PrivacyAkinator: Articulating Key Privacy Design Decisions by Answering LLM-Generated Multiple-choice Questions

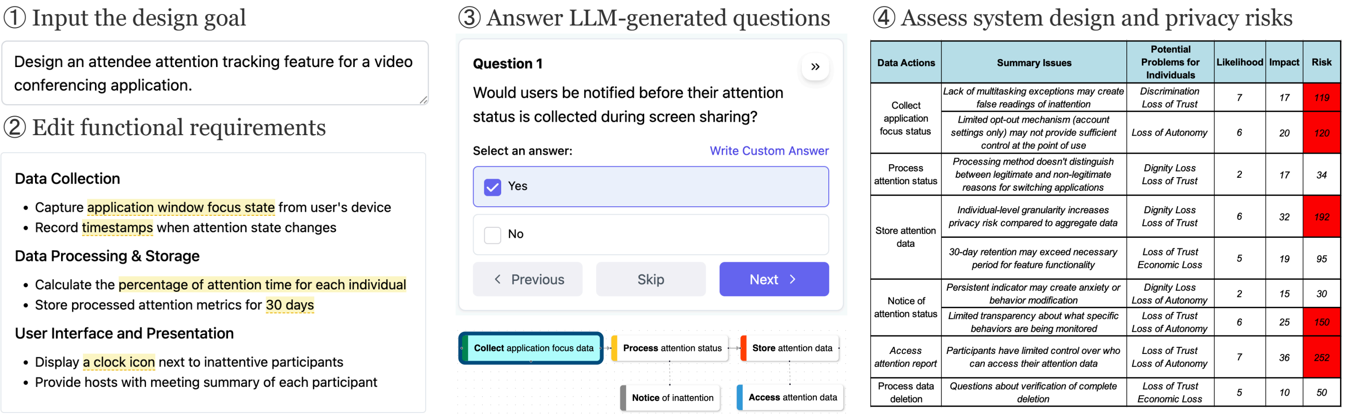
Qiyu Li
University of California San Diego
La Jolla, California, USA
qiyuli@ucsd.edu

Yuen Sum Wong
University of California San Diego
La Jolla, California, USA
y6wong@ucsd.edu

Yuen Kei Wong
University of California San Diego
La Jolla, California, USA
ykw001@ucsd.edu

Longxuan Yu
University of California Riverside
Riverside, California, USA
ylong030@ucr.edu

Haojian Jin
University of California Riverside
La Jolla, California, USA
haojian@ucsd.edu



Design an attendee attention tracking feature for a video conferencing application.

Data Collection

- Capture application window focus state from user's device
- Record timestamps when attention state changes

Data Processing & Storage

- Calculate the percentage of attention time for each individual
- Store processed attention metrics for 30 days

User Interface and Presentation

- Display a clock icon next to inattentive participants
- Provide hosts with meeting summary of each participant

Question 1

Would users be notified before their attention status is collected during screen sharing?

Select an answer:

☑ Yes

□ No

Data Actions	Summary Issues	Potential Problems for Individuals	Likelihood	Impact	Risk
Collect application focus status	*Lack of multitasking exceptions may create false readings of inattention*	*Discrimination* *Loss of Trust*	*7*	*17*	*119*
	Limited opt-out mechanism (account settings only) may not provide sufficient control at the point of use	*Loss of Autonomy*	*6*	*20*	*120*
Process attention status	*Processing method doesn't distinguish between legitimate and non-legitimate reasons for switching applications*	*Dignity Loss* *Loss of Trust*	*2*	*17*	*34*
Store attention data	*Individual-level granularity increases privacy risk compared to aggregate data*	*Dignity Loss* *Loss of Trust*	*6*	*32*	*192*
	30-day retention may exceed necessary period for feature functionality	*Loss of Trust* *Economic Loss*	*5*	*19*	*95*
Notice of attention status	*Persistent indicator may create anxiety or behavior modification*	*Dignity Loss* *Loss of Autonomy*	*2*	*15*	*30*
	Limited transparency about what specific behaviors are being monitored	*Loss of Trust* *Loss of Autonomy*	*6*	*25*	*150*
Access attention report	*Participants have limited control over who can access their attention data*	*Loss of Trust* *Loss of Autonomy*	*7*	*36*	*252*
Process data deletion	*Questions about verification of complete deletion*	*Loss of Trust* *Economic Loss*	*5*	*10*	*50*

Figure 1: PrivacyAkinator helps developers articulate key privacy design decisions by answering LLM-generated multiple-choice questions. After (1) developers provide a high-level description of their system, (2) PrivacyAkinator expands it into detailed, editable functional requirements with adaptable design choices highlighted. (3) PrivacyAkinator guides developers through key design decisions with contextualized, specific questions. As developers make choices, PrivacyAkinator organizes and documents these decisions within a structured privacy representation. (4) PrivacyAkinator maps these design decisions to NIST PRAM worksheets [75] to further identify potential issues and prioritize privacy risks. Note that we only enumerate data actions and relevant design decisions; analysts still need to manually evaluate risk likelihood and impact.

Abstract

NIST's Privacy Risk Assessment Methodology (PRAM) provides a structured framework for privacy experts to assess privacy risks. However, its complexity and reliance on expert knowledge make it difficult for novice developers to use effectively. This paper explores methods to lower these barriers. We first performed an observational study with 12 participants using PRAM in real-world scenarios, and found that novice developers struggled most with articulating privacy-related design decisions. We then developed PrivacyAkinator, an interactive tool that helps developers articulate key privacy decisions by answering LLM-generated multiple-choice questions. PrivacyAkinator introduces three innovations: a universal privacy representation that abstracts privacy-related design decisions into data flows and stakeholder interactions; a domain-aware design space mined from 10K privacy-related news articles; and a dynamic question-generation workflow to prioritize relevant questions. Our user study with 24 participants suggests that developers using PrivacyAkinator identified 47% more key decisions in 73% less time compared to PRAM.

CCS Concepts

• **Security and privacy** → **Usability in security and privacy**; • **Human-centered computing** → **Interactive systems and tools**; **User studies**.

Keywords

Privacy Risk Assessment, Privacy Engineering, Privacy by Design, Developer Support Tools

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3790408

ACM Reference Format:
Qiyu Li, Yuen Sum Wong, Yuen Kei Wong, Longxuan Yu, and Haojian Jin. 2026. PrivacyAkinator: Articulating Key Privacy Design Decisions by Answering LLM-Generated Multiple-choice Questions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3772318.3790408

1 Introduction

The U.S. National Institute of Standards and Technology (NIST) released the Privacy Risk Assessment Methodology (PRAM) in 2019 to help organizations analyze, assess, and prioritize privacy risks [75]. PRAM provides step-by-step guidance on risk assessments through four worksheets (WS): WS1 defines business objectives and organizational privacy governance; WS2 assesses system design; WS3 helps prioritize risks based on likelihood and impact; and WS4 selects appropriate controls to mitigate risks. Despite the institutional endorsement, there is little empirical evidence on the effectiveness of NIST's PRAM in practice [46, 104].

In addition to NIST's PRAM, several other frameworks offer complementary strategies to assess privacy risks, such as Privacy Impact Assessment (PIA) [109] and LINDDUN [24]. These methods typically follow a similar process: defining system objectives, system mapping, threat enumeration, impact/risk analysis, and mitigation [52]. However, prior studies found that these frameworks can be cumbersome and overly complex [8, 112], often requiring significant expertise to use effectively [111]. As a result, many organizations, particularly startups without dedicated privacy professionals, perceive privacy risk assessments as onerous burdens rather than practical tools for managing privacy risks [56, 101].

In this paper, we investigate methods to lower the barrier for novice developers to apply privacy risk assessments. We first performed an observational study to explore the challenges they face when applying PRAM. We recruited 12 participants, including students and junior software engineers, and asked them to conduct privacy risk assessments for four task scenarios using PRAM worksheets. We began with a brief training session that walked participants through the four worksheets and demonstrated their use with an example from PRAM materials. Participants then selected one scenario they were familiar with, and followed the worksheet instructions to complete each one. We found that participants struggled most with articulating privacy design decisions within the system, which revealed three key challenges: (1) participants struggled to distribute their attention across numerous privacy design decisions; (2) PRAM's vague terminology and open-ended structure created confusion about what constitutes appropriate responses; and (3) participants often missed important design decisions without explicit prompting.

We then designed PrivacyAkinator (Figure 1), an interactive tool that helps developers articulate key privacy-related design decisions by answering multiple-choice questions generated by large language models (LLMs). The tool specifically targets PRAM's WS2 (Assess System Design), as novice developers struggled most with this stage in our studies. PrivacyAkinator transforms this open-ended task into a series of multiple-choice questions with concrete answer options. By decomposing system design into discrete design choices, PrivacyAkinator allows developers to focus on one decision at a time. The system leverages LLMs to transform abstract privacy concepts into specific, contextual questions with concrete options. By dynamically generating questions based on previous responses, PrivacyAkinator guides developers' attention toward key design decisions that they might otherwise overlook.

While prior studies have explored using question answering to communicate privacy concepts or explain system behaviors [29, 39, 81], these efforts primarily focus on explaining existing content such as privacy policies. In contrast, the unique design challenge for PrivacyAkinator is to guide developers through a vast, underexplored privacy design space. PrivacyAkinator has three key innovations.

First, we design a novel universal representation for privacy design (Figure 4) that captures key privacy design decisions to help developers concentrate upon the essentials without distraction from irrelevancies. Unlike existing models (e.g., Data Flow Diagrams, Component Diagrams [3]) that focus solely on data flows, such as who is collecting the data, how it is collected, when, and why, our representation also captures interaction designs with different stakeholders. We organize privacy-related information about a data practice into three levels: (1) a high-level data flow modeling information movement through the system, (2) nodes that represent specific data actions (e.g., collect, process) and stakeholder interactions (e.g., obtaining consent, requesting data deletion), and (3) node properties that specify individual design choices (e.g., opt-in/out).

Second, we introduce a data-driven approach to construct a domain-aware privacy design space (Figure 5). In contrast to prior work that creates taxonomies through manual labeling [105], we mine domain-specific taxonomies of privacy design decisions by analyzing privacy-relevant documents, such as public privacy news reports. Building on our privacy representation, we map each decision to a key-value pair, organized into three categories: (1) universal keys with universal values (e.g., *consent_mode: opt-in/opt-out*), (2) universal keys with domain-specific values (e.g., *data_type: medical records*), and (3) domain-specific keys with domain-specific values (e.g., *voice_masking: enabled*). We first reviewed the literature on privacy design [30, 36, 83, 85] and manually created an initial design space based on this structure. To enrich this design space, we text-mined 10K privacy-related news using LLMs, annotating domain labels and extracting relevant design choices. We then analyzed their co-occurrence statistics to uncover associations of privacy design decisions across different contexts.

Third, we devise a question generation workflow that dynamically prioritizes key privacy design decisions. PrivacyAkinator maintains an underlying representation of the current design and formulates two types of questions: exploratory questions that add new nodes to the data flow, and exploitative questions that leverage the design space to elicit specific design decisions for existing nodes. Since presenting all potential decisions at once would overwhelm developers, PrivacyAkinator strategically determines which decisions to present first. In each round, PrivacyAkinator selects the question type to balance exploration and exploitation, retrieves relevant design decisions, and prioritizes those strongly correlated with prior choices.

We conducted two experiments to validate the effectiveness of our design. We first evaluated the coverage of key design decisions using our system. To establish ground truth, we organized brainstorming sessions with privacy experts to examine 30 real-world data practices. We provided initial descriptions of each data practice and answered questions based on their actual implementations. Our results show that PrivacyAkinator identified 94% of key design decisions recognized by privacy experts, and its options covered 77% of design choices made in practice. We then performed a user study with 24 participants and asked them to apply the NIST PRAM framework to three real-world application scenarios with or without our tool. We found that developers using PrivacyAkinator were able to articulate 47% more key privacy decisions with 73% less time.

In this paper, we make the following contributions:

- An empirical study identifying three key challenges of applying PRAM for novice developers.
- PrivacyAkinator, a novel tool that uses LLM-generated multiple-choice questions to help developers articulate key privacy design decisions.
- A universal representation for privacy design that abstracts privacy-related design decisions into data flows and stakeholder interactions.
- A data-driven approach to construct a domain-aware privacy design space by mining privacy-related news.

2 Related Work

The main objective of this paper is to lower barriers for novice developers to conduct privacy assessments by (1) transforming privacy design into a structured task and (2) guiding developers through the process step-by-step using a series of multiple-choice questions. We organize related work into three categories: structuring privacy design, developer support for privacy, and privacy question answering.

2.1 Structuring Privacy Design

A central challenge in supporting developers with privacy engineering is that privacy design decisions are often diffuse, implicit, and difficult to articulate. We conceptualize structuring privacy design as systematically identifying and organizing privacy-related design decisions. Prior work has approached this problem by developing privacy representations and mapping out the privacy design space.

Privacy Representation. A fundamental challenge in engineering privacy into systems is to articulate privacy [10, 36, 43, 44]. Solove proposed a taxonomy focused on privacy violations [89]. Contextual Integrity [74] defines privacy as appropriate information flows. However, these representations are too high-level for practitioners to think about concrete system design decisions.

Recent studies found that practitioners often combine general-purpose diagrams to articulate a privacy design [12, 62, 96, 102], such as Data Flow Diagram (DFD) [57] and Unified Modeling Language (UML) [77]. Privacy nutrition labels offer a standardized format for disclosing what data is collected and how it is used [50, 51]. These representations help describe data practices, but they do not explicitly surface the underlying design decisions.

We hypothesize that not all design decisions are equally important; a common subset (e.g., retention periods, consent options) accounts for a substantial share of decisions developers face in practice. Our goal is to design an alternative privacy representation that makes these key decisions explicit and decomposes system design into discrete decisions.

Privacy Design Space. The complexity of privacy design has driven researchers to develop structured approaches for mapping out the privacy design space, particularly for notice and choice mechanisms [30, 85]. For example, Schaub et al. proposed a design space for privacy notices by identifying key dimensions such as timing, modality, and channel [85]. While these efforts provide usable taxonomies and vocabularies for categorizing and communicating different privacy designs, they focus narrowly on notice and user control rather than the broader privacy design landscape.

In parallel, there has been extensive work on the analysis of privacy policies [22, 38, 87, 94, 115]. Wilson et al. developed the OPP-115 taxonomy [105] that categorizes privacy practices by manually annotating a corpus of website privacy policies. Building on such efforts, natural language processing (NLP) techniques have been widely used to automate the extraction and summarization of salient information from privacy policies [6, 9, 64, 81, 115]. For example, ToS;DR extracts key points from privacy policies and presents them as easy-to-read summaries, improving accessibility and user understanding [82, 90]. However, these approaches offer little support for developers articulating design decisions because privacy policies focus on external-facing high-level statements, rather than implementation-level decisions developers need to consider when designing a system.

We hypothesize that a data-driven approach to constructing the privacy design space can better scale across diverse domains and surface recurring design decisions that matter in practice. In this work, we explore the feasibility of developing a privacy design space by mining design decisions from real-world data practices.

2.2 Developer Support for Privacy Design

Prior work has explored various ways to support developers in making privacy design decisions [58, 59]. Prior studies have identified common challenges developers face in understanding and implementing privacy principles [7, 60, 86, 93]. Developers often reported limited privacy awareness and knowledge, particularly in startups and small organizations that lack dedicated privacy experts [49, 79]. Without formal privacy training or institutional support, many developers rely on ad-hoc knowledge sources, which may be inconsistent or incomplete [79]. Many developers think of privacy in terms of security, missing broader privacy concerns such as data retention and internal misuse [37].

Several tools have been developed to embed privacy practices into software development workflows. For example, Coconut is an Android Studio plugin that helps developers manage privacy through required annotations [58]. PARROT supports privacy-aware IoT development with interactive guidance [3]. However, existing tools focus on supporting experienced developers with prior privacy knowledge. Little attention has been given to helping novice developers conduct privacy risk assessments. Our work aims to bridge this gap by decomposing the privacy assessment task into an interactive, step-by-step process that guides novices through concrete design decisions.

2.3 Privacy Question Answering

Prior research has explored the use of privacy question answering as an effective approach to improve the usability of privacy policies [29, 39, 80, 81]. For example, PriBots developed conversational assistants that answer users' questions about privacy practices to enhance readability and user comprehension [39]. Recent work in explainable AI explored using LLMs to build dialogue-based privacy tools [17, 32, 92]. For example, CLEAR used contextual, LLM-powered assistants to analyze privacy policies in real-time, highlight possible risks, and generate explanations as users interact with LLM applications [17]. Sun et al. built an open-ended QA agent to help users understand privacy policies [92]. In contrast, PrivacyAkinator is unique in (1) helping developers rather than users; (2) generating closed-ended questions to elicit design decisions rather than answering open-ended questions by interpreting existing policies.

Privacy Nutrition Labels require developers to answer questions about their app's data practices [50, 51]. For example, Apple Privacy Nutrition Labels ask developers to specify which data types their app collects, how this data is used, and whether it links to user identity or is used for tracking purposes [54, 114]. However, prior studies have found that filling out these forms is time-consuming and requires significant expertise [31, 61]. The labels often use vague terms to accommodate broad use cases, which may confuse developers [61]. In contrast, our work aims to help novice developers articulate privacy design decisions using multiple-choice questions similar to the Akinator game [84]. By leveraging LLMs, we generate specific, contextualized questions with concrete options to reduce ambiguity and potential confusion.

3 Background

In this section, we review existing techniques for privacy risk assessments and motivate our choice of PRAM.

Privacy Impact Assessment (PIA) [109] is a process used to evaluate the potential effects of information systems on individual privacy. PIAs are increasingly adopted by government agencies and organizations to encourage early integration of privacy considerations into the system development lifecycle [46, 108]. However, prior research found that existing PIA processes typically lack clear guidelines or methodologies to sufficiently support privacy risk assessments (PRA)–the technical evaluation of privacy risks [4, 23]. Consequently, PIAs often rely on ad-hoc, checklist-like approaches that serve more as compliance rituals than actually addressing privacy risks [8, 19, 101].

Several efforts have attempted to address these gaps [4, 23, 24, 78]. For example, PRIAM introduces a risk model that formalizes key risk factors: privacy harms, feared events, privacy weaknesses, and risk sources [23]. LINDDUN develops a taxonomy of privacy threats for identifying and mitigating privacy risks [24]. However, these frameworks are often overly abstract or complex, requiring significant privacy expertise to implement effectively [88, 104, 112].

Privacy Risk Assessment Methodology (PRAM) [75] stands as a notable effort to provide systematic guidance for conducting PRA. PRAM instantiates the risk model from NISTIR 8062 [12], which defines risks as the product of likelihood and impact of adverse privacy effects for individuals. PRAM guides practitioners through PRA using four worksheets (WS), each containing a set of tasks:

- *WS1: Framing Business Objectives and Organizational Privacy Governance.* WS1 requires analysts to elicit system requirements, including functional requirements (e.g., business goals) and privacy-related non-functional requirements (e.g., legal obligations).
- *WS2: Assessing System Design.* WS2 focuses on privacy threat modeling. *Task 2 (Supporting Data Map)* instructs analysts to create a system model or data flow diagram. *Task 3 (Contextual Factors)* encourages analysts to consider contextual factors such as the nature of the organizations and privacy expectations of these organizations. *Task 4 (Data Action Analysis)* requires the analyst to fill out a table by enumerating data actions in the system, the data being processed, relevant contextual factors, and a summary of issues.
- *WS3: Prioritizing Risk.* WS3 comprises four tasks: (1) *Assess Likelihood*: estimating the probability that a data action will become problematic for individuals, (2) *Assess Impact*: estimating the effects of potential problems for individuals using the organizational impact factors, (3) *Calculate Risk*: multiplying likelihood and impact to produce a risk score, and (4) *Prioritize Risks*: providing suggestions to help the organization prioritize risks.
- *WS4: Selecting Controls.* WS4 supports the selection of controls to mitigate privacy risks identified in WS3.

With the proliferation of digital services (e.g., mini-programs and microservices) and the rise of AI programming tools, software is increasingly being developed by individuals or small teams who lack privacy expertise [13, 25, 49, 79]. However, existing privacy risk assessment frameworks are mostly not designed for novices, and can be challenging for them to apply effectively. Based on these gaps, we propose the following research questions:

RQ1: How can privacy design be turned into a structured task?

RQ2: How can the structured privacy-assessment task be unfolded for novice developers?

4 Understanding Challenges of Using PRAM for Novice Developers

We conducted a study with 12 participants to explore the challenges of applying privacy risk assessments for novice developers. We used PRAM as a need-finding tool to uncover developers' underlying needs and pain points. We selected PRAM over other frameworks since (1) NIST offers well-structured worksheets and instructional materials; (2) PRAM is more lightweight than LINDDUN [24] and LINDDUN Go [112], which better suits novices.

4.1 Method

4.1.1 Recruitment.

Since our use case envisions startups hiring new graduates without dedicated privacy professionals, we primarily recruited students and junior professionals. We posted the recruitment advertisements through social media, including our institution's subreddit/Slack/Discord channels, and student mail lists, and sent targeted invitations to junior software developers.

Participants who expressed interest in the study were asked to fill out screening forms, which we used to determine their eligibility for the study. Selection criteria were based on age (over 18 years) and software development experience (e.g., coursework, internships, or

Table 1: Task scenarios used in our study. We selected four task scenarios based on participant familiarity, scenario complexity, and domain diversity. For each scenario, we provided participants with a high-level description of the system.

#ID	Scenario	Description	Participants
#1	Zoom Attendee Attention Tracking [5, 62]	A feature that monitored attendees' attention during video conferences by tracking whether Zoom was the active application on a participant's computer. It showed a clock icon next to inattentive participants and generated a post-meeting report with attention percentage scores.	P1, P2, P3, P8
#2	Alexa's Smart Home Voice Assistant [68]	A cloud-based voice service for smart home devices that allows users to control their home environment through voice commands. It captures audio when activated by a wake word, processes requests in the cloud, and executes actions through connected devices.	P4, P7, P11
#3	Target Retail Recommendation System [41]	An analytics system that identified pregnant customers based on purchase patterns of approximately 25 indicator products, allowing Target to predict pregnancy and estimate due dates. This enabled targeted marketing of pregnancy and baby-related products.	P5, P9, P12
#4	23andMe Direct-to-consumer Genetic Testing [71]	A service that collects and analyzes genetic information from customer saliva samples for ancestry, health traits, and disease risks. The service stores comprehensive genetic profiles that can reveal sensitive information about individuals and their biological relatives.	P6, P10

professional work). During screening, we excluded participants who had formal industry or research experience in privacy engineering, as the studies targeted novices lacking privacy expertise.

4.1.2 Participants. We recruited 12 participants (five identified as female, seven as male, aged 22 to 24) through social media and mailing lists. Each participant was compensated for their time with a $20 gift card. The sample included 2 graduate students, 9 undergraduates, and 1 software engineer. To assess participants' privacy knowledge, we asked them to rate their familiarity with privacy principles or "Privacy by Design" practices on a scale from 1 (not familiar) to 5 (very familiar). Reported knowledge was low (M = 1.58, SD = 1.11), with 10 out of 12 scoring 2 or below.

4.1.3 Study Procedure. We assumed participants to be the developers responsible for designing a specific app or service and asked them to use PRAM worksheets [75] to assess the privacy risks of one data practice. We first gathered a broad range of privacy-related scenarios, then selected four to cover diverse privacy-related design dimensions (e.g., interface, cloud, AI, and physical) across different domains (e.g., video conferencing, IoT, advertising, and health). We also chose scenarios with moderate complexity, so they do not impose a huge understanding overhead on participants. We asked participants to select one of four data practices (Table 1). For each scenario, we provided only a high-level description of its functional requirements to reflect real-world conditions. We imposed a 2-hour time limit for the study. We captured participants' responses through a hybrid format: hand-written format for WS1 and WS2 to facilitate brainstorming and drawing tasks, and using the NIST open source digital version (Excel) of WS3 to calculate risk. The choice among four data practices was selected based on participants' interests and familiarity with the cases, allowing them to engage with scenarios they found relevant to their experience. After the study, three authors reviewed the participants' assessments and asked additional questions to clarify details in their assessments. We started with broad questions to ask about participants' overall experiences and challenges with PRAM, followed by more targeted questions about their decision-making on specific worksheet tasks. We conducted all study sessions in person to allow for detailed observation and immediate feedback.

4.1.4 Data Analysis. We conducted a thematic analysis [103] of post-task interview transcripts and notes we took during the study. Two researchers independently coded 8 transcripts, then discussed their codes and developed a codebook that the two researchers agreed on. Using this codebook, they divided the remaining transcripts, with each coding two transcripts. Finally, the two researchers discussed and resolved coding conflicts. As our focus was on identifying emerging themes, we did not calculate interrater reliability (IRR) to measure theoretical agreement.

4.1.5 Ethical Considerations. The study received approval from the Institutional Review Board (IRB) of our institution. We obtained participants' consent electronically via Google Forms at the beginning of each study, and reserved their rights to withdraw at any time in the study. We informed participants that the information collected during the interview will be anonymized and only the results obtained from analyzing the transcribed interviews will be disclosed. To ensure participants' privacy, we used Zoom (approved by our institute) for transcription. After each interview, the research team checked the transcription using the recording, then immediately and permanently deleted the recording.

4.2 Findings

We identified four main challenges of conducting privacy risk assessments for novice developers based on participants' behaviors and feedback. The first three challenges relate to how they articulate system design, while the fourth involves estimating risk scores for privacy risks. We provide the codebook in Appendix E.

4.2.1 Attention Allocation. Participants struggled to manage their mental resources to focus on important design considerations. We identified two main causes of this challenge.

Information Overload. The complexity of PRAM led to cognitive overload for all participants, who had to juggle privacy concepts, technical details, and organizational factors simultaneously. As sessions progressed, this overload led to noticeable fatigue. Several participants (P1, P2, P4, P5, P7) described struggling to "keep all of this in [their] head at once" despite repeated exposure.

This cognitive strain was reflected in their task performance. Several participants offered detailed responses in WS1, but failed to articulate the same level of detail in WS2. In post-task interviews, participants attributed this drop-off to mental overload, with P4 noting that it was "*too hard to recall and include all the details...in one graph*."

Ineffective Prioritization (P1, P3, P4, P5, P7, P8, P11). Many participants struggled to prioritize privacy considerations effectively, often focusing on technically familiar aspects rather than the most critical privacy risks. Without clear guidance, they misallocated attention, spending too much time on early, familiar tasks and giving insufficient attention to later, equally important ones.

This was particularly evident in WS2 (*Assessing System Design*). Several participants concentrated heavily on the first data action (e.g., collection), but included only brief or incomplete responses for later actions like retention, transformation, or disclosure. P5 reflected that "*I spent too much focus and time on the first data action*." In task #1, P8 spent 10 minutes on collection but rushed through the rest, missing key issues like third-party disclosure. P8 explained that "*By the time I reached disclosure, my focus was fragmented*."

Participants tended to focus on privacy design decisions that aligned with their technical background, often giving less attention to unfamiliar but equally important aspects. For example, P4, a software engineer, provided detailed input on storage but overlooked the frequency of wake word detection, an important decision related to household surveillance. P4 explained that "*I wasn't confident about the low-level detail, so I chose not to extend that option*." In other cases, participants defaulted to whatever came to mind, with P11 noting that "*There are too many variables...I just couldn't think that more*."

4.2.2 *Inaccessible Privacy Knowledge.*

Participants often lacked the knowledge of what privacy aspects to consider or failed to recall relevant concepts when needed.

Knowledge Blindspot (P3, P8, P11). Several participants missed important design decisions because they did not know what to consider. In the Zoom attention tracking scenario, P8 initially overlooked issues such as data deletion and access control. However, during the post-study interview, when these considerations were discussed, he immediately recognized their importance and suggested improvements like allowing users to control deletion timing and limiting access to attention reports. P11 noted that "*It is not intuitive to do a systematic analysis of this task if I don't have too much privacy knowledge*."

Knowledge Recall Failure (P1, P2, P4, P7, P8). In several cases, participants had relevant privacy knowledge but didn't apply it when needed, often due to memory lapses or misdirected attention. Many Participants (P1, P2, P4, P5, and P7) struggled to carry earlier insights into later tasks. P5 noted that they had "*focused too much on the first part*," leaving little mental capacity for later stages. In the Zoom scenario, some participants overlooked user consent mechanisms, noting that they had considered them earlier but forgot to include them during the assessment because "[PRAM] *never prompted [them] to consider it*". P8 shared a similar experience, explaining that he forgot to include a notification "*not because he didn't know it, but because it slipped my mind*" during the task. He only recalled the information when the researcher brought it up in the post-study discussion.

4.2.3 *Insufficient Guidance.*

Participants found PRAM's instructions vague, with unclear expectations and ambiguous terms that left them uncertain about the level of detail and what to include.

Unclear Expectations for Detail and Scope (P2, P4, P5, P7). Participants were often unsure how much technical depth was expected in their responses. The framework lacked examples to guide the level of specificity, leading to confusion for participants. P4 directly expressed this uncertainty: "*Am I supposed to break down this system to the level of specific algorithms and data structures, or keep it high-level?*" This lack of clarity led participants to over-focus on their familiar areas, while overlooking critical issues such as user consent.

Beyond ambiguity in granularity required, participants also felt uncertain about what to include and how comprehensive their assessments should be. Many participants felt "*unsure what's expected*," and questioned whether they had done "*enough*", which led to hesitation and second-guessing. P7 mentioned that "*I just wish there are more specific instructions on what to fill out*."

Vague Terms (P3, P4, P5, P7, P9, P12). Participants felt that the terminology in PRAM WS2 was unclear. Terms used in PRAM (e.g., "generation/transformation" and "disclosure/transfer") were perceived as "*overly technical*" or "*ambiguous*". As P12 stated, "*The format is unclear and needs clearer definitions, for example, the difference between 'transformation' and 'transfer'*."

This confusion extended to the definitions and structure of the worksheets. Several participants noted inconsistencies in how concepts were presented. For example, the term "contextual factors" is initially defined as "circumstances surrounding the data by the system or individuals that influence whether a data action may be problematic", with examples like data sensitivity, data collection frequency. However, in the Contextual Factors tab, participants encountered a different framing–subjective factors from organizational, system, and user perspectives, such as user perceptions of data sensitivity. This mismatch left the participants unsure how to respond. As P12 noted, "*I wasn't sure what to fill out because the examples didn't match.*"

4.2.4 *Uncertainty in Risk Evaluation.*

Participants encountered two main challenges when applying PRAM's risk assessment approach in WS3: difficulty assigning risk scores and uncertainty about how to interpret them.

Inconsistent Risk Scoring (P3, P8). PRAM required analysts to independently evaluate each potential privacy problem across multiple organizational impact dimensions (e.g., compliance, business, reputation, culture) and assign 1-10 ratings to each factor. We observed significant variation in how participants rated similar privacy issues due to unclear guidance on how to weigh impact categories. For the same scenario, P3 gave a low score for surveillance-related autonomy loss (7), while P8 rated it much higher (26). A similar gap appeared in their dignity-related assessments (6 vs. 17).

As P3 noted, "*I wasn't sure how to weigh different kinds of organizational impacts against each other. Does a regulatory fine matter more than losing customers' trust? The framework doesn't tell you.*"

Unclear Meaning of Risk Scores (P5, P11, P12). Even after assigning scores, participants struggled to interpret their meaning. PRAM does not define thresholds for high, medium, or low privacy risks, leaving participants to set arbitrary cutoffs. In task #3, participants calculated significantly different risk scores for different data actions (ranging from 152 to 739), but struggled to determine which risks required action due to the lack of guidance on how to interpret these scores (P5, P11, P12). P5 expressed confusion about the resulting score, "*Without some kind of baseline, these numbers don't really tell me what that means to my scenario*". Without clear thresholds, participants had difficulty prioritizing risks confidently or consistently.

5 PrivacyAkinator Design Overview

This section presents the high-level design of PrivacyAkinator, including design goals, architecture and system scope.

Design Goals. Motivated by the challenges identified in our observational study, we derive the following design decisions to address the most pressing user needs (i.e., articulating system design).

- *Information overload*: Drawing on distributed cognition theory [42, 76], we design a structured representation as an external artifact [76] to support thinking and offload cognitive load. A good representation captures the key privacy design decisions while deliberately leaving out the rest, allowing developers to only focus on what matters.
- *Ineffective prioritization*: Novices often stick to the first thing they notice, so we explicitly prompt them to consider different privacy aspects and use a mix of explorative and exploitative questions to balance breadth and depth.
- *Knowledge blindspots*: We mine privacy-related news to build a knowledge base of key privacy design decisions.
- *Knowledge retrieval failure*: Our representation enables navigation across detailed privacy design decisions while maintaining a global view. We also use proactive prompts to help recall relevant privacy knowledge.
- *Insufficient guidance*: We transform the open-ended risk assessment process into a series of multiple-choice question answering to produce a more structured and guided workflow.
- *Vague terms*: We leverage LLMs to ground questions in the developer's context and provide concrete answer options.

System Architecture. A key design of PrivacyAkinator is to use multiple-choice question answering to articulate privacy-relevant design decisions. Instead of filling out open-ended forms, developers respond to a sequence of contextual, specific questions that help identify and document critical privacy choices. In doing so, PrivacyAkinator guides developers in thinking through the privacy design space by breaking it down into manageable, focused steps tailored to their system context.

PrivacyAkinator comprises three components: (1) a **universal privacy design representation** that enables systematic organization and documentation of privacy-relevant design decisions (Section 6), (2) a **domain-aware privacy design space** that uncovers privacy design choices across domain-specific contexts (Section 7), and (3) a **dynamic question generation workflow** that prioritizes key design decisions based on the developer's previous answers and system context (Section 8).

These components work together to generate questions that progressively explore the privacy design space. PrivacyAkinator draws from the design space to create contextually appropriate questions, prioritizing key design decisions most relevant to the developer's system. As developers respond to the questions, their answers incrementally enrich the underlying privacy representation.

PrivacyAkinator employs LLMs to extract information from news and user inputs, generate choices from embedded knowledge, and translate structured representations into natural language. These capabilities—extraction, recall, and generation—are among the most mature across LLMs [16]. We intentionally minimize the use of LLM reasoning. We further discuss how PrivacyAkinator mitigates the risks of relying on LLM-generated content in Section 11.

System Scope. PrivacyAkinator is a human-in-the-loop tool that helps developers articulate the key design decisions. Specifically, PrivacyAkinator assists with PRAM's *WS2: Assess System Design*, as our previous studies (Section 4) suggest that novice developers struggled most with this system design articulation process. Additionally, developers cannot effectively engage with later PRA stages (e.g., prioritizing risks) without clearly identifying and describing the specific design decisions first. PrivacyAkinator provides hints to help novice developers consider decisions they may have overlooked, but it does not guarantee full coverage.

While our studies also reveal broader challenges in later risk assessment phases, such as subjectivity and a lack of standards in risk evaluation, we consider these issues beyond the scope of our focus. Several approaches have been proposed to address these challenges, including quantitative risk models [21, 66] and formal methods for risk calculation [99], which are complementary to our contribution. We discuss future directions for extending support for risk assessment in Section 11.

6 Privacy Representation Design

We drew on the theory of distributed cognition [42, 76] and iteratively designed a representation for low-level privacy-related design decisions to manage information overload and support effective prioritization. In distributed cognition theory, representations serve as *external artifacts* that offload cognitive work into the environment, so people don't need to hold all the information in mind simultaneously. Norman similarly defines representations as the way information is structured or presented so that it can be perceived, understood, and used effectively by people, emphasizing that representations are inherently **selective** in capturing essential elements while deliberately omitting the rest [76].

Guided by these principles, we conceptualize privacy design around two fundamental components of data systems: data flows and stakeholders. We organize privacy decisions into building blocks that articulate how data moves through a system and how various stakeholders interact with these flows. This representation helps novices focus their attention on key privacy design choices instead of being overwhelmed by the full design space.

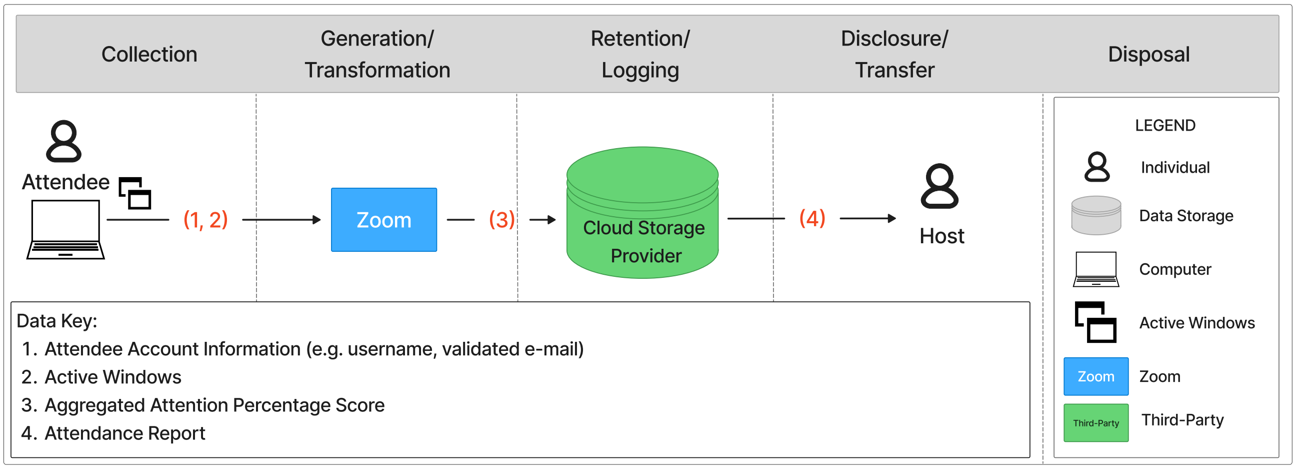


Figure 2: Data Flow Diagram defines data actions that occur across the data lifecycle, indicates the key entities involved in the system, and connects the data flows between them. However, the level of detail is insufficient to capture privacy nuances, and it omits interaction with different stakeholders, and may lead to deferred responsibility for privacy.

We postulate that an ideal representation for privacy design should satisfy the following properties.

- **Concise**: The representation should illustrate privacy-related design decisions concisely.
- **Expressive**: The representation should be expressive to encompass all critical aspects of privacy.
- **Simple**: The representation should be simple to minimize the learning curve and cognitive load.
- **Objective**: The representation should only present how stakeholders interact with the privacy design but not make any judgment. Evaluation should be independent of the representation.

6.1 Design Iterations

We used a bottom-up approach to guide the design of privacy representation. We collected 40 real-world data practices from news reports, privacy policies, and literature on privacy design [30, 36, 47, 83, 85, 105], and examined the privacy-relevant design decisions that emerge from these documents. We used similar selection criteria as in Section 4.1, but complexity was not considered. We then designed a representation to accommodate these data practices, iterated on the representations as we expanded the supported use cases, and collected early feedback from software developers through the authors' personal network.

Data Flow Diagram (DFD). We initially applied the data flow diagram from PRAM (*WS2: Supporting Data Map*), which visualizes how data moves through the system (Figure 2). PRAM defines a set of key information system operations, referred to as data actions, that occur across the data lifecycle: *collection*, *generation/transformation*, *retention/logging*, *disclosure/transfer*, and *disposal*. The diagram identifies the key entities involved in the system (e.g., user, third-party service) and maps the data flows between them. Each flow is annotated with the specific types of data being exchanged (e.g., names, addresses).

As we tested this representation on more use cases, we observed a few trade-offs:

+ This representation is intuitive and easy to understand.
– The level of detail is insufficient to capture privacy nuances in data practices (e.g., frequency of data collection).
– The representation design omits interactions with different stakeholders, such as showing privacy notices or providing user control mechanisms.
– Many system components fall outside developers' direct control, leading to deferred responsibility for privacy. For example, while external storage may seem out of reach, choosing third-party services versus self-hosting involves privacy implications that developers should consider.

Privacy Storyboard. We then made three changes to address the limitations of the data flow diagram: (1) shifting to a process-oriented rather than device-focused view to better align with developer workflows, (2) describing implementation of each data action to provide more fine-grained details, and (3) incorporating stakeholder roles to specify who is involved in or can control each part of the data flow. Figure 3 illustrates how participant screen activity is collected, processed, stored, and accessed in *Zoom Attendee Attention Tracking* scenario (Table 1), with each step outlining specific operations performed and the stakeholder roles (e.g., attendees giving consent to data collection, host accessing attention reports). We observed a few trade-offs in this iteration:

+ The representation design is developer-centric, closely reflecting how developers conceptualize and build systems.
– The representation presents too much information simultaneously, which may overwhelm readers and make it hard to maintain an overview of the entire privacy design.
– The representation often conflates multiple design decisions in a single description, making it difficult to extract specific privacy choices for analysis.

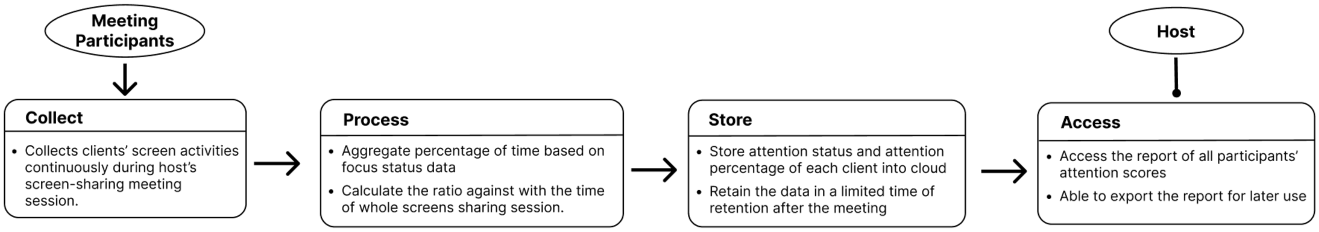
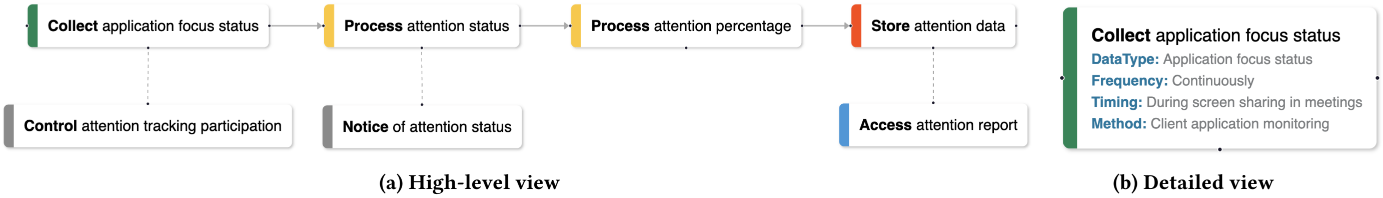

Meeting Participants
Collect
• Collects clients' screen activities continuously during host's screen-sharing meeting session.
Process
• Aggregate percentage of time based on focus status data
• Calculate the ratio against with the time of whole screens sharing session.
Store
• Store attention status and attention percentage of each client into cloud
• Retain the data in a limited time of retention after the meeting
Host
Access
• Access the report of all participants' attention scores
• Able to export the report for later use

Figure 3: Privacy Storyboard illustrates data actions across the data lifecycle with stakeholder roles. While more aligned with developer workflow, it can overwhelm users with too much information and often conflates multiple design decisions.


Collect application focus status
Process attention status
Process attention percentage
Store attention data
Control attention tracking participation
Notice of attention status
Access attention report
Collect application focus status
DataType: Application focus status
Frequency: Continuously
Timing: During screen sharing in meetings
Method: Client application monitoring

(a) High-level view

(b) Detailed view

Figure 4: Our final multi-layer graphical representation adopts a three-layer representation with data flow, stakeholder interaction, and individual design decision. The combination of both a high-level overview and a detailed view supports quick updates and maintaining awareness of the big picture.

– Merely mentioning stakeholders is too abstract; specific interaction types (e.g., consent or control) are critical for assessing privacy risks but remain inadequately defined.

6.2 Multi-layer Graphical Representation

As we iterated with more privacy representations, we explored the design space of the data practices. We found that stakeholders only have enumerable ways to interact with a data flow, and abstracted these into a common set of stakeholder interactions, summarized in Appendix Table 5.

We then developed a multi-layer graphical representation for privacy design decisions. Inspired by DENIM [63], we adopted a three-layer representation similar to website design [102]: data flow, stakeholder interactions, and individual design decisions (Figure 4). The data flow is organized as a sequence of data actions (e.g., collect, process, store) connected in a graph structure. Each data action and stakeholder interaction is modeled as an individual node, with stakeholder interactions linked to the relevant data actions. Specific design decisions are captured as node properties, such as frequency and timing of data collection, and collected data types.

The representation offers both a high-level overview and a detailed view. The overview includes only the top two layers (data actions and stakeholder interactions) but omits individual design choices. Users can zoom in on a specific node to see detailed design decisions as key-value pairs. This approach supports quick updates and helps maintain awareness of the big picture. It allows analysts to outline the overall privacy design before diving into specifics and refer back to the overview to maintain a clear sense of direction throughout their articulation process.

7 Mining Domain-aware Privacy Design Space

Inspired by Augur [28], we developed a data-driven approach to construct a domain-aware privacy design space by mining privacy-related news. In this section, we first describe the taxonomic structure of our privacy design space, then present our method for extracting privacy design decisions from privacy news.

7.1 Structuring Privacy Design Space

To understand the structure of the privacy design space, we examined the data practices collected in Section 6.1 using our privacy representation. We used an iterative, open-coding approach [100] to analyze the privacy design decisions. For each design decision, two authors independently annotated its key (the type of decision), the corresponding value (the specific choice made), and the associated data action or stakeholder interaction. We then collaboratively synthesized these openly generated annotations into a coding scheme by merging design decisions of the same type, and re-coded all design decisions using the finalized coding. This process yielded 96 unique privacy design decisions.

From this analysis, we make two key observations. First, we found that different data practices often involve common types of design decisions (e.g., `data_type`, `consent_mode`, `storage_location`). Second, while most decision types (or keys) are applicable across domains (e.g., health, smart homes), their specific values often depend on the context of data practices. For example, `data_type` may involve `medical_records` in a healthcare setting, but `GPS_coordinates` in a location-tracking app or `audio_commands` in a smart speaker. This demonstrates the need for a domain-aware design space that captures context-specific privacy nuances.

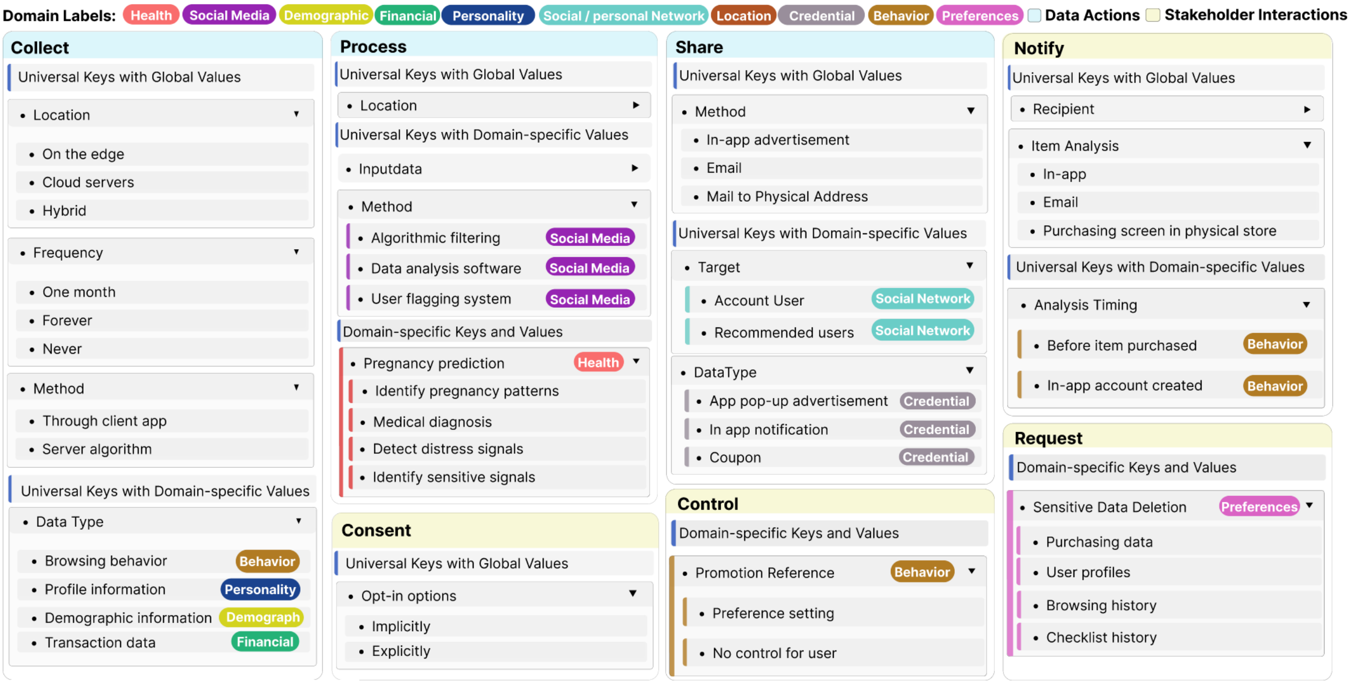


Figure 5: We organize our domain-aware privacy design space by data actions and stakeholder interactions, categorizing privacy design decisions based on the applicability of their keys and values across different domains.

Building on our privacy representation, we formulate the structure of the design space by organizing privacy design decisions into three categories (Figure 5): (1) universal keys with universal values that apply across domains (e.g., *consent_mode: opt-in/opt-out*); (2) universal keys with domain-specific values, where categories remain consistent but options vary by context (e.g., *data_type: medical records*); and (3) domain-specific keys for decisions unique to particular domains (e.g., *voice_masking: enabled*).

7.2 Mining Design Decisions from Privacy News

We first manually built an initial privacy design space of design decisions based on prior analysis results from Section 7.1 to establish the structure of the design space. We then extracted key privacy design decisions from the online corpus of privacy-related news using LLMs to populate and expand this design space. We hypothesize that important privacy design decisions are more likely to be reported in news and incident coverage because mistakes in these decisions tend to have severe consequences for individuals or organizations, triggering public scrutiny and media attention.

Dataset Preparation. We collected 10K news labeled with privacy category from popular technology news sources, including TechCrunch [95], Guardian [34], Wired [106], and The Verge [98]. We applied keyword-based filtering using terms like "privacy" and "data protection" to identify privacy-related news articles. To enable more precise filtering at scale, we sampled a small set of news articles and used the GPT-4o model to assess their relevance to privacy design. Using these LLM-generated labels, we trained a specialized classifier based on the RoBERTa model [65] as a cost-effective alternative to LLMs. We then applied this classifier to filter the entire corpus, which yielded 7,058 news articles related to privacy design.

Annotating Domain Labels. For each filtered news article, we segmented the news text into individual data practices to isolate distinct privacy scenarios. We then annotated each data practice with a set of domain labels derived from Privacy Contextual Domains in MITRE's PANOPTIC Taxonomy [48], which consists of 20 labels covering both data types and application domains (e.g., financial, health).

Design Space Expansion. Extracting design decisions from news text is challenging: open-ended prompts often lead LLMs to produce messy, inconsistent annotations, while being overly restrictive may limit their ability to uncover new design decisions. We developed an iterative two-step extraction process to balance flexibility and consistency. In the first step, we extracted values for design decisions covered in our current design space, allowing us to discover new domain-specific values for existing decision types. In the second step, we prompted the LLM to exclude the key-value pairs identified in the previous step and apply the 4W1H scheme (What/When/Who/Where/How) [2, 53] to systematically uncover new decision keys.

After each iteration, we enriched the design space by expanding value sets or adding new keys. For universal keys with domain-specific values, we incorporated new options uncovered across different domains. Additionally, we introduced new domain-specific keys and their associated values. We repeated the process until the design space reached saturation, with diminishing gains in new keys or values.

This iterative mining process not only extracts a broad set of privacy design decisions and their possible values, but also reveals how these decisions commonly appear together in real-world contexts. These co-occurrence patterns enable us to rank and prioritize design decisions during question generation (Section 8.2).

8 Question Generation Workflow

PrivacyAkinator leverages the underlying privacy representation to dynamically generate questions from the design space and prioritize questions based on developer's prior responses and domain context.

8.1 Generating Question Candidates

PrivacyAkinator generates two types of questions to systematically explore the privacy design space: (1) *exploratory* questions that add new nodes (data action or stakeholder interaction) to the data flow; (2) *exploitative* questions that draw on the existing set of design decisions to elicit more detailed information about specific nodes. As developers respond, PrivacyAkinator dynamically updates the underlying representation to reflect their answers.

Exploratory Questions. We prompt the LLM to propose new data actions (e.g., whether the data is stored, shared with external parties) or new stakeholder interactions for existing data action nodes (e.g., whether data collection obtains user consent, requesting deletion of stored data). We intentionally limit exploratory questions to a binary yes/no format to reduce cognitive load on developers. A "yes" response triggers the addition of a new node to the representation, while "no" responses do not make any changes. During question generation, LLM also annotates the related node in the representation and identifies potential new connections to enable automatic updates to the underlying representation.

Exploitative Questions. PrivacyAkinator maintains a pool of candidate design decisions. When adding a new node to the underlying representation, PrivacyAkinator retrieves relevant design decisions from our curated design space and enriches the pool. To ensure contextual relevance, PrivacyAkinator annotates domain labels for the current design practice based on user input, and filters out irrelevant data practices mined in Section 7.2 based on domain label similarity. We use the Jaccard similarity of two domain label sets to quantify the relevance of data practices:

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}$$

Only data practices with a similarity score above 0.4 are retained. From this filtered set, PrivacyAkinator selects design decisions that match the type of the current node and incorporates them into the decision pool.

Once an exploitative question is presented and answered by the user, PrivacyAkinator prompts the LLM to generate follow-up questions to probe corner cases and contextual specifics. For example, if the system asks, "*Do you limit API usage for third-party integrations?*", a follow-up question might be, "*What is the maximum number of API calls allowed per third party per day?*" In doing so, PrivacyAkinator progresses from broad privacy concepts to specific implementation details through exploration, exploitation, and follow-up questions with varying levels of granularity.

8.2 Prioritizing Key Questions

PrivacyAkinator prioritizes which questions to ask based on co-occurrence statistics and contextual relevance to efficiently explore the privacy design space.

Ranking Design Decisions by Co-Occurrence. We rank design decisions using their co-occurrence frequency in the filtered corpus of relevant data practices described in Section 8.1. The mutual information (MI) score [72] between two design decisions D_1 and D_2 is calculated as:

$$MI(D_1, D_2) = \log\left(\frac{P(D_1 \wedge D_2) \times N}{P(D_1) \times P(D_2)}\right)$$

where $P(D_1 \wedge D_2)$ is the probability of the two decisions co-occurring, $P(D_1)$ and $P(D_2)$ are their individual probabilities, and N is the total number of data practices involved.

PrivacyAkinator then uses the average co-occurrence statistics with all prior design choices to rank all design decisions in the pool. The hypothesis is that design decisions that frequently co-occur in privacy news or literature are more likely to be correlated and critical. At each step, PrivacyAkinator generates questions for the top-k decisions from the ranked list, as expanding too many decisions would significantly increase computational cost. For each prioritized decision, we provide the LLM with a key (the design decision) and its value set (possible choices), and prompt it to contextualize the question based on the developer's previous answers and the current underlying representation. To prevent biased suggestions, we prompt the LLM to present all possible design choices objectively and intentionally avoid framing that could prime developers. We used $k = 3$ in our setting.

Explore vs. Exploit. PrivacyAkinator maintains separate lists for exploratory and exploitative questions, and uses simple heuristics to switch between them. The choice depends on factors including how many questions have been answered, how the user has responded to earlier questions, and the current structure of the underlying representation. As more questions are answered, the system gradually increases the probability of selecting exploitative questions to refine existing components. If a user skips an exploratory question, it interprets this as a signal that the topic may be irrelevant or already addressed, and switches to exploitative questions. The system also checks for missing node types (e.g., notice, share), and asks exploratory questions to fill those gaps. In addition to these automated heuristics, users can manually switch between these two modes at any time.

To avoid asking repetitive questions, PrivacyAkinator prompts the LLM to check each new question against the question history and the current system representation. If a duplicate is detected, PrivacyAkinator simply fetches the next candidate question.

Termination of Question Generation. Inspired by the Akinator game [84], PrivacyAkinator provides three ways to end question generation. First, it imposes a hard limit of 25 questions to avoid overwhelming users. Second, it applies heuristics to stop early when enough information has been gathered or further questions seem unproductive (i.e., no new nodes are likely to be added and all potential design decisions have low co-occurrence statistics). Third, the user can end the session at any time if they feel their design is complete or no longer want to continue. When termination is triggered, PrivacyAkinator prompts the user to either proceed with more questions to refine details or start the assessment process.

9 Implementation

We implemented a prototype of PrivacyAkinator, comprising a user interface and a backend LLM service.

User Interface. We implemented a user interface for developers using React to streamline the process of conducting PRAM. Developers first provide a high-level design goal that describes the functional needs of their system (Figure 6). PrivacyAkinator then expands this into more detailed functional requirements, surfacing the privacy design decisions developers need to consider (Figure 7). Developers can review and modify these as needed. Here, the LLM is used to outline key data flows for initializing the privacy representation, rather than generate full system requirements.

The next stage involves a question-answering process (Figure 8). The interface displays the underlying system representation at the bottom of the screen, while the top section presents questions along with relevant contextual information. The system highlights the node associated with the current question in the representation and provides a detailed view of that node next to the question to aid understanding. Users can choose to hide the visual representation and focus solely on the question view. Users may select multiple answers they believe are correct. If none apply, they can provide a custom response or skip the question. They can also return to earlier questions and revise their answers at any point. As users progress, the interface also provides real-time updates of the underlying representation to help users understand how their responses influence the system model.

After answering several questions, users are prompted to either continue exploring more detailed questions or proceed to system assessment. In the assessment phase, the interface uses the LLM to generate a table summarizing data actions, types of data involved, and contextual factors (Figure 9), which users can edit directly. The LLM also suggests summary issues for each data action, which users can validate, discard, or supplement with their own insights.

Once this process is complete, users can export the results into a PRAM worksheet (Figure 10). We generate the worksheet using the ExcelJS library to ensure it follows the official PRAM format.

Backend LLM Service. The backend service handles requests from the user interface to fetch the next question. It generates questions asynchronously using the `claude-3-7-sonnet-20250219` model to reduce response latency. We selected this model for its strong instruction-following capabilities, stable output quality and cost-effectiveness. For each user session, the service maintains a separate question pool and selects the next question dynamically, following the process described in Section 8. To ensure question quality, we prompted the LLM to (1) ground each question in the underlying representation to ensure privacy relevance, (2) present design choices objectively to avoid judgment or biased framing (e.g., using "would" instead of "should"), (3) generate questions and options that are specific, concrete, and contextualized (e.g., replacing abstract terms with clear examples) while remaining concise, and (4) reference the current representation and question history to avoid duplicates. We set `temperature` = 0, `top_p` = 0.95 to minimize LLM output variability and maintain consistent question generation.

10 Evaluation

We conducted an experiment to evaluate our system's effectiveness in covering key privacy design decisions and a user study to assess the usability of our tool for developers.

10.1 Case Studies

We evaluated PrivacyAkinator's coverage of key privacy design decisions on real-world data practices.

10.1.1 Dataset. We curated a dataset of 30 data practices with problematic privacy design decisions from the literature on privacy incidents [62, 83] and news reports of practices that triggered significant user backlash [15, 18]. We provide detailed descriptions of the dataset in Table 6.

10.1.2 Ground Truth. We employed a multi-stage approach [110] to identify key design decisions underlying these data practices, which combined individual ideation, peer review and expert evaluation. We first conducted a structured brainstorming session [20] to enumerate privacy design decisions associated with these data practices. The session involved 11 students who had extensive research training in usable privacy, including advanced coursework, active research projects, and prior publications in privacy-related venues. We provided each participant with a general description of the practices and asked them to individually propose design decisions they felt significantly affected user privacy. They then worked in pairs to discuss their responses and identify additional design decisions.

The identification phase was followed by a ranking and selection stage conducted by more experienced researchers to identify the key design decisions. After the session, two graduate researchers with over three years of experience in usable privacy research independently coded the design decisions for each data practice based on participant input. We assessed inter-rater reliability using Cohen's kappa [69], which indicated strong agreement with $\kappa = 0.85$. The coders then discussed any discrepancies to resolve conflicts and reach a consensus on the codebook. After coding, they ranked the decisions by their frequency of occurrence across participants and perceived impact on user privacy, then selected the top-ranked decisions to serve as the ground truth.

10.1.3 Method. We manually defined a design goal for each data practice based on its functionality (e.g., "an attendee tracking feature for a video conferencing app" for Zoom's attention tracking), which served as the input to our system. We then answered the questions generated by PrivacyAkinator to align the responses with real-world implementations. For each data practice, we measured whether the resulting privacy representation could (1) capture design decisions by including relevant keys, and (2) accurately reflect actual choices. We used the GPT-4o model to compare the generated outputs against our ground truth, and supplemented this with human review and corrections to ensure the accuracy of evaluation. We also tested whether the effectiveness in capturing privacy-related design decisions primarily stems from our question-generation workflow or LLMs' inherent capabilities. For baselines, we directly prompted different LLMs to generate 20 questions. We set the following configurations for all models: `temperature` = 0, `top_p` = 0.95, and `max_tokens` = 4000. We tested each model three times to account for randomness and averaged the results.

10.1.4 Results. Table 2 compares the coverage of key privacy design decisions of PrivacyAkinator with three state-of-the-art LLMs, including GPT-4o, Claude 3.7, and Gemini 2.5 Flash. On average, PrivacyAkinator achieved a coverage rate of 93.67% for key decisions and 77.33% for actual choices, significantly outperforming the baseline LLM approaches, whose coverage ranged from 43.91% to 55.22% for key decisions and 20.83% to 42.71% for actual choices.

We further evaluated their performance across 10 categories of privacy decisions (4 data actions & 6 stakeholder interactions). We found that PrivacyAkinator consistently showed higher coverage on key design decisions and their actual choices across all categories. The performance advantages were more significant in data action categories (i.e., Collect, Process, Store, and Share), where PrivacyAkinator achieved coverage rates above 89% for key decisions, while baseline LLMs typically ranged between 38–65%.

Stakeholder interaction categories revealed more nuanced performance differences. While LLMs performed relatively well in the Consent category (75.72% for GPT-4o, 78.44% for Gemini 2.5), PrivacyAkinator still led with coverage of 96.15% for key decisions and 88.46% for actual choices. The most significant gaps appeared in Control and Access categories, where baseline LLMs struggled to identify relevant design decisions (40.65–62.16% coverage), while PrivacyAkinator maintained over 92% coverage.

These results suggest that PrivacyAkinator is more effective at systematically exploring the privacy design space. While general-purpose LLMs can generate contextual privacy-related questions, their outputs are often inconsistent and less effective due to limited domain-specific knowledge.

10.2 User Studies

We conducted a user study to evaluate how PrivacyAkinator could help developers articulate privacy design decisions.

10.2.1 Participants. We recruited 24 participants (16 identified as male and 8 as female; 18 aged 18–24, 6 aged 25–34) in the evaluation through the same approach as previous studies (Section 4.1). Each participant received a USD 20 Amazon gift card as compensation. Participants reported low levels of prior privacy knowledge (Mean = 2.4, SD = 1.3), with most rating themselves as beginners (ratings of 1–2). None of the participants had taken part in our earlier studies. The study received approval from our university's Institutional Review Board (IRB).

10.2.2 Study Procedure and Apparatus. The study used a within-subjects design, where participants used both PrivacyAkinator and NIST's PRAM to complete two tasks. Because this study focuses on novices, we anticipated that the individual differences in the between-subjects design could strongly influence outcomes. We randomized the presentation order of tools and tasks, and included warm-up tasks for each tool to mitigate the potential priming effect. We did not disclose that one of the tools was developed by the authors to prevent introducing potential bias. For PRAM, we only used *Worksheet 2 (Assessing System Design)* to ensure fair comparison, as it aligns with the focus of PrivacyAkinator on articulating system design decisions. We reused the task scenarios from previous studies (Table 1).

Each study included two sessions, one for each condition. Each session started with a 10-minute walk-through briefing on the study purpose and task overview, followed by a warm-up task that included a tutorial on the tool and the task using a sample scenario *ACME IDP service* drawn from the PRAM materials. Researchers observed the process and answered questions as needed.

Then we instructed each participant to complete one task using their assigned tool (PRAM or PrivacyAkinator) to assess the privacy design of one data practice scenario. We presented a high-level description of each scenario (Table 1) to remain consistent with previous studies (Section 4).

Upon completion of each task, we asked the participant to fill out a NASA-TLX survey [40]. After both sessions, participants took a semi-structured interview about their experiences with the two tools. In the post-study interview, we asked participants to compare their experiences with the two approaches and reflect on the strengths and weaknesses of each. The study was conducted in a lab space or via an online Zoom meeting. On average, our study took about 75 minutes for each participant.

10.2.3 Data Analysis. We measured the coverage of key privacy design decisions identified in participants' responses. Two researchers collaboratively annotated the design decisions described in participants' worksheets (written manually or generated by PrivacyAkinator), counted the **total number** of decisions included in each worksheet, and calculated the **coverage** of each design decision by comparing them to the ground truth for each scenario (Section 10.1). During annotation, we excluded content that did not represent design decisions, such as user privacy expectations (e.g., "*this data is highly sensitive for users*"), and removed duplicate entries. We performed Mann-Whitney U tests [70] to compare different scenario orders within each condition and found no evidence of ordering bias across all reported quantitative metrics. The qualitative analysis followed the same procedure as described in Section 4.1.

10.2.4 Quantitative Results. We found that PrivacyAkinator could help developers articulate system design more effectively and improve the quality of their design outputs.

Improved Efficiency of Articulating System Design. As shown in Table 3, participants completed the task 73% faster (average of 9.7 vs. 35.8 minutes) and covered 47% more privacy-related design decisions (average of 28.9 vs. 12.9) using PrivacyAkinator compared to the PRAM worksheet.

Additionally, participants reported more positive responses to the NASA TLX questions (Table 4) with PrivacyAkinator than the PRAM worksheet. These improvements include lower mental workload, effort, and frustration, along with increased self-perceived performance, which are all statistically significant under the Wilcoxon Signed Rank test [107] ($p < 0.01, r > 0.8$).

Broader Coverage of Key Privacy Design Decisions. We then compared the coverage of design decisions between conditions

Table 2: PrivacyAkinator achieved broader coverage of key privacy design decisions and their corresponding values than three baseline models (i.e., GPT-4o, Claude 3.7, and Gemini 2.5 Flash) across all categories.

Coverage	GPT-4o		Claude-3.7		Gemini-2.5 Flash		PrivacyAkinator	
	Decision	Choice	Decision	Choice	Decision	Choice	Decision	Choice
Collect	59.74 %	29.47 %	60.90 %	30.77 %	65.38 %	55.13 %	**98.08 %**	**80.77 %**
Process	38.38 %	18.84 %	55.10 %	28.57 %	54.42 %	40.14 %	**89.80 %**	**75.51 %**
Store	55.64 %	35.16 %	55.26 %	30.70 %	75.27 %	56.76 %	**100.00 %**	**92.11 %**
Share	43.54 %	18.62 %	42.67 %	22.67 %	48.50 %	36.44 %	**92.00 %**	**80.00 %**
Consent	75.72 %	44.06 %	63.28 %	41.13 %	78.44 %	62.28 %	**96.15 %**	**88.46 %**
Notice	54.01 %	26.41 %	68.69 %	32.28 %	60.00 %	44.78 %	**89.47 %**	**63.16 %**
Control	40.65 %	15.34 %	62.16 %	31.96 %	59.74 %	54.62 %	**92.00 %**	**64.00 %**
Access	45.68 %	27.16 %	53.53 %	22.13 %	60.49 %	48.15 %	**92.59 %**	**81.48 %**
Request	29.63 %	7.41 %	40.74 %	22.22 %	40.74 %	25.93 %	**88.89 %**	**66.67 %**
Audit	40.00 %	6.67 %	94.44 %	48.41 %	53.33 %	40.00 %	**100.00** %	**60.00 %**
Overall	43.91 %	20.83 %	54.25 %	28.26 %	55.22 %	42.71 %	**93.67 %**	**77.33 %**

Table 3: In our user study, participants spent less time and covered more key design decisions with PrivacyAkinator compared to PRAM Worksheet. Format: mean ± standard deviation. We also highlight the higher value between the two conditions.

Task ID	PRAM Worksheet			PrivacyAkinator		
	Time (min)	# Total Decisions	Key Decision Coverage	Time (min)	# Total Decisions	Key Decision Coverage
Task #1	36.4 ± 10.9	12.4 ± 3.5	36.9 ± 13.8 %	**14.2** ± 9.6	33.5 ± 5.9	**79.1** ± 3.8 %
Task #2	32.6 ± 11.1	11.2 ± 3.6	40.0 ± 16.9 %	**7.0** ± 1.9	29.8 ± 6.2	**93.3** ± 6.2 %
Task #3	31.5 ± 5.2	11.4 ± 3.2	38.2 ± 7.6 %	**7.5** ± 1.8	26.3 ± 4.3	**93.9** ± 4.1 %
Task #4	41.8 ± 9.3	13.3 ± 6.0	48.6 ± 9.8 %	**11.2** ± 2.8	24.8 ± 1.7	**87.5** ± 4.8 %
Average	35.8 ± 10.7	12.9 ± 4.3	42.6 ± 12.0 %	**9.7** ± 6.4	28.9 ± 5.5	**89.3** ± 7.0 %

(Table 3). On average, participants covered 89.3% of key privacy design decisions when using PrivacyAkinator, compared to 42.6% of the PRAM worksheet. This difference was statistically significant under the Wilcoxon Signed Rank test [107] ($p < 0.01$). Additionally, we observed that PrivacyAkinator produced more consistent results across participants, with lower variance in coverage (7.0%) than 12.0% for the PRAM worksheet.

10.2.5 *Qualitative findings.*

We made the following findings based on participants' task behaviors and post-study interview feedback.

Lower Barriers for Privacy Non-Experts. Participants appreciated PrivacyAkinator's guided question answering process, which enabled them to describe privacy-related decisions without specialized privacy background. For example, P16 mentioned that "*[PrivacyAkinator] is pretty intuitive and easy to use. It doesn't require any extra privacy knowledge.*" P18 added that "*Personally, I haven't seen that workflow before, and I haven't systematically studied privacy. But I could still get where each design was coming from and the kind of details they were trying to address. The questions made a lot of sense to me, and they also gave me a few solid options to think through.*"

Ease of Cognitive Load. Participants expressed that PrivacyAkinator streamlined the system design articulation process, reducing their perceived effort and workload. Many participants described PrivacyAkinator as "*much easier to use*", with P12 noting that "*the only mental effort was answering the questions which mainly had just 2 or 3 options*", rather than "*having to come up with a lot of ideas and also decide on which idea to focus on.*" P4 mentioned that "*[PrivacyAkinator] was much more efficient for my thought process, and it allowed me to generate more ideas and also not focus about manually drawing the diagram.*" Participants also recognized that PrivacyAkinator "*made the specification process easier and more manageable*" by "*narrowing the mental search space*" (P18), and offering a simplified workflow that avoided "*doing all things manually*" (P4) and "*switching back and forth between different tabs*" (P15).

Enhanced Coverage of Privacy Design Decisions. Participants appreciated that PrivacyAkinator helped surface design considerations they might have otherwise overlooked. P13 shared that "*questions can help me to find those details I missed, or even just didn't know*", while P1 noted that "*the set of questions presented gave a lot of insights into designing and addressing data privacy issues in the real world projects or applications.*" Participants also perceived better performance with PrivacyAkinator than designing from scratch. P11 explained that without guidance, they were "*just imagining things from scratch,*" and felt "*limited in what I was able to come up with.*" In contrast, PrivacyAkinator "*gives something to build on and [they] just needed to make some reasonable adjustments*", resulting in a design that they felt was more accurate and complete.

Reduced Ambiguity through Structured Guidance. Participants expressed that PrivacyAkinator reduced their confusion by providing clear, step-by-step guidance throughout the process. Most participants struggled with traditional worksheets. They explained

Table 4: Participants perceived significantly lower cognitive load using PrivacyAkinator compared to PRAM worksheet. We present NASA TLX results (scale of 1 to 7) as "median (mean ± standard deviation)". We annotate statistical significance based on the Wilcoxon Signed-Rank test (*: $p < 0.05$,** : $p < 0.01$,*** : $p < 0.001$). ↓ denotes that the lower is better.

Condition	Mental ↓	Physical ↓	Temporal ↓	Performance ↑	Effort ↓	Frustration ↓
PRAM	6 (5.4 ± 1.4)	2 (2.5 ± 1.8)	3 (3.6 ± 1.6)	5 (4.3 ± 1.1)	5 (5.3 ± 1.0)	3 (3.4 ± 2.0)
PrivacyAkinator	**3 (3.3 ± 1.3)** ***	**1 (1.7 ± 1.1)** *	**2 (2.1 ± 0.9)** ***	**6 (5.7 ± 0.6)** ***	**3 (3.1 ± 1.5)** ***	**1 (1.7 ± 1.0)** **

that "*the worksheet definitions and examples are vague and sometimes disagreeing*" (P7), and initially felt "*kind of lost*" and "*having no idea what you're gonna do*" (P19). In contrast, Participants appreciated how PrivacyAkinator used contextualized prompts, a visual workflow diagram, and follow-up questions to make the process "*easier to understand*". P1 appreciated that "*[PrivacyAkinator] infers based on the questions answered, which serves as a great way of summarizing and visualizing details while designing a system.*" P6 also noted that "*questions generated by PrivacyAkinator were easier to understand,*" because they avoided heavy use of technical jargon.

Over-Reliance on Generated Design Decisions. Participants also noted the potential downside of over-relying on system-generated questions. P11 observed that when answering the generated questions, they are "*kind of shut off the thinking part of your brain, and then don't explore beyond it.*" While participants recognized that PrivacyAkinator is especially useful for beginners with limited knowledge, they suggested a hybrid approach for experienced developers. As P5 explained, "*If you're more experienced, the tool is still good, because it'll give you different options. But you should still take time to sit down and come up with your own ideas, instead of just relying on what it's generating.*"

11 Discussion & Limitations

New Design Knowledge. Our studies suggest several design implications for developing tools that support privacy assessment. Existing privacy assessment frameworks often prompt developers to analyze risks before they have fully articulated the underlying data practices [75, 112]. Our studies show that this leads to incomplete or inconsistent system descriptions. Without structured articulation, novices rely on ad-hoc reasoning and tend to fixate on the first thing they notice, resulting in partial and biased assessments. Our findings suggest that privacy assessment tools should defer judgment and **separate articulation from evaluation** [*Design Implication* #1]. Articulating data practices before assessing risks supports a more complete understanding of the system and reduces premature evaluation.

RQ1: How to turn privacy design into a structured task? We address this through a structured representation that organizes privacy design decisions around data flows and stakeholder interactions. In our studies, novices were often distracted by surface-level implementation details, echoing prior findings that novices tend to focus on superficial details whereas experts attend to deeper structural cues [11, 14, 73]. Our representation helps bridge this gap through **selective attention** [*Design Implication* #2]: by determining what to include and what is omitted, it not only documents the system design, but also actively directs developers' attention and guides their thinking [76].

RQ2: How to unfold this structured task for novice developers? PrivacyAkinator **decomposes** PRAM's broad, open-ended prompts into **a series of multiple-choice questions** [*Design Implication* #3] to make privacy assessment more approachable and actionable for novices. In our studies, participants often did not know how to begin with open-ended worksheets, felt confused about what level of detail was expected and frequently omitted key privacy decisions. With open-ended worksheet, novices need to consider numerous decisions simultaneously, which imposes high cognitive load and can be overwhelming. In contrast, our evaluation shows that PrivacyAkinator's structured prompts improved coverage of privacy-relevant decisions and reduced cognitive load, consistent with prior work on scaffolding novices [33, 67, 91, 97]. By breaking the task into discrete, concrete questions, the tool allows novices to focus on one decision at a time.

A key design trade-off is reducing cognitive load without oversimplifying the task or limiting expressiveness. To address this, we decouple *what* to articulate (RQ1) from *how* to support articulation (RQ2): The structured representation directs attention to privacy-relevant aspects of a system, while the interactive questions guide novices through the design space. Our design implications (e.g., separating articulation from evaluation, guiding selective attention, and decomposing complex tasks) can generalize beyond PrivacyAkinator and inform the design of related systems. Designers can operationalize these principles in domain-specific contexts; for example, we use closed-ended questions to scaffold privacy risk assessment tasks.

Practical Deployment Limitations. While PrivacyAkinator helps developers articulate their design choices, privacy risk assessment involves additional challenges beyond the tool's current scope. Improved coverage and articulation may not guarantee correct privacy decisions. Judging the severity of privacy risks remains subjective and often requires domain-specific expertise that novices may lack. Nevertheless, by constructing a closed-ended, structured design space, our work lays the groundwork for creating a representation of privacy design that is easier to analyze with automation. Future work could build on this representation to produce machine-readable outputs that support risk identification and prioritization—for example, by training ML models to classify risk levels, flag high-risk combinations of design choices, or map design patterns to known risks—thereby enabling more objective and consistent privacy assessments.

One related concern is the potential over-reliance on automated suggestions. PrivacyAkinator is not intended to replace existing PRA workflows but to serve as a structured prompt or checklist to help developers surface key design decisions. Developers should combine the automated question generation with manual review and domain expertise, as the LLM-based approach may miss context-specific privacy considerations.

Potential Bias of Mining Privacy News. As with any data-driven application, the coverage of PrivacyAkinator is influenced by its data sources. Since we construct the privacy design space by mining privacy-related news articles, it effectively captures high-profile, media-reported incidents but may overlook under-reported or emerging domains, such as healthcare IoT or industrial surveillance. For example, it may over-represent issues that were widely reported rather than those that are actually more common in practice. Additionally, the news corpus may reflect geographic bias, as the sources are drawn primarily from media outlets in the United States and United Kingdom, which could underrepresent privacy issues more prominent in other regions or regulatory environments.

Additional sources (e.g., legal documents) can be incorporated to broaden the system's coverage and relevance to sensitive contexts. Organizations can also apply our approach to create customized taxonomy using internal or domain-specific materials, such as engineering requirement documents.

Risks of Relying on LLM-generated Content. PrivacyAkinator relies on LLMs to generate questions, which can sometimes result in vague or inconsistent wording. During our user study (Section 10.2), several participants noted that some questions were repetitive or unclear. To mitigate this, PrivacyAkinator allows users to skip ambiguous or duplicate questions and submit custom responses when the provided options do not apply. Repetitive questions may cause inconsistencies. Future work could analyze the representation and provide automatic or human-assisted conflict resolutions. For vague or unclear questions, a conversational assistant could be developed to provide explanations, or regenerate questions for clarity. With the recent rapid advancements in LLMs [16, 113], we expect future models to reduce these issues and further improve the clarity and quality of generated questions.

Unconstrained question generation carries a high risk of LLM hallucinations, potentially producing nonsensical or irrelevant questions that do not connect to the developer's system. Instead of allowing free-form brainstorming, we ground the questions on the closed-ended design space and provide contextual cues (e.g., highlighting related nodes, summarizing prior decisions) to help developers flag irrelevant questions.

Another concern is that LLMs' output is not necessarily deterministic and may vary across runs. In our user studies, however, we observed high question stability: repeated invocations for same system descriptions produced largely consistent question sets, with variation mostly confined to minor wording changes. We attribute this stability to low-temperature sampling and the system's prioritization logic, which limits randomness in question selection. Future work may consider alternative question-generation approaches (e.g., template-based methods), which offer higher stability but may lack the flexibility to capture diverse, context-specific privacy nuances. For example, it would be challenging for templates to surface questions tailored to a developer's specific data uses.

Sharing system descriptions with an external LLM may expose proprietary system logic and sensitive data-processing details. Developers can address these privacy risks by establishing contractual agreements with commercial LLM providers (e.g., zero data retention policies [35]) or by deploying self-hosted models.

LLM interpretations of system descriptions may be inaccurate, leading to misrepresented features, data uses, or processing steps. To help detect these misinterpretations, PrivacyAkinator immediately visualizes how each input updates the underlying representation, making the LLM's interpretation explicit and easier to review.

PrivacyAkinator uses an LLM to generate preliminary system requirements from a short description. Recent work in requirements engineering shows that modern LLMs (e.g., GPT-4) can reliably interpret natural-language descriptions and produce structured, coherent software requirements specifications comparable to those written by entry-level engineers [27, 45, 55]. For example, in the Zoom-attention scenario, the model recognizes that application-window focus and webcam-based inference are two distinct methods for tracking attention. Its output uses window focus as it is less privacy-invasive, given that much sensitive information beyond attention status can be inferred from webcam data (e.g., facial expressions, body movements). Alternative methods for requirement elicitation, such as a drag-and-drop interface, offer more direct control over feature and data selection but may lack generalizability. For example, a predefined list of data types may struggle to capture the breadth of real-world design choices. Such approaches are better suited when customized for a specific organization or domain, where relevant data types are fixed and well-defined.

Applicability beyond Novices. We intentionally focused our studies on developers who are novices in privacy. Prior work shows that experts and novices often exhibit markedly different usage patterns when engaging in complex design and analysis tasks [1, 26], so we expect privacy experts to have different user needs that require different forms of support. Focusing on a specific user group allows us to probe their specific challenges in depth and provide tailored scaffolding, rather than a one-size-fits-all system that serves neither group well. We target novices in particular because they constitute a large portion of the overall developer population, and their lack of privacy expertise necessitates additional support for effective privacy assessments.

While PrivacyAkinator is specifically designed for novices, we believe several of its components can also benefit privacy experts. For example, the structured privacy representation can help them keep track of numerous privacy-relevant decisions and more easily navigate the design space. The prompting questions can function as a lightweight checklist to surface missing decisions or reveal edge cases, or as prompts to support brainstorming of alternative design choices. However, experts may demand different workflows: they often do not need step-by-step guidance and instead may prefer to manually assemble the diagrams from the building blocks (i.e., data actions and interactions), similar to privacy storyboards [47]. They can then actively drive question generation, instructing the system to generate questions when needed to enrich the representation more deliberately. Because our studies primarily involved novices, our results may not fully reflect how experts would use or benefit from the system. Future work will test the tool with developers of varied expertise levels to incorporate expert perspectives for a more comprehensive assessment.

12 Conclusion

This paper explores methods to lower the barrier for novice developers to apply privacy risk assessments, particularly NIST's PRAM. Through an observational study (N=12), we identified three key challenges novice developers face when applying PRAM: (1) difficulty effectively distributing attention across numerous privacy design decisions, (2) confusion stemming from PRAM's vague terminology and open-ended structure regarding appropriate responses, and (3) frequently missing important design decisions without explicit prompting. To address these challenges, we developed PrivacyAkinator, a tool that helps developers articulate key privacy decisions by answering LLM-generated multiple-choice questions. Our user study (N=24) shows that PrivacyAkinator enables developers to identify 47% more key design decisions in 73% less time compared to using the PRAM worksheet.

References

[1] Saeema Ahmed, Ken M Wallace, and Lucienne T Blessing. 2003. Understanding the differences between how novice and experienced designers approach design tasks. *Research in engineering design* 14, 1 (2003), 1–11. doi:10.1007/s00163-002-0023-z

[2] Abdulrahman Alhazmi and Nalin Asanka Gamagedara Arachchilage. 2021. I'm all ears! Listening to software developers on putting GDPR principles into software development practice. *Personal and Ubiquitous Computing* 25, 5 (2021), 879–892. doi:10.1007/s00779-021-01544-1

[3] Nada Alhirabi, Stephanie Beaumont, Jose Tomas Llanos, Dulani Meedeniya, Omer Rana, and Charith Perera. 2023. PARROT: Interactive Privacy-Aware Internet of Things Application Design Tool. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 1 (March 2023), 37 pages. doi:10.1145/3580880

[4] Majed Alshammari and Andrew Simpson. 2018. Towards an effective privacy impact and risk assessment methodology: risk assessment. In *International Conference on Trust and Privacy in Digital Business*. Springer, 85–99. doi:10.1007/978-3-319-98385-1_7

[5] Jenna Amatulli. 2020. Zoom can track who's not paying attention in your video call. here's how. https://www.huffpost.com/entry/zoom-tracks-not-paying-attention-video-call_l_5e7b96b5c5b6b7d80959ea96. Accessed on 05/10/2025.

[6] Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A Smith. 2012. Automatic categorization of privacy policies: A pilot study. *School of Computer Science, Language Technology Institute, Technical Report CMU-LTI-12-019* (2012).

[7] Oshrat Ayalon, Eran Toch, Irit Hadar, and Michael Birnhack. 2017. How Developers Make Design Decisions about Users' Privacy: The Place of Professional Communities and Organizational Climate. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 135–138. doi:10.1145/3022198.3026326

[8] Ero Balsa. 2023. Technocracy, pseudoscience and performative compliance: the risks of privacy risk assessments. Lessons from NIST's Privacy Risk Assessment Methodology. arXiv:2310.05936 [cs.CR] https://arxiv.org/abs/2310.05936

[9] Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, Florian Schaub, and Norman Sadeh. 2020. Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1943–1954. doi:10.1145/3366423.3380262

[10] Mitra Bokaei Hosseini, Rocky Slavin, Travis Breaux, Xiaoyin Wang, and Jianwei Niu. 2020. Disambiguating requirements through syntax-driven semantic analysis of information types. In *Requirements Engineering: Foundation for Software Quality: 26th International Working Conference, REFSQ 2020, Pisa, Italy, March 24–27, 2020, Proceedings 26*. Springer, 97–115. doi:10.1007/978-3-030-44429-7_7

[11] John D Bransford, Ann L Brown, and Rodney R Cocking. 2000. How experts differ from novices. *How people learn: Brain, mind, experience, and school* (2000), 31–50. doi:10.17226/9853

[12] Sean Brooks, Michael Garcia, Naomi Lefkovitz, Suzanne Lightman, and Ellen Nadeau. 2017. An Introduction to Privacy Engineering and Risk Management in Federal Information Systems. doi:10.6028/NIST.IR.8062

[13] Lucas Brutschy, Pietro Ferrara, and Peter Müller. 2014. Static analysis for independent app developers. *SIGPLAN Not.* 49, 10 (Oct. 2014), 847–860. doi:10.1145/2714064.2660219

[14] Jean-Marie Burkhardt, Françoise Détienne, and Susan Wiedenbeck. 2002. Object-oriented program comprehension: Effect of expertise, task and phase. *Empirical Software Engineering* 7, 2 (2002), 115–156. doi:10.1023/A:1015297914742

[15] Nicholas Carlson. 2010. Warning: Google buzz has a huge privacy flaw-business insider. *Retrieved December* 18 (2010), 2018.

[16] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* 15, 3, Article 39 (March 2024), 45 pages. doi:10.1145/3641289

[17] Chaoran Chen, Daodao Zhou, Yanfang Ye, Toby Jia-Jun Li, and Yaxing Yao. 2025. CLEAR: Towards Contextual LLM-Empowered Privacy Policy Analysis and Risk Generation for Large Language Model Applications. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 277–297. doi:10.1145/3708359.3712156

[18] Amit Chowdhry. 2016. Uber: Users are more likely to pay surge pricing if their phone battery is low.

[19] Roger Clarke. 2009. Privacy impact assessment: Its origins and development. *Computer Law & Security Review* 25, 2 (2009), 123–135. doi:10.1016/j.clsr.2009.02.002

[20] Nancy J. Cooke. 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies* 41, 6 (1994), 801–849. doi:10.1006/ijhc.1994.1083

[21] R. Jason Cronk and Stuart S. Shapiro. 2021. Quantitative Privacy Risk Analysis. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 340–350. doi:10.1109/EuroSPW54576.2021.00043

[22] Hao Cui, Rahmadi Trimananda, Athina Markopoulou, and Scott Jordan. 2023. PoliGraph: Automated Privacy Policy Analysis using Knowledge Graphs. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 1037–1054. https://www.usenix.org/conference/usenixsecurity23/presentation/cui

[23] Sourya Joyee De and Daniel Le Métayer. 2016. PRIAM: a privacy risk analysis methodology. In *Data Privacy Management and Security Assurance: 11th International Workshop, DPM 2016 and 5th International Workshop, QASA 2016, Heraklion, Crete, Greece, September 26-27, 2016, Proceedings 11*. Springer, 221–229. doi:10.1007/978-3-319-47072-6_15

[24] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. 2011. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering* 16, 1 (2011), 3–32. doi:10.1007/s00766-010-0115-7

[25] Thomas Dohmke, Marco Iansiti, and Greg Richards. 2023. Sea Change in Software Development: Economic and Productivity Analysis of the AI-Powered Developer Lifecycle. arXiv:2306.15033 [econ.GN] https://arxiv.org/abs/2306.15033

[26] W Dou, DH Jeong, F Stukes, W Ribarsky, HR Lipford, and R Chang. 2009. Comparing Usage Patterns of Domain Experts and Novices in Visual Analytical Tasks. In *Sensemaking Workshop*.

[27] Moemen Ebrahim, Shawkat Guirguis, and Christine Basta. 2025. Enhancing Software Requirements Engineering with Language Models and Prompting Techniques: Insights from the Current Research and Future Directions. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*. Association for Computational Linguistics, Vienna, Austria, 486–496. doi:10.18653/v1/2025.acl-srw.31

[28] Ethan Fast, William McGrath, Pranav Rajpurkar, and Michael S. Bernstein. 2016. Augur: Mining Human Behaviors from Fiction to Power Interactive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 237–247. doi:10.1145/2858036.2858528

[29] Yuanyuan Feng, Abhilasha Ravichander, Yaxing Yao, Shikun Zhang, and Rex Chen. 2024. Understanding How to Inform Blind and Low-Vision Users about Data Privacy through Privacy Question Answering Assistants. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 2065–2082. https://www.usenix.org/conference/usenixsecurity24/presentation/feng-yuanyuan

[30] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 2021. A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 64, 16 pages. doi:10.1145/3411764.3445148

[31] David Flaherty. 2000. Privacy impact assessments: an essential tool for data protection. *Privacy Law & Policy Reporter* 5 (2000), 85. doi:au/journals/PrivLawPRpr/2000/45.html

[32] Vincent Freiberger, Arthur Fleig, and Erik Buchmann. 2025. "You Don't Need a University Degree to Comprehend Data Protection This Way": LLM-Powered Interactive Privacy Policy Assessment. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 36, 12 pages. doi:10.1145/3706599.3719816

[33] Xun Ge, Ching-Huei Chen, and Kendrick A. Davis. 2005. Scaffolding Novice Instructional Designers' Problem-Solving Processes Using Question Prompts in

a Web-Based Learning Environment. *Journal of Educational Computing Research* 33, 2 (2005), 219–248. doi:10.2190/5F6J-HHVF-2U2B-8T3G

[34] The Guardian. 2025. Privacy. https://www.theguardian.com/world/privacy. Accessed on 05/10/2025.

[35] Komal Gupta and Aditya Shrivastava. 2025. Zero Data Retention in LLM-based Enterprise AI Assistants: A Comparative Study of Market Leading Agentic AI Products. arXiv:2510.11558 [cs.AI] https://arxiv.org/abs/2510.11558

[36] Seda Gürses, Carmela Troncoso, and Claudia Diaz. 2011. Engineering privacy by design. *Computers, Privacy & Data Protection* 14, 3 (2011), 25.

[37] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. 2018. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering* 23, 1 (2018), 259–289. doi:10.1007/s10664-017-9517-1

[38] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 531–548. https://www.usenix.org/conference/usenixsecurity18/presentation/harkous

[39] Hamza Harkous, Kassem Fawaz, Kang G. Shin, and Karl Aberer. 2016. PriBots: Conversational Privacy with Chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO. https://www.usenix.org/conference/soups2016/workshop-program/wfpn/presentation/harkous

[40] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9

[41] Kashmir Hill. 2024. How target figured out a teen girl was pregnant before her father did. https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/. Accessed on 05/10/2025.

[42] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.* 7, 2 (June 2000), 174–196. doi:10.1145/353485.353487

[43] Mitra Bokaei Hosseini, Travis D. Breaux, Rocky Slavin, Jianwei Niu, and Xiaoyin Wang. 2021. Analyzing privacy policies through syntax-driven semantic analysis of information types. *Information and Software Technology* 138 (2021), 106608. doi:10.1016/j.infsof.2021.106608

[44] Mitra Bokaei Hosseini, John Heaps, Rocky Slavin, Jianwei Niu, and Travis Breaux. 2021. Ambiguity and Generality in Natural Language Privacy Policies. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*. 70–81. doi:10.1109/RE51729.2021.00014

[45] Cory Hymel and Hiroe Johnson. 2025. Analysis of LLMs vs Human Experts in Requirements Engineering. arXiv:2501.19297 [cs.SE] https://arxiv.org/abs/2501.19297

[46] Leonardo Horn Iwaya, Ala Sarah Alaqra, Marit Hansen, and Simone Fischer-Hübner. 2024. Privacy impact assessments in the wild: A scoping review. *Array* 23 (2024), 100356. doi:10.1016/j.array.2024.100356

[47] Haojian Jin, Hong Shen, Mayank Jain, Swarun Kumar, and Jason I. Hong. 2021. Lean Privacy Review: Collecting Users' Privacy Concerns of Data Practices at a Low Cost. *ACM Trans. Comput.-Hum. Interact.* 28, 5, Article 34 (Aug. 2021), 55 pages. doi:10.1145/3463910

[48] Samantha Katcher, Ben Ballard, Cara Bloom, Katie Isaacson, Julie McEwen, Stuart Shapiro, Shelby Slotter, Mark Paes, and Ryan Xu. 2024. The PANOPTIC™ Privacy Threat Model. In *Twentieth Symposium on Usable Privacy and Security (SOUPS)*.

[49] Dilara Keküllüoğlu and Yasemin Acar. 2023. "We are a startup to the core": A qualitative interview study on the security and privacy development practices in Turkish software startups. In *2023 IEEE Symposium on Security and Privacy (SP)*. 2015–2031. doi:10.1109/SP46215.2023.10179339

[50] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. doi:10.1145/1572532.1572538

[51] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1573–1582. doi:10.1145/1753326.1753561

[52] Shaymaa Mamdouh Khalil, Hayretdin Bahsi, and Tarmo Korõtko. 2024. Threat modeling of industrial control systems: A systematic literature review. *Computers & Security* 136 (2024), 103543. doi:10.1016/j.cose.2023.103543

[53] Jeong-Dong Kim, Jiseong Son, and Doo-Kwon Baik. 2012. CA5W1HOnto: Ontological Context-Aware Model Based on 5W1H. *International Journal of Distributed Sensor Networks* 8, 3 (2012), 247346. doi:10.1155/2012/247346

[54] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. 2022. Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 508–520. doi:10.1145/3531146.3533116

[55] Madhava Krishna, Bhagesh Gaur, Arsh Verma, and Pankaj Jalote. 2024. Using LLMs in Software Requirements Specifications: An Empirical Evaluation. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*. 475–483. doi:10.1109/RE59067.2024.00056

[56] Naomi Lefkovitz. 2020. NIST Privacy Framework: The Implementation Challenges. https://www.bankinfosecurity.com/interviews/nist-privacy-framework-implementation-challenges-i-4594. (Accessed: 2025-05-23).

[57] Qing Li and Yu-Liu Chen. 2009. *Data Flow Diagram*. Springer Berlin Heidelberg, Berlin, Heidelberg, 85–97. doi:10.1007/978-3-540-89556-5_4

[58] Tianshi Li, Yuvraj Agarwal, and Jason I. Hong. 2018. Coconut: An IDE Plugin for Developing Privacy-Friendly Apps. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 178 (Dec. 2018), 35 pages. doi:10.1145/3287056

[59] Tianshi Li, Lorrie Faith Cranor, Yuvraj Agarwal, and Jason I. Hong. 2024. Matcha: An IDE Plugin for Creating Accurate Privacy Nutrition Labels. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 33 (March 2024), 38 pages. doi:10.1145/3643544

[60] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I. Hong. 2021. How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 220 (Jan. 2021), 28 pages. doi:10.1145/3432919

[61] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 588, 24 pages. doi:10.1145/3491102.3502012

[62] Tony W Li, Arshia Arya, and Haojian Jin. 2024. Redesigning Privacy with User Feedback: The Case of Zoom Attendee Attention Tracking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 237, 14 pages. doi:10.1145/3613904.3642594

[63] James Lin, Mark W. Newman, Jason I. Hong, and James A. Landay. 2000. DENIM: finding a tighter fit between tools and practice for Web site design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 510–517. doi:10.1145/332040.332486

[64] Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A. Smith. 2014. A Step Towards Usable Privacy Policy: Automatic Alignment of Privacy Statements. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 884–894. https://aclanthology.org/C14-1084/

[65] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL] https://arxiv.org/abs/1907.11692

[66] Jesus Luna, Neeraj Suri, and Ioannis Krontiris. 2012. Privacy-by-design based on quantitative threat modeling. In *2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*. 1–8. doi:10.1109/CRISIS.2012.6378941

[67] Stephen MacNeil, Zijian Ding, Kexin Quan, Thomas j Parashos, Yajie Sun, and Steven P. Dow. 2021. Framing Creative Work: Helping Novices Frame Better Problems through Interactive Scaffolding. In *Proceedings of the 13th Conference on Creativity and Cognition* (Virtual Event, Italy) *(C&C '21)*. Association for Computing Machinery, New York, NY, USA, Article 30, 10 pages. doi:10.1145/3450741.3465261

[68] Giles Turner Matt Day and Natalia Drozdiak / Bloomberg. 2019. Thousands of Amazon workers listen to Alexa conversations. https://time.com/5568815/amazon-workers-listen-to-alexa/. Accessed on 05/10/2025.

[69] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[70] Patrick E. McKnight and Julius Najab. 2010. *Mann-Whitney U Test*. John Wiley & Sons, Ltd, 1–1. doi:10.1002/9780470479216.corpsy0524

[71] Liz Mineo. 2025. What happens to your data if 23andMe collapses? https://news.harvard.edu/gazette/story/2025/03/what-happens-to-your-genetic-data-if-23andme-collapses/. Accessed on 06/01/2025.

[72] Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. 2013. Metaphor Identification in Large Texts Corpora. *PLOS ONE* 8, 4 (04 2013), 1–9. doi:10.1371/journal.pone.0062343

[73] Tricia J. Ngoon, C. Ailie Fraser, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive Guidance Techniques for Improving Creative Feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3173574.3173629

[74] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

[75] NIST. 2020. NIST privacy framework: a tool for improving privacy through enterprise risk management. doi:10.6028/NIST.CSWP.01162020

[76] Donald A. Norman. 1993. *Things that make us smart: defending human attributes in the age of the machine.* Addison-Wesley Longman Publishing Co., Inc., USA. doi:10.5555/200550

[77] Object Management Group Standards Development Organization (OMG SDO). 2017. *Unified Modeling Language 2.5.1.* Object Management Group Standards Development Organization (OMG SDO). https://www.omg.org/spec/UML/2.5.1/PDF#page=681.07 OMG Document Number formal/2017-12-05, Chapter 18.

[78] Marie Caroline Oetzel and Sarah Spiekermann. 2014. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems* 23, 2 (2014), 126–150. doi:10.1057/ejis.2013.18

[79] Maxwell Prybylo, Sara Haghighi, Sai Teja Peddinti, and Sepideh Ghanavati. 2024. Evaluating Privacy Perceptions, Experience, and Behavior of Software Development Teams. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024).* USENIX Association, Philadelphia, PA, 101–120. https://www.usenix.org/conference/soups2024/presentation/prybylo

[80] Abhilasha Ravichander, Alan W Black, Thomas Norton, Shomir Wilson, and Norman Sadeh. 2021. Breaking Down Walls of Text: How Can NLP Benefit Consumer Privacy?. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).* Association for Computational Linguistics, Online, 4125–4140. doi:10.18653/v1/2021.acl-long.319

[81] Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. 2019. Question Answering for Privacy Policies: Combining Computational and Legal Perspectives. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* Association for Computational Linguistics, Hong Kong, China, 4947–4958. doi:10.18653/v1/D19-1500

[82] Hugo Roy, JC Borchardt, I McGowan, J Stout, and S Azmayesh. 2012. Terms of Service; Didn't Read. https://tosdr.org/. (Accessed: 2025-05-19).

[83] Ira S. Rubinstein and Nathaniel Good. 2013. Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents. *Berkeley Technology Law Journal* 28, 2 (2013), 1333–1413. https://btlj.org/data/articles2015/vol28/28_2/28-berkeley-tech-l-j-1333-1414.pdf

[84] Gal Sasson and Yoed N. Kenett. 2023. A Mirror to Human Question Asking: Analyzing the Akinator Online Question Game. *Big Data and Cognitive Computing* 7, 1 (2023). doi:10.3390/bdcc7010026

[85] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015).* USENIX Association, Ottawa, 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[86] Awanthika Senarath and Nalin A. G. Arachchilage. 2018. Why developers cannot embed privacy into software systems? An empirical investigation. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (Christchurch, New Zealand) *(EASE '18).* Association for Computing Machinery, New York, NY, USA, 211–216. doi:10.1145/3210459.3210484

[87] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. 2019. Going against the (Appropriate) Flow: A Contextual Integrity Approach to Privacy Policy Analysis. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7, 1 (Oct. 2019), 162–170. doi:10.1609/hcomp.v7i1.5266

[88] Laurens Sion, Dimitri Van Landuyt, Kim Wuyts, and Wouter Joosen. 2025. Robust and reusable LINDDUN privacy threat knowledge. *Computers & Security* 154 (2025), 104419. doi:10.1016/j.cose.2025.104419

[89] Daniel J. Solove. 2006. A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154, 3 (2006), 477–564. https://scholarship.law.upenn.edu/penn_law_review/vol154/iss3/1

[90] Mukund Srinath, Pranav Narayanan Venkit, Maria Badillo, Florian Schaub, C. Giles, and Shomir Wilson. 2024. Automated Detection and Analysis of Data Practices Using A Real-World Corpus. In *Findings of the Association for Computational Linguistics: ACL 2024,* Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 4567–4574. doi:10.18653/v1/2024.findings-acl.271

[91] Peter Stender and Gabriele Kaiser. 2015. Scaffolding in complex modelling situations. *Zdm* 47, 7 (2015), 1255–1267. doi:10.1007/s11858-015-0741-0

[92] BOLUN SUN, Yifan Zhou, and Haiyun Jiang. 2025. Empowering Users in Digital Privacy Management through Interactive LLM-Based Agents. In *The Thirteenth International Conference on Learning Representations.* https://openreview.net/forum?id=FEpAUnS7f7

[93] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. 2021. Privacy Champions in Software Teams: Understanding Their Motivations, Strategies, and Challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 693, 15 pages. doi:10.1145/3411764.3445768

[94] Chenhao Tang, Zhengliang Liu, Chong Ma, Zihao Wu, Yiwei Li, Wei Liu, Dajiang Zhu, Quanzheng Li, Xiang Li, Tianming Liu, and Lei Fan. 2023. PolicyGPT: Automated Analysis of Privacy Policies with Large Language Models. arXiv:2309.10238 [cs.CL] https://arxiv.org/abs/2309.10238

[95] TechCrunch. 2025. Privacy. https://techcrunch.com/category/privacy/. Accessed on 05/10/2025.

[96] Kirti Tiwari, Alpika Tripathi, Shipra Sharma, and Vandana Dubey. 2012. Merging of Data Flow Diagram with Unified Modeling. *International Journal of Scientific and Research Publications* 2, 8 (2012), 1–6. http://www.ijsrp.org/research-paper-0812.php?rp=P0772

[97] Christine CA van Nooijen, Bjorn B de Koning, Wichor M Bramer, Anna Isahakyan, Maryam Asoodar, Ellen Kok, Jeroen JG van Merrienboer, and Fred Paas. 2024. A cognitive load theory approach to understanding expert scaffolding of visual problem-solving tasks: A scoping review. *Educational Psychology Review* 36, 1 (2024), 12. doi:10.1007/s10648-024-09848-3

[98] The Verge. 2025. Privacy. https://www.theverge.com/privacy. Accessed on 05/10/2025.

[99] Marin Vukovic, Damjan Katusic, Pavle Skocir, Dragan Jevtic, Luka Delonga, and Daniela Trutin. 2014. User privacy risk calculator. In *2014 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM).* 211–216. doi:10.1109/SOFTCOM.2014.7039133

[100] Annie Waldherr, Lars-Ole Wehden, Daniela Stoltenberg, Peter Miltner, Sophia Ostner, and Barbara Pfetsch. 2019. Inductive Codebook Development for Content Analysis: Combining Automated and Manual Methods. *Forum: Qualitative Social Research* 20, 1 (Jan. 2019). doi:10.17169/fqs-20.1.3058

[101] Ari Ezra Waldman. 2021. *Industry unbound: The inside story of privacy, data, and corporate power.* Cambridge University Press. doi:10.1017/9781108591386

[102] Jinhe Wen, Yingxi Zhao, Wenqian Xu, Yaxing Yao, and Haojian Jin. 2025. Teaching Data Science Students to Sketch Privacy Designs Through Heuristics. In *2025 IEEE Symposium on Security and Privacy (SP).* 1251–1269. doi:10.1109/SP61157.2025.00147

[103] David Wicks. 2017. The Coding Manual for Qualitative Researchers (3rd edition). *Qualitative Research in Organizations and Management: An International Journal* 12, 2 (06 2017), 169–170. doi:10.1108/QROM-08-2016-1408

[104] Tharaka Wijesundara, Matthew Warren, and Nalin Asanka Gamagedara Arachchilage. 2025. SoK: Enhancing Privacy-Preserving Software Development from a Developers' Perspective. arXiv:2504.20350 [cs.SE] https://arxiv.org/abs/2504.20350

[105] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. 2016. The Creation and Analysis of a Website Privacy Policy Corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Association for Computational Linguistics, Berlin, Germany, 1330–1340. doi:10.18653/v1/P16-1126

[106] Wired. 2025. Privacy. https://www.wired.com/category/security/privacy/. Accessed on 05/10/2025.

[107] R. F. Woolson. 2005. *Wilcoxon Signed-Rank Test.* John Wiley & Sons, Ltd. doi:10.1002/0470011815.b2a15177

[108] David Wright. 2012. The state of the art in privacy impact assessment. *Computer Law & Security Review* 28, 1 (2012), 54–61. doi:10.1016/j.clsr.2011.11.007

[109] David Wright and Paul De Hert. 2012. *Introduction to Privacy Impact Assessment.* Springer Netherlands, Dordrecht, 3–32. doi:10.1007/978-94-007-2543-0_1

[110] Yuxi Wu, W. Keith Edwards, and Sauvik Das. 2022. "A Reasonable Thing to Ask For": Towards a Unified Voice in Privacy Collective Action. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 32, 17 pages. doi:10.1145/3491102.3517467

[111] Kim Wuyts, Riccardo Scandariato, and Wouter Joosen. 2014. Empirical evaluation of a privacy-focused threat modeling methodology. *Journal of Systems and Software* 96 (2014), 122–138. doi:10.1016/j.jss.2014.05.075

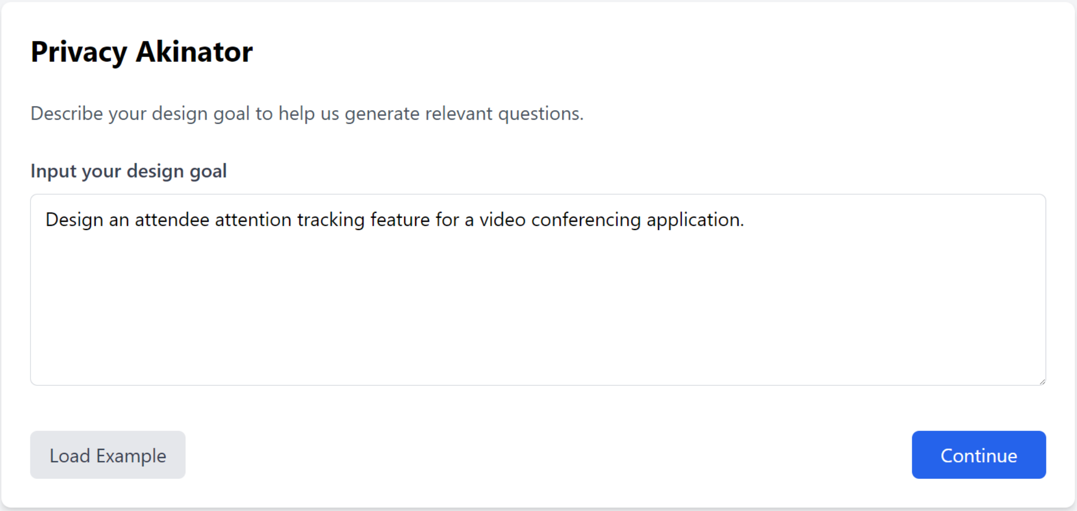
[112] Kim Wuyts, Laurens Sion, and Wouter Joosen. 2020. LINDDUN GO: A Lightweight Approach to Privacy Threat Modeling. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* 302–309. doi:10.1109/EuroSPW51379.2020.00047

[113] Zihao Yi, Jiarui Ouyang, Zhe Xu, Yuwen Liu, Tianhao Liao, Haohao Luo, and Ying Shen. 2025. A Survey on Recent Advances in LLM-Based Multi-turn Dialogue Systems. *ACM Comput. Surv.* (Oct. 2025). doi:10.1145/3771090

[114] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. 2022. How usable are ios app privacy labels? *Proceedings on Privacy Enhancing Technologies* (2022), 204–228. Issue 4. doi:10.56553/popets-2022-0106

[115] Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *23rd USENIX Security Symposium (USENIX Security 14).* USENIX Association, San Diego, CA, 1–16. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/zimmeck

A User Interface



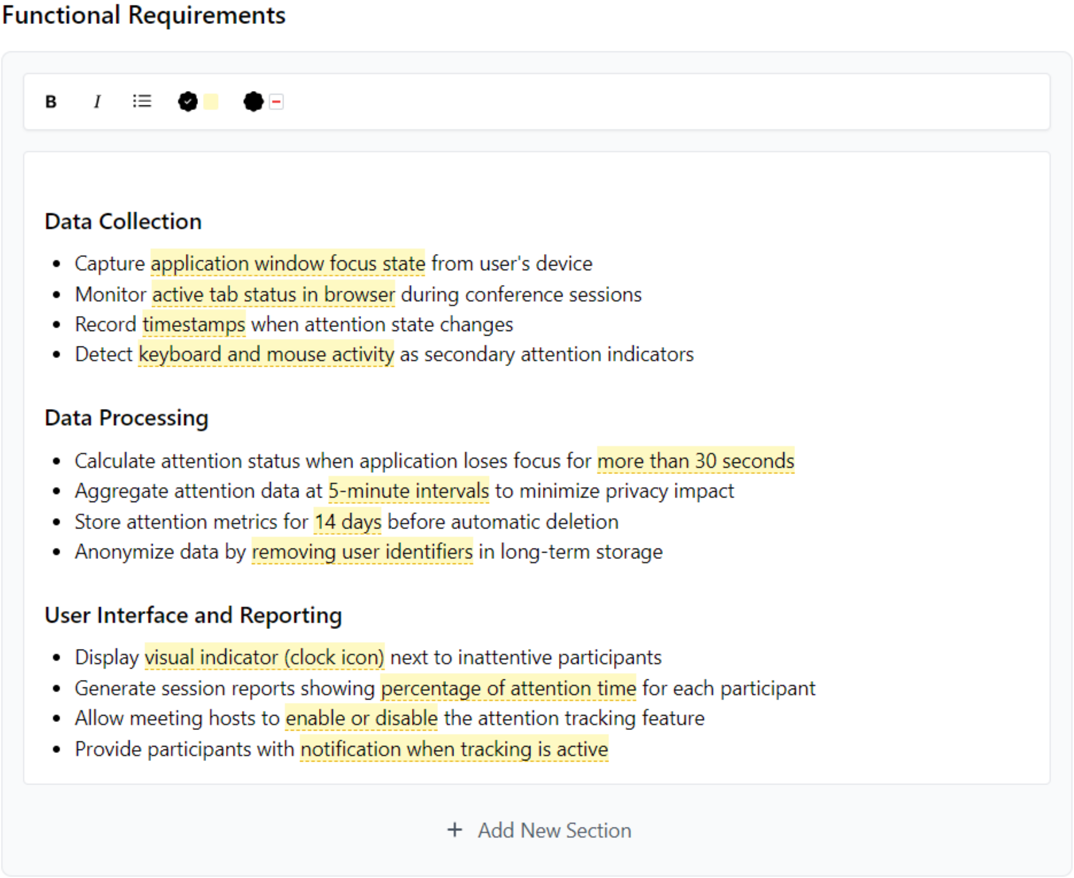
Privacy Akinator

Describe your design goal to help us generate relevant questions.

Input your design goal

Design an attendee attention tracking feature for a video conferencing application.

Figure 6: PrivacyAkinator Design Goal Panel



Functional Requirements

Data Collection

- Capture application window focus state from user's device
- Monitor active tab status in browser during conference sessions
- Record timestamps when attention state changes
- Detect keyboard and mouse activity as secondary attention indicators

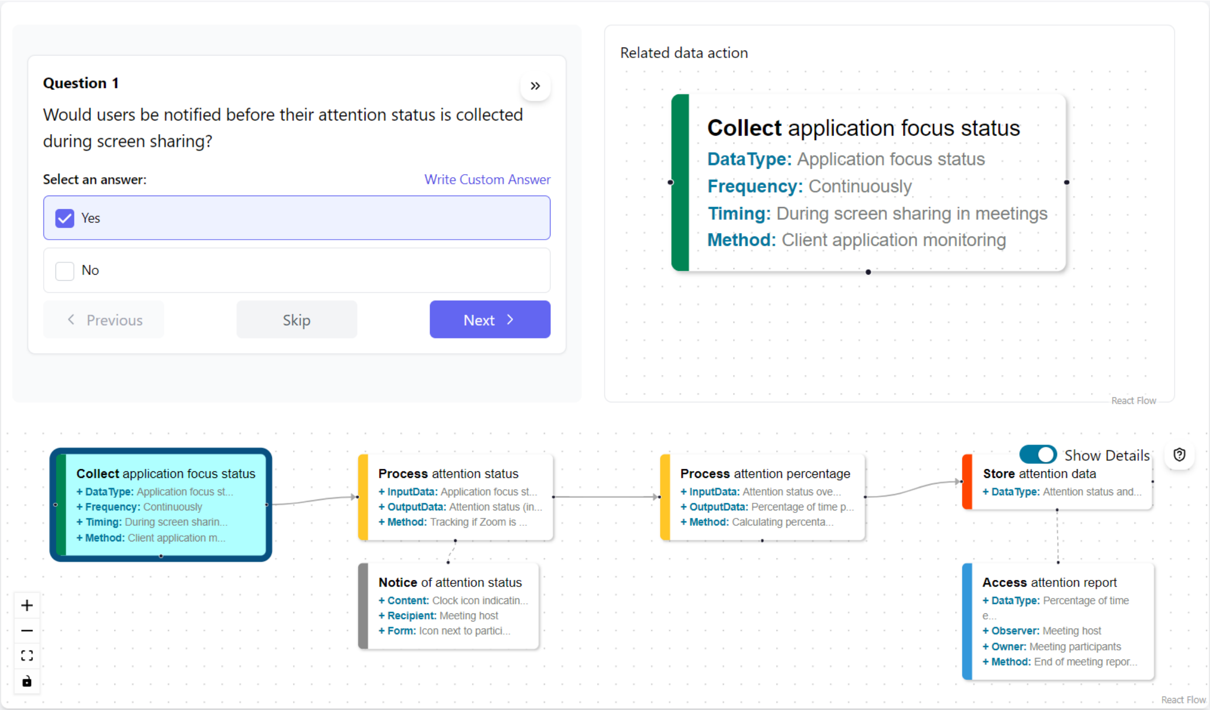
Data Processing

- Calculate attention status when application loses focus for more than 30 seconds
- Aggregate attention data at 5-minute intervals to minimize privacy impact
- Store attention metrics for 14 days before automatic deletion
- Anonymize data by removing user identifiers in long-term storage

User Interface and Reporting

- Display visual indicator (clock icon) next to inattentive participants
- Generate session reports showing percentage of attention time for each participant
- Allow meeting hosts to enable or disable the attention tracking feature
- Provide participants with notification when tracking is active

Figure 7: PrivacyAkinator functional Requirement Panel



Question 1

Would users be notified before their attention status is collected during screen sharing?

Select an answer:

☑ Yes

□ No

Related data action

Collect application focus status
DataType: Application focus status
Frequency: Continuously
Timing: During screen sharing in meetings
Method: Client application monitoring

Collect application focus status
+ **DataType:** Application focus st...
+ **Frequency:** Continuously
+ **Timing:** During screen sharin...
+ **Method:** Client application m...

Process attention status
+ **InputData:** Application focus st...
+ **OutputData:** Attention status (in...
+ **Method:** Tracking if Zoom is ...

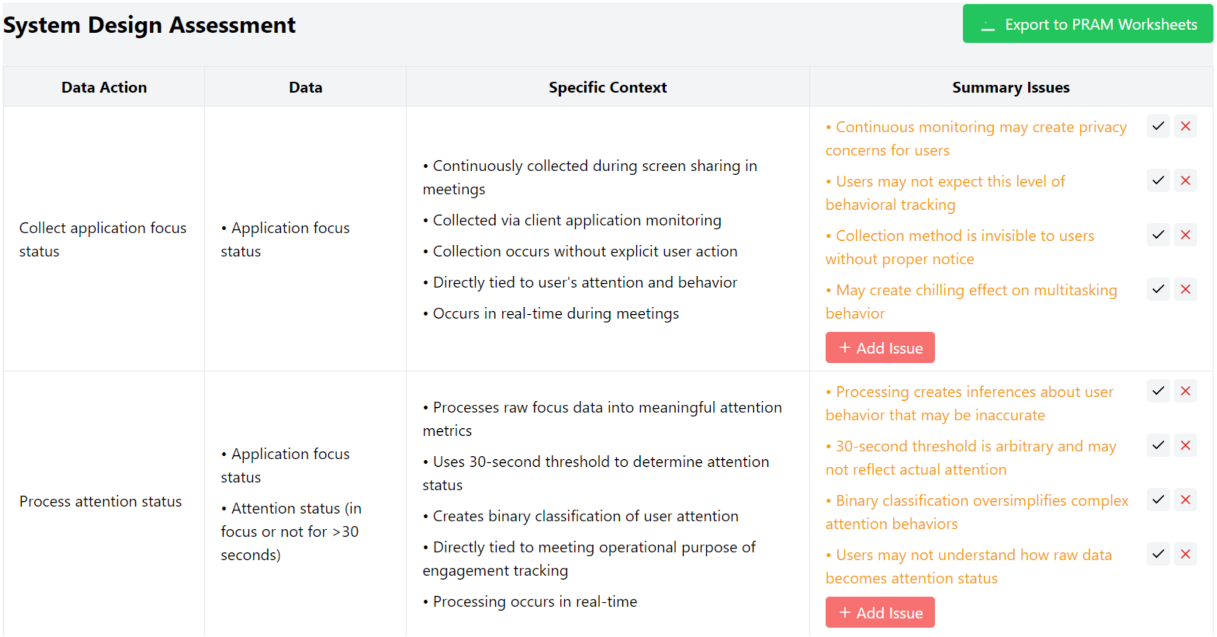
Notice of attention status
+ **Content:** Clock icon indicatin...
+ **Recipient:** Meeting host
+ **Form:** Icon next to partici...

Process attention percentage
+ **InputData:** Attention status ove...
+ **OutputData:** Percentage of time p...
+ **Method:** Calculating percenta...

Store attention data
+ **DataType:** Attention status and...

Access attention report
+ **DataType:** Percentage of time e...
+ **Observer:** Meeting host
+ **Owner:** Meeting participants
+ **Method:** End of meeting repor...

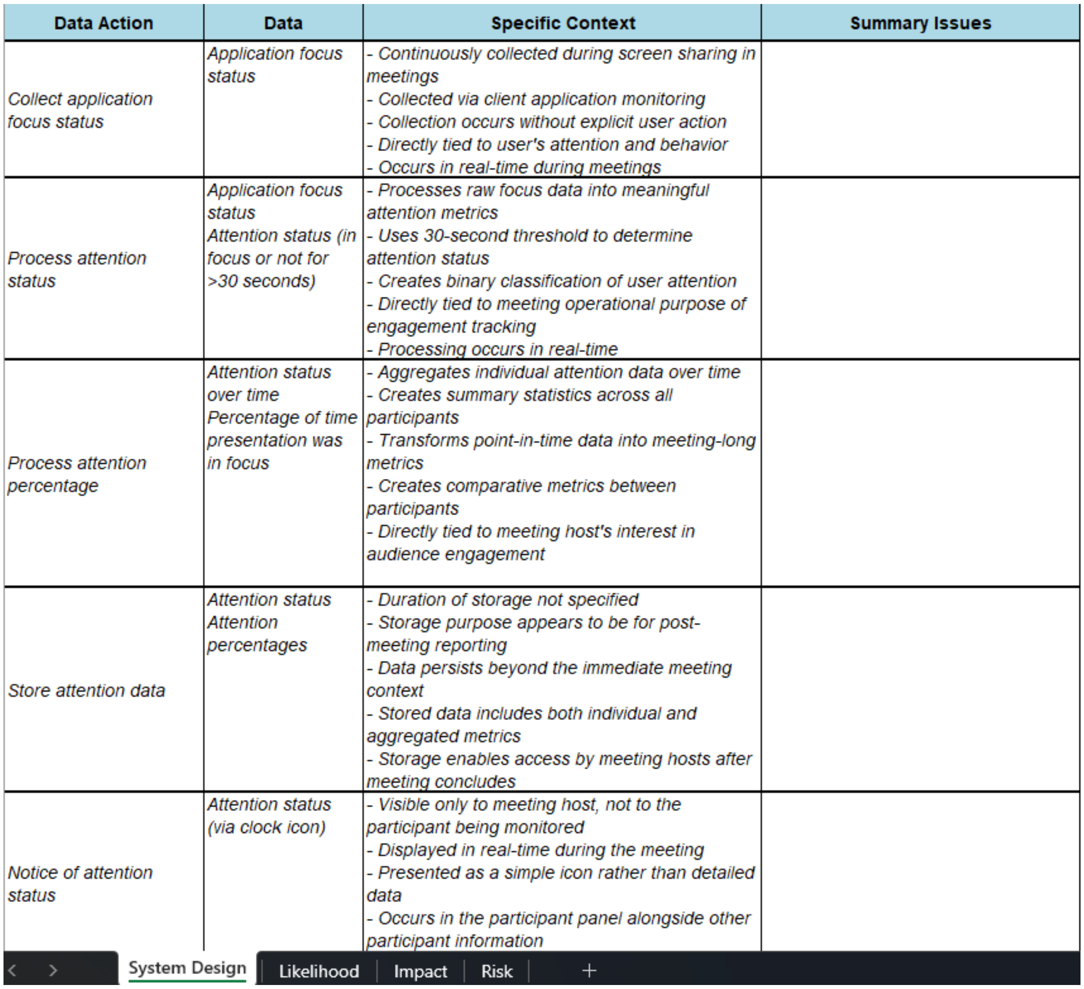
Figure 8: PrivacyAkinator Workflow & Questions Panel



System Design Assessment

Data Action	Data	Specific Context	Summary Issues
Collect application focus status	• Application focus status	• Continuously collected during screen sharing in meetings • Collected via client application monitoring • Collection occurs without explicit user action • Directly tied to user's attention and behavior • Occurs in real-time during meetings	• Continuous monitoring may create privacy concerns for users ✓ ✗ • Users may not expect this level of behavioral tracking ✓ ✗ • Collection method is invisible to users without proper notice ✓ ✗ • May create chilling effect on multitasking behavior ✓ ✗ + Add Issue
Process attention status	• Application focus status • Attention status (in focus or not for >30 seconds)	• Processes raw focus data into meaningful attention metrics • Uses 30-second threshold to determine attention status • Creates binary classification of user attention • Directly tied to meeting operational purpose of engagement tracking • Processing occurs in real-time	• Processing creates inferences about user behavior that may be inaccurate ✓ ✗ • 30-second threshold is arbitrary and may not reflect actual attention ✓ ✗ • Binary classification oversimplifies complex attention behaviors ✓ ✗ • Users may not understand how raw data becomes attention status ✓ ✗ + Add Issue

Figure 9: PrivacyAkinator Access System Design Panel



Data Action	Data	Specific Context	Summary Issues
Collect application focus status	*Application focus status*	*- Continuously collected during screen sharing in meetings* *- Collected via client application monitoring* *- Collection occurs without explicit user action* *- Directly tied to user's attention and behavior* *- Occurs in real-time during meetings*	
Process attention status	*Application focus status* *Attention status (in focus or not for >30 seconds)*	*- Processes raw focus data into meaningful attention metrics* *- Uses 30-second threshold to determine attention status* *- Creates binary classification of user attention* *- Directly tied to meeting operational purpose of engagement tracking* *- Processing occurs in real-time*	
Process attention percentage	*Attention status over time* *Percentage of time presentation was in focus*	*- Aggregates individual attention data over time* *- Creates summary statistics across all participants* *- Transforms point-in-time data into meeting-long metrics* *- Creates comparative metrics between participants* *- Directly tied to meeting host's interest in audience engagement*	
Store attention data	*Attention status* *Attention percentages*	*- Duration of storage not specified* *- Storage purpose appears to be for post-meeting reporting* *- Data persists beyond the immediate meeting context* *- Stored data includes both individual and aggregated metrics* *- Storage enables access by meeting hosts after meeting concludes*	
Notice of attention status	*Attention status (via clock icon)*	*- Visible only to meeting host, not to the participant being monitored* *- Displayed in real-time during the meeting* *- Presented as a simple icon rather than detailed data* *- Occurs in the participant panel alongside other participant information*	

Figure 10: PrivacyAkinator Generated Worksheet

B Data Actions and Stakeholder Interactions

Table 5: Data Actions and Stakeholder Interactions

Category	Operation	Definition
Data Action	Collect	Collect users' data or sensor inputs.
	Process	Process data to derive new information.
	Store	Keep data in a persistent storage system.
	Share	Share data to different parties.
Stakeholder Interaction	Consent	A data subject gives permission for specific data actions.
	Notice	A data observer informs data subjects about the data action or its results.
	Control	A data subject controls settings that determine how their data is stored or processed.
	Access	A data observer accesses and uses user data or derived data for a specific purpose.
	Request	A data subject asks to exercise their data rights.
	Audit	An auditor examines data actions for compliance with policies or regulations.
	Influence	A data beneficiary/victim is impacted by a data practice.

C Selected Data Practices for Evaluation

Table 6: A summary of 30 data practices collected to assess the coverage of key privacy design decisions.

#ID	Scenario	Description
#1	Zoom Attention Tracking	An attendee attention tracking feature for a video conferencing application
#2	Facebook Cambridge Analytica	A data platform that allows third-party apps to collect users' data through a social media platform
#3	AirTag	A real-time location tracking device for personal items
#4	Target Pregnancy Prediction	A retail analytics system that processes customer purchase histories to automatically generate personalized coupons and recommendations
#5	Facebook Emotional Contagion	A research experiment to study how exposure to emotional content affects users' own emotional expressions on a social media platform
#6	Google Buzz	A system that integrates email services with social networking functionality
#7	OKCupid Score manipulation	A research experiment to study how displayed compatibility scores influence user behavior and interactions on a dating platform
#8	Uber Price Discrimination	A dynamic pricing system for a ride-sharing service
#9	Staple Price Discrimination	An e-commerce pricing system that incorporates geographic and market data
#10	Expedia Price Discrimination	A dynamic pricing system for a travel booking platform
#11	Strava Fitbit Heatmap	A global visualization of fitness tracking data that unintentionally reveals military personnel's locations and movements in sensitive areas
#12	Alexa Smart Home	A voice-activated smart home assistant that minimizes data collection while maintaining functionality and user convenience
#13	Tesla Camera	A vehicle camera system that captures and processes environmental data to support autonomous driving features while establishing protocols for employee access to customer recordings
#14	23andMe Genetic Data Privacy	A direct-to-consumer genetic testing service that collects, analyzes, stores, and shares users' DNA data
#15	Starlink Satellite Surveillance	A satellite internet communication system for border security that enables continuous connectivity in remote areas, supports real-time data transmission from surveillance equipment, and facilitates agent communications while operating in regions with limited infrastructure
#16	Singapore TraceTogether	A nationwide contact tracing system that tracks citizens' movements and interpersonal contacts using smartphone applications and wearable tokens with data centrally stored in government databases
#17	Fog Reveal (Police)	A location data analytics platform that aggregates commercially available smartphone location information and provides law enforcement agencies with search capabilities to identify and track devices based on time, location, and movement patterns without requiring individual warrants
#18	Meta Facial Recognition	A social media photo tagging system that utilizes facial recognition technology to automatically identify individuals across a platform's user base and suggest tags based on facial geometry analysis from uploaded images and videos
#19	Ring Doorbell	A cloud-connected home security camera system that streams and stores video footage from users' homes, processes this data for motion detection alerts, and provides remote access capabilities through mobile applications with two-way audio communication features
#20	Reverse location search from photos	An AI-based image analysis system that can determine geographic location based on visual elements in photographs
#21	Dutch Child Care Fraudster Detection	An automated fraud detection system for government benefits programs that uses citizen data to create risk profiles and flag potentially fraudulent applications for investigation
#22	Telegram User Data Privacy Issues	A messaging platform that provides encrypted communications while managing legal compliance requirements regarding user data and content moderation
#23	League of Legends Chat-log Review	A system that analyzes employee behavior in company products during their personal time and incorporates this data into performance evaluations and employment decisions
#24	MiHoYo In-app Purchases	Mobile game with in-app purchases that appeals to users of all ages and maintains profitability
#25	Dark Patterns in Subscription Services	A subscription management system for digital services
#26	Olympic AI Surveillance	A video monitoring system that uses artificial intelligence to scan large crowds at major public events and automatically identify suspicious behaviors, unusual activities, and potential security threats
#27	Social Credit Score	A comprehensive monitoring and evaluation system for federal employees and contractors that tracks workplace performance, personal conduct, social media activity, and outside associations to generate trustworthiness scores influencing employment decisions
#28	Parental Control Applications	A parental control application for monitoring children's device usage
#29	Browser Fingerprint	A cross-site tracking system that identifies users without cookies by collecting technical information about their browsers and devices to create persistent identifiers that work even when privacy protections are enabled
#30	Government Email Transmission	An employee data management system for municipal governments that enables sharing workforce information with other government entities while managing sensitive personal details across organizational boundaries

D Interview Questions for the Observational and Evaluation Study

Table 7: The interview protocol used during the observational study.

Phase / Section	Prompts
Introduction	Thank you for participating. In this study, we are interested in understanding how developers perform privacy risk assessments. We will ask you to work through a task using a standard framework called PRAM. There are no right or wrong answers; we are interested in your natural thought process.
Instruction	As you work on the task, please try to think out loud. Tell us what you're thinking. This will help us understand your perspective.
Consent Script	Before we begin the study, I need to obtain your informed consent. This session will be recorded for the sole purpose of accurate transcription. The recording will be permanently deleted by the research team immediately after the transcript is verified. All data collected from the interview will be fully anonymized to protect your privacy. Your participation is completely voluntary. You have the right to skip any question you are not comfortable with, or to withdraw from the study at any time without penalty.
Post-study Questions	
Overall	How would you describe your overall experience of using the PRAM framework? Which of the PRAM worksheets or tasks did you find the most challenging, and why? Which part of the framework, if any, did you find helpful or easy to use? Why? If you could change one thing about this framework to make it easier for developers, what would it be?
WS1: Framing Business Objectives	How are you thinking about the main purpose of this system? What factors are you considering when describing the functional requirements?
WS2: Assessing System Design	As you map the data flows, could you explain: - How did you decide which system components were important to include? As you identify the data actions, could you explain: - What was your process for identifying all the data actions here? - How clear are the definitions for terms?
WS3: Prioritizing Risk	As you assign Likelihood/Impact scores: - Could you explain your reasoning for giving 'likelihood' that specific rating? As you calculate the final risk score: - What does that number tell you about the risk? - How do you decide if this score is high enough to require action?

Table 8: The interview questions used during the evaluation study.

Theme	Post-Study Interview Questions
Overall Experiences	How would you describe your overall experience using the PRAM worksheet versus PrivacyAkinator? Which approach did you find more intuitive or easier to use, and why?
Strengths & Weaknesses	What aspects of the PRAM worksheet did you find most helpful or effective? What aspects of PRAM worksheet were confusing, frustrating, or difficult to use? What challenges or limitations did you encounter with the PRAM worksheet? What aspects of PrivacyAkinator did you find most helpful or effective? What aspects of PrivacyAkinator were confusing, frustrating, or difficult to use? What challenges or limitations did you encounter with PrivacyAkinator?
Reflection & Preference	If you need to use one approach in a real project, which would you choose, and why? How would you improve either approach based on your experience?

E Qualitative Analysis Codebook from Observational Study

Table 9: The codebook developed during our qualitative analysis of the observational study

Theme	Code	Sub-code	Description
Attention Allocation	Information Overload	Cognitive Fatigue	Participants experienced heavy mental load from juggling numerous factors simultaneously. This led to growing fatigue as the tasks progressed.
		Context-Switching Overhead	Participants experienced additional cognitive load when transferring information across worksheets or repeatedly referring back and forth between different parts of the task.
	Ineffective Prioritization	Fixation on Familiar Concepts	Participants focused excessively on design aspects that aligned with their technical background, while neglecting less familiar but critical privacy risks.
		Fixation on Early Tasks	Participants spent a disproportionate amount of time and mental energy on the initial stages of a task, leading them to rush through or neglect later stages.
Inaccessible Privacy Knowledge	Knowledge Blindspot	Unaware of Key Concepts	Participants overlooked important privacy issues because they did not know these aspects needed to be considered.
	Knowledge Recall Failure	Missed Known Concepts	Participants knew about a relevant privacy concept but forgot to apply or include it during the assessment due to a slip of the mind.
		Memory challenges	Participants struggled to carry insights and considerations from one part of the assessment to a later stage, often due to being overwhelmed.
Insufficient Guidance	Unclear Expectations	Ambiguous Granularity	Participants felt unsure about the required level of technical detail for their responses, whether to be high-level or dive into specific implementations.
		Ambiguous Scope	Participants were unclear about how comprehensive their responses should be or what exactly to include.
	Vague Terms	Ambiguous Language	Key terms used in the framework were perceived as overly technical, unclear, or poorly defined, causing confusion for participants.
		Inconsistent Framing	Concepts were defined or presented in inconsistent ways in different parts of the framework.
Uncertainty in Risk Evaluation	Difficulty Assigning Scores	Inconsistent Scoring	Without guidance on how to weigh different organizational impacts, participants assigned widely varying scores to the same privacy risk.
	Difficulty Interpreting Scores	Arbitrary Risk Thresholds	Participants struggled to determine which calculated risk scores required action, as the framework lacked thresholds for what constitutes a high, medium, or low risk.