Privacy Needs Reflection: Conceptional Design Rationales for Privacy-Preserving Explanation User Interfaces

Peter Sörries*
Freie Universität Berlin
Berlin, Germany

Claudia Müller-Birn
Freie Universität Berlin
Berlin, Germany

Katrin Glinka
Freie Universität Berlin
Berlin, Germany

Franziska Boenisch
Fraunhofer AISEC
Berlin, Germany

Marian Margraf
Fraunhofer AISEC
Berlin, Germany

Sabine Sayegh-Jodehl
Charité — Department of
Psychosomatic Medicine
Berlin, Germany

Matthias Rose
Charité — Department of
Psychosomatic Medicine
Berlin, Germany

ABSTRACT

The application of machine learning (ML) in the medical domain has recently received a lot of attention. However, the constantly growing need for data in such ML-based approaches raises many privacy concerns, particularly when data originate from vulnerable groups, for example, people with a rare disease. In this context, a challenging but promising approach is the design of privacy-preserving computation technologies (e.g. differential privacy). However, design guidance on how to implement such approaches in practice has been lacking. In our research, we explore these challenges in the design process by involving stakeholders from medicine, security, ML, and human-computer interaction, as well as patients themselves. We emphasize the suitability of *reflective design* in this context by considering the concept of *privacy by design*. Based on a real-world use case situated in the healthcare domain, we explore the existing privacy needs of our main stakeholders, i.e. medical researchers or physicians and patients. Stakeholder needs are illustrated within two scenarios that help us to reflect on contradictory privacy needs. This reflection process informs conceptional design rationales and our proposal for *privacy-preserving explanation user interfaces*. We propose that the latter support both patients' privacy preferences for a meaningful data donation and experts' understanding of the privacy-preserving computation technology employed.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; *Scenario-based design*; • **Security and privacy** → *Human and societal aspects of security and privacy.*

KEYWORDS

Privacy preservation, machine learning, user interface, reflective design, conceptional design rationales

1 INTRODUCTION

The value of machine learning (ML) in healthcare is especially observable if patients have rare or highly complex diseases that cannot be studied in 'classical' randomized controlled trials [28]. However, the use of ML raises several privacy concerns (e.g. [14]) since such patients belong to a vulnerable group [20]. Previous research, for example, has highlighted privacy risks associated with the extraction of sensitive information about individuals from trained ML models (e.g. [9]) or the possible identification of specific individuals in the data used to build these models (e.g. [30]). These privacy risks emphasize the urgent need for protecting particularly sensitive data in medical contexts. One approach to mitigate these privacy risks is to apply privacy-preserving technologies (PPT) [15] and, more specifically, privacy-preserving ML (PP-ML) [16]. Specific methods (such as differential privacy (DP) [8]) are employed to generate ML models that provide high utility[1] while preserving privacy preferences of individuals whose data are used for training the ML model. Not only are ML and privacy highly complex concepts in human-computer interaction (HCI) but so is the design itself. When these concepts meet, i.e. when designing for PP-ML, HCI researchers and practitioners need to consider legal measures, technological capabilities, contextual requirements and social norms. One attempt to capture these complex requirements during the development of new PPT is the concept of *privacy by design* (PbD) (e.g. [33, 42]). PbD describes a 'dialogue' to "carry our core values into the future rather than letting them fade from the future" [7]. Researchers have adopted PbD in various settings and contexts. However, Wong and Mulligan note regarding to the HCI context that "the term 'design' and the roles it might play in protecting privacy remain under explored" [41]. Considering Wong and Mulligan's research, we adopt their perspectives[2] on PbD to highlight existing conceptional design approaches in privacy research and discuss how they can support our design of a novel PPT. A holistic understanding of privacy must also be applied in healthcare settings, where technologies should be understood as socio-technical systems that must accommodate multiple, possibly divergent privacy preferences. This situatedness of PP-ML within specific contexts (e.g. medical, educational, political) adds to the complexity of designing for PPT.

In this article, we propose a reflective approach (cf. [29]) when designing privacy-preserving computation technologies, i.e. PP-ML.

[*] peter.soerries@fu-berlin.de

[1] The utility in ML quantifies the correctness of the model's prediction. Typical functions to assess model utility rely on calculating the distance between correct or measurable outcomes for a data point and the model's prediction regarding it.

[2] These PbD perspectives should not be seen as distinct; instead, they overlap.

Mensch und Computer 2021, Workshopband, Workshop on 7. Usable Security und Privacy Workshop

https://doi.org/10.18420/muc2021-mci-ws14-389

This approach is motivated by our understanding of privacy as an individual right and societal value. Our research is motivated and informed by a real-world use case situated in a clinical setting. Here, ML is applied to build a decision-support system that recommends suitable treatments for diabetes patients based on their health conditions. We are collaborating for this use case with stakeholders from the security and ML domain as well as medical researchers or physicians and patients. Our overarching goal is to enable a sovereign data donation process for patients in a clinical setting. In Section 2, we discuss the possibilities offered by PbD in the design process from a software engineering, user-centered design (UCD) and reflective design perspective. In Section 4, we introduce our use case from the healthcare context. We illustrate our stakeholders and their privacy preferences by presenting the design situation in two scenarios. We use the latter (Section 4) as a valuable design technique to describe specific design contexts (cf. [5]) and use *contextual integrity* [22] as an analytical lens to reflect on the specific privacy preferences and explanation needs of our stakeholders. Based on this approach, we introduce conceptional design rationales for *privacy-preserving explanation user interfaces* (PP-XUI) in Section 5.

2 RELATED WORK

The concept of PbD emphasizes several perspectives that suggest incorporating privacy early in the design process [41]. In the following, we describe three perspectives on PbD: secure software engineering, UCD and reflective design. These perspectives are fluidly interconnected, yet, this subdivision supports us in gaining a more structured understanding of the respective research landscapes that equally inform our use case (cf. Section 3).

2.1 PbD from a Secure Software Engineering Perspective

The focus of PbD in software and system engineering often lies on solving a specific problem that relates to the users or the developer or engineers[3] [41]. Another option is for security experts to provide tools or methods in the form of privacy-enhancing technologies (PET) to help developers or engineers address privacy concerns in their development practices. Such PETs aim at “preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system” [38]. A new group of PETs are the so-called privacy-preserving computation technologies that ensure privacy by using the benefits of data processing, for example, ML. Existing PbD strategies in secure software engineering are interwoven into each phase of a cyclic development process[4], taking into account so-called *privacy design patterns*[5]. These strategies translate vague legal norms into concrete design requirements early in the design process. The underlying assumption of this line of research is that users carry out better privacy-enhancing decisions if they have the ‘right’ information and developers or engineers have the ‘right’ tools, for example, a suitable PET [41]. This PbD perspective considers predominantly developers or engineers. It is aimed at providing developers or designers with an adaptable ‘toolbox’ to consider privacy during system design and development and, thus, constitutes a valuable perspective on PP-ML for our use case (cf. Section 3). Nevertheless, it is also important to consider a UCD perspective that shifts the focus to users and actively involves them in the technology design process (e.g. [23]), which is presented in the next subsection.

2.2 PbD from a User-Centered Design Perspective

The UCD perspective often focuses on notice-and-consent mechanisms, which are regulatory mechanisms to legally ensure individuals’ accountability for their own privacy decisions (e.g. [2, 4]). Thus, the primary stakeholders in this perspective are users. The General Data Protection Regulation (GDPR) specifies that individuals, i.e. users, need to express their consent to the processing of personal data without coercion, in a specific situation, and in full knowledge of the facts, for example, by a written, electronic statement [25]. Such notice-and-consent mechanisms intend to bridge an assumed knowledge gap between an institution providing a service and a person using that service [19]. The GDPR deliberately does not provide any design templates or rules for how to bridge this gap effectively. Research has shown that current implementations of notice-and-consent mechanisms, i.e. privacy policies, especially on the web, are too lengthy and challenging to understand (e.g. [3, 32]). As a result, people are unlikely to read these policies [19].

The underlying assumption of a UCD perspective on privacy is that more ‘usable’ privacy information and tools help users to make better privacy-enhancing decisions [41]. Nonetheless, in UCD, it is the designer who ultimately ‘decides’ which (collected) user’s privacy preferences should be considered and are translated into user interface requirements. Thus, the specific world view of a developer or designer materializes in the interaction design or technology solution. As a result, it seems that individual choice is reduced to conformity [36]. By reducing the option of choice, individuals tend to feel disempowered, which also prevents them from becoming responsible actors. In the next section, we introduce *reflective design* as a valuable direction for realizing this objective.

2.3 PbD from a Reflective Design Perspective

For a socially responsible PPT design, developers or designers need to reflect on their “values, attitudes, and ways of looking at the world” [29]. This also evokes *critical* reflection that aims at bringing otherwise “unconscious aspects of experience to conscious awareness” [29]. Reflective design integrates insights from various approaches that are part of the third wave of HCI [11]. It adopts essential ideas, inter alia, from *participatory design* [26, 39], *value-sensitive design* [10], and *critical technical practice* [1]. Various stakeholders[6] have an active role in the design process in reflective design. Developers or designers and users reflect on their practice. Design is used to explore the needs of people and situations in group activities or workshops. During the latter two, developers

[3] With ‘developers or engineers’ we refer to researchers and/or practitioners who have a focus on the computational and engineering aspects of PbD. This intentionally contrasts this group to ‘developers or designers’ who focus more on the design aspects that relate directly to users and interactions.

[4] The phases are described in detail in Hoepmann’s “Privacy Design Strategies (The Little Blue Book)” [12].

[5] A privacy pattern catalog is provided at https://privacypatterns.org.

[6] We differentiate two stakeholder groups: ‘direct stakeholders’ interact directly with the technology and ‘indirect stakeholders’ are indirectly affected by the technology.

or designers can collect and reflect on stakeholders' values and expertise. In addition to efficiency and usability, other social values, such as justice, welfare and human rights, become the focus of the design. The close engagement with all stakeholders (direct and indirect) brings initially unconscious values to the fore. Thus, design is considered as process of inquiry, and developers or designers are researchers situated in a "context of practice" [27]. Through such an approach, the limits and holistic effects of a design are reflected from the beginning in the design process, during which technology is often seen as an approach to support skepticism about and reinterpretation of one's own working [29]. Wong and Mulligan highlight that technologies incorporating these ideas are sensitive to socio-cultural differences and specificities [41]. Such technologies do not rely on a predefined conception of privacy; instead, the concept of privacy emerges within the reflective design process and the engagement with the respective context [41]. Accordingly, privacy is considered to be "[...] a living, continually changing thing, a fluid concept, dependent on socio-cultural factors" [17], which is shaped by different stakeholders and social, cultural and institutional contexts.

In summary, by building on these three perspectives, we have brought to the surface essential ideas for a PPT design and highlight how each PbD perspective considers a specific stakeholder group. In the next section, we describe our real-world use case and how we implemented such a reflective design process therein.

3 USE CASE: PRIVACY-PRESERVING ML IN HEALTHCARE

Based on our discussion of the related work, we introduce our use case which is situated in the healthcare domain and focuses on enabling sovereign data donation for patients. At the start of the project, our diverse stakeholders (patients, medical researchers or physicians, security researchers) confronted us with a dilemma: Medical researchers or physicians emphasized their need for 'more' data to improve the utility of the ML model in order to support better medical decisions. Patients shared their concerns regarding their privacy preferences. Security researchers explained that the need for more privacy might impact the utility of the ML model. In alignment with our reflective design process that integrates the expertise and values of all involved stakeholders, we knew that we could not prioritize only one of these requirements and, thus, needed to find a compromise. Balancing this trade-off required us to reflect on our design process with our stakeholders. Consequently, we decided to use scenarios which finally informed our conceptional proposal of a PP-XUI.

In our use case, ML is applied to build a decision-support system that recommends suitable treatments for diabetes patients based on their personal health data. The ML model training relies on data from a long-term clinical research program based on randomized controlled trials. It includes patients' health conditions (e.g. blood and liver values) and contextual information (e.g. marital status, income). The ML-based decision-support systems have become increasingly important in healthcare. They enable the transfer of cutting-edge knowledge to hospitals which are not research-intensive where limited experiences in this field exist. However, the provision of such decision-support systems to external stakeholders accommodates the risk of personal data becoming public as research has shown that the process of turning training data into a model is reversible (e.g. [9]). This situation led to various propositions in the context of privacy-preserving computation. Cryptographic concepts, for example, can keep the training data secret, or private computing methods can alter the data randomly, thereby making them unrecognizable. One of these techniques is DP [8],[7] which builds on the idea of adding 'noise' to the results of a computation based on certain constraints. DP is implemented in ML libraries such as Pytorch[8] and TensorFlow[9]. We employ the latter library with the Private Aggregation of Teacher Ensembles (PATE) framework for realizing different notions of privacy guarantees [24].[10]

The information flows and stakeholders in the system environment based on PP-ML are shown in Figure 1. We differentiate two main stakeholders: Ⓐ patients (data owners) and Ⓑ medical researchers or physicians (data consumers). According to PATE, the system environment is separated into a private and a public part.[11] The latter contains the public Ⓒ PP-ML model outcome that is trained on anonymized data. From a HCI perspective, we focus on the question how we can support patients in making an informed and sovereign decision when donating their data for research purposes. We envision that patients donate their data based on their individual privacy preferences. The data become part of the model training and are combined with other patients' data. Based on DP, the data privacy is provided in the Ⓓ system based on PP-ML according to patients' privacy preferences. Given the patients' data and the privacy mechanism, the actual PP-ML model provides an outcome which guarantees a specific level of privacy. The resulting PP-ML model can be provided to external medical researchers or physicians without compromising the privacy of the patients. The informed decisions of the patients are materialized in a Ⓔ PP-XUI which conveys how patients' privacy preferences impact the utility of the privacy-preserving computation technology, i.e. it communicates the added value of individual data use. Thus, we accompany the process of data donation with additional information, namely, how patients' privacy preferences impact the utility or accuracy of the trained model, i.e. the PP-ML model outcome for medical research. We suggest relating the individual decision to not only the risks of re-identification but also to the added value of data use.

This proposed concept emerged during the design process, which we present in the next section (cf. Section 4). Building on that, we discuss our proposal of a Ⓔ PP-XUI in detail (cf. Section 5) and introduce open research questions.

[7] In addition to DP, there are methods such as Secure Multiparty Computation [44], Federated Learning [43], and Homomorphic Encryption [35]. An overview of privacy risks in ML and the corresponding protection methods is provided in Liu et al. [18].

[8] For further information, please see https://opacus.ai/.

[9] For further information, please see https://github.com/tensorflow/privacy.

[10] PATE was chosen as an algorithm since it inherently offers a separation between a private and a public data space. It thereby does not only offer high privacy protection but also exhibits potential for intuitive understanding of the underlying mechanism [24].

[11] We purposefully simplified the functionality of PATE, since we focus on the design process rather on the technical realization.

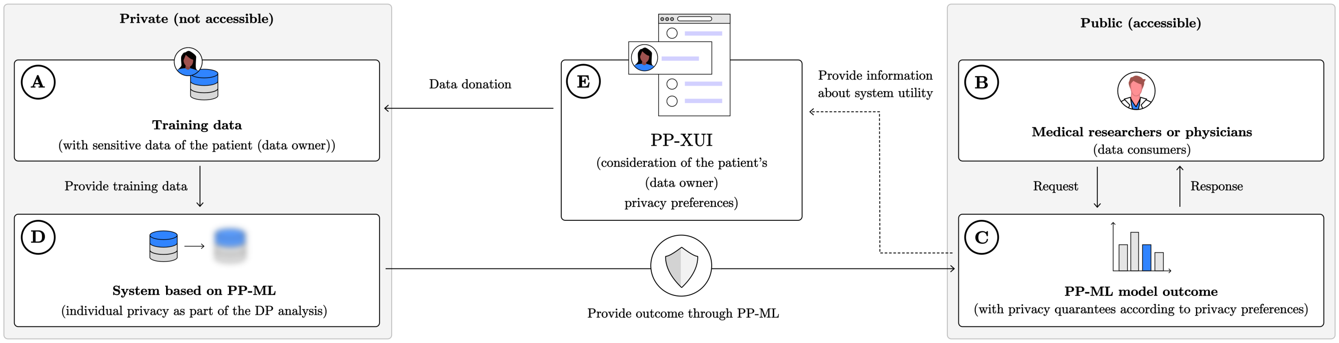
Private (not accessible)
A
Training data
(with sensitive data of the patient (data owner))
Provide training data
D
System based on PP-ML
(individual privacy as part of the DP analysis)
Data donation
E
PP-XUI
(consideration of the patient's
(data owner)
privacy preferences)
Provide outcome through PP-ML
Provide information
about system utility
Public (accessible)
B
Medical researchers or physicians
(data consumers)
Request
Response
C
PP-ML model outcome
(with privacy quarantees according to privacy preferences)

Figure 1: Information flows and stakeholders (Ⓐ patient and Ⓑ medical researchers or physicians) in a system environment based on PP-ML divided into a public and private part. The Ⓓ system based on PP-ML consists of a model that is trained by privatized data which leads to the Ⓒ PP-ML model outcome. Individual privacy preferences can be modified by the Ⓔ PP-XUI.

4 DESIGNING FOR PRIVACY-PRESERVING COMPUTATION

At the beginning of the project, we planned a typical human-centered design process, assuming that the problem was well-defined and that we only needed to explain the functionality of DP to the patients. Instead, we experienced that a lot of questions existed regarding the ML pipeline and the privacy-preserving computation technology used, namely DP, also in our interdisciplinary project team (consisting of experts in medicine, ML, security and HCI). Among them were questions such as "how does DP affect the utility of an ML model in general?" or "how do specific privacy preferences impact the utility of the model?" We realized that in order to design a PP-XUI that communicates the impact of patients' privacy preferences, we needed to consider not only the patients but also the medical researchers or physicians who need to understand the capabilities and limitations of PP-ML. Thus, we had to take a step back and decided to employ Carroll's proposal of a *scenario-based design* [5] to reflect on the needs of our stakeholders by specifying scenarios. A scenario-based design provides "a framework for managing design that considers the nature of design problem solving as it occurs in the context of technology development" [5]. Scenarios are concrete hypotheses about what people using a possible design will do, think and experience. Characteristic elements exhibit a specific setting and actors who typically have personal goals and preferences. Scenarios consist of a plot, i.e. a set of sequences of actions in which actors do something or something happens to them [5]. These scenarios allow us (developers or designers) to experience the design as a 'conversation' with a situation which is comprised of many interdependent components[12] [27]. Such a conversation allows for a reflection on the potential consequences of the design. At the same time, the process of constructing such scenarios has already evoked reflection [5], even though they only provide a partial view of the design.

In the following, we present two scenarios for PPT that went through multiple iterations to capture existing requirements from the perspective of multiple stakeholders comprehensibly. The protagonists of the following scenarios are fictional, yet, the circumstances described are based on well-documented discussions and conversations we have had with experts in the medical, ML, security and HCI fields since the beginning of our project, and privacy concerns discussed from the patient-user perspective. We use these scenarios to uncover a number of conceptional design rationales that resulted in our proposal of a PP-XUI in Section 5. Each scenario is situated in a different context of use. In the future, we plan to use these scenarios as a resource for reflection in co-creation workshops.

4.1 Scenario 1: Using Privacy-Preserving ML

Nicola is a senior physician specializing in nutritional medicine who is finalizing a long-term study about the nutrient habits of people and the likelihood of them developing type 2 diabetes mellitus. She conducts a randomized controlled trial with over 2000 patients and collects data on health conditions (e.g. blood and liver values) and contextual information (e.g. marital status, income). Her colleague Porter, who participates in a global health program, introduces the idea of building a ML-based decision support system on her data that recommends individualized treatments on new patient data. He wants to provide this system to other hospitals on a global scale. Nicola is hesitant because she has not had any experience with ML systems in medicine. Nonetheless, Porter somehow convinces her, and they set up a new research project (funded by the global health program) that needs to be reviewed by the hospital's ethics committee due to the use of personal data. The answer comes immediately; the ethics committee declines the proposal since it does not protect the personal data included sufficiently. Even more concerning for the ethics committee is the fact that data were collected for a specific study and not for building a piece of software. Nicola does not understand this decision; she wants to help people. In her opinion, healthcare is more important than privacy. After pondering about this, Porter contacts his colleague Anna from the medical informatics department and discusses the issue. She recommends the use of PPT. Porter is excited while Nicola is now perplexed. Anna uses words such as anonymization, utility, DP and noise and stresses:

[12] Carroll suggests that a scenario is like a 'soft' prototype since it provides claims on how a potential user will interact with the system and what the user will experience.

"There is an urgent need to apply these technologies. Privacy and medical research are not mutually exclusive." Nicola shakes her head: "Based on my experience, there is a high willingness of patients to donate their data. Why do we need all this security stuff?" After a long discussion with Anna and Porter, she finally agrees to submit the revised proposal to the ethics committee, but her reservation about such technology remains.

4.2 Scenario 2: Individual Privacy Needs

Harper and her family doctor, Toni, have an honest patient-doctor relationship. At her regular checkup, Harper learns about a new global diabetes study. Toni tells her that the study employs a novel application that allows patients to donate their data individually and securely. In the past, Harper has been offered the opportunity to participate in similar studies, but she always deliberately declined. One primary concern is her employer; if they learn about her severe diabetes, it might impact her career opportunities. On a digital rights blog platform, she reads about more extreme cases in which nonconsensual publication of mental health diagnoses led to people being dismissed. As a result, Harper has begun to be more careful about sharing data. She distrusts institutions, no matter how well-known they are or how many guarantees they might provide regarding protecting her data. In addition, the consent forms used by many institutions are overly long and she has difficulty understanding the wording. However, she can discuss every step of her diabetes therapy or ask questions about her data with Toni. He urges Harper to participate in the study because she has been diagnosed with type 2 diabetes mellitus. While considering this option, she wonders: "How do these people ensure that my data are safe?" Toni explains that the data would be anonymized and made available to other medical institutions in such a way that there would be no way for conclusions to be drawn about Harper's person based on her data donation. Toni cannot provide more suitable explanations because he has not received enough information about the study. However, he recommends that Harper takes a closer look at the study website: "Please, read through the study program more carefully. I can imagine that it could be extremely promising for you. Unfortunately, I am just a family doctor, and I do not have the resources to familiarize myself with recent research. Nevertheless, this research program can provide me with valuable insights into treating your disease even better. Think about it again."

5 CONCEPTIONAL DESIGN RATIONALES

Privacy in HCI is understood as both individually subjective and socially situated [20]. This notion of privacy is visible in our use case. In addition to the reflection on privacy concerns during the creation of each scenario (cf. Section 4), we also use an analytical lens to capture existing information gaps better. Accordingly, we use Nissenbaum's framework of *contextual integrity* to explore the needs of our stakeholders in terms of informational norms[13] which relate to a specific individual context [22]. Nissenbaum suggests the contextual integrity is a heuristic or "a model for understanding and predicting reactions to alterations in information practices, particularly those caused by the deployment of information technology" [22]. Thus, when an information technology employed (e.g. PP-ML) considers informational norms, contextual integrity is ensured; otherwise, it is not.

The first scenario is situated in a clinical setting, in which the protagonists aim to use a ML-based decision support system for knowledge transfer in a global health initiative. We can differentiate specific actors: patients are the senders of information and also the information subjects, as they provide their personal data. Nicola and her colleagues are the recipients of information. Connecting both information types, i.e. health conditions and contextual information, can violate the contextual integrity, as finally argued by the ethics committee which declines the proposal due to privacy concerns about inadequately protected patient data. The use of PP-ML is intended to resolve this situation. The second scenario is characterized by mutual trust in a family practice. Harper is the sender of information and also the information subject. However, the receiver of information is not clearly named in this scenario, i.e. who are the users of the system in the global health initiative? By whom is the ML model retrained? How is privacy ensured in the system? From Harper's perspective, the scenario describes a challenging situation since Toni, whom she trusts, tries to convince her to donate data. It is not clear from the scenario to what extent Harper is able to consider her personal values regarding her privacy when donating her data. We can observe different violations of the context-relative informational norms in both scenarios which channel into the same observation: information senders and receivers need to be able to interpret the technology used in order to be able to make an informed decision and understand its effects, i.e. in terms of individual and social impact. This is why we hypothesize that we should not only focus on the notice-and-consent mechanisms, as is often the case in UCD privacy research, but that we also need to consider the PP-ML used. We relate this hypothesis back to the starting point of this article. For this, we consider conceptional design rationales when providing information that explain why and how they are designed (e.g. [13]). Conceptional design rationales help one make an informed decision and build consensus, ensure quality, and elicit perspectives and trade-offs from our stakeholders. We consider techniques that provide transparency about privacy in direct relationship to donated data with the intention of supporting patients in interpreting their data donation in terms of individual privacy preferences.

With reference to Figure 1, our conceptional design rationales consider two main layers: first, the underlying system based on PP-ML and, secondly, the actual PP-XUI. We identify the primary explanation needs[14] of our stakeholders based on the scenarios. In order to balance the trade-off explained in the use case we suggest three conceptional design rationales that we use as the foundation for planned co-creation workshops and will then inform the future design of our PP-XUI: We need to (1) facilitate interpretability regarding the PPTs that are used to ensure data protection (e.g. [6, 21]), (2) communicate the privacy risks and opportunities of data usage through transparency (e.g. [34, 40, 45]) and (3) promote *reflective thinking* through meaningful decision support (e.g. [36]). Only in

[13] Four aspects characterize these norms: context, actors (senders of information, recipients of information and information subjects), attributes (information types) and transmission principles.

[14] 'Explanation needs' highlight different views of stakeholders on ML systems [37] that should be addressed in explanation user interfaces.

this way can we ensure that our stakeholders have an adequate understanding of legal norms, are aware of potential privacy risks and can understand their privacy decisions. These conceptional design rationales aim to accomplish the balancing act between our stakeholders and the underlying PPT. This can be done by promoting an understanding of how privacy-enhancing decisions affect the PP-ML model outcome and how these can be interpreted, and how individual decision making can be critically reflected within a PP-XUI.

6 CONCLUSION AND FUTURE WORK

Our paper explores how multiple stakeholder needs and PbD in the medical, ML, security and HCI domains can be critically reflected in a specific setting. We have developed two scenarios that illustrate different aspects from the perspective of our two main stakeholder groups using a real-world use case from the healthcare domain. We use these scenarios to highlight the consequences or contradictions when designing with privacy-preserving computation technologies. We evaluate our scenarios using contextual integrity, which allows us to understand which informational norms, whether in cultural or social contexts, ensure the expected information flow in terms of privacy [31]. This evaluation leads to the identification of concrete gaps in informational norms. Understanding these gaps in informational norms, in turn, informs a specific trade-off between individual stakeholder needs and their privacy concerns, thus, characterizing conceptional design rationales of PP-XUIs. As a next step, we will reassess the assumptions of our conceptional design rationales by using them as a basis for co-creation workshops. The results from those workshops and additional aspects of our ongoing research will finally inform the development and design of our future PP-XUI.

7 ACKNOWLEDGEMENTS

We thank the reviewers for their insightful comments. This work is supported by the Federal Ministry of Education and Research (grant 16SV8463: WerteRadar).

REFERENCES

[1] Philip Agre. 1997. *Computation and human experience.* Cambridge University Press, Cambridge.

[2] O. Ben-Shahar and C. E. Schneider. 2010. The Failure of Mandated Disclosure. *University of Pennsylvania Law Review* 159 (2010), 647–749.

[3] Mike Bergmann. 2008. Generic Predefined Privacy Preferences for Online Applications. In *The Future of Identity in the Information Society*, Simone Fischer-Hübner, Penny Duquenoy, Albin Zuccato, and Leonardo Martucci (Eds.). Springer US, Boston, MA, 259–273.

[4] Ryan Calo. 2013. Code, Nudge, or Notice? *Iowa Law Review* 99 (2013), 773 – 802.

[5] John M. Carroll. 1999. Five Reasons for Scenario-Based Design. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 3 - Volume 3 (HICSS '99).* IEEE Computer Society, USA, 3051.

[6] Diogo Carvalho, Eduardo Pereira, and Jaime Cardoso. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* 8 (07 2019), 832. https://doi.org/10.3390/electronics8080832

[7] Ann Cavoukian. 2010. Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D. *Identity in the Information Society* 3, 2 (2010), 247–251. https://doi.org/10.1007/s12394-010-0062-y

[8] Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II* (Venice, Italy) *(ICALP'06).* Springer-Verlag, Berlin, Heidelberg, 1–12. https://doi.org/10.1007/11787006_1

[9] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. 2015. Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15).* Association for Computing Machinery, New York, NY, USA, 1322–1333. https://doi.org/10.1145/2810103.2813677

[10] Batya Friedman. 1996. Value-Sensitive Design. *Interactions* 3, 6 (1996), 16–23. https://doi.org/10.1145/242485.242493

[11] Steve Harrison, Deborah Tatar, and Phoebe Sengers. 2007. The Three Paradigms of HCI. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM Press, New York, 1–18.

[12] Jaap-Henk Hoepman. 2018. Privacy Design Strategies (The Little Blue Book). Radboud University.

[13] John Horner and Michael E. Atwood. 2006. Design Rationale: The Rationale and the Barriers. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (Oslo, Norway) *(NordiCHI '06).* Association for Computing Machinery, New York, NY, USA, 341–350. https://doi.org/10.1145/1182475.1182511

[14] Eric Horvitz and D. Mulligan. 2015. Data, privacy, and the greater good. *Science* 349 (07 2015), 253–255.

[15] Peter Hustinx. 2010. Privacy by design: delivering the promises. *Identity in the Information Society* 3, 2 (2010), 253–255. https://doi.org/10.1007/s12394-010-0061-z

[16] Georgios A. Kaissis, Marcus R. Makowski, Daniel Rückert, and Rickmer F. Braren. 2020. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence* 2, 66 (Jun 2020), 305–311. https://doi.org/10.1038/s42256-020-0186-1

[17] Bert-Jaap Koops and Ronald E. Leenes. 2005. Code and the Slow Erosion of Privacy. *Michigan Telecommunications and Technology Law Review* 12, 1 (2005), 115.

[18] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–36.

[19] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4 (2008), 543.

[20] Nora McDonald and Andrea Forte. 2020. The Politics of Privacy Theories: Moving from Norms to Vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376167

[21] Christoph Molnar. 2019. Interpretable Machine Learning. https://christophm.github.io/interpretable-ml-book/.

[22] Helen Nissenbaum. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford University Press, USA.

[23] Donald A. Norman and Stephen W. Draper. 1986. *User Centered System Design; New Perspectives on Human-Computer Interaction.* L. Erlbaum Associates Inc., USA.

[24] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Úlfar Erlingsson. 2018. Scalable Private Learning with PATE. arXiv:1802.08908 [stat.ML]

[25] Oliver Radley-Gardner, H. G Beale, and Reinhard Zimmermann. 2016. *Fundamental texts on European private law / edited by Oliver Radley-Gardner, Hugh Beale and Reinhard Zimmermann.* (second edition. ed.). Hart Publishing, Oxford, UK ; Portland, Oregon.

[26] Douglas Schuler and Aki Namioka. 1993. *Participatory Design: Principles and Practices.* L. Erlbaum Associates Inc., USA.

[27] Donald A. Schön. 1983. *The reflective practitioner: how professionals think in action.* Basic Books, New York.

[28] Ian Scott, David Cook, and Enrico Coiera. 2020. Evidence-based medicine and machine learning: a partnership with a common purpose. *BMJ Evidence-Based Medicine, Online First* EBM Analysis, August 19 (2020), n.a.

[29] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility* (Aarhus, Denmark) *(CC '05).* Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1094562.1094569

[30] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership Inference Attacks Against Machine Learning Models. In *2017 IEEE Symposium on Security and Privacy (SP).* IEEE, San Jose, CA, USA, 3–18. https://doi.org/10.1109/SP.2017.41

[31] Patrick Skeba and Eric P. S. Baumer. 2020. Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1 – 22. https://doi.org/10.1145/3415172

[32] Robert H Sloan and Richard Warner. 2013. Beyond notice and choice: Privacy, norms, and consent. *Journal of High Technology Law* 14 (2013), 370.

[33] Sarah Spiekermann. 2012. The Challenges of Privacy by Design. *Commun. ACM* 55, 7 (2012), 38–40. https://doi.org/10.1145/2209249.2209263

[34] Simone Stumpf, Vidya Rajaram, Lida Li, Margaret Burnett, Thomas Dietterich, Erin Sullivan, Russell Drummond, and Jonathan Herlocker. 2007. Toward Harnessing User Feedback for Machine Learning. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) *(IUI '07).* Association for Computing Machinery, New York, NY, USA, 82–91.

https://doi.org/10.1145/1216295.1216316

[35] Xiaoqiang Sun, Peng Zhang, Joseph K Liu, Jianping Yu, and Weixin Xie. 2018. Private machine learning classification based on fully homomorphic encryption. *IEEE Transactions on Emerging Topics in Computing* 8, 2 (2018), 352–364.

[36] Arnout Terpstra, Alexander Schouten, Alwin Rooij, and Ronald Leenes. 2019. Improving privacy choice through design: How designing for reflection could support privacy self-management. *First Monday* 24 (07 2019). https://doi.org/10.5210/fm.v24i7.9358

[37] Richard Tomsett, Dave Braines, Dan Harborne, Alun D. Preece, and Supriyo Chakraborty. 2018. Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems. *CoRR* abs/1806.07552 (2018), 8–14. arXiv:1806.07552 http://arxiv.org/abs/1806.07552

[38] G. W. Van Blarkom, John J. Borking, and JG Eddy Olk. 2003. Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague* 198 (2003), 14.

[39] John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. 2013. Configuring Participation: On How We Involve People in Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 429–438. https://doi.org/10.1145/2470654.2470716

[40] Sandra Wachter, Brent D. Mittelstadt, and Chris Russell. 2017. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *CoRR* abs/1711.00399 (2017), 52. arXiv:1711.00399 http://arxiv.org/abs/1711.00399

[41] Richmond Y. Wong and Deirdre K. Mulligan. 2019. Bringing Design to the Privacy Table: Broadening "Design" in "Privacy by Design" Through the Lens of HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300492

[42] Heng Xu, Robert E. Crossler, and France Bélanger. 2012. A Value Sensitive Design Investigation of Privacy Enhancing Tools in Web Browsers. *Decision Support Systems* 54, 1 (2012), 424–433. https://doi.org/10.1016/j.dss.2012.06.003

[43] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 2 (2019), 1–19.

[44] Chuan Zhao, Shengnan Zhao, Minghao Zhao, Zhenxiang Chen, Chong-Zhi Gao, Hongwei Li, and Yu-an Tan. 2019. Secure multi-party computation: Theory, practice and applications. *Information Sciences* 476 (2019), 357–372.

[45] Verena Zimmermann and Karen Renaud. 2021. The Nudge Puzzle: Matching Nudge Interventions to Cybersecurity Decisions. *ACM Transactions on Computer-Human Interaction* 28, 1 (Jan 2021), 7:1–7:45. https://doi.org/10.1145/3429888