User Acceptance Criteria for Privacy Preserving Machine Learning Techniques

Sascha Löbner
sascha.loebner@m-chair.de
Goethe University
Frankfurt am Main, Germany

Sebastian Pape
sebastian.pape@m-chair.de
Goethe University
Frankfurt am Main, Germany

Vanessa Bracamonte
va-bracamonte@kddi-research.jp
KDDI Research, Inc.
Saitama, Japan

ABSTRACT

Users are confronted with a variety of different machine learning applications in many domains. To make this possible especially for applications relying on sensitive data, companies and developers are implementing Privacy Preserving Machine Learning (PPML) techniques what is already a challenge in itself. This study provides the first step for answering the question how to include the user's preferences for a PPML technique into the privacy by design process, when developing a new application. The goal is to support developers and AI service providers when choosing a PPML technique that best reflects the users' preferences. Based on discussions with privacy and PPML experts, we derived a framework that maps the characteristics of PPML to user acceptance criteria.

KEYWORDS

Privacy Preserving Machine Learning, User Acceptance, Privacy-by-design

1 INTRODUCTION

With increased implementation of machine learning (ML) in daily applications, such as self-driving cars or voice assistants, the user acceptance of ML systems has become an important criteria for developers and service providers.

However, a significant percentage of users is concerned about privacy [32] which has a negative impact on the user acceptance of the corresponsing ML system. Nowadays, companies can also benefit from meeting the privacy requirements of the users by using it as a unique selling point, thus gaining market advantage towards competitors. One way to address the users' privacy concerns is to make use of a privacy preserving machine learning (PPML) technique to provide the requested service in a privacy-friendly manner. In many cases, the existing PPML technique requires a trade-off between privacy and other properties such as performance [35, 42] or accuracy [54]. An additional challenged is that in general, users lack a deeper understanding about the technologies used. Thus, users can not be asked directly, which PPML they would prefer and the determination of ”the best“ PPML technique for a certain use case becomes challenging.

In general, investigations of end-user acceptance are not new. The Technology Acceptance Model (TAM) was introduced by Davis [15] already in 1985 and is based on the theory of reasoned action [20] and the theory of planned behavior [1]. With the evolution of TAM, privacy concerns were also considered to play an important role for some systems and can prevent users from adopting a system. However, the investigation of User Acceptance Criteria (UAC), in particular for Privacy Enhancing Technologies (PETs) is quite sparse [7, 24]. Even if a developer or service provider aims to deliver a privacy preserving service, it is not clear how it is related to the user acceptance [10]. Regarding ML services, the aim of PPML is to allow the training of such models while keeping at the same time the data of the input parties (data subject) private [2]. While it is already a challenge to identify the best suitable PPML technique from a technical point of view [29, 39], it is even harder to assess which technique has the best end-users acceptance. It is widely recognised that knowledge [11, 37] and privacy literacy [23] influence the users' privacy concerns, and thus the acceptance of the service. Since self reported knowledge of users oftentimes does not match their actual knowledge [25] and due to the complexity of PPML, educating users might be not appropriate. Therefore, it is important to provide some guidance to developers and service providers which UAC are influenced by different PPML characteristics. While also other criteria such as familiarity with PPML technology or maintanance costs exist, UAC can be (one of) the most important aspects to consider when deploying a new system.

Our contribution is a first step on answering the question how to include the users' PPML preferences into the privacy by design process. To achieve this we provide a mapping that allows developers and service providers to conclude about preferred PPML characteristics – even if the users are unfamiliar with the specific PPML techniques. Thus, we aim to support developers and AI service providers when choosing a PPML technique that best reflects the users' preferences. Many of the identified PPML characteristics will not only depend on a single technique, but may depend on a combination of them or on other characteristics of the architecture. Thus, we provide the first step for:

i) A structured discussion on the end-user perception of PPML techniques.
ii) A structured decision support for developers and service providers to identify the PPML techniques with a good (expected) user acceptance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ARES 2023, August 29-September 1, 2023, Benevento, Italy

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0772-8/23/08…$15.00
https://doi.org/10.1145/3600160.3605004

2 RELATED WORK

In this section, we provide an overview of related literature that investigates the users' concerns and behaviour when interacting with PETs/PPML techniques.

Research on the users' expectations towards Differential Privacy (DP), using vignette scenario surveys, identified that users are concerned about the risk of information disclosure. Moreover, the willingness to share data is increased if the risk of information disclosure is reduced [14]. Padyab and Ståhlbröst [36] found a reluctance of users to integrate PET-tools to protect their privacy in the internet because the users are not involved enough in the evaluation and design process of technologies. They conclude that further research on evaluation criteria for PET tools and user behaviour is required to increase the actual adoption of PETs. This emphasises the importance of analysing user concerns towards PPML. Regarding the attitude towards PET tools, PET users exhibit a higher level of surveillance concern by a privacy enhancing tool compared to a non-privacy tool [11]. User concerns in location obfuscation scenarios, where also a lack of privacy awareness was found might be counteracted by providing explanations and visualisations to the user [12]. Zhang et al. [53] investigated how comfortable people are with the presence of cameras in different video analytic scenarios. They found a lack of awareness by the participants and a need for improved transparency, enabling the user to decide about the collection and processing of such data. Moreover, they found a diverse set of preferences what underpins the need to empower users to manage their privacy decisions. Research on users' information processing of privacy risks in terms of design aspects of data exposure visualisation found that users' understanding of which type of personal data is shared by an app has still to be improved [50]. Boulemtafes et al. [9] proposed a multilevel taxonomy for developers to categorise PPML techniques based on privacy preserving tasks and key technological concepts. They find that a successful solution depends on several constraints and might require several PPML techniques.

To the best of our knowledge, no work focusing on the user acceptance of different PPML techniques has been brought up so far. While we could identify investigations of user acceptance and understanding for single technologies, or the comparison of PPML techniques from a technical point of view, an overall framework, comparing different technologies with regard to certain UAC is missing. Compared to existing frameworks that analyse privacy attitudes of users towards PET tools, we aim for a framework to support developers, deciding which PPML technique to use by mapping the UAC to PPML characteristics.

3 METHODOLOGY

In this section, we have a look at the overall PPML environment that we are investigating, describe our research questions and explain how we address them.

3.0.1 Research Questions: We investigate which PPML technique a user would prefer to be implemented for a certain ML service that requires the disclosure of personal data to the PPML service chain. Since all PPML techniques aim to improve privacy a distinction is made by analysing technical differences. However, it is difficult to directly obtain users' preferences for a technological application (PPML) where specific technical background is required. A solution would be to educate users about the technologies what is expected to be quite difficult due to the complexity of this topic. With our approach, we aim to provide a structured decision support for developers and service providers to identify the PPML techniques with good (expected) user acceptance. Our research questions are:

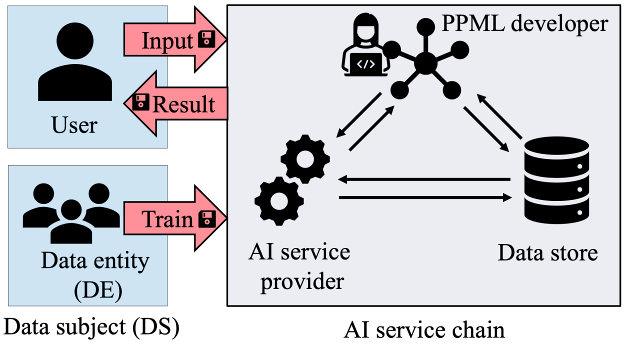
R1: What are the user acceptance criteria a user can notice?
R2: What are PPML characteristics that separate the different PPML techniques from each other?
R3: Which of the user acceptance criteria can be influenced by the PPML characteristics that separate the different PPML techniques from each other?

3.0.2 Defining the PPML Environment: Since different entities have a different perspective, a common understanding of the PPML environment is necessary for our later discussion (see Figure 1). The entities and their interaction were elicited from literature, e.g. [5, 18, 45], the GDPR, and the expert discussions. As potential *data subject* we aggregate two different types of natural persons. Both types can be a data subject as defined in the GDPR, if personal data is involved in the data sharing. First, a user[1] as a natural person using an AI service as a service customer of the AI service provider. The user provides input data for a service requested and receives a result that is calculated based on the input data. If the input data is used/shared with other entities we denote it as user data. Second, the people who provide data for the training of the PPML model are called data entities. A user can be a data entity if she is providing data for model training. We exclude a user providing private data about another user because this does not change the choice of the PPML technique. In general, data can be shared by another service. For the data subject, the AI service chain is a blackbox. An *AI service chain* can consist out of multiple tasks. We picked the most important three tasks and assigned them to three entities. First, the PPML developer was commissioned by the AI service provider to create a PPML model for their users. Creating a PPML model is not necessarily a one time task, but can happen continuously. The data required by the PPML developer can (a) be taken from the data store containing data from data entities not known to the AI service provider or (b) is provided/created by the AI service provider from user data. In the latter case, the user becomes also a data entity. Based on this, the PPML developer creates the model for the AI service provider who uses it in the user relationship. The data store can also be responsible for storing user data from the AI service provider or receive data from the PPML developer. The three tasks (1) providing the service, (2) developing the PPML model and (3) storing data/creating a database can be performed by a single entity.

3.0.3 Expert Discussions, Mapping, Evaluation, and Pruning: From the related work and the existing models, we created two initial groupings with collected attributes as a common starting point for the expert discussions (see Figure 2). In total, there were 32 participants consisting of 18 disjunct experts. All expert groups consisted out of 7-9 junior and senior researchers with minimum two senior

[1] Art. 3, AI Act [4] (proposed EU law) defines user as: "[...] any natural or legal person, public authority, agency or other body using an AI system under its authority [...]".

Figure 1: PPML environment including the existing entities and data relationships.



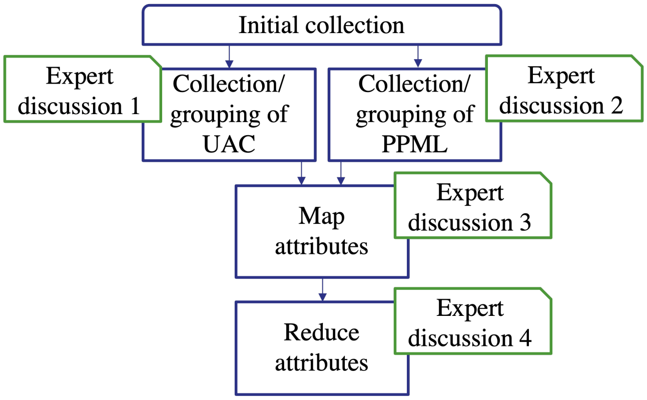
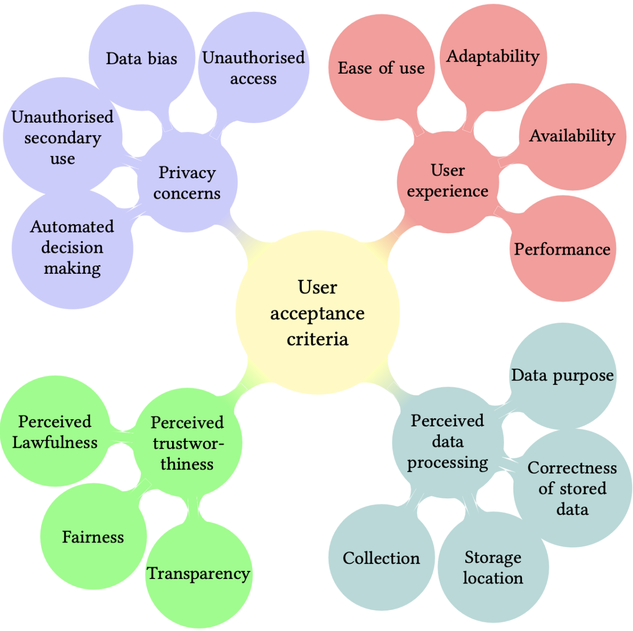


Figure 2: Flowchart of collection, mapping and reduction process

researchers. We had 9 experts in discussion (1), 7 experts in discussion (2). To become an expert in discussion (1), a background in privacy and user acceptance models, and for discussion (2), a background in ML and PPML development was required. Some of the participants from academia and enterprises were recruited from the project CyberSec4Europe under Agreement 830929 of European Union's Horizon 2020 research and innovation program. Discussion (1) and (2) were starting from the respective initial collection and regarding the experts independent without any overlap. In discussion (3) we had 7 experts and in discussion (4) we had 8 experts, from the previous groups. All discussions took approximately one hour. In the discussions we handed out the attributes with the respective descriptions. A brief introductory presentation explaining the scope of discussion and an attribute overview were provided. Then participants had time to read the attribute descriptions carefully. Shared notes and comments were collected during the reading time to not overlook any argument. This was followed by a discussion in the whole group that was structured by the collected arguments. As a starting point of discussion (3), all three authors independently mapped the attributes from the previous groupings. We put more weight on the justification for a connection then the amount of hits so that a connection can be included if one researcher has a valid argument and all three agreed on it. To evaluate the strength of our agreement, we calculated the Fleiss Kappa [21] that is corrected for the random match of evaluators for the user and for the data entity individually. For the user, we achieved a Fleiss Kappa of 0.5038 and for the data entity a Fleiss Kappa of 0.5336. Landis and Koch [27] define Kappa within the boundaries from $0 - 1$ with $0 - 0.2$ indicating a poor agreement and 0.81 an almost perfect agreement. For an agreement of $0.41 - 0.6$ they evaluate the strength of agreement as moderate. This highlights the difficulty of mapping UAC and PPML characteristics. Thus, the score still shows a good validation of our results. To prune the framework, we removed attributes with no link between UAC and PPML characteristics, then merged attributes where connections were highly similar. The resulting mapping was evaluated in two expert discussions (3) with the previous experts. We split the group in the 3rd round of discussions to increase the speaking time each participant. In discussion (4) we validated the reduced framework. The experts' arguments are included in the results and the discussion of the mapping. This research was reviewed by the ethics committee at our institution and was deemed to be out of scope. The project has been classified as ethically acceptable.

Figure 3: Grouping of User Acceptance Criteria



4 RESULTS

This section presents the attributes that were collected and sorted during the expert discussions: *User acceptance criteria* and *PPML characteristics*. Finally, we map and prune both mindmaps to further investigate research question R3.

4.1 User Acceptance Criteria

15 attributes out of the initial 21 UAC were selected as a result of the expert discussions (cf. Fig. 3, Tab. 1).

The expected effect for the user acceptance for each attribute is given: positive (+), negative (-) or scenario dependent (*). For better structure, UAC are organised into 4 classes.

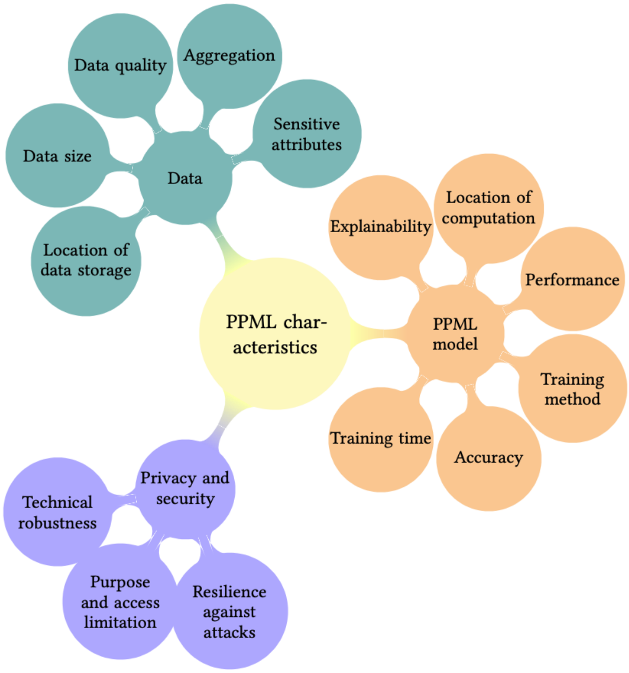
Privacy concerns (-) are found in many models, i. e., the IUIPC [31] or APCO [6] models. Also the personal privacy experience of a user has an impact on the individual privacy concerns [6]. *Automated decision making (ADM) (*)* describes the concern of the process getting out of hand with people treated rather as numbers than individuals. It comes along with the concern of insufficient ways to disagree to ADM also known as "reduced judgement" [43]. *Unauthorised secondary use (−)* describes the concern of a misuse of data initially collected for a certain purpose, for a secondary purpose without authorisation [43]. *Data bias (−)* is the concern of discriminating results by unrepresentative data [17]. According to the Ethics Guidelines for Trustworthy AI (in the following EGTAI) [13] biases can lead to unintended discrimination, prejudice or unfair competition. *Unauthorised access (-)* is the concern of access of personal data to unauthorised people [17, 43] leading to problems in confidentiality and integrity [49].

User Experience (+) according to ISO 9241-210 [16] is: "A person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service.". It is anti-proportional with the time a user requires to use the technology [47]. *Ease of use (+)* is the expected effort associated with the use of PPML [34, 47]. The assessment should consider demographic details such as age, gender and also focus on users with disabilities in all societal groups [13]. *Adaptability (−)* is the concern that a system cannot be adapted to a change in context, e.g., locations or time of the day [48]. *Availability (+)* is described as the level to which a user can successfully access a certain technology when required [49]. *Performance (+)* is the extent to which the use of a technology will benefit certain activities [34, 47].

Perceived data processing (*) is more focused on the perception of the technology compared to privacy concerns. *Collection*(*) is the concern that massive amounts of personal data are collected and stored [43, 52]. Users want to know when and how data was collected, and when the consent was given [17]. *Data purpose (*)* for personal data collection has to be legitimate, explicit and specified [19]. While generally providing the purpose has a positive effect, the effect of unrelated use is ambiguous [17, 44]. *Correctness of stored data (*)* is the concern that the stored data exhibits errors or incorrect user data, e.g., caused by insufficient protection [43] or incorrect data collection. *Storage Location (*):* A transfer of data across geographical borders might reveal personal data and is thus regulated in Article 44 ff. [19]. Physical storage location can be local or with a cloud provider who has full data control and can perform malicious tasks [40].

Perceived trustworthiness (+): is defined as the perceived level of a trustees' trustworthiness, influenced by ability, integrity and benevolence [33]. Additionally, trustworthy AI is the interaction of lawfulness, robustness and adherence to ethical principles [13]. *Perceived lawfulness* (+) is the concern of a user that data is not processed lawfully. Not lawful behaviour or even the suspicion thereof is assumed to have a negative effect on the user acceptance. *Fairness* (+) is mentioned in e.g. Article 5 and 14, GDPR. According to Malgieri [30] fairness is a substantial balancing of the involved parties. It is separated from lawfulness and transparency by not being a legal construct. Fairness aims to mitigate situations of unfair imbalances, where the data subject feels vulnerable. *Transparency* (+) is found in Article 14, GDPR includes the right to receive meaningful information about the logic when automated decision-making or profiling is used. Insufficient information about an AI service can lead to unsatisfied users [17].

Figure 4: Grouping of PPML Characteristics



4.2 PPML Characteristics:

We have elicited 14 attributes directly related to PPML, represent by 3 classes (cf. , Fig. 4, Tab. 1). Since PPML and the ML model are nested, we have enriched the collection with attributes that are used in ML literature to compare ML models [8, 26]. We focused on attributes that differentiate PPML techniques.

Data: This node, contains all data related attributes. *Data size* is defined by the number of rows and columns of a dataset. Some PPML techniques require different sizes of data to achieve a high accuracy. *Data quality* is an important factor that might restrict the choice of possible PPML techniques because not all algorithms can calculate with noisy or missing data. *Aggregation* of data during the de-identification or distribution between entities can have a negative impact on accuracy and utility [29]. *Sensitive attributes* can identify a person. Following Article 4, GDPR, a person can be identified (in)directly by reference to an identifier [19]. In our definition, an attribute is sensitive if it identifies a data subject as member of a certain group, e.g., gender, age or race [51]. Another derived risk is *Location of data storage.* Large amounts of data which cannot be stored within a company might be outsourced to a third party [5].

PPML model: *Explainability* from the technical perspective requires human understanding and traceability of the PPML model decisions [13]. Good *explainability* can only be achieved if considered in each step of the model design [28]. *Location of computation* describes the different model structures, e.g., collaborative or individual learning [9]. Learning structures can be derived from horizontally and vertically distributed data, [41], e.g., server-based or -assisted approaches [9]. *Training method* depends on the task and data structure. Frequently used methods are supervised learning and unsupervised learning. Further training problems are e.g., reinforcement or semi-supervised learning. Training tasks are, e.g., classification, regression, anomaly detection, or language processing [8]. *Accuracy* is the most important factor for measuring the effectiveness of a PPML model [9, 26]. The appropriate accuracy metric has to be chosen from a variety to prevent masking bad model utility [26]. *Training time* determines the adaptability of a PPML model, e.g., in FL [22] and depends on the model's computational complexity [5, 9]. *Performance* is influenced by many factors, e.g., computational complexity, communication overhead [9, 22, 55], and the fallback strategy [13]. Thus, we define performance as the overall run time of the PPML service from the user request to the final results.

Privacy and Security: This node collects attributes directly influencing privacy and security of the PPML model. *Resilience against attacks* should prevent model or data manipulation, corruption or leakage by malicious actors [13]. Although PPML techniques try to overcome possible weaknesses, specific attacks trying to undo the de-identification or shut down the whole system exist. *Purpose and access limitation* describe the probability of revealing personal data from the model. Tanuwidjaja et al. [45] differentiate between privacy of model, client and result. A combination of PPML techniques might be necessary to address all privacy risks [9]. *Technical robustness* has to be considered during the whole PPML life cycle and the model has to be evaluated in the respective context. Models must be reliable, robust against changes in hardware, software, and interacting parties and minimise unintentional or unexpected behaviour [13].

4.3 Mapping

We examined the influence of PPML characteristics on the UAC for single users and both user and data entity (see Table 1). The mapping was developed from 3 independent mappings by the authors. It was validated, pruned and re-validated in the expert discussions. We present the most interesting findings below:

Privacy concerns in *automated decision making* primarily affect the user of a PPML service. Two approaches can reduce the user's concerns: *explainability* increasing understanding of why a certain decision was made, and *accuracy* ensuring the prediction is always correct. Thus, we assessed *data quality* and *accuracy* as most important factors. The experts argued that a wrong classification based on bad input data can be circumvented with high *data quality*. They also suggested to extend *explainability* with a communication channel, enabling users to challenge results or report a bias. *Data bias* we evaluate as a user-specific concern. According to the experts *training methods* and pre-processing can overcome biased data. Another approach is training AI to make fair decisions despite a bias [46] without reducing the accuracy. However, this is difficult to implement. *Explainability* for early identification of data bias and high *data quality* are also important countermeasures. Regarding *unauthorised secondary use* and *unauthorised access*, the *location of data storage*, *purpose and access limitations*, besides others were linked as they address the users' concern of e.g, a cloud provider accessing private data. Further mitigation strategies are first, distributed or local model architectures, moving control to the users [40]. Second, *aggregation* of data during collection and third, occurrence of fewer *sensitive attributes*. *Resilience against attacks*, and *technical robustness* ensuring that the model behaves reliable, prevent *unauthorised access*[13]. We agreed that *unauthorised secondary use* is related to intentionally using data for another purpose, thus attributes like *resilience against attack* do not apply.

Regarding **user experience**, *ease of use* was highlighted. Especially high *accuracy*, low *training time* for on device services and fast *performance* have a huge impact[3]. Experts emphasised that *data quality* is important for mandatory services. A system which cannot deal with small errors in data submission/collection, e.g. spaces in IBANs, can cause frustration. *Adaptability* is only connected to *accuracy* since a good accuracy testing should cover all real world influences e.g. different locations [48]. This connection was also emphasised by the technical experts. Strongly connected to all PPML characteristics except **privacy and security** is *performance* indicating the speed at which a user receives the result. Regarding *availability*, this includes a fallback strategy and fault tolerance. Also the *location of computation* can be important in case of a broken cloud connection. Trade-offs in *performance*, e.g., high accuracy increasing training time and computational overhead and reducing achieved privacy, exist. Thus, developers should carefully weight these attributes to match the user's expectations.

Perceived data processing is strongly connected to the PPML characteristic **data**, especially *location of data storage* and *sensitive attributes*. Regarding *collection*, *explainability* can help the user to understand the origin of data. For the UAC *collection* and *data purpose* linked to *purpose and access limitation* most connections are overlapping. The experts emphasised that the principles of data minimisation and storage time limitation (*data purpose*) by limiting how long data is stored/used impacts the *training method* and thus also accuracy. It was also noticed that an overdependent on old data can mask recent developments. To ensure the *correctness of stored data*, *input data quality* is a key factor. In general, wrong data causes less harm if no *sensitive attributes* are included. For the *data purpose* we highlight the importance of the *training method* because not every method is suitable for every purpose.

All UAC in **perceived trustworthiness** were elicited from the GDPR. The experts stressed that an appropriate level of security is frequently mentioned in the GDPR, thus we connected *resilience against attacks* with *perceived lawfulness* for both, user and data entity. Strongly connected to this is also the *technical robustness*. But also *explainability*, *location of data storage*, and *computation* are essential to consider because they strongly influence the *lawfulness* of the service. Regarding *fairness*, especially *data biases* should be avoided, resulting in all attributes from *data bias* being included in *fairness*. To achieve *fairness*, reliability and minimisation of unintended outcomes are crucial, thus we included *technical robustness*. We incorporated *training methods* since those can deal with *biased data*. For the UAC of *transparency*, *explainability* is the most important factor increasing the users understanding of the computed

Table 1: Mapping of User Acceptance Criteria and PPML Characteristics

		PPML Characteristics													
		Data					Model						P & S		
User Acceptance Criteria		1. Location of data storage	2. Data size	3. Data quality	4. Aggregation	5. Sensitive attributes	1. Explainability	2. Location of computation	3. Training method	4. Accuracy	5. Training Time	6. Performance	1. Resilience against attacks	2. Purpose and access limit.	3. Technical robustness
Privacy concerns	PC1. Automated decision making			👤			👤			👤					
	PC2. Unauthorized secondary use	👤+		👤	👤	👤+		👤+						👤+	
	PC3. Data bias			👤			👤		👤						
	PC4. Unauthorized access	👤			👤	👤+		👤		👤			👤+	👤+	👤
User experience	UX1. Ease of use			👤						👤	👤	👤			
	UX2. Adaptability									👤					
	UX3. Availability							👤				👤			👤
	UX4. Performance	👤	👤	👤				👤	👤	👤	👤	👤			
Perc. data processing	DP1. Collection	👤+	👤		👤+	👤+	👤							👤+	
	DP2. Data purpose	👤+	👤		👤+	👤+			👤+					👤+	
	DP3. Storage location	👤+				👤+		👤+							
	DP4. Correctness of stored data	👤+		👤+		👤+									
Perc. trustw.	PT1. Perceived lawfulness	👤+			👤+	👤+	👤	👤+					👤+	👤+	👤
	PT2. Fairness			👤		👤+	👤		👤	👤					👤
	PT3. Transparency						👤	👤							

P & S - Privacy and Security 👤 - User 👤+ - User and Data Entity

result. Finally, the *location of computation* is relevant as users might not understand complicated architectures, e.g., FL.

5 DISCUSSION

In this section, we evaluate and discuss our research questions, discuss difficult or interfering connections, and give an outlook on the future roadmap.

Regarding **R1**, we collected 23 UAC that were structured into four groups and finally reduced to 15 criteria, iteratively including the experts attributes and arguments. While a variety of possible attributes exists, one of the key challenges is to pick the right definition for similar but slightly different terms. Thus, we focused on a clear distinction from other attributes. We find that the attributes in **privacy concerns** and **perceived data processing** are likely to negatively influence the UAC. Besides this, **user experience** can overall positively influence the user acceptance. Balancing decisions are required for inverse effects, e.g., between user **experience** and **privacy concerns** based on the scenario. Again, the representation for the criteria is at this step just a collection and not conclusive but might be extended with the future developments, e.g., the proposal of the Artificial Intelligence Act [4].

Regarding **R2**, 21 PPML attributes were collected that separate the different PPML techniques. Those were finally reduced to 14. The experts, clarified that the use of PPML is highly scenario-dependent. Thus, not only characteristics from PPML but also ML, e.g., the *training method* are included. Again, the importance of existing trade-offs between attributes was highlighted, e.g., the accuracy privacy trade-off. Although out of scope for this framework, budget constraints should be considered. Overall, the experts found the collection of PPML characteristics useful. The attributes *data utility* and *upkeep* were added in discussion (1) but removed in discussion (4), as well as *implementation effort*, since no mapping was made. Disagreements among experts emerged when defining *performance*. On the one hand, performance can be analysed per component. On the other hand, performance can be treated as the overall performance of the application; our agreed-on definition. The final framework was presented to the experts again in discussion (4), and all definition are agreed to be valid.

Regarding **R3**, UAC that can be influenced by the PPML techniques were identified. From a user's perspective, changes by the developers have to be comprehensible, thus too complex connections were removed. A challenge when aiming to increase the user acceptance are user trade-offs, e.g., between *availability*, *performance* and *accuracy*. Technically, the *availability* and *performance* of a service can often be increased by using approximations instead of complex models but this can cause wrong classifications or *data biases*. Moreover, the experts raised that users have different sensitivities of privacy. What works well to increase the users' acceptance in one use case does not need to work in another. E.g., for the *storage location* it is not clear, whether data is stored more securely at a cloud service or at the user's device since the result of the trade-off depends on the *trustworthiness* of the cloud service, the security, and control of the device [38]. Finally, this mapping builds the base for a structured decision support for developers and service providers to identify the PPML with a high (expected) user acceptance by providing user relevant and technically alterable attributes.

5.1 Evaluation

To evaluate the relevance we have compared the final, expert validated mapping with the EGTAI [13], GDPR [19], and proposed AI Act [4]. From 15 UAC, 3 are not covered by the EGTAI (UX3, DP1, DP3), 5 not by GDPR (PC3, UX1 - UX4) and 7 not by the AI Act (PC4, UX1 - UX3, DP2, PT1, PT2). Generally, the EGTAI has a more human centred approach, taking ease of use and adaptability into account. The GDPR has a strong focus on data protection including collection and processing, but ML related issues are not covered explicitly. The AI Act aims to fill this gap, covering specific AI related attributes. From the 14 PPML characteristics 6 are not covered in the EGTAI (Data1, Data2, Data4, Model2, Model5), 6 not by GDPR (Data2, Data4, Model3, Model4, Model5, Model6) and 3 not by the AI Act (Data2, Model3, Model5). Overall, the AI Act enriches the GDPR with a special focus on AI and biometrics. The EGTAI highlights the need for explainable AI but AI Act and GDPR mention it rarely. Although data size, training method and time are not covered, these attributes are highly relevant to the developers because they strongly influence the application. Since none of the regulations and guidelines covers all attributes, this emphasises the need to support developers with a structured mapping.

5.2 Impact

We contribute to the discussion of the user perception of different PPML techniques by providing an expert validated collection of UAC criteria that is mapped with PPML characteristics. The discussion on user perception of PPML is relevant for policy makers, future standardisation and is still an open question in research. Moreover, our framework helps developers understand how UAC can be influenced by PPML characteristics in an application, with the aspiration to best match user preferences. This is relevant because with accepted PPML applications, privacy can be used as a competitive advantage against competitors.

5.3 Limitations

A limitation of this work is the focus on attributes that can be perceived by users as an UAC and at the same time be influenced by the developer. With our approach we only take a look at the user perspective but for a complete picture, developers have to uphold companies' interests and comply with legal regulations. Thus, our mapping is not a standalone guideline but provides a first evidence what to consider when choosing a PPML technique for a certain scenario. Nevertheless, comparing PPML attributes in different scenarios, e.g., *resilience against attacks* is not trivial and complicated by the accuracy privacy trade-off. Our mapping is a starting point for a user oriented PPML design.

5.4 Future Work

In future we aim to provide with our framework a score for a certain PPML solution based on the users' acceptance criteria. To achieve this, the framework will be complemented by a comparison of the most common PPML techniques based on our elicited PPML characteristics. How to use the framework will also be showcased in the next step of our research roadmap. We also plan an evaluation of our framework by users and experts to validate usability and utility of our solution.

6 CONCLUSION

This work contributes to the discussion of user perception of PPML by providing an expert validated mapping of user acceptance criteria and the influencing characteristics of PPML. With this mapping we aim to provide developers and service provider with a better understanding of how they can address users' concerns when using PPML services that rely on users' sensitive data. Furthermore, we aim to increase the users' trust in ML applications and at the same time to meet their need for privacy preservation without overwhelming the users with technical terms and representations. Developers should always take the users' needs into account from the beginning of the ML lifecycle when designing a PPML model. This is relevant because with accepted PPML applications, privacy can be used as a competitive advantage against competitors. Our framework also contributes to the discussion on user perception of PPML and thus also has an impact on policy makers, future standardisation and current research. In this work, to validate our approach we have used discussions with privacy and security experts and an evaluation of key relevant regulations and guidelines. The next steps are an attribute ranking by users and the comparison of the most common PPML techniques based on our framework.

ACKNOWLEDGMENTS

This work was supported by the European Union's Horizon 2020 research and innovation program through the project CyberSec4Europe under Agreement 830929. We especially thank the partners Davy Preuveneers of Katholieke Universiteit Leuven and Liina Kamm from Cybernetica for their expertise and participation in several discussions.

REFERENCES

[1] Icek Ajzen. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2 (1991), 179–211.

[2] Mohammad Al-Rubaie and J Morris Chang. 2019. Privacy-preserving machine learning: Threats and solutions. *IEEE Security & Privacy* 17, 2 (2019), 49–58.

[3] Mohammed A Al-Sharafi, Ruzaini A Arshah, EA Abo-Shanab, and N Elayah. 2016. The effect of security and privacy perceptions on customers' trust to accept internet banking services: An extension of TAM. *JEAS* 11, 3 (2016), 545–552.

[4] Artificial Intelligence Act. 2021. Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence and amending certain Union legislative acts. *EUR-Lex* (2021).

[5] Md Momin Al Aziz, Md Nazmus Sadat, Dima Alhadidi, Shuang Wang, Xiaoqian Jiang, Cheryl L Brown, and Noman Mohammed. 2019. Privacy-preserving techniques of genomic data—a survey. *Briefings in bioinformatics* 20, 3 (2019), 887–895.

[6] John H Benamati, Zafer D Ozdemir, and H Jeff Smith. 2017. An empirical test of an Antecedents–Privacy Concerns–Outcomes model. *JIS* 43, 5 (2017), 583–600.

[7] Zinaida Benenson, Anna Girard, and Ioannis Krontiris. 2015. User Acceptance Factors for Anonymous Credentials: An Empirical Investigation.. In *WEIS*.

[8] Christopher M Bishop. 2006. Pattern recognition. *Machine learning* 128, 9 (2006).

[9] Amine Boulemtafes, Abdelouahid Derhab, and Yacine Challal. 2020. A review of privacy-preserving techniques for deep learning. *Neurocomputing* 384 (2020), 21–45.

[10] Vanessa Bracamonte, Sebastian Pape, and Shinsaku Kiyomoto. 2021. Investigating User Intention to Use a Privacy Sensitive Information Detection Tool. *SCIS* 8, 26 (2021), 27.

[11] Vanessa Bracamonte, Sebastian Pape, and Sascha Loebner. 2022. "All apps do this": Comparing Privacy Concerns Towards Privacy Tools and Non-Privacy Tools for Social Media Content. *PoPETs* 2022, 3 (2022), 57–78.

[12] AJ Bernheim Brush, John Krumm, and James Scott. 2010. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proceedings of the 12th ACM UbiComp*. 95–104.

[13] European Commission, Content Directorate-General for Communications Networks, and Technology. 2019. *Ethics guidelines for trustworthy AI*. Publications Office.

[14] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. 2021. " I need a better description": An Investigation Into User Expectations For Differential Privacy. In *ACM CCS 2021*. 3037–3052.

[15] Fred Davis. 1985. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Ph. D. Dissertation. MIT.

[16] DIN EN ISO 9241-210. 2011. Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.

[17] Karolina Drobotowicz, Marjo Kauppinen, and Sari Kujala. 2021. Trustworthy AI Services in the Public Sector: What Are Citizens Saying About It?. In *REFSQ 2021*. Springer, 99–115.

[18] John J Dudley and Per Ola Kristensson. 2018. A review of user interface design for interactive machine learning. *ACM TiiS* 8, 2 (2018), 1–37.

[19] European Parliament and Council of The European Union. 2016. REGULATION (EU) 2016/679 General Data Protection Regulation (GDPR).

[20] Martin Fishbein and Icek Ajzen. 1977. Belief, attitude, intention, and behavior: An introduction to theory and research. *Philosophy and Rhetoric* 10, 2 (1977).

[21] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[22] Manish Gawali, CS Arvind, Shriya Suryavanshi, Harshit Madaan, Ashrika Gaikwad, KN Bhanu Prakash, Viraj Kulkarni, and Aniruddha Pant. 2021. Comparison of privacy-preserving distributed deep learning methods in healthcare. In *Annual Conference on Medical Image Understanding and Analysis*. Springer, 457–471.

[23] David Harborth and Sebastian Pape. 2020. How Privacy Concerns, Trust and Risk Beliefs and Privacy Literacy Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor. *ACM SIGMIS Database* 51, 1 (2020), 51–69.

[24] David Harborth, Sebastian Pape, and Kai Rannenberg. 2020. Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. *PoPETs* 2020, 2 (2020), 111–128.

[25] Carlos Jensen, Colin Potts, and Christian Jensen. 2005. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies* 63, 1-2 (2005), 203–227.

[26] John D Kelleher, Brian Mac Namee, and Aoife D'arcy. 2020. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.

[27] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[28] Sascha Löbner, Welderufael B Tesfay, Toru Nakamura, and Sebastian Pape. 2021. Explainable machine learning for default privacy setting prediction. *IEEE Access* 9 (2021).

[29] Sascha Löbner, Frédéric Tronnier, Sebastian Pape, and Kai Rannenberg. 2021. Comparison of De-Identification Techniques for Privacy Preserving Data Analysis in Vehicular Data Sharing. In *Computer Science in Cars Symposium*. 1–11.

[30] Gianclaudio Malgieri. 2020. The concept of fairness in the GDPR: a linguistic and contextual interpretation. In *ACM FAccT 2020*. 154–166.

[31] Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.

[32] Lydia Manikonda, Aditya Deotale, and Subbarao Kambhampati. 2018. What's up with privacy? User preferences and privacy concerns in intelligent personal assistants. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 229–235.

[33] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[34] Mohamed Merhi, Kate Hone, and Ali Tarhini. 2019. A cross-cultural study of the intention to use mobile banking between Lebanese and British consumers: Extending UTAUT2 with security, privacy and trust. *Technology in Society* 59 (2019).

[35] Payman Mohassel and Yupeng Zhang. 2017. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*. IEEE, 19–38.

[36] Ali Padyab and Anna Ståhlbröst. 2017. Privacy Enhancing Tools: A Literature Review on End-User Role and Evaluation. In *HAISA 2017*.

[37] Sebastian Pape, David Harborth, and Jacob Leon Kröger. 2021. Privacy concerns go hand in hand with lack of knowledge: The case of the German corona-warn-app. In *IFIP SEC 2021*. Springer, 256–269.

[38] Sebastian Pape and Kai Rannenberg. 2019. Applying Privacy Patterns to the Internet of Things' (IoT) Architecture. *MONET* 24, 3 (2019), 925–933.

[39] Kai Rannenberg, Sebastian Pape, Frederic Tronnier, and Sascha Löbner. 2021. *Study on the Technical Evaluation of De-Identification Procedures for Personal Data in the Automotive Sector*. Technical Report. Goethe University Frankfurt.

[40] B Thirumala Rao et al. 2016. A study on data storage security issues in cloud computing. *Procedia Computer Science* 92 (2016), 128–135.

[41] P Ram Mohan Rao, S Murali Krishna, and AP Siva Kumar. 2018. Privacy preservation techniques in big data analytics: a survey. *Journal of Big Data* 5, 1 (2018), 1–12.

[42] Theo Ryffel, Andrew Trask, Morten Dahl, Bobby Wagner, Jason Mancuso, Daniel Rueckert, and Jonathan Passerat-Palmbach. 2018. A generic framework for privacy preserving deep learning. *arXiv preprint arXiv:1811.04017* (2018).

[43] H Smith, Sandra Milberg, and Sandra Burke. 1996. Information privacy: Measuring individuals' concerns about organizational practices. *MIS quarterly* (1996), 167–196.

[44] Eugene F Stone and Dianna L Stone. 1990. Privacy in organizations: Theoretical issues, research findings, and protection mechanisms. *Research in personnel and human resources management* 8, 3 (1990), 349–411.

[45] Harry Chandra Tanuwidjaja, Rakyong Choi, Seunggeun Baek, and Kwangjo Kim. 2020. Privacy-preserving deep learning on machine learning as a service—a comprehensive survey. *IEEE Access* 8 (2020), 167425–167447.

[46] Frédéric Tronnier, Sebastian Pape, Sascha Löbner, and Kai Rannenberg. 2022. A Discussion on Ethical Cybersecurity Issues in Digital Service Chains. In *Cybersecurity of Digital Service Chains*. Springer, Cham, 222–256.

[47] Viswanath Venkatesh, James YL Thong, and Xin Xu. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly* (2012), 157–178.

[48] Michael Vierhauser and Manuel Wimmer. 2020. Towards integrating data-driven requirements engineering into the software development process: A vision paper. In *REFSQ 2020*, Vol. 12045. Springer Nature, 135.

[49] Hugo Villamizar, Amadeu Anderlin Neto, Marcos Kalinowski, Alessandro Garcia, and Daniel Méndez. 2019. An approach for reviewing security-related aspects in agile requirements specifications of web applications. In *RE 2019*. 86–97.

[50] Daricia Wilkinson, Paritosh Bahirat, Moses Namara, Jing Lyu, Arwa Alsubhi, Jessica Qiu, Pamela Wisniewski, and Bart Knijnenburg. 2020. Privacy at a glance: the user-centric design of glanceable data exposure visualizations. *PoPETs* 2020, 2 (2020), 416–435.

[51] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. 2017. Fairness constraints: Mechanisms for fair classification. In *AISTATS*. 962–970.

[52] Miaoyi Zeng, Shuaifu Lin, and Deborah Armstrong. 2020. Are All Internet Users' Information Privacy Concerns (IUIPC) Created Equal? *AIS TRR* 6, 1 (2020), 3.

[53] Shikun Zhang, Yuanyuan Feng, Lujo Bauer, Lorrie Faith Cranor, Anupam Das, and Norman Sadeh. 2021. did you know this camera tracks your mood?": Understanding privacy expectations and preferences in the age of video analytics. *PoPETs* (2021).

[54] Tianwei Zhang, Zecheng He, and Ruby B Lee. 2018. Privacy-preserving machine learning through data obfuscation. *arXiv preprint arXiv:1807.01860* (2018).

[55] Huadi Zheng, Haibo Hu, and Ziyang Han. 2020. Preserving User Privacy for Machine Learning: Local Differential Privacy or Federated Machine Learning? *IEEE Intelligent Systems* 35, 4 (2020), 5–14.