19, pages 1–7, 2019.

[L99] G. Ras, M. van Gerven, and P. Haselager. Explanation methods in deep learning: Users, values, concerns and challenges. In *Explainable and interpretable models in computer vision and machine learning*, pages 19–36. Springer, 2018.

[L100] J. R. Reidenberg, N. C. Russell, A. J. Callen, S. Qasir, and T. B. Norton. Privacy harms and the effectiveness of the notice and choice framework. *ISJLP*, 11:485, 2015.

[L101] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[L102] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, 62(3):381–422, 2005.

[L103] S. Robbins. A misdirected principle with a catch: explicability for ai. *Minds and Machines*, 29(4):495–514, 2019.

[L104] A. Rosenfeld and A. Richardson. Explainability in human–agent systems. *Autonomous Agents and Multi-Agent Systems*, 33(6):673–705, 2019.

[L105] J. Schneider and J. Handali. Personalized explanation in machine learning: A conceptualization. *arXiv preprint arXiv:1901.00770*, 2019.

[L106] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2347–2356, 2014.

[L107] R. H. Sloan and R. Warner. Beyond notice and choice: Privacy, norms, and consent. *J. High Tech. L.*, 14:370, 2014.

[L108] K. Sokol and P. Flach. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 56–67, 2020.

[L109] K. Sokol and P. Flach. One explanation does not fit all. *KI-Künstliche Intelligenz*, 34(2):235–250, 2020.

[L110] A. Springer and S. Whittaker. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th international conference on intelligent user interfaces*, pages 107–120, 2019.

[L111] A. Sunyaev, T. Dehling, P. L. Taylor, and K. D. Mandl. Availability and quality of mobile health app privacy policies. *Journal of the American Medical Informatics Association*, 22(e1):e28–e33, 2015.

[L112] S. Thiebes, S. Lins, and A. Sunyaev. Trustworthy artificial intelligence. *Electronic Markets*, 31(2):447–464, 2021.

[L113] S. Thomsen. Corporate values and corporate governance. *Corporate Governance: International Journal of Business in Society*, 4(4):29–46, 2004.

[L114] N. Tintarev and J. Masthoff. Effective explanations of recommendations: user-centered design. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 153–156, 2007.

[L115] N. Tintarev and J. Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4):399–439, 2012.

[L116] C.-H. Tsai and P. Brusilovsky. Explaining recommendations in an interactive hybrid social recommender. In *Proceedings of the 24th international conference on intelligent user interfaces*, pages 391–396, 2019.

[L117] K. Vaccaro, D. Huang, M. Eslami, C. Sandvig, K. Hamilton, and K. Karahalios. The illusion of control: Placebo effects of control settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[L118] A. E. Waldman. Privacy, notice, and design. *Stanford Technology Law Review*, 21:74, 2018.

[L119] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[L120] A. Weller. Transparency: motivations and challenges. In *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 23–40. Springer, 2019.

[L121] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[L122] J. F. Wolfswinkel, E. Furtmueller, and C. P. Wilderom. Using grounded theory as a method for rigorously reviewing literature. *European journal of information systems*, 22(1):45–55, 2013.

[L123] V. M. Wottrich, E. A. van Reijmersdal, and E. G. Smit. App users unwittingly in the spotlight: a model of privacy protection in mobile apps. *Journal of Consumer Affairs*, 53(3):1056–1083, 2019.

[L124] K.-P. Yee. Aligning security and usability. *IEEE Security & Privacy*, 2(5):48–55, 2004.

[L125] E. N. Zalta, U. Nodelman, and C. Allen. *Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Center for the Study of Language and Information, 1995.

[L126] M. Zanker. The influence of knowledgeable explanations on users' perception of a recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 269–272, 2012.

[L127] J. Zhou and F. Chen. Towards trustworthy human-ai teaming under uncertainty. In *IJCAI 2019 Workshop on Explainable AI (XAI)*, 2019.

[L128] J. Zhou, H. Hu, Z. Li, K. Yu, and F. Chen. Physiological indicators for user trust in machine learning with influence enhanced fact-checking. In *International cross-domain conference for machine learning and knowledge extraction*, pages 94–113. Springer, 2019.

[L129] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Mobile app recommendations with security and privacy awareness. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 951–960, 2014.