AI-driven Personalized Privacy Assistants: a Systematic Literature Review

Citation for the original published paper (version of record):
Morel, V., Iwaya, L., Fischer-Hübner, S. (2025). AI-driven Personalized Privacy Assistants: a Systematic Literature Review. IEEE Access, 13: 160982-161002.
http://dx.doi.org/10.1109/ACCESS.2025.3609188

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 25 July 2025, accepted 5 September 2025, date of publication 11 September 2025, date of current version 18 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3609188

SURVEY

AI-Driven Personalized Privacy Assistants: A Systematic Literature Review

VICTOR MOREL[1,2], **LEONARDO HORN IWAYA**[3], **AND SIMONE FISCHER-HÜBNER**[1,2,3], **(Senior Member, IEEE)**
[1]Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]University of Gothenburg, 405 30 Gothenburg, Sweden
[3]Department of Mathematics and Computer Science, Faculty of Health, Science, and Technology, Karlstad University, 651 88 Karlstad, Sweden

Corresponding author: Victor Morel (morelv@chalmers.se)

This work was supported in part by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The work of Leonardo Horn Iwaya was supported in part by the Knowledge Foundation of Sweden (KKS); in part by Region Värmland under Grant RUN/230445 and the European Regional Development Fund (ERDF) under Grant 20365177 through the Digital Health Innovation (DHINO 2) Project; and in part by Vinnova through the DigitalWell Arena Project under Grant 2018-03025.

ABSTRACT In recent years, several personalized assistants based on AI have been researched and developed to help users make privacy-related decisions. These AI-driven Personalized Privacy Assistants (AI-driven PPAs) can provide significant benefits for users, who might otherwise struggle with making decisions about their personal data in online environments that often overload them with different privacy decision requests. So far, no studies have systematically investigated the emerging topic of AI-driven PPAs, classifying their underlying technologies, architecture and features, including decision types or the accuracy of their decisions. To fill this gap, we present a Systematic Literature Review (SLR) to map the existing solutions found in the scientific literature, which allows reasoning about existing approaches and open challenges for this research field. We screened several hundred unique research papers over the recent years (2013-2025), constructing a classification from 41 included papers. As a result, this SLR reviews several aspects of existing research on AI-driven PPAs in terms of types of publications, contributions, methodological quality, and other quantitative insights. Furthermore, we provide a comprehensive classification for AI-driven PPAs, delving into their architectural choices, system contexts, types of AI used, data sources, types of decisions, and control over decisions, among other facets. Based on our SLR, we further underline the research gaps and challenges and formulate recommendations for the design and development of AI-driven PPAs as well as avenues for future research.

INDEX TERMS Artificial intelligence, data protection, machine learning, privacy, privacy assistant, systematic review.

I. INTRODUCTION

As the world becomes increasingly digitalized, people are faced with a growing number of requests for decisions related to their online privacy. Nowadays, individuals are using several apps every day, visiting different websites, and the number of smart gadgets and Internet of Things (IoT) devices they use continues to grow [1]. Furthermore, to comply with privacy laws such as the General Data Protection Regulation (GDPR) [2], software systems frequently demand from us to make privacy-related decisions regarding our personal data: *Do you grant this permission? Do you want to accept the cookies? Should this sensor be left on when you host friends?* Consequently, the cognitive burden increases, leaving users in disarray, tired, and unable to decide in their best interests [3].

During the last decade, researchers have been building privacy assistants to alleviate this burden and support users in their decisions. One of the first research work in that field has resulted in the patent on a Personalized Privacy Assistants (PPAs), registered in 2023 in the United States by Sadeh et al. [4]). With the progress made in Artificial Intelligence (AI), it comes at no surprise that many of the

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks.

privacy assistants proposed or developed in recent years leverage AI technology, notably to enable better personalized support.

This personalization can enhance the quality of decision support, adapted to the individuals' needs, preferences and current context. However, the extent to which AI drives these AI-driven PPAs, their efficiency, privacy-friendliness, functionality, and how legal requirements are eventually addressed remains unclear. In fact, to the best of our knowledge, there have been no surveys or systematic reviews on the topic of AI-driven PPAs, despite the substantial amount of work published on the topic in the recent years. This lack of systematization of knowledge makes it more difficult for other researchers and developers to reason about the opportunities that current AI-driven PPAs may offer. It also makes it more difficult to identify limitations and existing gaps to be addressed, as well as open research challenges that remain for the future.

For this reason, we present a Systematic Literature Review (SLR) of the body of knowledge to provide a common vocabulary and better compare, categorize and analyze the different AI-driven PPA solutions. In doing so, we aim to draw insights and lessons for future assistants and to formulate better recommendations for research, design, and development of AI-driven PPAs. Therefore, this SLR addresses the following Research Questions (RQs):

- **RQ1:** *What is the current state of the literature on AI-driven PPAs for automated support of end-users privacy decisions in IT systems?*
- **RQ2:** *What are the key attributes and properties of the proposed AI-driven PPAs in the literature?*

Here, we consider **agents** and **assistants** in a broad sense (any logical entity able to support users, including unimplemented theoretical models, see our selection criteria in Table 1); **AI** in a generic sense as well (see Section II-C); and, **privacy decisions** as the individual's decisions regarding their personal information management (see Section II-A).

To address our RQs, we performed an SLR on research papers providing technical solutions, published between 2013 and 2025 in peer-reviewed venues, including a snowballing process until early February 2025. We screened several hundred papers from IEEE, ACM, Scopus, and Web of Science, resulting in 41 selected papers after several rounds of snowballing. We extensively read and analyzed all included papers, and the information extracted forms the basis of our work.

Our SLR results in the following contributions:

- **A Classification for AI-driven PPAs** – We propose the first classification for AI-driven PPAs, providing a common vocabulary for designers of such systems.
- **Data Charting & Quantification** – We charted and quantified several aspects of AI-driven PPAs based on the aforementioned classification.
- **Research Gaps & Challenges** – We underline the current gaps in the state of the art and highlight challenges for designing AI-driven PPAs based on our data.
- **Recommendations & Research Avenues** – We formulate recommendations for improving AI-driven PPAs, and propose several avenues for future research.

In the following sections, we present the background and related work in Section II. The study's methodology is detailed in Section III. The results and classification are organized and presented in Sections IV and V. Based on the findings, we present our discussion of research gaps and future work in Section VI. Lastly, Section VIII concludes our work.

II. BACKGROUND

As a background, this section first provides an overview of different types of privacy decisions for which individuals could receive support from AI-driven PPAs. Then, it summarizes legal requirements for transparency that AI-driven PPAs should meet, and refers to a classification scheme of explainable AI that we are using for our classification.

When discussing legal requirements, we will primarily refer in the section and for the rest of this paper to the European Legal Framework, including the GDPR and the AI Act [5], since the study was conducted in Europe with the support of a European funding foundation. Moreover, the GDPR has been regarded as the "gold standard" for data protection with a territorial scope that goes beyond Europe, and is therefore also used as a point of reference.

A. PRIVACY DECISIONS

Among the most notable definitions, Westin [6] has defined privacy as the right to informational self-determination, meaning that individuals should have the *right to decide* for themselves when, how, and what information about them is communicated to others. As mentioned, in the EU, the GDPR emphasizes that individuals should have control of their personal data (Recital 7), and thus should be empowered to make decisions about their data as one prerequisite for exercising such control. Delving deeper into this notion of *privacy decisions*, we further elaborate on this concept in the following subsections.

1) INDIVIDUAL PRIVACY DECISIONS REGULATED BY LAWS

Some privacy decisions individuals can make to exercise control over their data is regulated under the GDPR and other privacy laws. These decisions notably include, but are not limited to, the *decisions to grant or to withdraw consent* to data collection and processing. Art. 4 (11) of the GDPR defines 'consent' of data subjects as any freely given, specific, informed, and unambiguous indication of the data subject's wishes by which they, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to them.

Moreover, the GDPR and most other privacy laws regulate further *decisions to exercise data subject rights* granted by the respective laws. For instance, according to Art. 15-22 GDPR, data subjects have the rights to access data, request rectification or deletion of data, export data, and object to direct marketing and profiling. Data subjects can also object in cases where the legal ground for the processing is public interest or legitimate interest, or exercise their right not to be subject to automated decision-making.

Also the AI act (Art. 14) provides individuals with the right to Human Oversight for critical decision-making processes for high-risk AI systems, including the right to human review and potential override of automated decisions.

2) FURTHER TYPES OF PRIVACY DECISIONS

Further types of privacy decisions concerning users' choices regarding the use of their data by others, which are not directly mentioned or regulated by the GDPR, include *decisions of individuals to publish or share data on their own initiative*, e.g., in social networks. In these cases, data sharing has typically not been formally triggered by a consent request to allow data sharing with another party.

Moreover, privacy decisions encompass *privacy permission* settings (or access control rights), which grant others certain rights for using their data and are, for instance, typically used for permission systems of mobile phone operating systems, such as Android or iOS. Setting privacy permissions on mobile operating systems often requires consent at installation or during runtime. However, instead of consent, other legal grounds – such as a contract (Art. 6 (1)(b) GDPR) –, can be used, e.g., for a banking app to forward account information when transferring money [7]. Let us also note the peculiar case of Global Privacy Control (GPC), a unary signal that permits or prohibits third-party tracking on the browser [8]. Due to its enforceability under the California Consumer Privacy Act (CCPA) [9], it is regulated by a privacy law but is technically more akin to a privacy permission.

Additionally, some privacy-enhancing technologies (PETs) and protocols allow users to decide and set *privacy preferences*, which are simply indications of the users' privacy wishes of how their data should be used without actually granting any rights to others, and thus without legal mandate. Privacy preferences have, for instance, been used earlier by the Platform for Privacy Preferences (P3P) [10] or Do Not Track (DNT) [11], as an example for signals that can be set manually in browser settings for allowing users to specify their privacy choices.

B. REQUIREMENTS FOR TRANSPARENCY

Transparency of data processing is an important prerequisite for users for making well-informed decisions, and should therefore be provided by any privacy assistants that should support users in decision-making. In cases where the data controllers of the AI-driven PPAs are not the data subjects themselves, the controllers should provide the data subjects with privacy policy information *ex-ante* at the time when data is obtained from them according to Art. 13 GDPR, and *ex-post* through the right to access granted in Art. 15 GDPR. This should particularly include information about purposes of processing, data categories concerned, but also information about the logic involved and significance, and envisioned consequences of automated decision-making and profiling performed by the AI-driven PPAs.

The EU AI Act also includes obligations for transparency for the producers and deployers of limited-risk and high-risk AI systems (Art. 50). While the providers of limited-risk AI systems have to mainly ensure that humans are informed that AI systems are used, high-risk AI systems require that further clear, comprehensible and adequate information is given to the deployer (Art. 13), traceability of results via logging (Art. 12) and appropriate human oversight (Art. 14). However, AI-driven PPAs are typically not in the high-risk category, since they are used for users' own personal privacy management, which should typically not interfere with the fundamental rights of others. Exceptions could, however, be AI-driven PPAs that are, for example, used for setting permissions for safety-critical applications impacting the safety of the users or others.

Also, Ethics Guidelines for Trustworthy AI, promoted by the EU Commission,[1] emphasize the requirement for transparency and explainability for AI systems to be deemed trustworthy.

C. AI FOR DECISION-MAKING

AI is a generic term for various strategies and techniques enabling computers and machines to simulate human intelligence and problem-solving capabilities [12]. Machine learning (ML) is a field of AI (we subsume the former under the latter in the rest of the document) that develops and studies statistical algorithms and models, draws inferences from patterns in data, and learns and adapts without following explicit instructions. AI-powered tools can particularly lighten the user's cognitive load and thereby improve their decision-making, e.g., by decision support, augmentation, or automation.

While there are different ways to categorize AI systems, we refer in the present work to the survey paper on eXplainable AI (XAI) by Arrieta et al. [13]. They distinguish between transparent models and those requiring post-hoc explainability[2] (denoted *non-inherently transparent* in this paper). We use this reference because AI-supported decisions must be explained under specific circumstances according to the GDPR and the AI Act [14].

[1] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[2] According to them, "Post-hoc explainability targets models that are not readily interpretable by design by resorting to diverse means to enhance their interpretability, such as *text explanations, visual explanations, local explanations, explanations by example, explanations by simplification* and *feature relevance explanations* techniques [13]."

1) TRANSPARENT AI MODELS

In their words: "A model is considered to be transparent if by itself it is understandable." [13]. An overview with short definitions of transparent AI models, which were used by the solutions surveyed in this paper, is provided below: **Decision trees** are a method used for classification and regression tasks, which model decisions and their possible consequences as a tree-like structure of conditions and actions. **K-nearest neighbors** is a simple algorithm used for classification and regression, where the output is based on the majority class or average of the k-nearest data points in the feature space. **Bayesian models** are statistical models that apply Bayes' theorem to update the probability of a hypothesis as more evidence or information becomes available. **Rule-based learning** identifies and utilizes a set of relational rules to make predictions or classifications based on input data. Rules can take the form of simple conditional if-then rules or more complex combinations of simple rules to form their knowledge. A **Generalized Additive Model** is a statistical framework that extends generalized linear models by allowing the linear predictor to depend on smooth functions of the predictor variables, enabling more flexible modeling of non-linear relationships. **Hierarchical clustering** is a method of cluster analysis that builds a hierarchy of clusters either through a bottom-up approach (agglomerative) or a top-down approach (divisive), creating a dendrogram to represent the nested grouping of data points.

2) NON-INTRINSICALLY TRANSPARENT AI MODELS

Parts of the surveyed AI-driven PPAs used non-inherently transparent AI models, including neural networks (especially deep and convoluted) and Support Vector Machines (SVM), as well as reinforcement learning [15]. A short overview with definitions of these non-inherently transparent AI models follows below: **Neural networks** are a class of machine learning models inspired by the structure and function of biological neural networks, consisting of interconnected layers of artificial neurons that process input data to produce outputs for tasks such as classification, regression, and pattern recognition, and a **deep neural network** is an artificial neural network with multiple layers between the input and output layers. **Random forests** are ensemble learning methods for classification, regression and other tasks that works by creating a multitude of decision trees during training. **AdaBoost** is a machine learning ensemble meta-algorithm that combines multiple weak classifiers to create a strong classifier, where each new weak learner focuses on the errors of the previous ones to improve overall accuracy. **Support Vector Machines** are supervised learning models that analyze data for classification and regression tasks by finding the optimal hyperplane that best separates different classes in the feature space. **Reinforcement learning** is a type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize cumulative reward. **Game theory** is a mathematical framework that studies strategic interactions between rational decision-makers, analyzing how their choices influence outcomes. **Large language models** are advanced AI systems designed to understand and generate human-like text based on vast amounts of training data.

Nonetheless, models that are not deemed intrinsically transparent can be made explainable through the use of *post-hoc* techniques.

III. METHODOLOGY

This study adopts the widely known methodology for systematic literature reviews (SLRs) proposed by Kitchenham [16]. The SLR methodology offers us a well-defined and rigorous sequence of methodological steps consisting of three main phases: (1) planning, (2) conducting, and (3) reporting the review. An SLR Protocol that describes the entire research process has been written for this study. Furthermore, we make our research data openly available in a GitHub repository for reproducibility.[3] Our material comprises the citation files of each query, the Data Extraction Forms (DEFs) of the selected papers, and the charting spreadsheet used to compile our data. We refer to these documents for methodological details.

A. PLANNING THE REVIEW

The first activity of the planning phase was to determine the need for this SLR. Several databases were searched to verify if any surveys or reviews had been conducted on AI-driven PPAs. Search terms such as privacy, data protection, assistant, agent, artificial intelligence, and machine learning were used. However, we could not identify any survey or systematic reviews on the topic, reassuring the need for an SLR. The research questions, presented in Section I, guided the remaining phases of this SLR with respect to the search process, selection criteria, and data synthesis.

B. CONDUCTING THE REVIEW

1) SEARCH STRATEGY

Based on our RQs and previous preliminary searches when designing the SLR Protocol, we identified a list of nine relevant keywords, i.e., *privacy, data protection, assistant, agent, artificial intelligence, machine learning, intelligent, automatic*, and *personalized*. These keywords were used to construct the search query in Listing 1. As such, the search query targets papers working on three joint topics: 1) privacy (*or* data protection), using either 2) an assistant *or* an agent, *and* leveraging 3) artificial intelligence *or* personalization.

Four scientific databases were selected, i.e., Scopus, Web of Science, IEEE Xplore, and ACM Digital Library, due to their high relevance to the areas of computer science and engineering, comprising the vast majority of published research in the field. We also specified inclusion and exclusion criteria (see Table 1) used during the screening of publications retrieved from the databases. Marky et al. [17]

[3]https://github.com/Victor-Morel/SLR_AI_PPA

LISTING 1 Composition of Search Query for Literature Search.

```
Search Query = {
   (privacy OR "data protection") AND
   (assistant* OR agent*) AND ("artificial
   intelligence" OR "machine*learning"
   OR intelligent OR automat* OR personali*ed)
}
```

TABLE 1. Criteria for the inclusion and exclusion of studies.

Inclusion Criteria
- Provides a technical solution (implemented or theoretical) to help end-users automate personal (and personalized) privacy decisions with an assistant (or artificial agent) in IT systems. - Papers from 2013 onward to concentrate on the state-of-the-art. - The concept of AI needs to be explicitly stated in the papers.

Exclusion Criteria
- Papers with solutions that are purely theoretical without substantial explanations on how they could be implemented in practice. - Papers with solutions that solely automate the analysis of privacy policies but without any type of personalization. - Papers with poor scientific quality (e.g., lack objectives or research questions, the methodology is not described, the solution is insufficiently/vaguely described, etc.).

is a good illustration of a relevant paper not meeting our selection criteria. In spite of providing a technical solution to automate privacy decisions (explicitly called a PPA), the paper does not use AI and hence is not included in our list of papers. Bollinger et al. [18] provides another example of an excluded yet relevant paper. The paper provides a technical solution for automating privacy decisions and uses AI, but does not personalize the decisions.

Before starting the search process, two authors piloted the searches on all databases and ran a *calibration exercise* to verify the consistency of the inclusion criteria. For that, the authors independently screened 10% of the results and discussed their decisions. The conflicts were all discussed and solved, sometimes with the help of the third author. This process was repeated a second time, screening another 10% of the papers at a point that the authors fully agreed with the consistency of the selection process.

2) DATA MANAGEMENT

To manage the screening process, we exported search results from each database and imported them to the RAYYAN software (https://rayyan.ai/), allowing two reviewers to independently select papers (i.e., double-blinded) and to manage conflicts by a third reviewer. Duplicated publications were also removed using RAYYAN during the selection process. Bibliographies of final results were exported to Zotero (for citing and sharing research).

3) SELECTION PROCESS

Figure 1 presents an overview of the selection process. The querying of the databases mentioned above on October 19, 2023, yielded 2386 papers and 1697 unique entries after removing duplicates. The screening phase lasted until November 23, 2023, and resulted in the selection of 33 papers. Two authors then read 10% of these 33 papers (3) and adjusted the DEF based on mutual feedback. This step helped us add new important fields to the DEF and consistently extract data from the papers.

TABLE 2. Summary of information included in the DEFs.

Extracted Information
- Bibliographic information, such as title, abstract, authors and affiliations, venues, year of publication, etc. - Key information of the AI-driven PPA, such as its source(s) of data, its eventual architecture, its system context, the type of privacy decision considered, the accuracy of the decisions, the type of AI used, etc. - The presence of a user study, and performed its critical appraisal (see below). - Extent of evaluation, a scale of validation activity that is measured. - Quality assessment and critical appraisal of the studies that have validated or evaluated the AI-driven PPA. - Features of user's control over decisions (initially guided by EU consent requirements).

The first data extraction phase, consisting of a full reading of each of the 33 papers, was performed over weeks 4 to 7 (included) in 2024. Fifteen papers were excluded after full reading for different reasons: they were duplicates (i.e., same work published in different venues); they did not provide any technical solution; the automated decisions were not personalized to an end user; AI was not used for automating decisions; or they are of poor scientific quality (see our criteria in Table 1); one paper was not available for download, we could not access it even after reaching out the authors.

We then proceeded to several snowballing phases [19], during which we checked the abstracts of all seemingly relevant[4] papers cited (backward snowballing), and screened citing papers (forward snowballing). The snowballing process lasted from week 8 to week 19 of 2024 and resulted in 21 additional papers after exclusion, for a total of 39 papers (33-15+21). We performed an update of our results in February 2025, repeating the whole process in a 2nd round of searches. This update led us to include one additional paper through the database searches and another one through the snowballing process, for a total of 41 papers.

4) DATA EXTRACTION AND ANALYSIS

The data extracted in the DEFs was compiled and further organized in spreadsheets during weeks 20-21 in 2024. This process also included the initial aggregation of data and the creation of frequency charts across several data categories (e.g., studies per year, types of publications, authors and affiliations, etc.). Table 2 shows a summary of the components extracted from the included studies.

It is also worth noting that we classified publications by their types of contributions according to the following

[4]We only assessed papers cited in relevant sections, e.g., related work.

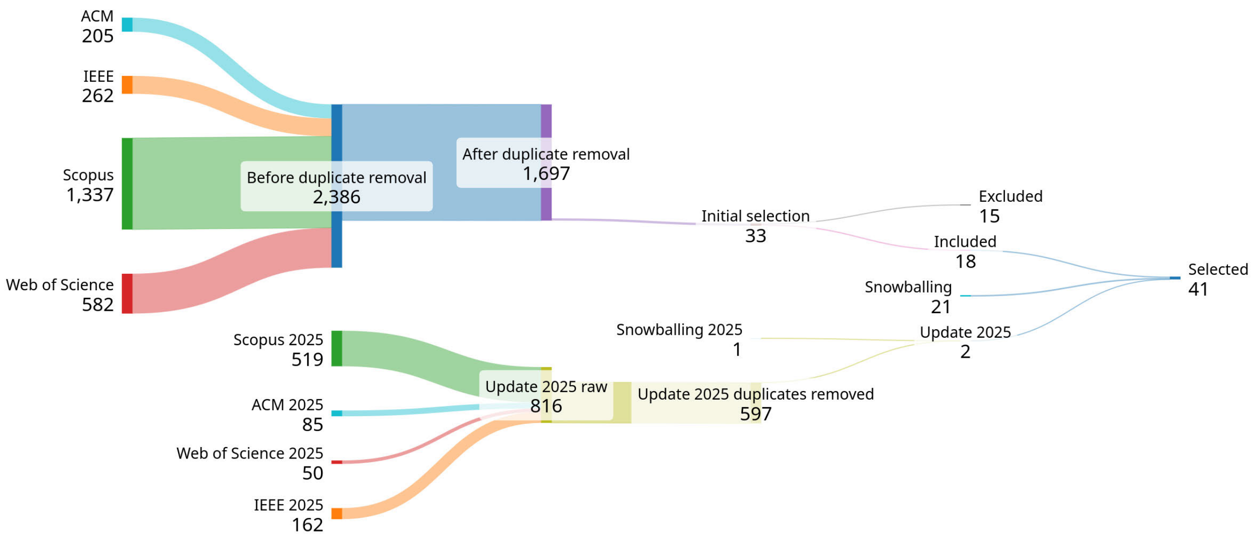

FIGURE 1. **Sankey chart of the selection process.**

items proposed by Kuhrmann et al. [20] and Shaw [21] (i.e., all that apply): *(i) model*, as a representation of observed reality by concepts after conceptualization; *(ii) theory*, as a construct of cause-effect relationships; *(iii) framework*, including frameworks/methods (related to automated privacy decisions); *(iv) guidelines*, as a list of advice; *(v) lessons learned*, as a set of outcomes from obtained results; *(vi) advice*, as recommendations (from opinion); and, *(vii) tool*, as tools to automate privacy decisions.

Although we attempted to extract as much data from the studies as possible using a DEF, we found that, during the data analysis process, there was a need to further categorize studies across other *facets*. For instance, additional information was compiled in the spreadsheets, such as a high-level categorization of certain fields (i.e., the type of AI used) or a critical appraisal of the user studies presented in the selected papers, although only when those user studies were used to evaluate the AI-driven PPA, and not when they were used for data collection to build datasets.

This collection of facets created during the study design and data analysis processes forms the basis of the work's final classification scheme, presented as part of the main results. All authors were involved in the data analysis process and the definition of facets that further classify studies on the topic.

For the *critical appraisal*, we used the CAT (Critically Appraised Topic) Manager App of CEBMA (the Center for Evidence-Based Management) [63], which provides a practical yet rigorous approach to evaluate studies based on objective criteria. It helps determine a study's trustworthiness regarding cause and effect questions. Once a study is evaluated, the possible outcomes are: Very high (A+), High (A), Moderate (B), Limited (C), Low (D), or Very low (D-).

By definition, an AI-driven PPA leverages AI techniques. We therefore collected information about the *Type of AI used*. AI models rely on data for training and decision-making. As such, we extracted the *source of data*. During the adjustment of the DEF, we observed that AI-driven PPAs are usually designed for a specific *system context*, and for one or several *types of decision*. Connected to the system context, we extracted the *choice of architecture* of the implementation (if any) to analyze the trust implications.

We also collected the methods for an *empirical assessment*, presence, and quality of user studies, or the means used to measure the accuracy, to gain insights on eventual benchmarks of AI-driven PPAs. Studies can be classified as evaluation or validation research, as proposed by Wieringa et al. [64]. An evaluation works in real-world practice and is implementing/deploying the solution or testing in an actual project with real test users, such as real case studies and realistic user testing of prototypes/systems. A validation is a limited illustrative or hypothetical "case study" or "use case" performed as a lab experiment. In general practice, prototypes are often validated by cross-sectional studies.

Finally, initially guided by legal requirements for consent and the exercise of data subject rights under the GDPR (although eventually, no paper considers consent), we extracted what became *user control over decisions*.

C. REPORTING THE REVIEW

Based on the data analysis, a whole coherent narrative was written by the research team, i.e., this SLR article, conveying all the results, our interpretation of the main findings, and identifying research gaps. This synthesis on AI-driven PPAs is thus reported in the following sections.

TABLE 3. This table informs the type of contribution, informing whether the surveyed solution presents a *framework*, a *tool* (i.e. with an implementation), a *model*, *lessons learned*, or *advice*.

Year	Publication	Type of contribution: Framework	Tool	Model	Theory	Lessons learned	Advice
2014	Xie et al. [22]	•				•	
2015	Apolinarski et al. [23]	•	•				
2015	Hirschprung et al. [24]	•		•			
2015	Squicciarini et al. [25]	•		•			
2016	Liu et al. [26]		•			•	
2016	Albertini et al. [27]		•				
2016	Dong et al. [28]			•		•	
2017	Baarslag et al. [29]		•	•		•	
2017	Fogues et al. [30]		•				
2017	Zhong et al. [31]			•			
2017	Misra et al. [32]		•				
2017	Nakamura et al. [33]			•			
2017	Olejnik et al. [34]	•	•				
2018	Das et al. [35]		•				
2018	Tan et al. [36]		•				
2018	Wijesekera et al. [37]		•			•	
2018	Yu et al. [38]			•	•		
2018	Bahirat et al. [39]			•			
2018	Raber et al. [40]		•	•		•	
2019	Klingensmith et al. [41]		•				
2019	Barbosa et al. [42]			•			•
2019	Alom et al. [43]	•					
2019	Alom et al. [44]	•					
2020	Kasaraneni et al. [45]		•	•			
2020	Kaur et al. [46]			•			
2020	Botti-Cebria et al. [47]		•				
2020	Kökciyan et al. [48]			•			
2020	Sanchez et al. [49]			•			
2021	Kaur et al. [50]			•			
2021	Lobner et al. [51]			•			
2022	Filipczuk et al. [52]	•	•				
2022	Hirschprung et al. [53]	•		•			
2022	Kökciyan et al. [54]			•			
2022	Ulusoy et al. [55]			•			
2022	Zhan et al. [56]			•			
2022	Brandão et al. [57]		•				
2022	Mendes et al. [58]		•	•		•	
2022	Shanmugarasa et al. [59]		•	•			
2023	Ayci et al. [60]		•	•			
2023	Serramia et al. [61]		•	•			
2024	Wang et al. [62]		•				

IV. SUMMARY OF DATA CHARTING RESULTS

This section provides an overview of the quantitative insights generated through the data charting process (e.g., publications per year, citations, types of decisions). The main findings related to the critical appraisals are also introduced in this section. It is worth noting, nonetheless, that the classification features are further detailed in the following Section V.

Among the 41 papers surveyed, we tallied 15 different countries for the authors' affiliations (see Table 4), with the USA and UK leading in numbers. About 55% ($n = 22$) of the selected publications were published from 2019 to 2024, with the year 2021 being the most productive with 8 publications

TABLE 4. Countries of affiliation of authors of selected papers.

Countries	Total	Countries	Total
United States	14	China	2
United Kingdom	6	Israel	2
Japan	4	Portugal	2
Netherlands	4	Switzerland	1
Italy	4	Turkey	1
Germany	4	Canada	1
Spain	3	Australia	1
India	2		

TABLE 5. Number of publications per year.

Year	N. of Publications	Year	N. of Publications
2013	0	2019	4
2014	1	2020	5
2015	3	2021	2
2016	3	2022	8
2017	6	2023	2
2018	5	2024	1

(see Table 5). At the time of data collection, papers were cited between 0 and 275 times with an average of 38.56 citations, a median of 14, and a standard deviation of 59, indicating a power law distribution of the citation count. The most cited papers are Liu et al. [26] ($n = 275$), Yu et al. [38] ($n = 199$), and Squicciarini et al. [25] ($n = 130$) (numbers at the time of data collection).

As shown in Table 6, regarding the sources of data used by the AI-driven PPAs, context data, attitudinal data, and metadata were the most prevalent. We observed a relatively balanced distribution when it comes to the types of decisions (between 12 and 15 for each type) and the system contexts (between 11 and 13, with two outliers for *Cloud* and *Intelligent retail store*). For the types of AI systems that we were able to classify, most models were deemed non-intrinsically transparent (NIT, $n = 14$), followed by transparent (T, $n = 8$) and partially transparent (PT, $n = 4$) models. Note also that Das et al. [35] did not specify the type of AI used in their paper, we were therefore unable to categorize their solution in that respect (under *Type of AI used*).

In Table 3, the publications were also classified by their types of contributions, according to the categories proposed by Kuhrmann et al. [20] and Shaw [21]. We observed a prevalence of models ($n = 24$) and tools ($n = 21$), followed by frameworks ($n = 9$). Nonetheless, these models, tools, and frameworks lack empirical assessment, an issue further analyzed in Section V-F.

Finally, the results of our critical appraisal can be found in Table 7. Out of the 41 publications, only 15 presented a user study, i.e., qualitative research that is suitable to be critically appraised. In terms of quality, they mostly scored "low" or "very low" ($n = 9$) according to the CEBMA checklist. Exceptionally, only the studies of Liu et al. [26] and Baarslag et al. [29] were appraised as of high quality.

This suggests that the empirical evidence around AI-driven PPAs remains incipient, and existing solutions can be further tested in real-world settings, a challenge (or opportunity) that is discussed in Section VI.

V. CLASSIFICATION FOR AI-DRIVEN PPAS

We provide in this section a classification for AI-driven PPAs as the main contribution of this SLR. Summarized in Figure 2, the classification comprises several dimensions, i.e., features typically considered in the design of such an assistant (see also Tables 6 and 7). These dimensions are the *type of decision* (Section V-A), the *type of AI* (Section V-B) and the *source of data* (Section V-C) used in the decision, the *system context* (Section V-D), the *choice architecture* of its eventual implementation (Section V-E), the *empirical assessment* (Section V-F), and the extent to which *users have control over the decisions* (Section V-G).

The classification and its dimensions are **data-driven**, in the sense that they were derived based on what is described in the papers, reflecting the current state of the literature. For example, considering the category of system contexts, more dimensions could be envisioned, but we limited it to the five dimensions (i.e., mobile apps, social media, IoT, cloud, and intelligent retail stores) that were found in the papers. Each feature will be explored in more detail in this section, and substantiated with non-exhaustive examples for each possible option, while an overview is provided in Figure 2.

Note that not all dimensions are necessary for composing an AI-driven PPA. The dimensions for the type of AI, source of data, type of decision, and system context are *"mandatory,"* consisting of essential requisites that an AI-driven PPA needs to consider (solid boxes in Figure 2). Other dimensions such as the empirical assessment, choice architecture, and user control over decisions are *"optional"* since not all the identified AI-driven PPAs were evaluated, some do not have an implementation (and therefore an architecture), and some (regrettably) do not empower users with much control for various reasons (dashed boxes in Figure 2).

Furthermore, note that most but not all dimensions are non-exclusive. For instance, it is possible to combine different types of data and/or AI models (non-exclusive), but the system context is often exclusive in the sense that solutions are often designed for a specific system context.

A. TYPE OF DECISION

Decisions taken by an AI-driven PPA can be of different types, and it is essential to distinguish them to assess the possibilities they offer. Indeed, some decisions – such as permissions – have a binding character, i.e., they constrain the system to act according to the user's choice, while others do not, such as preferences. Note that it may not always be possible to distinguish between each type of decision clearly (as discussed in Section II-A2). Other types of decisions with different implications regarding their enforcement can be envisioned by an AI-driven PPA (such as consent or deletion requests, see Section II-A).

1) PERMISSIONS

The first type of decisions that many AI-driven PPAs assist the users with is *permissions*, which, as discussed in Section II-A2, correspond to access control settings. Permissions are system-specific and binding, as the underlying operating system should enforce them.

We typically find mobile app permissions (e.g., in Baarslag et al. [29], mobile apps are addressed in 11 papers), but they are not restricted to the mobile environment. AI-driven PPAs can deal with permissions in IoT environments (see, e.g., [41], IoT is covered by 13 papers) or in the cloud [24].

2) PREFERENCES

The second type of decision covered by the literature is *preferences*, which, unlike permission settings, should be understood as expressions of will. Several works refer to preferences while they actually deal with permissions [24], [26], [37], [52], [59]. It is indeed common to talk about preferences imprecisely, but they should not be confused with permissions that have a binding property.

3) DATA SHARING

Data sharing is the third type of privacy decision of AI-driven PPAs encountered in the reviewed literature, for which the binding character is uncertain for users. For instance, assessing whether a limitation in the audience is enforced is not always possible from a user point of view because the underlying technical system is inaccessible to them, see, e.g., Ulusoy and Yolum [55]. Typically, it can be difficult or even impossible to assess whether most social media platforms strictly account for the user's privacy decisions, or merely welcome them as recommendations to be applied only if possible. Papers classified under this type of decision usually do not mention the binding character of their solution (or the lack thereof).

B. AI TECHNOLOGY USED

Another significant characteristic of AI-driven PPAs is the type of AI used. Many solutions are based on machine learning models, such as supervised ML (classification), non-supervised ML (clustering), and reinforcement learning, sometimes combined. It is, however, also possible to find older AI techniques grouped under the umbrella of expert or rule-based systems.

We also classified the different AI technologies used by the AI-driven PPAs reviewed regarding their explainability, or their inherent transparency. However, XAI is only explicitly addressed by one work [51]; the other models are therefore categorized based on Arrieta et al. [13]'s taxonomy, which defines non-ML based systems as AI (including rule-based). For classification models, we annotated T for Transparent in Table 6, NIT for Not-Inherently Transparent, and PT for Partially Transparent when the solution relies on

TABLE 6. **Summary table of our classification, part 1. It presents the mandatory features of AI-driven PPAs, namely the type of decision, the type of AI used (for which we specified whether the classification model is Transparent (T), Not-Inherently Transparent (NIT), or Partially Transparent (PT) because several models are used), the type of source of data, and the system context (note that IRS stands for Intelligent Retail Store). An empty field signifies that the solution does not exhibit the characteristic (e.g., does not consider Y type of decision).**

Year	Publication	Type of decision			Type of AI used						Type of source of data						System context				
		Permissions	Preferences	Data sharing	Classification	Clustering	Rule-based	Logic-based	Reinforcement	LLM	Context	Attitudinal data	Metadata	Data type	Content of data	Behavioral data	Mobile apps	Social media	IoT	Cloud	IRS
2014	Xie et al. [22]		•		NIT						•										
2015	Apolinarski et al. [23]	•			NIT						•						•				
2015	Hirschprung et al. [24]	•				•						•								•	
2015	Squicciarini et al. [25]			•			•				•		•		•						
2016	Liu et al. [26]	•			T	•						•					•				
2016	Albertini et al. [27]			•			•					•						•			
2016	Dong et al. [28]			•	T						•				•			•			
2017	Baarslag et al. [29]	•						•				•		•			•				
2017	Fogues et al. [30]			•	PT						•	•						•			
2017	Zhong et al. [31]			•	NIT							•		•				•			
2017	Misra et al. [32]			•	NIT						•				•			•			
2017	Nakamura et al. [33]		•		NIT									•		•					
2017	Olejnik et al. [34]	•			T						•	•					•				
2018	Das et al. [35]		•									•							•		
2018	Tan et al. [36]	•			T								•				•				
2018	Wijesekera et al. [37]	•			NIT						•		•	•			•				
2018	Yu et al. [38]			•	NIT						•				•			•			
2018	Bahirat et al. [39]		•		T							•							•		
2018	Raber et al. [40]	•			T							•				•					•
2019	Klingensmith et al. [41]	•			NIT								•			•			•		
2019	Barbosa et al. [42]		•		PT							•	•						•		
2019	Alom et al. [43]		•		PT						•	•									
2019	Alom et al. [44]		•		NIT							•				•			•		
2020	Kasaraneni et al. [45]			•	T	•							•					•			
2020	Kaur et al. [46]		•		NIT						•		•				•		•		
2020	Botti-Cebria et al. [47]			•	PT										•			•			
2020	Kökciyan et al. [48]	•					•				•		•						•		
2020	Sanchez et al. [49]	•				•						•							•		
2021	Kaur et al. [50]	•							•		•		•				•		•		
2021	Lobner et al. [51]		•		T								•	•		•		•			
2022	Filipczuk et al. [52]	•						•				•		•			•				
2022	Hirschprung et al. [53]			•				•			•							•			
2022	Kökciyan et al. [54]		•					•			•					•			•		
2022	Ulusoy et al. [55]			•					•		•					•		•			
2022	Zhan et al. [56]	•						•	•			•							•		
2022	Brandão et al. [57]		•		NIT	•					•						•				
2022	Mendes et al. [58]	•			NIT						•						•				
2022	Shanmugarasa et al. [59]	•				•					•	•	•	•					•		
2023	Ayci et al. [60]			•	NIT							•	•					•			
2023	Serramia et al. [61]		•				•					•	•						•		
2024	Wang et al. [62]	•								•	•	•	•	•	•				•		

models with different levels of transparency. Note that LLMs were not considered in Arrieta et al., we nonetheless classify them as non-inherently transparent.

1) TRANSPARENT

a: CLASSIFICATION

Supervised machine learning, also called classification models, is a common set of techniques deployed in AI-driven PPAs. In this context, a model is trained to classify an object of decision into a choice tailored to the users' desires.

Transparent classification models [13] (used in 8 papers) are composed of decision trees (used for instance in Bahirat et al. [39]), k-nearest neighbors (leveraged in Botti-Cebria et al. [47]), and Bayesian models (see Olejnik et al. [34]).

b: CLUSTERING

Several works use clustering techniques for their AI-driven PPA. In this context, clustering is classically used to create a set of *privacy profiles*, i.e., an archetypal ensemble of default parameters (for preferences or permissions) to which a user is then assigned. Clustering algorithms (leveraged in 6 papers) used are hierarchical clustering [26], k-means [57], k-modes [59], although several papers did not disclose the exact method used, such as Hirschprung et al. [24].

c: RULE-BASED

AI-driven PPAs can be powered by non-machine-learning algorithms, based instead on rules (e.g., Albertini et al. [27] implement association rules). This comprises theoretical as

TABLE 7. Summary table of our classification, part 2. It presents the optional features of AI-driven PPAs, such the Architecture, under which we denote with "–" when the criterion is not applicable (no implementation/tool is presented) and when the solution presents an implementation, but the paper did not specify enough information to infer its architecture. For user control over decisions, we specify the elements present to inform users under *Informed* (type of Data, Purpose, Controller). It also presents the accuracy (if any) of the predictions (see Section V-F); the type of user control over decisions; the presence or not of a user study, and the type of user study if applicable; and the results of our critical appraisal (see Section III-B4).

Year	Publication	Architecture			User control over decision				Accuracy	User study	Critical appraisal
		Local	Remote	Federated	Informed	Semi-automated	Specific	Revoke			
2014	Xie et al. [22]	–	–	–	No	Yes	Yes [a]	No	68%	Online user experiment α	–
2015	Apolinarski et al. [23]	●			D	Yes	Yes	No	–	No	–
2015	Hirschprung et al. [24]	–	–	–	D	No [b]	Yes	No	–	Online qualitative survey	D-, very low (55%)
2015	Squicciarini et al. [25]	–	–	–	D	Yes	Yes	No	92.53%	Cross sectional study [c]	D-, very low (55%)
2016	Liu et al. [26]	?	?		D, P	Yes	Yes	Yes	78.7%	Randomized controlled studies [d]	A, high (90%)
2016	Albertini et al. [27]		●		D	Yes	No	No	–	Cross-sectional study	D, very low (55%)
2016	Dong et al. [28]	–	–	–	–	–	–	–	89,8% F1	Case studies α	–
2017	Baarslag et al. [29]	●	●		Unclear	Yes	Yes	No	–	Randomized controlled study [e]	A, high (90%)
2017	Fogues et al. [30]		●		No	No	No	No	Around 50%	Online survey α	–
2017	Zhong et al. [31]	–	–	–	–	–	–	–	79%	Survey α	–
2017	Misra et al. [32]		●		D	Yes	Yes	No	91.8%	Non-controlled before-after study [f]	C, limited (70%)
2017	Nakamura et al. [33]	–	–	–	–	–	–	–	85%	Cross-sectional study [g]	D, very low (55%)
2017	Olejnik et al. [34]	●			No	Yes	Yes	No	More than 80%	Yes, for data collection α	–
2018	Das et al. [35]		●		Yes	It depends	Yes	No	–	No	–
2018	Tan et al. [36]		●		No	No [h]	Yes	No	95% [i]	No	–
2018	Wijesekera et al. [37]	●	●		D, C	Yes	Yes	Yes	95%	Interrupted time series study (ESM)	B, moderate (80%)
2018	Yu et al. [38]	–	–	–	–	–	–	–	–	Cross-sectional study [j]	D, very low (55%)
2018	Bahirat et al. [39]	–	–	–	D, P [k]	It depends	It depends	No	81.54%	No	–
2018	Raber et al. [40]	–	–	–	D	Yes [l]	Yes	Yes	70%	Non-controlled before-after study [m]	C, limited (70%)
2019	Klingensmith et al. [41]	●	●		D	Not always	Yes	No	–	No	–
2019	Barbosa et al. [42]	–	–	–	–	–	–	–	86.8% [n]	Survey α	–
2019	Alom et al. [43]	–	–	–	–	–	–	–	Up to 72.2% (satisfaction)	Cross-sectional study [o]	D, very low (55%)
2019	Alom et al. [44]	–	–	–	–	–	–	–	96.4% and 94.5% [p]	Yes, for labeling and evaluation α	–
2020	Kasaraneni et al. [45]		●		D	Yes	Yes	No	–	No	–
2020	Kaur et al. [46]	–	–	–	–	–	–	–	–	No	–
2020	Botti-Cebria et al. [47]		●		D	Yes	Yes	No	– [q]	No	–
2020	Kökciyan et al. [48]	–	–	–	–	–	–	–	–	No	–
2020	Sanchez et al. [49]	–	–	–	Unclear	Yes	Yes	No	84.74%	Online survey to build their dataset α	–
2021	Kaur et al. [50]	–	–	–	–	–	–	–	–	No	–
2021	Lobner et al. [51]	–	–	–	–	–	–	–	83.33% [r]	Survey α	–
2022	Filipczuk et al. [52]	●	●		D	Yes	Yes	No	65% [s]	Non-controlled before-after study [t]	C, limited (70%)
2022	Hirschprung et al. [53]	–	–	–	–	–	–	–	–	Cross-sectional study	D, low (60%)
2022	Kökciyan et al. [54]	–	–	–	No	It depends	Yes	No	Between 41 and 92% [u]	No	–
2022	Ulusoy et al. [55]	–	–	–	–	–	–	–	Around 75% [v]	No	–
2022	Zhan et al. [56]	–	–	–	–	–	–	–	74%	No	–
2022	Brandão et al. [57]			●	–	–	–	–	Between 82 and 88%	Field study α	–
2022	Mendes et al. [58]		●		No	No	No	No	92%	Field study α	–
2022	Shanmugarasa et al. [59]	●			No	Yes	Yes	No	92.62%	Cross-sectional study [w]	D, very low (55%)
2023	Ayci et al. [60]		●		No	Yes	Yes	No	89%	No	–
2023	Serramia et al. [61]		●		No	Yes	No	No	3.78/5 [x]	Cross-sectional study [y]	D, very low (55%)
2024	Wang et al. [62]		●		No	No	Yes	No	–	No	–

[a] Only location
[b] Not necessarily, depends on what they call the Configuration Options
[c] A survey-based study and a direct user evaluation
[d] Two surveys
[e] Between-participants design
[f] Online survey
[g] Online questionnaire
[h] Not by default, they have a sort of 'user settings' for expert users
[i] For privacy leakage detection (not to be confused with preferences detection)
[j] To measure the interpretability of the approaches
[k] Not consistently
[l] Based on the current data collection
[m] Online survey
[n] AUC of binary allow/deny for a given scenario
[o] User satisfaction
[p] Accuracy based on appreciation of evaluators
[q] The accuracy presented is for the right category of data
[r] With interpretability of the results
[s] On average, but seems higher in specific case
[t] Between-subject experimental design
[u] Depends on several parameters
[v] Difficult to assess because they measure utility of decisions in a simulated setting
[w] Online survey
[x] Acceptability rate, not accuracy
[y] To measure the level of comfort of the norms inferred
α Alpha means that the user study is not meant to assess the solution, but only meant to collect data

well as practical works, with two out of four providing a tool ([27] and [61]).

2) NOT-INHERENTLY TRANSPARENT

a: CLASSIFICATION

Non-transparent classification models (found in 14 papers) typically encompass classic neural networks (as in Klingensmith et al. [41]) and deep neural networks (see for instance Yu et al. [38]); random forests [32], Ada Boost [58] and Support Vector Machines (used in Wijesekera et al. [37]) complete the picture. Post-hoc explanations must complement these models, as they are not easily understandable by themselves.

b: REINFORCEMENT

Reinforcement learning is the least used family of machine-learning techniques in AI-driven PPAs. It is implemented in Kaur et al. [50] and Ulusoy and Yolum [55], both used to adapt users' feedback to their preferences, and in Zhan et al. [56]. The first paper uses it to disclose information (using permissions), while the second uses it to learn bidding preferences in a negotiation context.

c: LOGIC-BASED

AI-driven PPAs can be based on logic (5 papers), for instance, expert systems (Kökciyan et al. [54] uses an agent-based model) or game theory (such as Hirschprung and Alkoby [53]). These works, albeit few, span various system contexts and types of decisions.

d: LLM

We selected only one paper using Large-Language Models (LLMs) [62] during an update of the survey. Here, the authors leverage GPT-3.5 and GPT-4.0 to produce *privacy rules*,

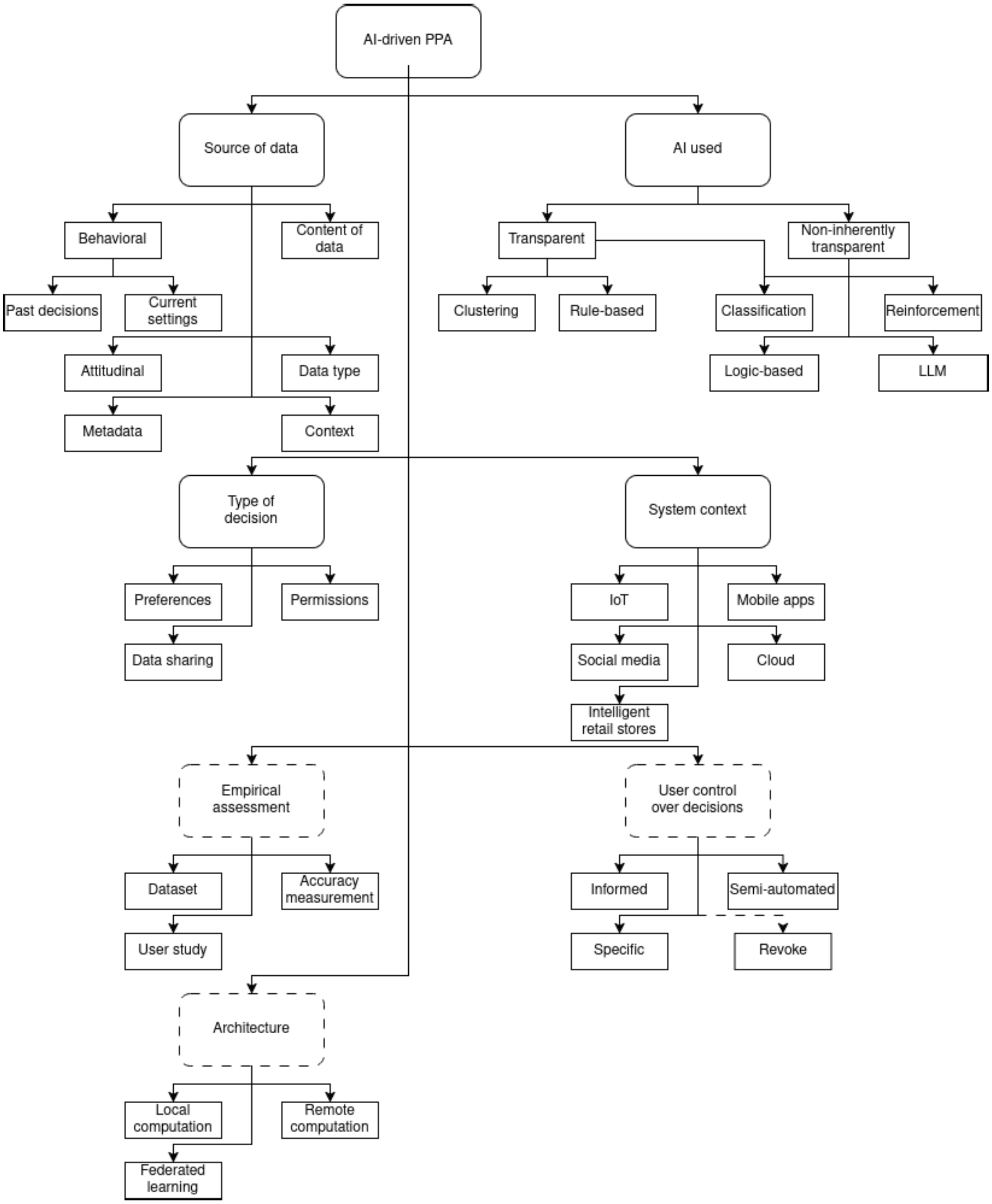

FIGURE 2. A schematic representation of the classification presented in Section V. Each facet is represented as a *rounded* box, *solid* for the mandatory features, and *dashed* for the optional ones. For User control over decisions (see Section V-G), we distinguish between *qualities* of control (solid arrows) and *instruments* of control (dashed arrows).

based on sensors outputs and user preferences. Note that LLMs are based on deep neural networks and are, therefore, difficult to explain and even prone to hallucinations.

C. SOURCE OF DATA

An AI-driven PPA can rely on various *sources of data* when using AI to help with a privacy decision. These data sources are very often combined, and a careful choice is necessary to fully exploit the potential of the models described in the previous section.

1) CONTEXT

Context is an often-used data source, yet not always well-defined. However, when it is defined, it is composed of the location [22], the time, relationships with other individuals [30], or the activity performed [43].

External data provided by third parties or other unrelated entities is sometimes used to predict privacy decisions, and this external data can arguably be considered context. For instance, under this term, we find risk factors [60] or information related to other applications in the background [57].

Context is usually a crucial component for an effective AI-driven PPA because, as has been argued under Nissembaum's theory of privacy as contextual integrity [65], context is paramount to designing appropriate information flows and respecting privacy norms.

2) ATTITUDINAL DATA

A few AI-driven PPAs ask users questions to elicit so-called *attitudinal data* about stated practices, or preferences regarding privacy recommendations to avoid the so-called cold-start problem, which arises when no past data is available to provide a recommendation. For example, Nakamura et al. [33] focuses on asking a minimal set of questions while keeping accuracy as high as possible, or Alom et al. [44] asks *"a reasonable [sic] number of questions (50) to the users."*

3) BEHAVIORAL DATA

Another common source of data is *behavioral data*. Behavioral data has the advantage of reflecting the *actual* users' privacy decisions to predict the next ones, as it does not simply rely on stated practices (unlike attitudinal data). While it can be a powerful tool, it can also create a feedback loop, reinforcing the same decisions.

Behavioral data can encompass past decisions, such as in Zhan et al. [56], which leverage past choices to fill a knowledge base and then use them to predict privacy decisions. It can also comprise current settings or preferences on a specific type of data to infer a decision for another type [24]. The system can also use these preferences to match users to a particular privacy profile, using for instance clustering techniques (see Section V-B1b).

4) METADATA

Metadata is data that provides information about other data, for example, the name of an application used [37], network requests [36], the purpose associated with processing [42], the usage frequency of certain permissions (such as location) by an app [46], or tags associated with images [25]. To some extent, metadata can overlap with context, for instance, when considering time or location. However, the articles surveyed more often refer to the time and location of collection *of a certain data point* for metadata, and to the *current time and location* when a decision has to be made for context. Metadata can provide peripheral information to make decisions, although it is rarely used as a sole source of data (out of 14 leveraging metadata, only 4 papers [36], [41], [45] rely solely on it).

5) DATA TYPE

The *data type* refers to the category of data concerned by the decision, such as whether it is an image to share on social media [31], the location requested by an app [52], or various sensor data by an IoT device [59]. The type of data can provide accurate information about the sensitiveness of a decision (location data can, for instance, provide sensitive information regarding the users' context, e.g., from location data that reveals that a user visits a clinic or church, medical, or religious information could be inferred), yet only a relatively low number of solutions rely on the data type to build an AI-driven PPA [29], [31], [33], [37], [51], [52], [59].

6) CONTENT OF DATA

The *content of data* refers to the specific content of a data point, as the name indicates. However, we also include data that can be directly inferred from the content of data under this category. For example, Botti-Cebria et al. [47] and Dong et al. [28] estimate the sensitivity of the content of the information to be shared to help make a decision. Indeed, content can be leveraged to tailor decisions: a picture deemed private should not receive the same treatment as one deemed public, and a geolocation trace that may potentially allow inferring religious practice should cautiously be dealt with.

D. SYSTEM CONTEXT

Most AI-driven PPAs target a specific *system context*, that is, a set of technologies with distinct characteristics. Indeed, each system context has specific requirements that one must consider when designing an AI-driven PPA. System contexts differ by the availability of an **interface**, **computational power**, and control over the **architecture**.

1) MOBILE APPS

Several works focus on mobile applications, and often on Android [23], [37]. Mobile ecosystems have the advantage of being well-defined ecosystems, enabling the possibility to strictly enforce privacy decisions (i.e., it is often addressed with permissions, see Section V-A1).

Mobile phones also possess reasonable computational power (in the sense that they can run an AI-driven PPA) and a screen enabling direct user interactions. Hence, an AI-driven PPA can be implemented directly on a smartphone (see Baarslag et al. [29]), and it can interact with and even regulate mobile apps, all of which make mobile ecosystems suitable candidates for AI-driven PPAs under the users' control.

2) INTERNET OF THINGS

Another widely used system context for AI-driven PPAs is the Internet of Things (IoT). We understand IoT as a network of devices, including sensors, mechanical and digital machines, as well as consumer devices, all connected to the Internet. In practice, AI-driven PPAs have been developed for smart homes [42], [59], on campuses [35], or for wearables such as fitness devices [49] for instance.

Most IoT devices are usually not equipped with proper interfaces and lack computational power. These characteristics make it challenging to build AI-driven PPAs assisting with permission settings, yet not impossible (see Klingensmith et al. [41] for instance, who manage to do so with an AI-driven PPA located on end devices).

3) SOCIAL MEDIA
According to our classification of the literature, the third major system context is social media, for which several AI-driven PPAs have been designed to help make privacy decisions. In this case, neither the interface nor the computational power are usually limiting factors. However, the design and implementation of social media platforms (that are usually not published openly) make it difficult to assess the binding character of privacy decisions supported by AI-driven PPAs running on social media platforms. AI-driven PPA solutions are rather designed to support data sharing, i.e., whether a specific post should be shared on social media and with whom, than focusing on assisting users with privacy decision-making.

4) CLOUD
A less prevalent system context is cloud environments, with only one of the reviewed articles proposing an AI-driven PPA targeting cloud environments [24]. Their solution offers a method to simplify information disclosure in cloud environments such as Google Drive. However, this work is thus a lone example and contrasts with the otherwise balanced distribution of works among other system contexts.

5) INTELLIGENT RETAIL STORE
Similar to cloud environments, only one work, Raber et al. [40], was captured under this category. Their AI-driven PPA provides a solution to automate decisions in intelligent retail stores, combining pervasive computing and online applications.

E. ARCHITECTURE
By architecture, we refer here to where the computation happens, i.e., the decision-making, and not necessarily the pre-processing steps such as building privacy profiles. Directly connected to the architecture is the trust model of the AI-driven PPA. While this term is usually reserved for security-oriented research, describing whether one has to trust the different entities or not provide relevant information for understanding the privacy boundaries.

Note that the location of the computation is only relevant for implemented AI-driven PPAs, and not for theoretical models. Similarly, most solutions surveyed do not explicitly describe a threat or trust model in their paper. Nonetheless, it is possible to infer that trusted parties are required in some solutions. For instance, Tan et al. [36] describe an architecture comprising a remote classifier (in which one has to place trust), yet no trust model is described.

1) LOCAL COMPUTATION
The processing can happen locally on the user device, such as on a smartphone (see, e.g., Olejnik et al. [34]), but this device can also be a home pod in an IoT context (see, e.g., Shanmugarasa et al. [59]).

Creating and processing user profiles, using local AI models, and locally deriving privacy decisions have the advantage that the user can keep control over the locally processed data, including their profiles and AI models, which usually can include sensitive information about the user's preferences or behavior. However, local data processing also puts more responsibilities on the user to secure the devices properly against malware or other attacks.

2) REMOTE COMPUTATION
The AI-driven PPA could also be based on remote data processing (according to the user's point of view), involving a central server that processes personal privacy decisions and contextual data, including, e.g., location data or another type of data. Remote computation raises the question of the trust placed in the party performing this computation to protect the data properly, to enforce the data subject's rights (e.g., to access or to delete their data and computed profiles or models), and not to use the data for any unintended purposes [36].

Several solutions rely on a remote third party that has to be trusted, e.g., Baarslag et al. [29], or Tan et al. [36]'s solution that places trust on their own remote classifier. The solution developed by Wang et al. [62] relies on ChatGPT, a closed-source chatbot on which all trust has to be placed with little or no accountability. In contrast, others only require trusting the operating system (OS) on which the AI-driven PPA is implemented [34], or require trusting both the OS and mobile applications [23].

3) FEDERATED LEARNING
Only one article, by Brandão et al. [57], presented an AI-driven PPA based on federated learning. In this work, the processing of user data for the computation of locally trained neural network models happens on the user devices. These devices only share the neural network weights with a central server, which will, in turn, average all the local weights and send back the results to the clients, which can use these new weights to continue the training process. Federated learning is a privacy-enhancing approach for processing the users' raw data only locally, which can achieve a performance comparable to the centralized approach (remote computation). Nonetheless, federated learning could still be attacked, e.g., with membership inference attacks, to leak personal data from locally trained models [66].

F. EMPIRICAL ASSESSMENT
AI-driven PPAs' performance can be measured in terms of accuracy, but because several solutions are meant to be usable tools, assessing an AI-driven PPA encompasses more than a mere measurement of how well a privacy decision is predicted.

As mentioned in Section III-B4, an empirical assessment can be an evaluation (see, e.g., [26]) or a validation (e.g., [53]).

1) USER STUDY

A classical way to validate a tool or a method is to conduct a user study, and we found 16 papers reporting a user study to validate usability. A user study can have various interpretations, ranging from a simple questionnaire to rate satisfaction (such as Alom et al. [43]) to a large-scale randomized controlled study (see, e.g., Liu et al. [26]) – the former being more akin to a mere validation, the latter a full-fledged evaluation.

Note that several works elicited data to build a dataset through a user study, which was therefore not meant as a means of assessment (annotated as α in Table 7).

2) PURELY STATISTICAL (DATASET)

Several works provide a validation without a user study, that is, only based on a purely statistical analysis based on a dataset [39], [45], [47], [48], [54], [56], [60],. Such a measure, although potentially subject to a higher degree of statistical rigor, cannot necessarily capture users' expectations and may even fall into the pitfall of Goodhart's law.[5]

3) ACCURACY MEASUREMENT

Accuracy can measure the capacity of an AI-driven PPA to predict a privacy decision, but not all papers measure the same type of accuracy. Tan et al. [36] measure privacy leakage detection, Botti-Cebria et al. [47] whether the correct category of data is predicted or not, Amoros et al. [61] the acceptability rate, Barbosa et al. [42] the Area Under the Curve (AUC) of a binary allow/deny for a given scenario, etc.

Other works, while they do measure the accuracy of their solution to predict a privacy decision, present their work with limited rigor or precision. For example, Fogues et al. [30] only present their results in plots. In contrast, others, such as Olejnik et al. [34], dedicate an entire subsection to explaining accuracy measurements.

G. USER CONTROL OVER DECISIONS

Finally, AI-driven PPAs should not only assist users with making privacy decisions but should at the same time also empower users with various options to improve *control over their decisions*. These options span over **qualities** of control (solid arrows in Figure 2) and **instruments** of control (dashed arrows). The former denotes adjectives that can be appended to control (akin to non-functional requirements in software engineering [67]), and the latter denotes concrete possibilities or actions for users (similar to functional requirements).

These options are partly related to GDPR requirements for consent (introduced in Section II-A), which are thus relevant for privacy decisions that constitute consent. Note, however, that only a handful of papers specifically refer to legal considerations. Filipczuk et al. [52] refer to the GDPR, Mendes et al. [58] acknowledge that an automated response to a permission request might not constitute legal consent, Lobner et al. [51] base the rationale of explainability on legal requirements, and Sanchez et al. [49] even claim GDPR compliance. Nonetheless, decisions for setting permissions for mobile operating systems, for instance, still require consent at installation or run time. Thus, legal requirements for consent remain relevant for these types of decisions.

1) EX-ANTE TRANSPARENCY

Under Art 13 GDPR, data subjects should receive information if data is collected from them, and informing users is also an integral part of the dominant transparency paradigm in the US (the *notice* of the notice and choice approach). Informing data subjects with intelligible notices arguably improves their control over decisions. Several AI-driven PPAs only inform about the type of data concerned by the privacy decision [27], some inform in addition about the controller [64] or of the purpose of processing [39]. In theory, meeting this criterion should not be difficult, although providing intelligible notices requires significant expertise in practice (as illustrated in Schaub et al. [68]).

2) SEMI-AUTOMATED

The semi-automated character of a decision refers to including an affirmative action of the user to confirm the decision, which is therefore not fully automated [69]. Most solutions provide a semi-automated decision process, although not systematically (e.g., Das et al. [35] mention that only opt-out is possible for facial recognition), or not always (e.g., Klingensmith et al. [41] offers different types of "privacy profiles", one of which – *Laissez-Faire* – enables full automation). Tan et al. [36] do not leave users in the loop by default, but the system allows the possibility to change the settings for "experienced users," while it depends on the Configuration Option for Hirschprung et al. [24].

3) SPECIFIC

The specificity of a decision refers to the presentation and the possibility for users to decide on the granularity of each data type, purpose, and controller separately. For an AI-driven PPA, it means having a fine-grained selection process, during which users should not be presented with bundled decisions. For instance, Shanmugarasa et al. [59]'s solution works per "situational context": who (is requesting data), data type, purpose, and re-sharability (to third parties); while the solution of Xie et al. [22] only works for one type of data (location), therefore only meeting this option in a restricted sense.

4) REVOKE

Finally, we observed that some AI-driven PPAs enable users to withdraw decisions. Here, rather than denying a decision or a recommendation, revoking operates after a given decision to withdraw it. This feature has rarely been observed in practice – at least explicitly – although the solution of Liu et al. [26] allows revoking previously granted decisions.

[5] According to which "When a measure becomes a target, it ceases to be a good measure."

Revoking previously made decisions, such as sharing data on social media, can be challenging to enforce. Also, note that certain operating systems – such as mobile OSes – will still allow users to revoke their decisions manually, although we stress that this action is performed outside the AI-driven PPA.

VI. DISCUSSION

This systematic literature review provides unique insights into how state-of-the-art research has designed AI-driven PPAs in recent years. For instance, IoT became a system context of interest only in 2018, and we observed a similar late adoption trend for reinforcement learning after 2021.[6] However, AI techniques have been used in every system context for all types of decisions throughout the years without any apparent pattern. While this lack of a clear pattern is not the most informative in itself, we ought to look instead at the **gaps** this survey highlights, the **challenges** AI-driven PPAs raise, then to inform better **design and development recommendations** based on these analyses.

We acknowledge that our survey of scientific articles reveals primarily gaps in the state-of-the-art research on AI-driven PPAs, not gaps in AI-driven PPAs that are already used in practice. Nonetheless, the best practice recommendations for addressing identified gaps also target developers of AI-driven PPAs and may, in these cases, not be appropriate for research projects (as opposed to deployed systems). However, knowledge and awareness of these best practice recommendations may still be helpful for researchers nonetheless.

Based on our main findings, this section provides a detailed discussion organized in seven parts: the issues of properly *evaluating AI-driven PPAs* in Section VI-A; AI-driven PPAs not sufficiently addressing *Privacy-by-Design* in Section VI-B; the (lack of) *explanations and explainability* in Section VI-C; the concerns surrounding *system contexts* in Section VI-D; the relationship with *legal considerations* in Section VI-E; the challenges in leveraging different *sources of data* in Section VI-F; and finally, potential *research avenues* are introduced in Section VI-G.

A. EVALUATING AI-DRIVEN PPAS

The problem of the evaluation of AI-driven PPAs is two-fold. First, we observe that **the evaluations of AI-driven PPAs are not based on the same or comparable accuracy metrics or measurements**. As presented in Section V-F, accuracy is measured regarding a privacy decision, but also a privacy leakage, acceptability rate, etc. Second, **our data shows a lack of user study evaluations**, and our critical appraisal shows a trend toward "low" and "very low" scores to assess cause and effect. Only 15 out of 39 papers mentioned that they performed a user study to evaluate their solution,[7],[8] but only six studies scored above (or equal to) 70% based on the CEBMA critical appraisal we performed. We acknowledge that user studies may go beyond the scope of strictly theoretical papers (e.g., models or frameworks without prototype implementation). Yet, we contend that the validation offered by these theoretical papers, often cross-validated on a dataset, is far from being able to reflect reality. Any proposed AI-driven PPAs must be validated and evaluated to substantiate empirical evidence of their value and feasibility. Without setting unrealistic standards for research, it is still essential that academics and developers strive to put their proposed solutions to the test in real-world settings.

To address the issue of disparate indicators, it is crucial to establish standardized accuracy metrics that can be uniformly applied across studies, facilitating more meaningful cross-study comparisons. Additionally, there should be a greater emphasis on conducting high-quality user trials, as these are essential for providing rigorous empirical validation and ensuring the practical feasibility of AI-driven PPAs.

Recommendation: *Based on the current lack of empirical evidence, we propose that the usability of AI-driven PPAs should be evaluated through user studies following high-quality standards for qualitative and quantitative research, and such evaluations should notably encompass the accuracy of the privacy decision taken.*

B. LACK OF PRIVACY-BY-DESIGN

Since AI-driven PPAs typically analyze the users' attitudinal or behavioral privacy preferences, metadata or content, or other data types for personalized assistance, they need for this purpose to process personal data and user profiles, which could be considered sensitive data. We identified, however, a gap regarding following a privacy-by-design approach for AI-driven PPAs, since hardly any of the papers we surveyed focus on, or mention how, the AI-driven PPAs themselves can be designed in a privacy-preserving manner. More specifically, among papers describing technical architectures,[9] **only one uses federated learning as a privacy-enhancing approach** [57], however, without discussing that federated learning is still vulnerable to privacy attacks, such as model reconstruction and member inference attacks (see e.g. Shaw et al. [21] and Mothukuri et al. [70]).

Therefore, federated learning needs to be complemented with other PETs, such as differential privacy, which reduces the risk of re-identification attacks by adding random noise. Besides such data obfuscating PETs, also other PETs are available that protect the confidentiality of personal data when creating and/or using trustworthy AI models, including encrypted data processing technologies such as homomorphic encryption, functional encryption or

[6] Some papers may have been published on the topic earlier than in 2013, the year from which we started to include papers in our SLR.

[7] Some papers include a user study for collecting data, which is not focused on their proposed solution.

[8] We verified whether the authors of all submitted papers published follow-up articles, and did not find any.

[9] Recall that theoretical papers are excluded from this analysis.

multi-party computation and the use of a trusted execution environment (TEE) (see also OECD [71] or Canard et al. [72] for overviews). Nonetheless, since the implementation of trustworthy and privacy-preserving AI models may also (in the case of obfuscation PETs) reduce the model accuracy, or may come with computational or communication costs, PETs need to be implemented and configured with care for achieving suitable trade-offs between privacy protection and accuracy, performance costs.

Many presented AI-driven PPAs require trust in a central server, where the data processing is performed, while data processing on the users' local device may be preferable from a privacy perspective, as it does not require trusting another (central) party. To this end, Wijesekera et al. [37] provides an insightful analysis of the trade-off of having either a local (offline) or a remote computation, concluding that offline learning still performs well (almost 95% accuracy). Also note that the privacy threat models are rarely described, making it difficult to evaluate security and privacy assumptions critically.

Recommendation: *We contend that AI-driven PPAs must embrace stronger privacy-by-design principles, including better design strategies but also better integration of Privacy Enhancing Technologies, with suitable privacy – accuracy and – performance trade-offs. In particular, PETS for achieving data minimization, such as federated learning combined with differential privacy, or PETs for enabling privacy-preserving data analytics by AI-driven PPAs based for instance on multi-party computation, homomorphic or functional encryption, should be implemented and deployed.*

C. UNEXPLAINABLE AI

Another pitfall identified is the lack of explanations provided by most AI-driven PPAs, combined with the lack of explainability/interpretability offered by the AI models used. **Only one of the surveyed papers explicitly addresses explainability of the generated decisions** [51], and only 8 use transparent models (see Section V-B) to make predictions.

The growing trend to use deep learning architectures may not facilitate the explainability of decisions, but this challenge is not insurmountable. It is indeed possible to devise *post hoc* explanations, and to take inspiration from other existing work on usable explanations for AI-made decisions. Note, however, that inherent transparency can come at the expense of other quality aspects (e.g., accuracy, security, safety, ethical and social considerations [62]) of decisions – trade-offs must be considered case-by-case. Deep neural networks tend to outperform their simpler counterparts, although this statement does not seem to generalize to all kinds of decisions, such as decisions made in highly unpredictable settings like social predictions [73].

As discussed in Section II-B, transparency of AI can be a legal requirement in some specific use cases related to AI-driven PPAs. For instance, transparency is required for the data controller according to the GDPR, or for the provider or deployer according to the AI Act, even though this will not apply to most AI-driven PPAs and use cases. In fact, none of the surveyed papers related to high-risk AI applications. Transparency can in general also foster trust in technology [74].

To enhance transparency, future research should prioritize the integration of post-hoc explanation tools and ensure that decision rationales are clearly presented in the user interface. Balancing accuracy and transparency is essential, and adopting inherently interpretable models or supplementary explanation tools can help achieve this equilibrium.

Recommendation: *Based on this analysis, we recommend 1) considering the use of inherently explainable AI techniques, such as decision trees, for the classification, whenever this implies that the potential quality loss will be appropriate for the specific use case, or 2) the integration of ad-hoc explanations otherwise, e.g., for neural networks and SVMs.*

D. MISSING SYSTEM CONTEXT

Our SLR covered five system contexts: mobile applications, social media, IoT, the cloud, and online retail stores. We are, however, surprised **not to find other contexts, such as web browsers or Trigger-Action Platforms (TAPs)**. The former because cookie notices are notoriously a "hassle" for users in modern web experience; we therefore expected to encounter solutions tackling this issue.[10] The latter refers to platforms offering applications for connecting otherwise unconnected devices and services using simple recipes, such as "Every morning at 7 am, send a Slack message with the first meeting of the day from Google Calendar." Trigger-action programming has gained a lot of traction in the last years (IFTTT, the most important TAP, boasts over 27 million users [75]), yet no AI-driven PPAs specifically addresses it.

Both these system contexts possess their specific features: many controllers with non-standard interfaces for cookie notices, and numerous actors mediated through a single centralizing entity for TAPs. They therefore require targeted efforts from designers to offer adequate technological solutions to manage privacy decisions.

Recommendation: *We encourage researchers and developers of AI-driven PPAs to expand their efforts into a broader range of system contexts, also encompassing but not limited to web browsers and TAPs.*

E. FEW LEGAL CONSIDERATIONS

As some privacy decisions made by AI-driven PPAs have legal privacy implications or issues, legal requirements, e.g., under the GDPR, the AI Act, or other national legislation, should be discussed and considered for the design and use of AI-driven PPAs. However, a couple of reasons could

[10] Recall that Bollinger et al. [18] does not personalize decisions.

explain the lack of discussion of legal requirements and implications according to the EU legal framework. Firstly, the geographic distribution of the solutions surveyed, given that only 13 papers have authors with EU affiliations. Secondly, the timing of the publications, as 17 papers were published before the GDPR was enacted and none before the AI Act came into force. Nonetheless, it is still surprising to find **only 4 papers mentioning (but not even addressing) legal considerations**.

In the more general case, we contend that even when legal privacy principles do not apply for a particular use case or context, **they can still provide valuable guidelines for the design of AI-driven PPAs**. For instance, assisting users with making informed, unambiguous, and explicit privacy decisions (as required for consent) may foster trust in AI-driven PPAs even when the privacy decision does not formally constitute consent. Also, usable explanations of the risks and implications when using an AI-driven PPA can in general help raise awareness among users.

Recommendation: *We recommend a deeper consideration of legal requirements for the design of AI-driven PPAs. Considering GDPR legal requirements could particularly amount to: 1) meeting consent requirements when assisting on decisions related to consent, such as permission settings; 2) the introduction of AI-driven PPAs assisting and enabling users to exercise their data subject rights; and, 3) the incorporation of usable explanations for the logic behind the AI-based proposed decisions, as well as information about the significance and the envisaged consequences of such automated processing for the data subject. Moreover, for considering legal requirements of the AI act, we recommend the design of AI-driven PPAs for assisting users with exercising their right to human oversight of critical decision-making processes involving their personal data, including guidance on potential overrides of automated decisions*

F. USE OF VARIED SOURCES OF DATA, ACCOUNTING FOR BOTH CONTEXT AND PERSONAL PREFERENCES

Lastly, our study yielded that AI-driven PPAs leverage various sources of data (context, attitudinal data, behavioral data, type of data, content of data, and metadata), but not necessarily within the same solution. However, utilizing several of these data sources can be a challenge in itself, as it requires careful curation of the datasets and adequate use of the AI models. The benefits harnessed can be high, resulting potentially in higher prediction accuracy and, thus, in higher quality of privacy guidance and assistance.

We also acknowledge the difficulty of determining certain sources of data – such as context –, or the problem of the sensitivity of data. As mentioned in Section V-C1, context is rarely defined. It is, however, possible to draw inspiration for a rigorous definition from the seminal paper by Barth et al. [76] on the formalization in a logical framework of the concept of contextual integrity coined by Nissenbaum [65]. As for the sensitivity of data, it is notably incumbent on context when, for instance, the same type of data (e.g., location) can be deemed non-sensitive in one context (e.g., at a workplace in the middle of the week), but sensitive in another (e.g., Sunday morning near a church, thereby disclosing potential religious beliefs).

Recommendation: *Based on the relative singularity of data sources, we advocate for a plurality of data sources, encompassing context as much as personal preferences.*

G. RESEARCH AVENUES

In this final section of our Discussion, we explore prospective research paths on AI-driven PPAs, informed by the results of our study and the current social, technical, and legal landscape.

1) THE FUTURE OF AI-DRIVEN PPAS AND GENERATIVE AI

While the uptake of generative AI, such as Large Language Models (LLMs), is undeniable, their application to AI-driven PPAs is not yet prevalent in the literature. **Only one of the included studies leverages LLMs to enable automated privacy decisions.** This solution, referred to as PrivacyOracle, allows users to have a "privacy firewall" for filtering and managing personal data flows in the context of smart buildings [62]. Nonetheless, we can also acknowledge other research not included in this SLR, such as the work of Hamid et al. [77] that provided a benchmark for evaluating Generative AI-based Privacy Assistants to simplify and make privacy policies more user-friendly. Note that this work was not included in the SLR, as it only enhanced explanations but did not automate any privacy decisions.

We are perhaps one step away from having a new wave of LLM-powered AI-driven PPAs. LLMs are great for summarizing and capturing insights from a large amount of text (e.g., privacy policies, logs, traffic data, and system documentation). We envision that if such insights become reliable enough, the user's privacy preferences could be automatically matched with a given system's privacy configurations, semi-automating decisions, providing allow/deny rules based on previous privacy settings for similar systems, etc. In the work of Wang et al. [62], their PrivacyOracle was already achieving 98% accuracy in identifying privacy-sensitive states from sensor data and 75% accuracy in measuring the social acceptability of information flows. However, such opportunities also raise a series of risks.

LLMs are intrinsically challenging to explain and lack transparency and interpretability. Furthermore, due to the risk that they respond with false or misleading information presented as facts (or "hallucinate"), **they conflict directly with compliance requirements such as the GDPR data accuracy principle,[11] and should therefore be**

[11] Art 5 (i) (e) GDPR indeed stipulates that data needs to be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy')".

incorporated into AI-driven PPAs with caution. Still, future research should address opportunities and challenges of designing and using LLM-based AI-driven PPAs, as well as technical and legal requirements for involving LLMs in assisting users with privacy decisions.

This SLR shows that integrating generative AI and LLMs into AI-driven PPAs remains largely unexplored. It is worth mentioning that this finding was rather unexpected to our research team, as the SLR's search scope has included publications up to 2025, and generative AI has already made its mark in the privacy research area [78], [79]. In particular, recently LLMs have been developed and tested for assessing privacy policies [80], and it has been demonstrated that LLMs can be very helpful for analyzing and extracting privacy practices from privacy policies [81], and thus provide valuable input for users' decision-making. Therefore, despite the risks that LLMs bring, it is foreseeable that they can be further enhanced and meaningfully used to automate users' privacy tasks, such as in AI-driven PPAs. With that in mind, this SLR benefits from being inherently extendable, and in a few more years, researchers can update the review to confirm this finding.

2) DESIGNING GENUINELY PRIVACY-FRIENDLY AI-DRIVEN PPAS

A promising yet critical avenue for future research remains **to design a genuinely privacy-friendly AI-driven PPA, with the right amount of notice** to empower users and avoid the so-called "consent fatigue." This right amount of notice can be a difficult balance to achieve – some users favor more notice than others – but it is a crucial step for the uptake of such assistants.

The design should naturally be informed by the latest results in the academic literature [82]; it should carefully consider the number of notices, their content, their timing, etc. However, it should also be complemented by usability studies conducted in the early stages of the prototype, as iterations over the design of the assistant are likely to be required to fine-tune it.

3) TRUST IN THE AI ASSISTANTS AND AUTOMATION BIAS

Individuals tend to overly trust AI systems and favor AI-based decision-making while ignoring contradictory information made without automation, a phenomenon known as automation bias [83], which is a problem also related to the Elisa effect first described by Weizenbaum [84].

If the user's decisions are biased towards following a privacy recommendation proposed by a AI-driven PPA, the users' autonomy may be negatively impacted in practice. Hence, future research should examine if users may too easily trust and rely on proposed or nudged decisions by AI-driven PPAs without critically judging or adapting proposed decisions and how such a problem could be addressed by Human-Computer Interaction research.

VII. THREATS TO VALIDITY

A. THREAT I–PLANNING LIMITATIONS OF THE SLR

The first threat relates to the planning of the SLR in terms of identifying the need and justification for this study. Here, we were concerned with identifying existing reviews (systematic and non-systematic) on the topic of AI-driven PPAs. The initial searches did not reveal any review studies on the topic, as described in Section III-A, pointing to a significant gap in secondary research AI PPAs. The planning phase of the SLR is also critical to outline the research questions and provide the basis for an objective investigation of the studies that are being reviewed. If the RQs are not explicitly stated or omit the key topics, the literature review results can be flawed, overlooking the key information. To mitigate this threat, we outlined two RQs and objectives (Section III-A). In summary, we seek to minimize any bias or limitations during the planning phase when defining the scope and objectives of this SLR. As a last step in the planning phase, the team finalized and cross-checked the study protocol to minimize the limitations of the SLR plan before proceeding to the subsequent phases.

B. THREAT II–VALIDITY OF THE SEARCH PROCESS

Identifying and selecting the studies reviewed in the SLR are also significant processes to be observed. Selecting studies is a critical step; if any relevant papers are missed, the results of the SLR may be flawed. Therefore, we followed a stepwise process (Section III-B), starting with a literature screening and followed by a complete reading of papers. This selection process was carried out independently by two reviewers. We also performed forward and backward snowballing, looking for references to other potentially relevant studies. Also, this SLR restricts the selection of publications to four scientific databases: Scopus, Web of Science, IEEE Xplore, and ACM Digital Library. These databases were used due to their high relevancy to computer science, privacy, and data protection, as well as to maintain a feasible search space. This search process gives us confidence that we minimized limitations related to (i) excluding or overlooking relevant studies or (ii) including irrelevant studies that could impact the results and their reporting in the SLR.

C. THREAT III–POTENTIAL BIAS IN THE SYNTHESIS PROCESS

Some threats should also be considered regarding the potential bias in synthesizing the data from the review and documenting the results. This means that if there are some limitations in the data synthesis, they directly impact the results of this SLR. Typical examples of such limitations could be a flawed research taxonomy and a mismatch of potential research gaps. To minimize the bias in synthesizing and reporting the results, we have created a data extraction form that uses well-known classification schemes, such as the ones proposed by Wieringa et al. [64] and Creswell and Creswell [85], or Arrieta et al. [13] for the classification of AI.

Three researchers independently reviewed this data extraction form while revising the research protocol. While one of the researchers led the data extraction step, two other authors helped by cross-checking the work throughout the process for consistency. Three authors were involved in the creation of the classification scheme derived from the literature (i.e., shown in Section 2), actively working on reviewing the list of categories for consistency through a series of meetings. Furthermore, this SLR also offers a complete replication package, enabling researchers to reproduce or extend this review (https://github.com/Victor-Morel/SLR_AI_PPA).

VIII. CONCLUSION

With the SLR presented in this article, we provide a classification and common vocabulary to compare and discuss AI-driven PPAs. Although many papers (41 in our selection) have already been published on AI-driven PPAs in the last decade, they do not yet form a coherent body of literature. AI-driven PPAs can be improved by performing standard evaluations (including their usability), integrating privacy by design in their design process, providing additional explanations for their decisions, and considering a broader range of system context and larger variety of data sources. We hope this survey and its classification allow users and developers of AI-driven PPAs to compare different solutions and understand their pros and cons. Moreover, the recommendations should help improve AI-driven PPAs in different ways, addressing the challenges raised by AI's latest developments (including LLMs), data collection, and modern regulations.

REFERENCES

[1] *State of IoT 2024: Number of Connected IoT Devices Growing 13% to 18.8 Billion Globally*, IoT Analytics, 2024. [Online]. Available: https://iot-analytics.com/number-connected-iot-devices

[2] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*, document 32016R0679, European Commission, Official Journal of the European Union, Apr. 2016.

[3] H. Choi, J. Park, and Y. Jung, "The role of privacy fatigue in online privacy behavior," *Comput. Hum. Behav.*, vol. 81, pp. 42–51, Apr. 2018.

[4] N. Sadeh, B. Liu, A. Das, M. Degeling, and F. Schaub, "Personalized privacy assistant," U.S. Patent 10 956 586, Mar. 23, 2021.

[5] *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC), no. 300/2008, (EU), no. 167/2013, (EU), no. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text With EEA Relevance)*, document 32024R1689, European Commission, Official Journal of the European Union, Jun. 2024.

[6] A. F. Westin, "Privacy and freedom," *Washington Lee Law Rev.*, vol. 25, no. 1, p. 166, 1970.

[7] *Opinion 2/2013: Apps on Smart Devices*, document Article 29 Working Party, 2013.

[8] S. Human, H. J. Pandit, V. Morel, C. Santos, M. Degeling, A. Rossi, W. Botes, V. Jesus, and I. Kamara, "Data protection and consenting communication mechanisms: Current open proposals and challenges," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Jun. 2022, pp. 231–239.

[9] Cal. Civ. Code §1798.100 et Seq. (2018). *California Consumer Privacy Act of 2018*. [Online]. Available: https://oag.ca.gov/privacy/ccpa

[10] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle, *The Platform for Privacy Preferences 1.0 (P3P1. 0) Specification*, vol. 16. Wakefield, MA, USA: W3C Recommendation, 2002.

[11] W3C Recommendation. (2019). *Tracking Preference Expression (DNT)*. [Online]. Available: https://www.w3.org/TR/tracking-dnt/

[12] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. London, U.K.: Pearson, 2002.

[13] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[14] C. Panigutti, R. Hamon, I. Hupont, D. Fernandez Llorca, D. Fano Yela, H. Junklewitz, S. Scalzo, G. Mazzini, I. Sanchez, and J. S. Garrido, "The role of explainable AI in the context of the AI Act," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, 2023, pp. 1139–1150.

[15] E. Puiutta and E. M. Veith, "Explainable reinforcement learning: A survey," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction* Springer, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-57321-8_5puiutta_explainable_2020

[16] B. Kitchenham, "Procedures for performing systematic reviews," Keele Univ., Keele, U.K., NICTA Tech. Rep. 0400011T.1, 2004.

[17] K. Marky, A. Stöver, S. Prange, K. Bleck, P. Gerber, V. Zimmermann, F. Müller, F. Alt, and M. Mühlhäuser, "Decide yourself or delegate–user preferences regarding the autonomy of personal privacy assistants in private IoT-equipped environments," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–20.

[18] D. Bollinger, K. Kubicek, C. Cotrini, and D. Basin, "Automating cookie consent and GDPR violation detection," in *Proc. 31st USENIX Secur. Symp. (USENIX Secur.)*, 2022, pp. 2893–2910.

[19] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proc. 18th Int. Conf. Eval. Assessment Softw. Eng.*, May 2014, pp. 1–10.

[20] M. Kuhrmann, P. Diebold, and J. Münch, "Software process improvement: A systematic mapping study on the state of the art," *PeerJ Comput. Sci.*, vol. 2, p. e62, May 2016.

[21] M. Shaw, "Writing good software engineering research papers," in *Proc. 25th Int. Conf. Softw. Eng.*, May 2003, pp. 726–736.

[22] J. Xie, B. P. Knijnenburg, and H. Jin, "Location sharing privacy preference: Analysis and personalized recommendation," in *Proc. 19th Int. Conf. Intell. User Interface*, Feb. 2014, pp. 189–198.

[23] W. Apolinarski, M. Handte, and P. J. Marron, "Automating the generation of privacy policies for context-sharing applications," in *Proc. Int. Conf. Intell. Environ.*, Jul. 2015, pp. 73–80.

[24] R. Hirschprung, E. Toch, and O. Maimon, "Simplifying data disclosure configurations in a cloud computing environment," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–26, May 2015.

[25] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede, "Privacy policy inference of user-uploaded images on content sharing sites," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 1, pp. 193–206, Jan. 2015.

[26] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. A. Zhang, N. Sadeh, Y. Agarwal, and A. Acquisti, "Follow my recommendations: A personalized privacy assistant for mobile app permissions," in *Proc. 12th Symp. Usable Privacy Secur. (SOUPS)*, 2016, pp. 27–41.

[27] D. A. Albertini, B. Carminati, and E. Ferrari, "Privacy settings recommender for online social network," in *Proc. IEEE 2nd Int. Conf. Collaboration Internet Comput. (CIC)*, Nov. 2016, pp. 514–521.

[28] C. Dong, H. Jin, and B. P. Knijnenburg, "PPM: A privacy prediction model for online social networks," in *Proc. Int. Conf. Social Inform.* Springer, 2016. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-47874-6_28dong_ppm_2016

[29] T. Baarslag, A. T. Alan, R. Gomer, M. Alam, C. Perera, E. H. Gerding, and M. schraefel, "An automated negotiation agent for permission management," in *Proc. 16th Int. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 2017, pp. 380–390. [Online]. Available: https://eprints.soton.ac.uk/405608/1/An_Automated_Negotiation_Agent_for_Permission_Management.pdf baarslag_automated_2017

[30] R. L. Fogues, P. K. Murukannaiah, J. M. Such, and M. P. Singh, "SoSharP: Recommending sharing policies in multiuser privacy scenarios," *IEEE Internet Comput.*, vol. 21, no. 6, pp. 28–36, Nov. 2017.

[31] H. Zhong, A. Squicciarini, D. Miller, and C. Caragea, "A group-based personalized model for image privacy classification and labeling," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3952–3958.

[32] G. Misra and J. M. Such, "PACMAN: Personal agent for access control in social media," *IEEE Internet Comput.*, vol. 21, no. 6, pp. 18–26, Nov. 2017.

[33] T. Nakamura, S. Kiyomoto, W. B. Tesfay, and J. Serna, "Easing the burden of setting privacy preferences: A machine learning approach," in *Information Systems Security and Privacy*, vol. 691, O. Camp, S. Furnell, and P. Mori, Eds., Cham, Switzerland: Springer, 2017, pp. 44–63.

[34] K. Olejnik, I. Dacosta, J. S. Machado, K. Huguenin, M. E. Khan, and J.-P. Hubaux, "SmarPer: Context-aware and automatic runtime-permissions for mobile devices," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 1058–1076.

[35] A. Das, M. Degeling, D. Smullen, and N. Sadeh, "Personalized privacy assistants for the Internet of Things: Providing users with notice and choice," *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 35–46, Jul. 2018.

[36] H.-Z. Tan, W. Zhao, and H.-H. Shen, "A context-perceptual privacy protection approach on Android devices," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.

[37] P. Wijesekera, J. Reardon, I. Reyes, L. Tsai, J.-W. Chen, N. Good, D. Wagner, K. Beznosov, and S. Egelman, "Contextualizing privacy decisions for better prediction (and Protection)," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–13.

[38] J. Yu, Z. Kuang, B. Zhang, W. Zhang, D. Lin, and J. Fan, "Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1317–1332, May 2018.

[39] P. Bahirat, Y. He, A. Menon, and B. Knijnenburg, "A data-driven approach to developing IoT privacy-setting interfaces," in *Proc. 23rd Int. Conf. Intell. User Interface*, Mar. 2018, pp. 165–176.

[40] F. Raber, D. Ziemann, and A. Krueger, "The 'Retailio' privacy wizard: Assisting users with privacy settings for intelligent retail stores," in *Proc. 3rd Eur. Workshop Usable Secur.*, 2018, p. 23.

[41] N. Klingensmith, Y. Kim, and S. Banerjee, "A hypervisor-based privacy agent for mobile and IoT systems," in *Proc. 20th Int. Workshop Mobile Comput. Syst. Appl.*, Feb. 2019, pp. 21–26.

[42] N. M. Barbosa, J. S. Park, Y. Yao, and Y. Wang, "'What if?' predicting individual users' smart home privacy preferences and their changes," *Proc. Privacy Enhancing Technol.*, vol. 2019, no. 4, pp. 211–231, 2019.

[43] M. Z. Alom, B. Carminati, and E. Ferrari, "Helping users managing context-based privacy preferences," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jul. 2019, pp. 100–107.

[44] M. Z. Alom, B. Carminati, and E. Ferrari, "Adapting Users' privacy preferences in smart environments," in *Proc. IEEE Int. Congr. Internet Things (ICIOT)*, Jul. 2019, pp. 165–172.

[45] M. Kasaraneni and J. P. Thomas, "A self–learning personal privacy assistant," in *Advanced Information Networking and Applications*, vol. 1151. Cham, Switzerland: Springer, 2020.

[46] H. Kaur, I. Echizen, and R. Kumar, "Smart data agent for preserving location privacy," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2020, pp. 2567–2575.

[47] V. Botti-Cebriá, E. Del Val, and A. García-Fornes, "Automatic detection of sensitive information in educative social networks," in *Proc. 13th Int. Conf. Comput. Intell. Secur. Inf. Syst. (CISIS)*, 2021, pp. 184–194.

[48] N. Kökciyan and P. Yolum, "TURP: Managing trust for regulating privacy in Internet of Things," *IEEE Internet Comput.*, vol. 24, no. 6, pp. 9–16, Nov. 2020.

[49] O. R. Sanchez, I. Torre, Y. He, and B. P. Knijnenburg, "A recommendation approach for user privacy preferences in the fitness domain," *User Model. User-Adapted Interact.*, vol. 30, no. 3, pp. 513–565, Jul. 2020.

[50] H. Kaur, R. Kumar, and I. Echizen, "Reinforcement learning based smart data agent for location privacy," in *Proc. Int. Conf. Adv. Inf. Netw. Appl.*, 2021, pp. 657–671.

[51] S. Löbner, W. B. Tesfay, T. Nakamura, and S. Pape, "Explainable machine learning for default privacy setting prediction," *IEEE Access*, vol. 9, pp. 63700–63717, 2021.

[52] D. Filipczuk, T. Baarslag, E. H. Gerding, and M. C. Schraefel, "Automated privacy negotiations with preference uncertainty," *Auto. Agents Multi-Agent Syst.*, vol. 36, no. 2, p. 49, Oct. 2022.

[53] R. S. Hirschprung and S. Alkoby, "A game theory approach for assisting humans in online information-sharing," *Information*, vol. 13, no. 4, p. 183, Apr. 2022.

[54] N. Kökciyan and P. Yolum, "Taking situation-based privacy decisions: Privacy assistants working with humans," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 703–709.

[55] O. Ulusoy and P. Yolum, "PANOLA: A personal assistant for supporting users in preserving privacy," *ACM Trans. Internet Technol.*, vol. 22, no. 1, pp. 1–32, Feb. 2022.

[56] X. Zhan, S. Sarkadi, N. Criado, and J. Such, "A model for governing information sharing in smart assistants," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Jul. 2022, pp. 845–855.

[57] A. Brandão, R. Mendes, and J. P. Vilela, "Prediction of mobile app privacy preferences with user profiles via federated learning," in *Proc. 12th ACM Conf. Data Appl. Secur. Privacy*, Apr. 2022, pp. 89–100.

[58] R. Mendes, M. Cunha, J. P. Vilela, and A. R. Beresford, "Enhancing user privacy in mobile devices through prediction of privacy preferences," in *Proc. Eur. Symp. Res. Comput. Secur.*, 2022, pp. 153–172.

[59] Y. Shanmugarasa, H.-y. Paik, S. S. Kanhere, and L. Zhu, "Automated privacy preferences for smart home data sharing using personal data stores," *IEEE Secur. Privacy*, vol. 20, no. 1, pp. 12–22, Jan. 2022.

[60] G. Ayci, M. Sensoy, A. Özgür, and P. Yolum, "Uncertainty-aware personal assistant for making personalized privacy decisions," *ACM Trans. Internet Technol.*, vol. 23, no. 1, pp. 1–24, Feb. 2023.

[61] M. Serramia, W. Seymour, N. Criado, and G. Flucke, "Predicting privacy preferences for smart devices as norms," in *Proc. 22nd Int. Conf. Auto. Agents Multiagent Syst.*, 2023, pp. 2262–2270.

[62] B. Wang, L. A. Garcia, and M. Srivastava, "PrivacyOracle: Configuring sensor privacy firewalls with large language models in smart built environments," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2024, pp. 239–245.

[63] *CAT Manager App*, CEBMa, Leiden, The Netherlands. Accessed: Feb. 11, 2025. [Online]. Available: https://cebma.org/resources/tools/cat-manager-app/

[64] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: A proposal and a discussion," *Requirements Eng.*, vol. 11, no. 1, pp. 102–107, Mar. 2006.

[65] H. Nissenbaum, "Privacy as contextual integrity," *Wash. L. Rev.*, vol. 79, no. 1, pp. 119–157, 2004.

[66] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 3–18.

[67] F. Pallas, K. Koerner, I. Barberá, J.-H. Hoepman, M. Jensen, N. R. Narla, N. Samarin, M.-R. Ulbricht, I. Wagner, K. Wuyts, and C. Zimmermann, "Privacy engineering from principles to practice: A roadmap," *IEEE Secur. Privacy*, vol. 22, no. 2, pp. 86–92, Mar. 2024.

[68] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor, "A design space for effective privacy notices," in *Proc. 11th Symp. Usable Privacy Secur. (SOUPS)*, 2018, pp. 365–393.

[69] V. Morel and S. Fischer-Hübner, "Automating privacy decisions—Where to draw the line?" in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Jul. 2023, pp. 108–116.

[70] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, Feb. 2020.

[71] OECD, "Sharing trustworthy AI models with privacy-enhancing technologies," OECD Publishing, Paris, France, Tech. Rep. 38, Jun. 2025.

[72] S. Canard, B. Vialla, B. Bozdemir, O. Ermis, M. Önen, M. Barham, B. Rozenberg, R. Shmelkin, and M. Azraoui, "Papaya project-D3.3—Complete specification and implementation of privacy preserving data analytics," Eur. Res. Project, Papaya, Tech. Rep., 2020. [Online]. Available: https://www.papaya-project.eu/node/157.html

[73] A. Narayanan and S. Kapoor, *AI Snake Oil: What Artificial Intelligence Can do, What it Can't, and How to Tell the Difference*. Princeton, NJ, USA: Princeton Univ. Press, 2024.

[74] S. Crane, H. Lacohée, and S. Zaba, "Trustguide—Trust in ICT," *BT Technol. J.*, vol. 24, no. 4, pp. 69–80, 2006.

[75] *IFTTT—Automate Business & Home*, IFTTT, San Francisco, CA, USA. Accessed: Nov. 7, 2024. [Online]. Available: https://ifttt.com/

[76] A. Barth, A. Datta, J. C. Mitchell, and H. Nissenbaum, "Privacy and contextual integrity: Framework and applications," in *Proc. IEEE Symp. Secur. Privacy (S&P)*, May 2006, pp. 15–pp.

[77] A. Hamid, H. R. Samidi, P. Pappachan, T. Finin, and R. Yus, "GenAIPABench: A benchmark for generative AI-based privacy assistants," *Privacy Enhancing Technol.*, vol. 2024, no. 3, pp. 336–352, Jul. 2024.

[78] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy," *IEEE Access*, vol. 11, pp. 80218–80245, 2023.

[79] A. Golda, K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola, and B. Sikdar, "Privacy and security concerns in generative AI: A comprehensive survey," *IEEE Access*, vol. 12, pp. 48126–48144, 2024.

[80] V. Freiberger, A. Fleig, and E. Buchmann, "'You don't need a university degree to comprehend data protection this way': LLM-powered interactive privacy policy assessment," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2025, pp. 1–12.

[81] D. Rodriguez, I. Yang, J. M. Del Alamo, and N. Sadeh, "Large language models: A new approach for privacy policy analysis at scale," *Computing*, vol. 106, no. 12, pp. 3879–3903, Dec. 2024.

[82] Y. Feng, Y. Yao, and N. Sadeh, "A design space for privacy choices: Towards meaningful privacy control in the Internet of Things," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–16.

[83] M. L. Cummings, "Automation bias in intelligent time critical decision support systems," in *Decision Making in Aviation*. Evanston, IL, USA: Routledge, 2017, pp. 289–294.

[84] J. Agassi and J. Wiezenbaum, "Computer power and human reason: From judgment to calculation," *Technol. Culture*, vol. 17, no. 4, p. 813, Oct. 1976.

[85] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Newbury Park, CA, USA: Sage, 2017.

VICTOR MOREL received the Ph.D. degree in computer science from Inria-Privatics, and the French National Engineering degree from Polytech' Lyon, France. He is currently a Postdoctoral Researcher with the Security and Privacy Laboratory, Chalmers University of Technology, on usable privacy for IoT applications. His research interests include privacy and data protection, networks security, usability and human–computer interactions, applied cryptography, and ethics of technology in a broad manner. He appreciates sober, hackable, and accessible technology that works. He is a Selected Member of the EDPB's Support Pool of Experts.

LEONARDO HORN IWAYA received the B.Sc. degree in computer science from Santa Catarina State University, Brazil, the M.Sc. degree in electrical engineering from the University of São Paulo, and the Ph.D. degree in computer science from Karlstad University, Sweden. From 2011 to 2014, he was a Research Assistant with the Laboratory of Computer Networks and Architecture (LARC), PCS-EPUSP. From 2019 to 2021, he was a Postdoctoral Researcher with the School of Computer Science, The University of Adelaide, Australia, as part of the Cyber Security Cooperative Research Centre (CSCRC). From 2021 to 2022, he was a Postdoctoral Researcher with the Interdisciplinary Research Group on Knowledge, Learning and Organizational Memory (KLOM), Federal University of Santa Catarina. He is currently an Associate Senior Lecturer with the Department of Mathematics and Computer Science, Karlstad University. He is also with the Privacy and Security (PriSec) Research Group, Karlstad University, contributing to projects, such as DHINO and DigitalWell Arena. His research interests include privacy engineering, cybersecurity, human factors, mobile and ubiquitous health systems, and the privacy impacts of new technologies.

SIMONE FISCHER-HÜBNER (Senior Member, IEEE) has been a Full Professor with Karlstad University, since June 2000, where she is the Head of the Privacy and Security (PriSec) Research Group. Since April 2022, she has been a part-time Guest Professor with Chalmers University of Technology. She has been conducting research in privacy, cyber security, and privacy-enhancing technologies for more than 35 years. She is the Coordinator of the Swedish IT Security Network for Ph.D. Students (SWITS) and the Swedish Industrial Graduate School for Cybersecurity (SIGS-CyberSec). She is the Swedish Representative of the International Federation for Information Processing (IFIP) TC 11 (Technical Committee on Information Security and Privacy Protection), the Vice Chair of IFIP TC 11, and a Board Member of the Privacy Enhancing Technology Symposia (PETS) and the NordSec conferences. She received the IFIP Silver Core Award, in 2001, the IFIP William Winsborough Award, in 2016, and an Honorary Doctorate from Chalmers University of Technology, in 2020.