

# Teaching Data Science Students to Sketch Privacy Designs through Heuristics

Jinhe Wen<sup>◊†</sup>, Yingxi Zhao<sup>◊†</sup>, Wenqian Xu<sup>◊†</sup>, Yaxing Yao<sup>§</sup>, Haojian Jin<sup>†</sup>

<sup>†</sup>University of California, San Diego <sup>§</sup>Virginia Tech

**Abstract**—Recent studies reveal that experienced data practitioners often draw sketches to facilitate communication around privacy design concepts. However, there is limited understanding of how we can help novice students develop such communication skills. This paper studies methods for lowering novice data science students’ barriers to creating high-quality privacy sketches. We first conducted a need-finding study (N=12) to identify barriers students face when sketching privacy designs. We then used a human-centered design approach to guide the method development, culminating in three simple, text-based heuristics. Our user studies with 24 data science students revealed that simply presenting three heuristics to the participants at the beginning of the study can enhance the coverage of privacy-related design decisions in sketches, reduce the mental effort required for creating sketches, and improve the readability of the final sketches.

## 1. Introduction

Designing privacy in data practices is a collaborative process where data practitioners need to frequently communicate key privacy design concepts to others, such as how data is collected, used, shared, and how these data flows interact with users and other stakeholders [1], [2]. A few recent studies reveal that experienced data practitioners often use sketches, either on paper or with digital tools, to facilitate communication around privacy design concepts [3], [4].

Currently, there is no standardized approach to privacy sketching; instead, practitioners tend to improvise in an ad-hoc manner, combining multiple types of diagrams—such as Data Flow Diagrams, Use Case Diagrams, and Component Diagrams [4]. This ad-hoc privacy sketching resembles user experience (UX) sketching decades ago, where only a few UX designers could sketch the UX design concepts effectively based on unstructured intuitions and artistic skills [5]. Since then, HCI researchers have developed multiple structured frameworks to lower the barriers for UX designers to sketch user experiences [6].

In this paper, we draw inspiration from UX sketching and hypothesize that structured frameworks can help lower novice data science students’ barriers to creating high-quality privacy sketches for given data practice scenarios. To explore the potential framework, we conducted three studies with 54 unique participants across three U.S. universities, with no participant overlap between studies.

We began with a need-finding study to identify the challenges novice students face in communicating privacy designs (Section 3). We recruited 12 participants, asking them to consider two data practice scenarios, sketch their privacy design solutions on paper, and explain their designs using their sketches. This study revealed three key challenges: (1) novice students lack the appropriate vocabulary to sketch privacy designs, (2) participants struggle with planning and organizing the sketching space, and (3) the constraints of the sketching process can limit their ability to think expansively about design solutions.

We then adopted a human-centered design approach [7], [8] to guide the method’s development (Section 4). We iteratively prototyped a teaching method, tested it with a new group of users, gathered feedback, and refined the method iteratively. We explored three methods, ranging from a heavy-lifting approach with detailed digital diagramming to a lightweight method offering only three heuristics. Interestingly, the most lightweight method, which simply presents participants with a table of three heuristics (Table 1), proved highly effective in improving the quality of novice students’ sketches despite its simplicity.

We conducted a between-subjects experiment to validate the effectiveness of the heuristic-based approach (Section 5). We recruited a total of 24 participants, with each participant creating a few sketches and interpreting the sketches from other participants. We found that participants who received heuristic-based instructions were able to cover 30% more of the privacy-related decisions in their sketches than the participants without exposure to these instructions. The final sketches from the participants who received heuristic-based instruction are also more readable, with an increase of 77% in interpretation accuracy.

Our primary contribution is a heuristic-based approach that lowers the barriers for novice students to create high-quality privacy sketches, along with insights gained from the iterative development process. Our exploration will help researchers understand how to teach data science students to communicate privacy designs. To the best of our knowledge, this is the first study to systematically explore methods for teaching novice students how to sketch privacy designs.

## 2. Related Work

This project builds on ideas from four key areas: (1) user experience and software design sketches, (2) education, (3) diagrams, and (4) developer support for privacy design.

<sup>◊</sup>Equal contributions.

TABLE 1: We found that simply presenting the table below (i.e., three heuristics along with explanations and examples) to participants at the beginning of the study can enhance the coverage of privacy-related design decisions in sketches, reduce the mental effort required for creating sketches, and improve the readability of the final sketches.

Heuristic	Explanation	Examples
Device-Based Data Flow	Indicate the devices involved at each stage of the data flow to show how data moves from one point to another	Capturing photos on a camera; saving user profiles on the server...
Stakeholder Interactions with Data Flow	Show each individual or group’s interactions with the data flow, including involvement and privacy-related choices	Manager authorizes AI modeling with user data; data scientist analyzes user’s activities...
Multi-Layered Representation	Provide an overview of the privacy design, and then separate this from more detailed privacy considerations	Displaying “storing data” in the overview, with details like “for 5 years with encryption” positioned in a separate layer (e.g., a different area within the sketch)

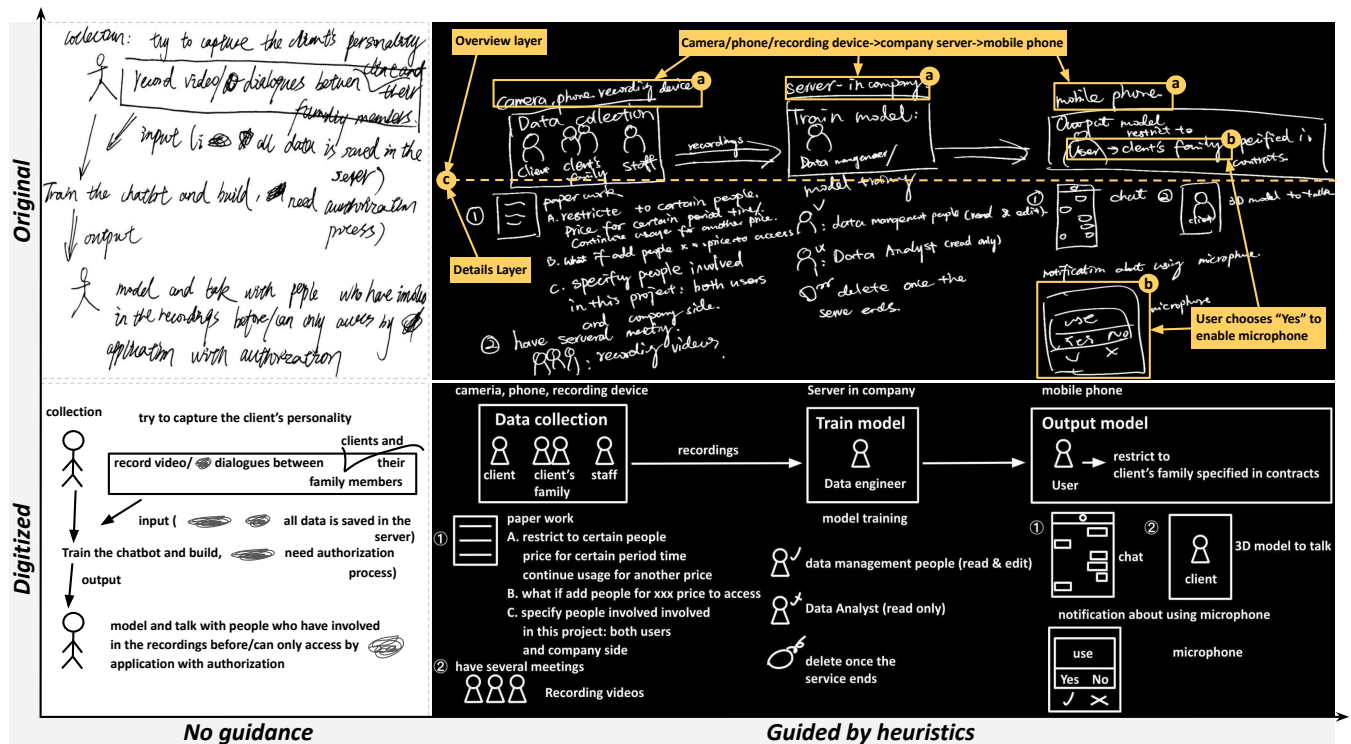


Figure 1: Example sketches for the *Afterlife Chatbot* scenario (see Table 2), with sketches from participants exposed to heuristics (Table 1) on the right and those without on the left. Heuristics-guided sketches (right) are more structured and comprehensive. The raw sketches (top) have been digitized (bottom) for improved readability.

## 2.1. User Experience and Software Design Sketches

Sketching has proven to be an efficient approach for exploring, communicating, and iterating on UX design ideas [9] and software design concepts [10], [11]. Sketching allows designers to rapidly visualize and refine user interactions while intentionally ignoring detailed specifications in early stages [5], [12]. This abstraction facilitates a smooth progression from initial concepts to functional design [13]. These findings have informed the development of structured sketching practices, where designers use specific frameworks and conventions to ensure clarity and consistency in

visual communication [14], [15].

We hypothesize that an analogous approach can be applied to privacy design. Similar to how UX designers use sketches to map user journeys and identify pain points [5], [9], privacy practitioners can use sketching to illustrate how data is collected, processed, and shared and how the data flows interact with stakeholders.

## 2.2. Education for Privacy Design

Current privacy design education primarily focuses on privacy concepts [16] and design methodologies [17], [18],

TABLE 2: Task scenarios used in our study. We initially crafted 14 privacy-related scenarios from media reports [25], [26], [27], [28], [29], [30], [31], [32], research papers [3], [33], [34], [35], and product introductions [36], [37], [38], [39], [40], [41], [42], [43], then selected four scenarios based on participant familiarity, scenario complexity, and domain diversity.

#	Scenario	Description
1	Online Meeting Attention Tracking [3], [26]	Designing an Attention Tracking feature for an online meeting app. The feature captures attendees’ focus and generates attention scores, enabling hosts to monitor engagement levels during meetings.
2	Financial Risk Management [40], [41], [42]	Designing a feature for a bank’s mobile app to enhance risk management, including capabilities like anomaly detection and client credit assessments.
3	Afterlife Chatbot [25], [37], [38]	Designing an “Afterlife Chatbot” service that allows clients to record video biographies. After the client passes away, their family can interact with the chatbot to preserve their memory.
4	Sensitive Image Detection [27], [39], [43]	Designing an automated content moderation system for an image management service to detect inappropriate or harmful content in users’ uploaded pictures.

[19], with most educators using case studies [17], [20] as the main form of instruction. For instance, some classes include a privacy review procedure [20], [21], where students assess smartphone privacy notices, identify issues, and propose design improvements in brief essays. Organizations such as the National Institute of Standards and Technology (NIST) [22] and the International Association of Privacy Professionals (IAPP) [23] also provide privacy design education by offering standards, certifications, and online learning resources, which provide practical guidance and structure for both academic and professional learning. Additionally, educators have expanded the instruction of privacy design to a broader audience, including industry professionals and policymakers [24]. In contrast, our work aims to complement existing curricula by exploring methods to teach students how to sketch privacy designs, equipping them with privacy design communication skills.

### 2.3. Diagrams for Privacy Design

Practitioners often use a combination of diagrams to communicate privacy considerations in system designs [3], [4], [44], [45]. Research has shown that these diagrams help developers visualize system architecture by simplifying complex design elements, making it easier to communicate and design privacy-aware applications [4]. For example, when tasked with communicating the privacy design of an IoT application for diabetes treatment and monitoring, developers used Data Flow Diagrams (DFDs) [46] and Unified Modeling Language (UML) [47] to break down the flow of sensitive health data into manageable components, with annotations clarifying the encryption mechanism [4].

While existing studies shed light on the use of diagrams for communicating privacy considerations in system design, their focus has been on professional developers. *How* to teach these diagramming techniques to novice data science students remains unclear. Our work aims to address this gap by using sketches as the lens to investigate the specific challenges students face when communicating privacy design through diagrams and provide insights on targeted guidance to address these difficulties.

### 2.4. Developer Support for Privacy Design

Prior research has explored many approaches to support experienced developers in privacy design [48], [49]. For instance, some studies have offered insights into how to help developers make privacy design choices by identifying the common challenges in understanding and implementing privacy considerations [50], [51], [52]. Additionally, various tools have been developed to support developers by embedding privacy practices directly into development workflows. For example, Coconut, an Android Studio plugin, helps developers create privacy-friendly apps by requiring privacy annotations and assisting them in organizing privacy-related details [53]. Similarly, PARROT helps developers create privacy-aware IoT applications by providing interactive privacy annotations and guidance [4].

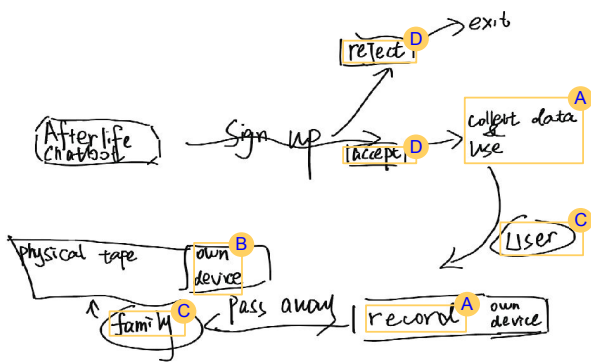
However, existing studies focus on supporting experienced developers with existing knowledge of privacy. Little attention has been given to helping novice data science students learn to communicate privacy design. Our work bridges this gap by providing targeted support to help **novices** develop the skills needed to effectively communicate privacy design.

## 3. Why is Sketching Privacy Challenging?

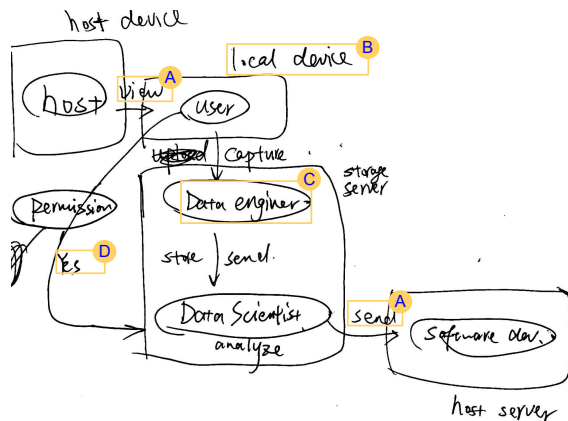
We began with a need-finding study [54] to identify barriers for data science students in sketching privacy designs.

**Participants.** To avoid priming, we advertised the study as a “data science experiment design study” rather than one specifically about privacy through social media and mailing listings across three U.S. universities. We recruited twelve students (seven identified as female, five as male) aged 19 to 24 (Mean = 21.7 years, SD = 2.1 years). Each participant was compensated for their time with a \$10 gift card. The sample included five graduates and seven undergraduates. We used the ACM Data Science Task Force’s definition of Data Science [55] to determine participants’ eligibility.

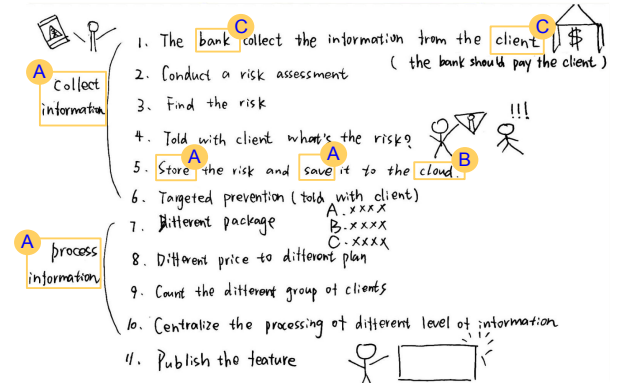
We asked participants, “Do you have any knowledge of privacy-related concepts? Please list” in a pre-screening survey to collect information about their privacy knowledge background. Among the 12 participants, 6 answered “No”; 3



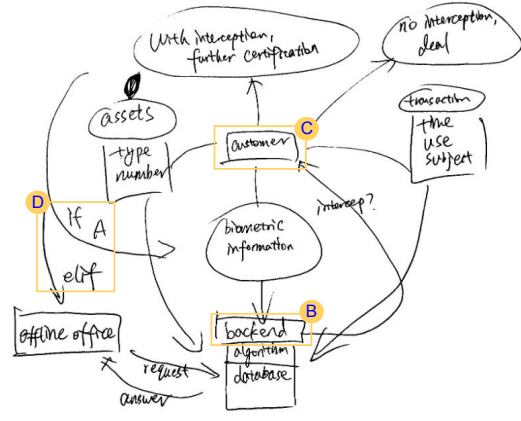
(a) Afterlife Chatbot (by N3).



(c) Online Meeting Attention Tracking (by N10).



(b) Financial Risk Management (by N6).



(d) Financial Risk Management (by N11).

Figure 2: Many sketches share some common components, which we annotate with letters: (A) Data action (e.g., “collect”); (B) Device (e.g., “server”); (C) Stakeholder (e.g., “user”); and (D) Choice (e.g., “accept”). For readability, we annotate only a few examples for each component type.

listed relevant privacy design terms (e.g., privacy by design) they learned through research or course but lacked hands-on experience with privacy design; 3 provided responses unrelated to privacy.

**Method.** In each study, we randomly presented students with brief descriptions of two data practice scenarios (Table 2). For each scenario, we asked them to consider the data practice, sketch relevant privacy designs on paper, and explain their designs to the researchers using their sketches. To understand the fundamental barriers they face, we did not impose time limits, allowing participants to sketch freely until satisfied. After the sketching session, three authors reviewed the participants’ sketches and asked additional questions to clarify concepts in their sketches. The authors also inquired whether they utilized any strategies or encountered any difficulties throughout their sketching process.

On average, participants took approximately 25 minutes (SD = 5.1 minutes) to complete a sketch. We summarized key insights after each study and identified data saturation [56] by the eighth study, with no new insights emerging. We then concluded with four more participants [57].

**Findings.** We made three main observations. **Concepts in**

**privacy design are abstract, and most students lack the appropriate vocabulary to sketch these concepts effectively.** The process of sketching privacy is the process of discovering reusable visual vocabularies. Interestingly, most final sketches shared common components (Figure 2), such as stakeholder and device, but participants were unaware of these initially. Participants often started with an arbitrary line drawing and slowly recognized the basic vocabularies through iterations, and often re-drawn the sketch multiple times near the end using a few visual components discovered in the process. For example, when designing the Afterlife Chatbot service (the sketch without guidance in Figure 1), N10 initially focused on “client,” “record,” “video” and “family” to represent an outline. However, she quickly felt lost, unsure who would manage and process these videos to build the chatbot. She also tried adding the term “server” to indicate where the data would be stored and used. Besides, recognizing that not everyone could view the original videos and that only family members could use the service, she included “authorization” to suggest access control. In the end, her sketch became a mix of arrows, lengthy text, and icons as placeholders for unspecified roles.

TABLE 3: We iteratively experimented with three methods to teach data science students sketching privacy designs. This table summarizes the design changes and experimental observations of each method.

Teaching Method	Participants	Design Changes	Interface	Observations
Object-Oriented Diagramming	T1-T6	(1) Reusable components ( <code>DataAction</code> and <code>Stakeholder</code> classes); (2) Pre-defined attributes with preset options	Web-based application	(1) Learning curve associated with the tool structure; (2) Constrained design options due to incomplete attributes; (3) Appreciation for the re-usable components; (4) Preference for hand-sketch tools with copy-paste support
Vocabulary-Based Sketching	T7-T11	(1) Visual vocabulary (e.g., “ellipse” for stakeholder); (2) Free text descriptions	Tablet-based sketchboard	(1) Appreciation for tablet-supported features; (2) Excessive attention to visual symbols; (3) Preference for flexible design guidelines over strict vocabulary; (4) Distraction from the overall design due to focus on details
Heuristic-Based Sketching	T12-T18	Three heuristics: (1) Device-based data flow; (2) Stakeholder interaction with data flow; (3) Multi-layered representation	Tablet-based sketchboard	Preference for sketching with heuristics rather than vocabulary

### Sketching privacy requires many iterations, and participants have difficulty planning the space in advance.

A simple data practice can involve many low-level privacy-related design decisions, and many of them are interdependent. Participants often complained that the compact positioning (e.g., Figure 2d) makes it hard to “*read through the sketch*” (N11). To make the problem worse, most participants need to iterate multiple rounds to have a decent design since it is hard to have a clear big picture of the privacy design before starting to sketch and think about the data practice. Participants reported that it is hard to “*make targeted edits*” (N2) and complained that sketching on paper lacked the flexibility to drag-and-drop for content rearrangement. In the end, participants either re-started a sketch to “*re-plan the layout*” (N6), squeezed updates into “*any available blank space*” (N2), or even “*abandoned the idea to iterate*” (N4). These challenges highlighted the need for better guidance and prompted us to explore teaching methods using digital tools (e.g., PC or iPad) to edit their already sketched content.

**The limits of the sketches mean the limits of privacy designs.** We encouraged participants to sketch while exploring the design space, but an unintended consequence emerged: they often stopped considering new design possibilities once they ran out of paper space. As a result, many sketches included surprisingly detailed components but overlooked the broader context of data practices. In the wrap-up interview, participants realized they “*should include them but forgot*” (N1). Others found it physically demanding to “*resketch the components*” (N2) already created in the previous steps. They expressed a desire for a “*template*” (N8) to guide them on how to add details to their designs and allow for “*copy-paste*” (N2) to reuse.

## 4. Teaching Methods Experimentation

**Design Goal.** Our goal is to develop a method that enables data science students to create high-quality privacy sketches with minimal training. A high-quality sketch should meet two criteria: (1) it includes sufficient detail while covering the key components of the privacy design, and (2) it is clear and easy to interpret, allowing the audience to quickly understand the privacy concepts presented.

**Experimentation Method.** We employed an interactive pro-

totyping method [7], [8] to guide the development of the teaching method. Interactive prototyping is a design method where users engage with evolving, often low-fidelity prototypes—such as sketches, paper interfaces, or mockups—in realistic scenarios while designers simulate system behavior, observe user interactions, offer feedback, and iteratively refine the design before the actual system is built. We iteratively prototyped a teaching method, tested it with a new group of users, gathered feedback, and refined the method. Since previous research found that 80% of usability problems are detected with four or five subjects and the most severe problems are likely to have been detected in the first few subjects [58], we experimented with each method using a relatively small participant sample (5-7 for each iteration).

Table 3 summarizes the process and the insights we obtained through the process. In this section, we will describe the three methods we explored, our observations, and the trade-offs associated with each. We recruited all participants in this experiment through the same approach as the need-finding study in Section 3. To maintain the integrity and independence of our findings, we also ensured there was no participant overlap across any of our studies, including future ones. Each participant was compensated for their time with a \$10 gift card.

### 4.1. Object-Oriented Diagramming

Our first attempt, Object-Oriented Diagramming (OOD), is inspired by the Object-Oriented Programming (OOP) principle in software engineering [59], [60].

**4.1.1. Design.** OOD has two key design ideas. First, we organized the privacy sketch based on the concept of objects, which can contain properties and methods (Figure 3) and support inheritance, similar to objects in OOP. We defined two base classes, `DataAction` and `Stakeholder`, along with their attributes. `DataAction` has four sub-classes (`Collect`, `Store`, `Process`, and `Access`). These sub-classes share two common attributes, *data* and *device*, which indicate the direction of data flow (e.g., a photo moving from a camera to a server). Each subclass also has its specific attributes. Figure 3a presents the other class `Stakeholder`, which includes a name attribute for free text input. It consists of two methods: `Involve` and `Decide`. `Involve`



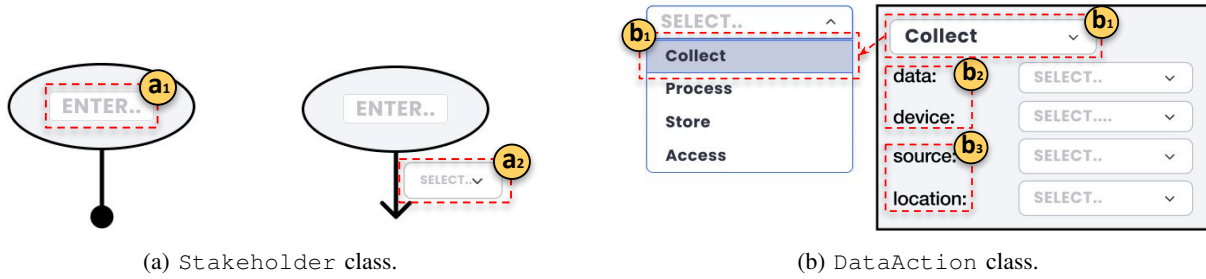


Figure 3: Object-Oriented Diagramming (OOD) is inspired by the Object-Oriented Programming principle in software engineering. Each basic element in OOD has methods and properties. For example, the *Stakeholder* class includes a  $a_1$  name attribute for free input. Its *Involve* method (shown on the left) represents the stakeholder’s involvement, while the *Decide* method (right) includes a  $a_2$  binary parameter, *choice*, to indicate whether a stakeholder enables or disables a *DataAction*. The *DataAction* class first requires specifying a  $b_1$  specific subclass. After this, students can select from preset options for both  $b_2$  inherited and  $b_3$  subclass-specific attributes.

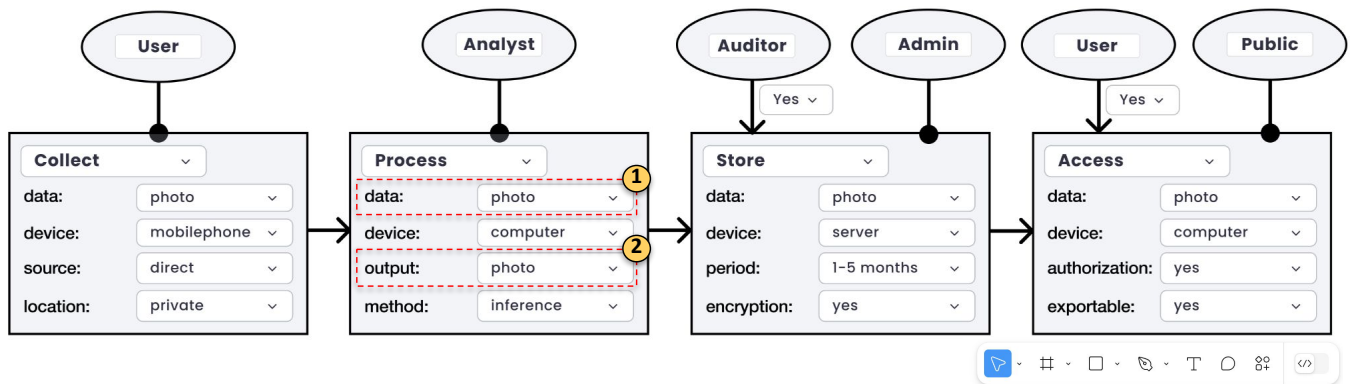


Figure 4: When conducting Object-Oriented Diagramming (on Figma), participants found the reusable components helpful but experienced significant cognitive load due to the tool’s learning curve and limited design options. For instance, in participant T2’s design of the *Sensitive Image Detection* scenario, she wanted to specify ① input and ② output data as “raw photo” and “labeled photo” for the *Process* object, but the closest option available was “photo,” making it challenging to communicate the design to others.

denotes a stakeholder performing a *DataAction* (e.g., an admin storing data), while *Decide* represents making a *choice* (a binary parameter of the method, such as giving consent on data collection or revoking access).

Second, we offered preset attribute options to help novice students explore the design space systematically (Figure 3 and 4). Instead of describing each *DataAction* in natural language, students could select predefined attribute options. For example, when specifying data *source* to *Collect*, students could choose “direct” for newly generated data or “secondary” for pre-existing data.






**4.1.2. Tool Support.** We implemented our solution using Figma [61], a web application for UI design. This diagrammatic design employs three primary visual elements (Figure 3): **rectangle**, **ellipse**, and **edge**. Rectangles prompt students to choose from four predefined subclasses (*Collect*, *Store*, *Process*, and *Access*) with preset attribute options. Ellipses allow students to specify the stakeholder’s role with custom text. Edges depict relationships: a horizontal arrow between two rectangles indicates a sequential

process, while a vertical connection from an ellipse to a rectangle represents either the stakeholder’s involvement (line) or a choice (arrow) that impacts the data flow.

**4.1.3. Experiment.** We first tested this approach with two undergraduate and two graduate data science students. We aimed to compare the Object-Oriented Diagrams’ usability and users’ cognitive load with the hand-sketch method used in the pilot study (Section 3). Each participant created a privacy design using each approach, and we counterbalanced the order of approaches and randomly assigned two scenarios to each participant (Table 2).

Because the prototype is low-fidelity—for example, it does not comprehensively enumerate all properties and methods for each object—we encouraged participants to interact with the experimenters to ask questions and clarify any uncertainties during the session. We observed that all four students experienced significant cognitive burdens when using OOD, which may have been partly due to the limited fidelity of our prototype. To address this, we recruited two additional privacy experts through personal

TABLE 4: In Vocabulary-Based Sketching, we incorporate visual components and use free text descriptions to replace the previous attribute structure. For testing, we presented participants with this worksheet, including each component’s name, symbol, and examples. Additionally, we specify the source and target nodes applicable to each edge component.

Vocabulary	Symbol	(Source, Target)	Examples
Stakeholder		—	User; admin; data scientist...
Data Action (on a Device)		—	Developer captures photos on a (camera); engineer saves user profiles on the (server); data scientist analyzes user activities on their (computer); admin views logs on an (internal workstation)...
Involvement in the Data Flow		(Stakeholder, Data Action)	See examples above (i.e., performing the action indicates the stakeholder is involved in the data flow)
Choice on the Data Flow		(Stakeholder, Data Action)	User provides consent for data collection; manager authorizes AI modeling with user data...
Step Procedure between Data Actions		(Data Action, Data Action)	After collecting the data, proceed to save it...

connections to compensate for the limited fidelity—a common practice in interactive prototyping [7].

**4.1.4. Observations.** All participants completed their hand sketches (Mean = 25.4 minutes, SD = 4.3 minutes), and our observations aligned with from the pilot study (Section 3). However, participants experienced a significant cognitive burden when using the diagramming approach. Only one privacy expert completed the privacy design in 23.1 minutes, whereas the remaining five participants required more time, ranging from 35 minutes to over an hour. Most participants felt their privacy designs were incomplete and eventually reached a point where they couldn’t make further progress. Four participants chose to terminate the task without satisfaction after it extended beyond an hour, explaining that they “became lost” (T3) and were “unsure how to improve the design” (T5), for several reasons:

Participants expressed the need for more time to become familiar with the web interface and the Object-Oriented concepts. For example, one user mentioned spending a significant amount of time “navigating each component and its attributes” (T1) to understand how it fits into their design, which they “wouldn’t need to do with a sketch” (T4).

Next, participants raised concerns about the **constrained design options** provided to them. For instance, all participants found the preset options of attributes limiting, including T2, who completed her diagram (Figure 4). “I wanted to indicate the input as ‘original photo’ and output as ‘labeled photo with detection results’ in the processing step. However, the only option for me was ‘photo,’ which is hard to infer if someone was reading my design.” Others also expressed that they prefer to “use plain text to describe” (T1) their designs. T5 complained, “This rigid structure made me feel like I was designing within a box.”

On the other hand, participants appreciated the **clarity of visual shapes** and the **efficiency offered by reusable components**. Many noted that preset elements significantly streamline the sketching process, as they found such elements “impractical for on-paper sketching” (T3). Participants also expressed a strong desire to use tablets for editing efficiency. “It would be more natural to work on an iPad - I could easily duplicate and reuse components” (T2).

## 4.2. Vocabulary-Based Sketching

Building on the previous observations, we refined our teaching method in design and tool support.

**4.2.1. Design. A Vocabulary of Privacy Design Components.** We relaxed the hierarchy requirement of the object-oriented design and simplified the vocabulary set (Table 4). For example, we eliminated the subclasses of `DataAction` and defined a `Data Action` node represented by a rectangle. Specifically, as one of the common components in the pilot study (Section 3), `<DEVICE>` was designed as a free text handwriting input to specify where the data action occurs. Additionally, the two methods (Figure 3a) of the `Stakeholder` class and the process between data actions were represented as three types of edges connecting nodes.

**Free Text Descriptions.** To address the concern of constrained design space, we removed the design of predefined attributes. Instead, we allowed participants to use free text to describe privacy-related details.

**4.2.2. Tool Support.** As our last experiment suggested, rather than a web-based interface, users preferred hand-sketched tools with copy-paste functionality, where they could reuse the created components. We adapted our teaching method to a tablet-based sketch board. For example, students could use an iPad with a stylus to create their privacy designs on a sketch board app such as Notability [62].

**4.2.3. Experiment.** We tested this approach with three undergraduate and two graduate data science students. Due to frequent difficulties in completing the diagrams (Section 4.1.4), we did not ask them to test the Object-Oriented Diagramming in this round. We also asked each participant to sketch only one randomly selected scenario (Table 2). To support participants’ learning process, we provided a worksheet (Table 4) with vocabulary, symbols, and usage examples. We guided them through its content before they started the design tasks and encouraged them to add free-text privacy details. For example, they may specify whether encryption was applied when saving user data.

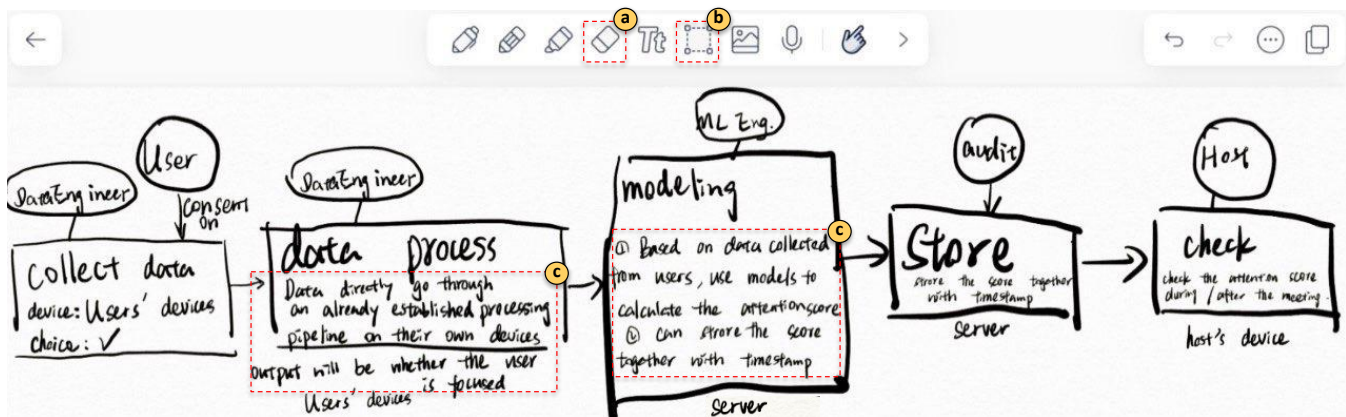


Figure 5: During Vocabulary-Based Sketching, participants could use sketchboard-supported features like the (a) eraser and (b) selection tools (for copy-pasting and resizing) to aid their creations. However, they often became overly focused on visual symbols and were distracted by the finer details in their designs. For instance, in T11’s design for the *Online Meeting Attention Tracking* scenario, she included (c) extensive details within each data action component, which made it challenging for her to maintain an overview of the entire design.

**4.2.4. Observations.** All participants finished their privacy design sketches (Mean = 18.6 minutes, SD = 2.7 minutes) and took advantage of the sketch board’s editing features to erase, copy-paste, drag, or resize their created content (Figure 5). However, they raised several concerns about this vocabulary-guided approach.

First, participants reported **focusing too much on the visual components**, which they felt was “*not very intuitive*” (T7). They expressed needing to “*constantly stop and think about how to use each element*” (T10). Four participants wanted **flexible guidance**, with T8 saying, “*Just give me a simple guideline, like including the stakeholder’s engagement, so I can focus on the design without worrying about using rectangles or ellipses.*” (T8).

Second, participants struggled to **balance** detailed privacy considerations while maintaining a design overview (Figure 5). For instance, T10 noted, “*When I was deciding how long data should be retained on the server, I lost track of the next data action I wanted to sketch.*” Three participants desired an “*extra page*” to capture such details, with T9 suggesting, “*An extra page for a Data Action node could include data retention periods, so I could simply edit that page if I wanted to update it from 1 month to 1 year.*”

### 4.3. Heuristic-Based Sketching

Based on the previous findings and teaching method designs (Section 3 and Table 3), we moved toward a more lightweight solution that explores heuristics as the main approach to teach students to sketch privacy designs.

**4.3.1. Design & Tool Support.** We further simplified our method by removing constraints on vocabulary and visual representations, allowing participants to focus on the content of the design. This modification led to three heuristics (Table 1) for sketching privacy designs.

**Heuristic①: Device-Based Data Flow.** As one of the com-

monly used components in the pilot study (Section 3), the device enables students to outline their designs in a modular manner. Students can articulate data movement between devices as data actions (another common component) progress in their privacy designs. For instance, a student might sketch data flowing from a smartphone to a cloud server and then to a third-party server. This device-based communication could support further reflection on the design [63], such as evaluating risks associated with cloud storage or potential exposure when data is shared with third-party servers.

**Heuristic②: Stakeholder Interactions with Data Flow.** This heuristic encourages students to approach their design from a role-based perspective, incorporating each party’s roles in the privacy design. Here, “interaction” could include any privacy-related behavior of stakeholders, such as their involvement and decisions about the data flow. For example, a student might sketch an admin accessing user data, which is then shared with a data scientist for further analysis. By articulating stakeholders within their design, designers and their collaborators could trace accountability [64], ensuring stakeholders are held responsible for unauthorized access or data use beyond its intended purpose.

**Heuristic③: Multi-Layered Representation.** Inspired by DENIM [65], a multi-layered approach to website design that supports quick updates and helps maintain awareness of the overall structure, this heuristic encourages students to work with both high-level and detailed perspectives in their privacy designs. By using this multi-layered approach, students can plan their privacy design before diving into specifics and refer back to the overview to maintain a clear sense of direction throughout their design process.

**4.3.2. Experiment.** We tested the Heuristic-Based Sketching against Vocabulary-Based Sketching with four undergraduate and three graduate data science students. Each participant sketched privacy designs using both approaches,



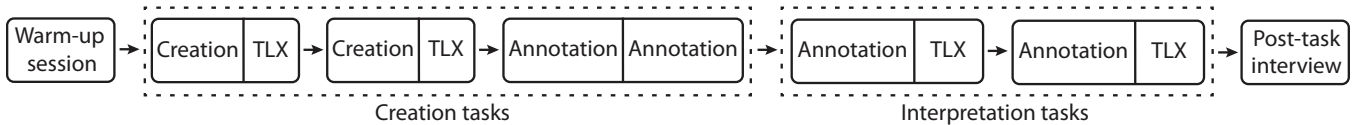


Figure 6: Each evaluation study included a warm-up session, two creation tasks (followed by an annotation process of privacy-related design decisions), two interpretation tasks, and a post-task interview. During annotation, participants also orally explained the content of each design decision. They completed a NASA TLX survey after each task.

and we counterbalanced the order of approaches and two scenarios presented to them (*Afterlife Chatbot* and *Financial Risk Management*, see Table 2). We chose these two scenarios because previous participants found these topics interesting, which enhances participants’ engagement.

To support their learning process, we included a warm-up session before each task based on feedback from previous participants (Section 4.2.4). We introduced an example scenario (*Online Meeting Attention Tracking*, see Table 2) and guided them through the corresponding worksheet (Table 1 or 4). Participants then sketched the example scenario and asked any questions only during warm-up.

**4.3.3. Observations.** All participants completed the warm-up within 20 minutes (Mean = 18.5 minutes, SD = 1.2 minutes) and reached satisfaction with each design within 15 minutes (Mean = 13.6 minutes, SD = 2.9 minutes). First, we observed a significant **reduction in completion time** with the vocabulary-based approach (14.3 vs. 18.6 minutes in Section 4.2.4), as participants reported that the warm-up session helped them become familiar with the method. Additionally, there was a clear **preference for heuristics over vocabulary-based design**. For example, participants who started with heuristic-guided sketching often intentionally applied the multi-layered heuristic when sketching vocabulary-based designs. However, we did not observe a tendency to re-apply visual vocabulary to heuristic-based tasks. T13 noted, “*I felt freer to move beyond ellipses and boxes and to organize my ideas in a more trackable way.*”

## 5. Evaluation

This section presents a detailed experimental evaluation of the heuristic-based approach. Our results indicate that this approach helps data science students create high-quality privacy sketches quickly with reduced mental effort. Participants also reported that the guided privacy sketches were more readable than the baseline sketches.

### 5.1. Study Overview

**Participants.** We recruited 24 participants in the evaluation through the same approach as previous studies (Sections 3, 4). Our final group included 10 undergraduates and 14 graduate students. In the pre-screening privacy knowledge survey (same as Section 3), 14 participants answered “No” to the privacy knowledge question; six gave responses unrelated

to privacy, and only four provided relevant privacy design terms but lacked experience in privacy design.

Among our participants, 10 (41.7%) identified as female, and others identified as male. Twenty (83.3%) participants were aged between 18 and 24, while others were between 25 and 30 (Mean = 22.0 years, SD = 2.3 years). Each participant received a US \$15 Amazon gift card as compensation.

**Study Procedure and Apparatus.** We followed a between-subjects study design. Participants received either vocabulary-based or heuristic-based guidance for their tasks. They completed tasks on an iPad with a stylus using a sketch board application (e.g., Notability [62]).

Figure 6 illustrates the study procedure. Each study began with a warm-up session (20-minute limit) that included an explanation of the worksheet (Table 1 or 4) and sketching privacy design for a sample scenario *Online Meeting Attention Tracking* (Table 2). Researchers observed the process and answered questions as needed.

Then, the study continued with two creation tasks in which participants sketched designs for two scenarios (i.e., *Afterlife Chatbot* and *Financial Risk Management*, as shown in Table 2) and annotated their privacy design decisions in the sketches. The participants were randomly assigned to a scenario (Table 5), and upon completion, they would complete another. Each task has a 15-minute time limit. These time limits were decided based on the task completion time in the last experiment (Section 4.3.3).

Then, each participant was asked to interpret two sketches created by previous participants (Table 9). Finally, we interviewed participants about what hindered or assisted them in their creation and interpretation tasks.

We took notes during each study session and summarized the key insights after each session. We observed saturation (i.e., no new insights emerged) [56] after the twentieth study. We then stopped participant recruitment and concluded the evaluation study with four more participants [57]. On average, the study session lasted 90 minutes.

During analysis, we performed Mann-Whitney U tests [66] to compare different scenario orders within each condition and found no evidence of ordering bias across all measures reported in Sections 5.2.3 and 5.3.3. For example, within the vocabulary group, we assessed the first creation task’s mental demand scores between participants who began with scenario F and those who began with scenario A.

In the following sections, we will denote the Vocabulary-Based Sketching as “Vocabulary” and the Heuristic-Based Sketching as “Heuristic.”

**Annotating Visual Sketches.** Comparing high-dimensional

TABLE 5: Creation task schedule. We counterbalanced the presentation order of the task scenarios (F: *Financial Risk Management*; A: *Afterlife Chatbot*) for both conditions (Vocabulary-guided and Heuristic-guided sketching).

Condition	Scenario		Creator					
	1st	2nd						
Heuristic	F	A	P1	P3	P5	P13	P15	P17
	A	F	P7	P9	P11	P19	P21	P23
Vocabulary	F	A	P2	P4	P6	P14	P16	P18
	A	F	P8	P10	P12	P20	P22	P24

sketches is challenging. To address this, we developed a codebook (Appendix Table 11) to annotate privacy-related design decisions covered in the sketches. We then used the annotated decisions to assess the design coverage in sketch creation and the communication effectiveness between the sketch creators and interpreters (Sections 5.2, 5.3).

Since participants often cannot broadly explore the design space (Section 3), we cannot derive the potential design decisions entirely based on empirical observations. Instead, we began by collecting privacy-related design decisions from prior literature [33], [67], [68], [69], [70], [71] to systematically capture the unique design decisions reflected in the sketches. Two researchers then collaboratively created an initial codebook for analysis. They then independently applied the codebook to annotate the sketches from previous studies (Sections 3, 4) and discussed potential modifications, deletions, and extensions to the coding scheme. After refining the scheme, all these previous sketches were re-coded using the updated framework. Given the subjective nature of interpreting sketches, we considered only codes independently validated by both researchers to ensure reliability.

## 5.2. Creation Tasks

**5.2.1. Data Collection.** We counterbalanced the presentation order of scenarios (Table 5). In each task, we asked participants to read the scenario description (Table 2) and sketch a privacy design. They could inform us if they felt satisfied with their sketch, i.e., having included as many privacy-related details as possible, before the time was up. We recorded their completion time and asked them to fill out a NASA TLX survey [72] after each creation.

We presented participants with a list of privacy-related design decisions (Table 11) and asked them to describe each decision they had included, annotating the corresponding content on their sketches. We didn’t set time limits for this process. Participants could skip any decisions they felt were not covered but could not modify their original designs.

**5.2.2. Data Analysis.** There is no perfect sketch for the creation task, as participants may have different designs for the same scenario, which may evolve during sketching. So, we use the codebook (Table 11) to systematically capture the unique design decisions reflected in each sketch. To assess the quality of the sketches, we documented the privacy design decisions annotated by participants, counted the **total number** of decisions included in each sketch, and calculated

TABLE 6: In creation tasks, heuristic-guided participants spent less time sketching until satisfaction and included more privacy-related design decisions in their sketches. Format: mean  $\pm$  standard deviation. We also highlight the higher value between two conditions.

Task Order	Vocabulary		Heuristic	
	Time (min)	# Decisions	Time (min)	# Decisions
First	14.89 $\pm$ 2.86	8.4 $\pm$ 3.0	<b>12.70</b> $\pm$ 0.86	<b>10.7</b> $\pm$ 2.8
Second	14.70 $\pm$ 3.71	9.3 $\pm$ 3.1	<b>10.36</b> $\pm$ 0.98	<b>12.3</b> $\pm$ 1.8
All	14.80 $\pm$ 3.23	8.9 $\pm$ 3.0	<b>11.53</b> $\pm$ 0.94	<b>11.5</b> $\pm$ 2.4

the **coverage percentage** of each design decision across all sketches in each condition.

Next, we conducted a thematic analysis [73] of post-task interview transcripts and notes we took during the study. First, two researchers independently coded eight transcripts (four from the heuristic group and four from the vocabulary group). After discussing and incorporating codes, we created a codebook that the two researchers agreed on. Using this codebook, the researchers divided the remaining transcripts, each coding eight transcripts per condition. Like many other qualitative analyses in S&P research [74], [75], the two researchers discussed and resolved coding conflicts in several weekly meetings. In this analysis, as we prioritized identifying emerging themes, we did not calculate the inter-rater reliability (IRR) to seek theoretical agreement [76].

**5.2.3. Quantitative Results.** We found that heuristics could help students sketch privacy designs more effectively. Their designs also demonstrated higher quality.

**Improved Sketching Efficiency.** As shown in Table 6, heuristic-guided participants reached their satisfied designs faster (average of 11.53 vs. 14.80 minutes) and covered more privacy-related design decisions (average of 11.5 vs. 8.9 per sketch), compared to the vocabulary group.

Besides, as they progressed to the second creation task, the heuristic-guided participants reported more positive responses to the NASA TLX questions (Table 7) than the first task. These improvements included a significantly reduced mental workload shown by the Wilcoxon Signed-Rank test [77] ( $p < 0.05, r = 0.70$ ), as well as decreases in physical, temporal, and effort demands. The heuristic group also noted increased self-perceived performance and reduced frustration, a trend not observed in the vocabulary group.

**Broader Coverage of Privacy-Related Design Decisions.** We then compared the coverage percentage of design decisions between conditions (Table 8). On average, each decision was included in 76.7% of the heuristic-guided sketches, compared to 59.2% of the vocabulary group. Specifically, heuristic group could more frequently cover decisions such as *Stored Data*, *Processing Input*, *Processing Approach*, *Accessed Output*, *Access Approach*, and *Choice Impacts*. This difference was statistically significant ( $p < 0.01, r = 0.69$ ) under the Wilcoxon Signed-Rank test [77].

**5.2.4. Qualitative Findings.** Based on the analysis of participants’ creation task behaviors and their feedback in post-

TABLE 7: In the second creation task, heuristic-guided participants reported more positive responses than they did in the first task, a trend not observed in the vocabulary group. We present NASA TLX results (scale of 1 to 5) as “median (mean  $\pm$  standard deviation),” and highlight second task responses if they are more positive than the first. “\*” indicates statistical significance ( $p < 0.05$ ) under Wilcoxon Signed-Rank test. “ $\downarrow$ ” denotes that a lower value is a more positive outcome.

Condition	Task Order	Mental Demand $\downarrow$	Physical Demand $\downarrow$	Temporal Demand $\downarrow$	Performance $\uparrow$	Effort $\downarrow$	Frustration $\downarrow$
Vocabulary	First	3.5 (3.17 $\pm$ 1.03)	4.0 (3.65 $\pm$ 1.23)	4.0 (3.33 $\pm$ 0.98)	3.0 (3.00 $\pm$ 0.74)	4.0 (3.83 $\pm$ 0.39)	2.0 (2.50 $\pm$ 0.90)
	Second	3.5 (3.33 $\pm$ 0.78)	4.0 (3.81 $\pm$ 1.31)	3.5 (3.17 $\pm$ 0.94)	3.0 (2.92 $\pm$ 1.08)	3.5 (3.50 $\pm$ 0.80)	3.0 (2.67 $\pm$ 0.89)
Heuristic	First	4.0 (3.67 $\pm$ 0.89)*	3.5 (3.48 $\pm$ 0.97)	4.0 (3.58 $\pm$ 0.79)	3.0 (2.92 $\pm$ 0.67)	3.5 (3.58 $\pm$ 0.67)	3.0 (3.00 $\pm$ 1.04)
	Second	2.5 (2.92 $\pm$ 1.24)*	3.0 (2.97 $\pm$ 0.82)	3.0 (3.00 $\pm$ 1.41)	3.5 (3.25 $\pm$ 0.87)	3.0 (3.17 $\pm$ 1.19)	2.0 (2.33 $\pm$ 1.07)

task interviews, we identified following main findings.

**Reduced Learning Curve.** The guidelines provide a starting point and ease the challenge of beginning a sketch from scratch. For example, P15 shared that “*The heuristic helps me to know what the design should roughly look like when I started to construct my ideas.*” P7 added that “*Including stakeholders prompts me to consider who is involved and who is taking action in data cases, which I believe is important when forming my ideas for sketches.*”

**Flexibility in Planning.** Participants appreciate the **Multi-layered Representation** heuristic for helping them in planning their sketches in a structured manner. P7 explained, “*The overview layer allows me to lay down the general ideas so I can worry about the details later.*” The device-based data flow and multi-layered approach also gave them a “*big-picture perspective*” (P3), which helped P5 to “*organize the system’s overall logic and focus on how data flows between different modules without getting distracted by other details.*” Participants also highlighted that the separate layers enable them to “*quickly update on details*” (P9) and “*easily to replicate existing content*” (P11). Furthermore, as they were familiar with the heuristics after the first design, participants in the experimental group were able to quickly grasp and re-apply the workflow to the second task.

**Ease of Cognitive Load.** After applying the heuristics, participants recognized that sketching was more manageable than anticipated, reducing their perceived effort and workload. P7 expressed that using the device annotation made “*each part more distinctive and save effort if I want to adding new ideas to my sketch.*” P19 initially worried about “*how to show all the privacy details*” but later found that “*the layered approach helped me break down complex ideas into simple parts.*” Despite the initial time pressure, P5 noted the benefit of sketching the overview layer first, explaining that it allows for “*managing time more effectively by following the outline I planned out from the start.*”

**Enhanced Coverage of Privacy Details.** The heuristics prompt participants to extend the scope of privacy details in their sketches, echoing the increased design decision coverage shown in Table 8. First, the **Stakeholder Interactions with Data Flow** heuristic encouraged the participants to “*consider human decisions and accountability with freedom*” in a data practice (P11), which is “*important for ensuring responsible use of data*” (P7). P23 explained that the heuristic enabled thinking “*beyond where the data travels in the system*” and encouraged consideration of “*what each*

*party is doing with the data.*”

Second, the **Multi-Layered Representation** heuristic encourages more detailed privacy designs. Participants noted that the separation of layers “*makes space for me to fill in details*” (P9) and that the flexibility to update individual layers enabled easy extension of their sketches, encouraging them “*to iterate more times without friction*” (P3).

As a result, compared to the vocabulary group, heuristic-guided sketches incorporated more interaction details beyond merely including “Choice” and “Involvement.” For example, as illustrated in Figure 1, these sketches featured interactive elements such as a chat window for access or a pop-up for making choices. This corresponded to the increased coverage (Table 8) of the three decisions related to *Choice & Notice—Choice Options, Choice Impacts, and Choice Notification*.

**Bias and Negative Consequence.** Participants noted certain limitations with the heuristics during the sketching process. P1 remarked that the Multi-layered Representation “*seems to be repetitive because I am filling similar information in both layers.*” P7 observed that creating a sketch on the given case was “*a systematic design*” and expressed a desire for a more “*complete structured framework to fill or follow.*”

### 5.3. Interpretation Tasks

**5.3.1. Data Collection.** As shown in Figure 6, each participant completed two interpretation tasks, each with a 7-minute time limit. In each task, they received a randomized list of privacy-related design decisions (excluding the “Procedure” column in Table 11) and an unannotated copy of a sketch created by a prior participant. We then asked them to annotate and orally describe the privacy-related design decisions they identified. This annotation is similar to what they have done during the creation tasks (Section 5.2.1). Table 9 outlines the interpretation task order, where each participant interpreted two sketches of different scenarios: one created by a heuristic-guided participant and another from the vocabulary group. We maintained scenario order consistent with the creation tasks (Table 5) to avoid biases from the most recent scenario they had sketched. For example, after sketching for *Afterlife Chatbot* as his second creation, P3 would first interpret a sketch of another scenario. After each task, we recorded their interpretation time and asked them to complete a NASA TLX survey [72].

**5.3.2. Data Analysis.** To assess whether heuristic-guided sketches could better facilitate communication of privacy

TABLE 8: Heuristic-guided sketches could include privacy design decisions more frequently. Additionally, the design decisions within the heuristic-guided sketches were interpreted more accurately than those within the vocabulary group. For each design decision, we report its coverage frequency (“Coverage”) across all sketches in each group. For communication, we report the frequency of being interpreted (“Interpret,” including misinterpretations), along with precision, recall, and F1-score, all measured at the design decision level. For each metric, we highlight the higher value and indicate the statistical significance between the two conditions using the Wilcoxon Signed-Rank tests ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ), all showing large effect sizes ( $r > 0.5$ ).

Design Decision	Vocabulary-Guided Sketches					Heuristic-Guided Sketches				
	Coverage	Interpret	Precision	Recall	F1	Coverage	Interpret	Precision	Recall	F1
Collected Personal Data	<b>100.0%</b>	86.4%	83.3%	92.1%	87.5%	95.8%	<b>87.0%</b>	<b>90.9%</b>	<b>100.0%</b>	<b>95.2%</b>
Data Provider	87.5%	95.5%	70.5%	73.8%	72.1%	<b>91.7%</b>	<b>95.7%</b>	<b>91.3%</b>	<b>95.5%</b>	<b>93.3%</b>
Collection Purpose	<b>87.5%</b>	<b>77.3%</b>	27.8%	29.4%	28.6%	83.3%	73.9%	<b>65.0%</b>	<b>76.5%</b>	<b>70.3%</b>
Stored Data	62.5%	77.3%	34.4%	32.4%	33.3%	<b>91.7%</b>	<b>78.3%</b>	<b>77.8%</b>	<b>77.8%</b>	<b>77.8%</b>
Storage Approach	79.2%	63.6%	33.3%	42.9%	37.5%	<b>91.7%</b>	<b>73.9%</b>	<b>78.9%</b>	<b>88.2%</b>	<b>83.3%</b>
Post-Storage Action	58.3%	31.8%	10.0%	21.4%	13.6%	<b>62.5%</b>	<b>60.9%</b>	<b>56.3%</b>	<b>64.3%</b>	<b>60.0%</b>
Processed Input	54.2%	<b>81.8%</b>	29.4%	27.8%	28.6%	<b>100.0%</b>	78.3%	<b>67.5%</b>	<b>75.0%</b>	<b>71.1%</b>
Processing Output	79.2%	<b>81.8%</b>	55.3%	58.3%	56.8%	<b>91.7%</b>	65.2%	<b>85.3%</b>	<b>96.7%</b>	<b>90.6%</b>
Processing Approach	45.8%	50.0%	29.2%	31.8%	30.4%	<b>83.3%</b>	<b>60.9%</b>	<b>66.7%</b>	<b>71.4%</b>	<b>69.0%</b>
Accessed Raw Data	<b>66.7%</b>	<b>54.5%</b>	21.4%	25.0%	23.1%	50.0%	52.2%	<b>59.1%</b>	<b>54.2%</b>	<b>56.5%</b>
Accessed Output	41.7%	<b>63.6%</b>	<b>58.3%</b>	50.0%	53.8%	<b>75.0%</b>	60.9%	52.8%	<b>67.9%</b>	<b>59.4%</b>
Access Approach	50.0%	59.1%	31.8%	26.9%	29.2%	<b>83.3%</b>	<b>69.6%</b>	<b>66.7%</b>	<b>62.5%</b>	<b>64.5%</b>
Choice Options	37.5%	50.0%	45.5%	45.5%	45.5%	<b>58.3%</b>	<b>52.2%</b>	<b>64.3%</b>	<b>75.0%</b>	<b>69.2%</b>
Choice Impacts	16.7%	<b>36.4%</b>	14.3%	12.5%	13.3%	<b>58.3%</b>	30.4%	<b>37.5%</b>	<b>64.3%</b>	<b>47.4%</b>
Choice Notification	20.8%	18.2%	16.7%	25.0%	20.0%	<b>33.3%</b>	<b>26.1%</b>	<b>75.0%</b>	<b>75.0%</b>	<b>75.0%</b>
Mean	59.2%	61.8%	37.4%	39.7%	38.2%	<b>76.7%</b>	<b>64.4%</b>	<b>69.0%</b>	<b>76.3%</b>	<b>72.2%</b>
<i>p</i> -value	**	( <i>n.s.</i> )	***	***	***	**	( <i>n.s.</i> )	***	***	***

TABLE 9: Interpretation task schedule. For instance, P3 first interpreted a sketch of *Financial Risk Management*, which was created by P1 (heuristic group), and then interpreted a sketch of another scenario by P2 (vocabulary group). The “-” means no interpretation task due to the lack of a prior sketch, and three sketches (one by P23 and two by P24) were not interpreted because of no subsequent interpreters.

Scenario Condition	F Heuris.	A Vocab.	Interpreter	Scenario Condition	A Vocab.	F Heuris.	Interpreter
	-	-	P1		P6	P5	P7
Creator	P1	P2	P3	Creator	P8	P7	P9
	P3	P4	P5		P10	P9	P11
	P11	P12	P13		P18	P17	P19
	P13	P14	P15		P20	P19	P21
	P15	P16	P17		P22	P21	P23
Scenario Condition	F Vocab.	A Heuris.	Interpreter	Scenario Condition	A Heuris.	F Vocab.	Interpreter
	-	P1	P2		P7	P6	P8
Creator	P2	P3	P4	Creator	P9	P8	P10
	P4	P5	P6		P11	P10	P12
	P12	P13	P14		P19	P18	P20
	P14	P15	P16		P21	P20	P22
	P16	P17	P18		P23	P22	P24

design, we measured the alignment between the creator’s and interpreter’s responses (including annotations and explanations) regarding each privacy-related design decision within a sketch. For each design decision, two researchers collaboratively compared both response versions and assigned a code to describe whether they aligned. They open-coded a subset of sketches (N=10, i.e., 15 codes per sketch  $\times$  10 sketches), conducted three rounds of discussions to reach a complete agreement, and generated an initial codebook. Then, two researchers coded the rest of the sketches independently. To ensure consistency, two coders discussed their codes regularly to reach an agreement and iteratively refined

the codebook (e.g., an initial code “only one person gave a response” was decoupled into two codes “only the creator responded” and “only interpreter responded”). Finally, two researchers coded all the sketches using the final codebook (Appendix Table 12), with a Cohen’s kappa [78] of 90.1%, which reflects an “almost perfect agreement” [79].

We then quantified the effectiveness of communication on two levels. First, to measure whether heuristic-guided sketches lead to high-quality interpretations for each privacy-related *design decision*, we used the creators’ responses as groundtruth and calculated *design decision-level precision* (i.e., the proportion of interpretations that matched the groundtruth) as well as **recall** (i.e., the proportion of groundtruth responses that were accurately interpreted).

Second, to determine whether sketching with heuristics aids in identifying correct privacy design decisions across an *entire sketch*, we computed *sketch-level precision* (the proportion of interpretations that matched the creator’s responses within each sketch) and **recall** (the proportion of design decisions covered by the creator that was accurately interpreted in each sketch). Appendix Table 13 provides further clarification of these metrics.

The qualitative analysis shared the same procedure as described in Section 5.2.2, and we will report the findings of interpretation tasks in Section 5.3.4.

**5.3.3. Quantitative Results.** We found that heuristic-guided sketches are easier to interpret and could better facilitate communication between creators and interpreters.

**Eased Workload of Interpretation.** All participants finished their interpretation within the time limit. There was no significant difference in interpretation time (4.9 minutes

TABLE 10: Participants perceived significantly lower workloads when interpreting heuristic-guided sketches than vocabulary-guided ones. We present their responses to NASA TLX survey as median (mean  $\pm$  standard deviation), with results for heuristic-guided sketches highlighted. We annotate statistically significant improvements based on the Wilcoxon Signed-Rank tests ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ), all with medium to near large effect sizes ( $r : 0.3 \sim 0.5$ ). “ $\downarrow$ ” indicates that a lower value is more positive.

Condition	Mental Demand $\downarrow$	Physical Demand $\downarrow$	Temporal Demand $\downarrow$	Performance $\uparrow$	Effort $\downarrow$	Frustration $\downarrow$
Vocabulary	3.0 (3.41 $\pm$ 1.10) $***$	3.5 (3.66 $\pm$ 0.92) $**$	3.0 (2.82 $\pm$ 1.18) $*$	3.0 (2.91 $\pm$ 1.15) $**$	3.5 (3.32 $\pm$ 1.04) $***$	3.0 (3.23 $\pm$ 1.07) $***$
Heuristic	<b>2.0 (2.29 <math>\pm</math> 0.86)<math>***</math></b>	<b>2.5 (2.61 <math>\pm</math> 0.88)<math>**</math></b>	<b>2.0 (2.17 <math>\pm</math> 0.87)<math>*</math></b>	<b>4.0 (3.75 <math>\pm</math> 0.79)<math>**</math></b>	<b>2.0 (2.25 <math>\pm</math> 0.85)<math>***</math></b>	<b>2.0 (2.04 <math>\pm</math> 0.81)<math>***</math></b>

for heuristic-guided sketches vs. 5.1 minutes for vocabulary-guided sketches). However, participants consistently reported significantly lower mental, physical, temporal, and effort-related demands and reduced frustration when working with heuristic-guided sketches. They also perceived improved performance, as indicated by Wilcoxon Signed-Rank tests (Table 10). Furthermore, participants identified more privacy design decisions in heuristic-guided sketches, with an average of 9.4 elements per sketch compared to 8.1 from vocabulary-guided ones. These results suggest that heuristic-guided sketches are more straightforward to interpret, decreasing workload while enhancing outcomes.

**Improved Communication Efficiency.** Table 8 presents the result of interpretation evaluation metrics at the privacy design decision level, which provides initial evidence for heuristic guidance’s capability to enable better creator-interpreter communication of privacy designs. First, heuristic-guided sketches achieved higher interpretation performance across metrics, with an increase of 31.6% in precision (i.e., increased fraction of correct interpretations) and 36.6% in recall (i.e., increased fraction of creator’s design decisions that were accurately interpreted).

At the level of a complete privacy design, heuristic-guided sketches demonstrated higher per-sketch interpretation performance, again with a higher precision (79.1% vs. 44.6% for the vocabulary group), reflecting the proportion of matches among all interpretations in a sketch, and higher recall (71.1% vs. 40.5%), representing the percentage of alignments among the creator’s all design decisions. These differences are all statistically significant according to the Wilcoxon Signed-Rank tests [77], highlighting the improved communication efficiency facilitated by heuristic guidance.

**5.3.4. Qualitative Findings.** We made the following findings based on participants’ behaviors and feedback about their interpretation tasks.

**Heuristics streamline the interpretation process.** Participants from both conditions found heuristic-guided sketches to be more structured and easier to understand, largely due to the enhanced readability provided by the **Multi-Layered Representation** heuristic. P17, a heuristic-guided participant, appreciated how the multi-layered design allowed him to quickly grasp the overall idea of the privacy design at first glance, noting that “[reading] this sketch felt quite similar to [creating] my own sketch work.”

Similarly, P10, who did not receive heuristic instruction, also found the sketch (Figure 7) easier to follow, explaining that “the overview layer explains a clear logic of the design,

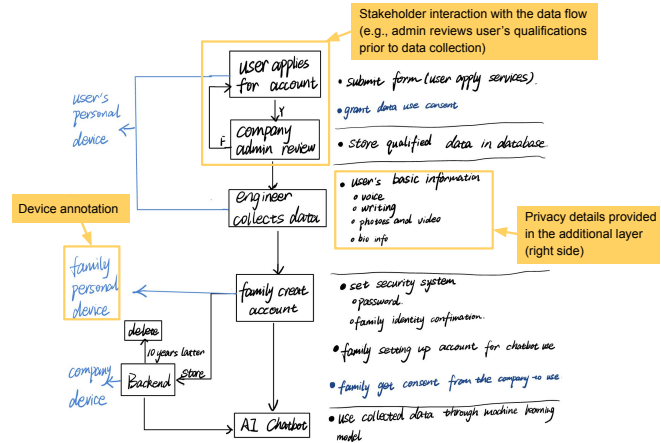


Figure 7: A heuristic-guided privacy design sketch for the Afterlife Chatbot scenario. It presents privacy details in parallel with the corresponding devices, stakeholders, and the steps taken, offering a clear logic for readers to follow.

such as where the privacy-sensitive procedure happens and who were involved in. However, the other [vocabulary-guided] sketch obviously falls short in this clarity.” She further expressed interest in “adopt[ing] this layered style in my future design sketches.” P4 also emphasized that separating the overview from details “enables a progressive manner of reading,” while P18 highlighted how “this multi-level sketch prevents my visual overload from digesting too much content in one place.”

**Heuristics offer varying content density.** Participants showed different preferences when retrieving information: some preferred more text for detailed understanding, while others favored less text for a smoother reading experience. The heuristic-guided sketches effectively accommodated both needs. P6 praised the multi-layered sketch for its support of effective information retrieval, saying, “first, I skim through the overview of the data flow to understand the big picture. If I want specific details, like the data involved in each step, I go to the next layer using the annotation labels.” P10 described the device box and overview layer as “pretty concise and simple,” while preferring more text for detailed understanding. P11 appreciated that the “detail layer has enough information for me to understand creator’s thought,” but felt that excessive text made interpretation time-consuming.



## 6. Discussion & Future Work

**Connections to Established Practices.** Professionals often draw on their existing skill sets when sketching for privacy. For instance, UX designers may use Customer Journey Maps, while software engineers rely on UML diagrams. We chose to develop a tailored method from scratch for two reasons. First, these skill adaptations only partially address key aspects of privacy design. For example, Customer Journey Maps focuses on stakeholder interactions. Second, most of our target audience—data science students—lacks formal training in both usability research and software engineering.

We have incorporated multiple relevant ideas in software engineering and UX research into our solutions, such as Data Flow Diagrams, Customer Journey Maps, and UX storyboards [80]. Future work may investigate whether our proposed heuristics can also enhance the practices of experienced professionals and how these heuristics complement their existing techniques.

**Why do heuristics work in sketching privacy?** Unlike rigid frameworks that limit creativity and restrict exploration [4], [46], the heuristic-based approach emphasizes flexibility and provides simple guidance, enabling students to deeply engage with privacy contexts while effectively organizing their ideas. Furthermore, the **Multi-layer Representation** heuristic helps students manage their sketches from a macro perspective, allowing them to approach designs strategically.

**Sketches vs. Diagrams.** We experimented with both sketches and diagrams in this project. We found that complex diagram structures hindered the learning process (Section 4.1.4). For instance, when students used diagrams to illustrate interactions between stakeholders and data flows, they needed a thorough understanding of diagram vocabulary. This steep learning curve increased their cognitive load, making it harder to complete the privacy design.

In contrast, sketches mitigated these challenges by enabling more intuitive idea generation (Section 5.2) and communication (Section 5.3). This aligns with prior research highlighting sketching as a flexible and intuitive tool for ideation and communication [5], [81]. However, sketches lack the structured detail of diagrams, which is crucial for tasks requiring precision and depth. While sketches reduce cognitive demands, they may limit students' ability to achieve the detailed understanding offered by structured diagrams. Future work could explore methods to help students effectively use diagrams to communicate privacy designs.

**Sketching for Privacy Literacy.** In this paper, our educational audience is data science students who are future data practitioners. One potential future direction is to push the technique to layperson [82], and explore if sketching privacy can help them form a more precise mental model [83] of the data practices.

**High-fidelity Privacy Prototypes.** Two students with UX/UI backgrounds noted that, although they were generally satisfied with their privacy designs, they felt uneasy about their inability to fully depict the visual aspects of

the product's interface. They also found that messy hand sketches contributed to difficulties in interpretation. Future work could explore higher-fidelity privacy prototypes, akin to high-fidelity UX prototypes.

**Sketches and Threat Modeling.** Our current study focuses on facilitating the communication of privacy design concepts through sketches. Future research may explore the benefits of sketches in other security and privacy applications. For example, participants may develop a stronger awareness of privacy risks during sketching [84], [85], [86], [87]. The improved communication enabled by privacy sketches may enhance the collaborative decision-making process [1]. Conventional threat modeling approaches [88], [89] often rely on static tables or Data Flow Diagrams (DFDs) to define technical scope and decompose applications. Compared to tables and DFDs, sketching may better capture dynamic aspects of a system, such as stakeholder interactions and contextual nuances that are difficult to convey textually.

## 7. Limitations

**Ground Truth for the Creation Task.** The creation task has no ground truth. While our codebook (Table 11) provided a structured approach for labeling and interpreting data, it may have constrained the granularity of our analysis. By relying on predefined categories, we may have overlooked subtle contextual differences or nuances in participant responses.

**Participant Pool Limitations.** Our participant pool consisted primarily of students from four-year U.S. universities, which introduced limitations in both sample diversity and size. In particular, some participants had prior exposure to privacy-related concepts, which may not reflect the perspectives of novice users who engage with privacy features with little to no foundational knowledge. Future research may expand recruitment to include participants from community colleges and technical bootcamps, capturing a broader range of educational backgrounds and user experiences.

## 8. Conclusion

This paper explores methods for teaching data science students to sketch privacy designs. Through a need-finding study (N=12), we identified three challenges students encounter when sketching privacy designs: (1) difficulty sketching a complete data flow, (2) omission of stakeholder interactions with the data flow, such as user consent, and (3) failure to make quick updates without disrupting the overall design. To address these challenges, we iteratively developed our teaching approach, ultimately leading to three heuristics: (1) Device Annotation, (2) Stakeholder Interaction with Data Flow, and (3) Multi-Layered Representation. Our between-subjects experiment (N=24) demonstrates the effectiveness of these heuristics in improving sketching quality and facilitating communication of privacy design. Future privacy education and privacy design toolkits could take our findings into consideration.

## 9. Research Ethics

This research received approval from the Institutional Review Board (IRB) of our institution. We conducted all studies using Zoom (approved by our institute) and used Zoom’s speech recognition feature to transcribe the audio. To enable the transcription, we had to turn on the audio recording. Participants were informed about Zoom’s transcription and recording features, and a pop-up notification appeared when these features were activated. After each interview, the research team checked the transcription using the recording, then immediately and permanently deleted the recording. As such, we did not store any audio recordings. In this process, we have removed all personal information (e.g., names) from the transcription, ensuring that none of the transcriptions are personally identifiable. Participants provided informed consent before the study and retained the right to withdraw at any stage. Throughout all phases of data collection (Sections 3, 4, 5), participants had the right to withdraw at any time.

## Acknowledgement

We thank the anonymous reviewers and the shepherd for their invaluable feedback and the participants for their kind involvement in our studies. This research is in part supported by a gift from Solix Technologies, the National Science Foundation CNS-2341187, CNS-2426397, a Google PSS Faculty Award, and a Meta Research Award.

## References

- [1] M. Degeling, C. Lentzsch, A. Nolte, T. Herrmann, and K.-U. Loser, “Privacy by socio-technical design: A collaborative approach for privacy friendly system design,” in *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2016, pp. 502–505.
- [2] P. Shetty, “Data privacy and risk management, collaboration is key on tackling privacy risks/issues,” *Journal of Artificial Intelligence & Cloud Computing*, vol. 224, pp. 2–4, 2023.
- [3] T. W. Li, A. Arya, and H. Jin, “Redesigning privacy with user feedback: The case of zoom attendee attention tracking,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–14.
- [4] N. Alhirabi, S. Beaumont, J. T. Llanos, D. Meedeniya, O. Rana, and C. Perera, “Parrot: Interactive privacy-aware internet of things application design tool,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–37, 2023.
- [5] J. A. Landay and B. A. Myers, “Interactive sketching for the early stages of user interface design,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 43–50.
- [6] B. Buxton, *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann, 2010.
- [7] D. Fitton, K. Cheverst, C. Kray, A. Dix, M. Rouncefield, and G. Saslis-Lagoudakis, “Rapid prototyping and user-centered design of interactive display-based systems,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 58–66, 2005.
- [8] R. Harte, L. Glynn, A. Rodríguez-Molinero, P. M. Baker, T. Scharf, L. R. Quinlan, G. ÓLaighin *et al.*, “A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology,” *JMIR human factors*, vol. 4, no. 1, p. e5443, 2017.
- [9] S. Greenberg, S. Carpendale, N. Marquardt, and B. Buxton, *Sketching user experiences: The workbook*. Elsevier, 2012.
- [10] M. Cherubini, G. Venolia, R. DeLine, and A. J. Ko, “Let’s go to the whiteboard: how and why software developers use drawings,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, pp. 557–566.
- [11] S. Branham, G. Golovchinsky, S. Carter, and J. T. Biehl, “Let’s go from the whiteboard: supporting transitions in work through whiteboard capture and reuse,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 75–84.
- [12] J. A. Landay, “Silk: sketching interfaces like crazy,” in *Conference companion on Human factors in computing systems*, 1996, pp. 398–399.
- [13] J. A. Landay and B. A. Myers, “Sketching interfaces: Toward more human interface design,” *Computer*, vol. 34, no. 3, pp. 56–64, 2001.
- [14] Q. Chen, J. Grundy, and J. Hosking, “An e-whiteboard application to support early design-stage sketching of uml diagrams,” in *IEEE Symposium on Human Centric Computing Languages and Environments, 2003. Proceedings. 2003*. IEEE, 2003, pp. 219–226.
- [15] M. Lewis, M. Sturdee, and N. Marquardt, “Sketching in hci: Hands-on course of sketching techniques,” in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–5.
- [16] TeachPrivacy, “Training Privacy and Data Protection by Design,” <https://teachprivacy.com/training-privacy-data-protection-by-design/>, 2024.
- [17] Carnegie Mellon University, “17-334: Usable Privacy and Security,” <https://cups.cs.cmu.edu/courses/ups.html>, 2024.
- [18] UC Berkeley, “Cyber 215: Usable Privacy and Security,” <https://www.ischool.berkeley.edu/courses/cyber/215>, 2024.
- [19] Y. Tian, S.-Y. Chu, Y. Liu, and H. Jin, “Panopticon: The design and evaluation of a game that teaches data science students designing privacy,” *Proceedings on Privacy Enhancing Technologies*, 2025.
- [20] University of Washington, “ECE P 595: Introduction to Privacy Engineering,” <https://peden.ece.uw.edu/pmp/wp-content/uploads/sites/2/2022/04/SPR22-Introduction-to-Privacy-Engineering-Bonaci.pdf>, 2020.
- [21] Univ. of Washington, “CSE 564: Computer Security and Privacy,” <https://courses.cs.washington.edu/courses/cse564/24wi/syllabus/>, 2024.
- [22] E. McCallister, T. Grance, and K. Scarfone, “Guide to protecting the confidentiality of personally identifiable information (pii),” *NIST Special Publication*, vol. 800, p. 122, 2010.
- [23] IAPP (International Association of Privacy Professionals), “CIPT Certification,” <https://iapp.org/certify/cipt/>, 2024.
- [24] Harvard University, “Data Privacy and Technology,” [https://www.harvardonline.harvard.edu/course/data-privacy-technology?utm\\_medium=referral&utm\\_source=pll&utm\\_campaign=DPAT&utm\\_content=pll-dpat-page](https://www.harvardonline.harvard.edu/course/data-privacy-technology?utm_medium=referral&utm_source=pll&utm_campaign=DPAT&utm_content=pll-dpat-page), 2024.
- [25] “How Ray Kurzweil and His Daughter Brought A Relative Back From The Dead,” <https://www.pcmag.com/articles/how-ray-kurzweil-and-his-daughter-brought-a-relative-back-from-the-dead>.
- [26] “Zoom Can Track Who’s Not Paying Attention In Your Video Call. Here’s How. | HuffPost Life,” [https://www.huffpost.com/entry/zoom-tracks-not-paying-attention-video-call\\_1\\_5e7b96b5c5b6b7d80959ea96](https://www.huffpost.com/entry/zoom-tracks-not-paying-attention-video-call_1_5e7b96b5c5b6b7d80959ea96), 2020, (Accessed: 2024-08-22).
- [27] D. Nelson, “Chatroulette Returns With the Help of AI-Driven Content Moderation,” <https://www.unite.ai/chatroulette-returns-with-the-help-of-ai-driven-content-moderation/>, Jan. 2021.

- [28] “Cheating husband sues Apple after wife discovered ‘deleted’ messages sent to sex workers,” <https://www.telegraph.co.uk/news/2024/06/13/cheating-husband-sues-apple-sex-messages/>, Jun. 2024.
- [29] J. Albright, “The Graph API: Key Points in the Facebook and Cambridge Analytica Debacle,” Mar. 2018.
- [30] “Glow Pregnancy App Exposed Women to Privacy Threats, Consumer Reports Finds,” <https://www.consumerreports.org/electronics-computers/mobile-security-software/glow-pregnancy-app-exposed-women-to-privacy-threats-a1100919965/>, Sep. 2020.
- [31] “South Korean Telecom Company Attacks Customers with Malware — over 600,000 Torrent Users Report Missing Files, Strange Folders, and Disabled PCs | Tom’s Hardware,” <https://www.tomshardware.com/tech-industry/cyber-security/south-korean-telecom-company-attacks-torrent-users-with-malware-over-600000-people-report-missing-files-strange-folders-and-disabled-pcs>.
- [32] “Uber: Users Are More Likely To Pay Surge Pricing If Their Phone Battery Is Low,” <https://www.forbes.com/sites/amitchowdhry/2016/05/25/uber-low-battery/>.
- [33] H. Jin, H. Shen, M. Jain, S. Kumar, and J. I. Hong, “Lean privacy review: Collecting users’ privacy concerns of data practices at a low cost,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 28, no. 5, pp. 1–55, 2021.
- [34] A. Woodruff, V. Pihur, S. Consolvo, L. Brandimarte, and A. Acquisti, “Would a Privacy Fundamentalist Sell Their DNA for \$1000... if Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences,” in *10th Symposium on Usable Privacy and Security (SOUPS 2014)*, 2014, pp. 1–18.
- [35] L. Wu, S. Shah, S. Choi, M. Tiwari, and C. Posse, “The browsemap: Collaborative filtering at linkedin.” *RWeb@ RecSys*, vol. 1271, 2014.
- [36] “Set up iCloud for Messages on all your devices,” <https://support.apple.com/guide/icloud/set-up-messages-mm0de0d4528d/icloud>.
- [37] “Re;memory - Immortality AI,” <https://www.deepbrain.io/rememory>, 2022, (Accessed: 2024-08-22).
- [38] “HereAfter AI — Interactive Memory App — Try Free,” <https://hereafter.ai/>, 2022, (Accessed: 2024-08-22).
- [39] Apple, “Csam Detection- Technical Summary,” [https://www.apple.com/child-safety/pdf/CSAM\\_Detection\\_Technical\\_Summary.pdf](https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf), 2021, (Accessed: 2024-07-08).
- [40] H2O.ai, “Mobile Transaction Forecasting and Anomaly Detection,” <https://h2o.ai/case-studies/mobile-transaction-forecasting-and-anomaly-detection/>, 2024.
- [41] “Complete Guide to Data Anomaly Detection in Financial Transactions,” <https://www.highradius.com/resources/Blog/transaction-data-anomaly-detection/>, 2024, (Accessed: 2024-08-22).
- [42] Investopedia, “Credit Risk: Definition, Role of Ratings, and Examples,” <https://www.investopedia.com/terms/c/creditrisk.asp>, 2024, (Accessed: 2024-08-22).
- [43] “CoStar Group delivers efficient Content Moderation and Image Processing for Commercial Real Estate with AWS | CoStar Video | AWS,” <https://aws.amazon.com/solutions/case-studies/costar-video-case-study/>.
- [44] S. Brooks, M. Garcia, N. Lefkowitz, S. Lightman, and E. Nadeau, “An introduction to privacy engineering and risk management in federal systems,” National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. NIST IR 8062, Jan. 2017.
- [45] H. Jin, B. Guo, R. Roychoudhury, Y. Yao, S. Kumar, Y. Agarwal, and J. I. Hong, “Exploring the needs of users for supporting privacy-protective behaviors in smart homes,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–19.
- [46] Lucidchart, “How to Make a Data Flow Diagram,” <https://www.lucidchart.com/pages/data-flow-diagram/how-to-make-a-dfd>, 2024.
- [47] Francesco Mandrelli, “UML Use Case Diagram,” <https://www.figma.com/community/file/986330591099819762/uml-use-case-diagram>, 2021, (Accessed: 2024-08-22).
- [48] M. Arciniegas-Mendez, A. Zagalsky, M.-A. Storey, and A. F. Hadwin, “Using the model of regulation to understand software development collaboration practices and tool support,” in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2017, pp. 1049–1065.
- [49] S. Komanduri, R. Shay, G. Norcie, and B. Ur, “Adchoices-compliance with online behavioral advertising notice and choice requirements,” *ISJLP*, vol. 7, p. 603, 2011.
- [50] O. Ayalon, E. Toch, I. Hadar, and M. Birnhack, “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*, 2017, pp. 135–138.
- [51] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor, “What matters to users? factors that affect users’ willingness to share information with online advertisers,” in *Proceedings of the ninth symposium on usable privacy and security*, 2013, pp. 1–12.
- [52] H. Jin, G. Liu, D. Hwang, S. Kumar, Y. Agarwal, and J. I. Hong, “Peekaboo: A hub-based approach to enable transparency in data processing within smart homes,” in *2022 IEEE symposium on security and privacy (SP)*. IEEE, 2022, pp. 303–320.
- [53] T. Li, Y. Agarwal, and J. I. Hong, “Coconut: An IDE plugin for developing privacy-friendly apps,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–35, 2018.
- [54] A. Oulasvirta and K. Hornbæk, “HCI research as problem-solving,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4956–4967.
- [55] ACM Data Science Task Force, *Computing Competencies for Undergraduate Data Science Curricula*. New York, NY, USA: Association for Computing Machinery, 2021.
- [56] B. Saunders, J. Sim, T. Kingstone, S. Baker, J. Waterfield, B. Bartlam, H. Burroughs, and C. Jinks, “Saturation in qualitative research: exploring its conceptualization and operationalization,” *Quality & quantity*, vol. 52, pp. 1893–1907, 2018.
- [57] J. J. Francis, M. Johnston, C. Robertson, L. Glidewell, V. Entwistle, M. P. Eccles, and J. M. Grimshaw, “What is an adequate sample size? operationalising data saturation for theory-based interview studies,” *Psychology and health*, vol. 25, no. 10, pp. 1229–1245, 2010.
- [58] R. A. Virzi, “Refining the test phase of usability evaluation: how many subjects is enough?” *Human factors*, vol. 34, no. 4, pp. 457–468, 1992.
- [59] M. Abadi and L. Cardelli, *A theory of objects*. Springer Science & Business Media, 2012.
- [60] R. Wirfs-Brock and B. Wilkerson, “Object-oriented design: A responsibility-driven approach,” *ACM sigplan notices*, vol. 24, no. 10, pp. 71–75, 1989.
- [61] “Guide to components in Figma,” <https://help.figma.com/hc/en-us/articles/360038662654-Guide-to-components-in-Figma>, 2024.
- [62] “Notability,” <https://notability.com/>, 2025.
- [63] A. Frik, J. Kim, J. R. Sanchez, and J. Ma, “Users’ expectations about and use of smartphone privacy and security settings,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–24.
- [64] H. Nissenbaum, “Accountability in a computerized society,” *Science and engineering ethics*, vol. 2, pp. 25–42, 1996.

- [65] M. W. Newman, J. Lin, J. I. Hong, and J. A. Landay, "Denim: An informal web site design tool inspired by observations of practice," *Human-computer interaction*, vol. 18, no. 3, pp. 259–324, 2003.
- [66] P. E. McKnight and J. Najab, "Mann-Whitney U Test," *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.
- [67] D. J. Solove, "A taxonomy of privacy," *University of Pennsylvania Law Review*, vol. 154, p. 477, 2005.
- [68] National Institute of Standards and Technology, "Nist privacy risk assessment methodology (pram)," 2019. [Online]. Available: <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources>
- [69] S. Al-Fedaghi, "On information lifecycle management," in *2008 IEEE Asia-Pacific Services Computing Conference*. IEEE, 2008, pp. 335–342.
- [70] Y. Feng, Y. Yao, and N. Sadeh, "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*, 2021, pp. 1–16.
- [71] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor, "A design space for effective privacy notices," in *Eleventh symposium on usable privacy and security (SOUPS 2015)*, 2015, pp. 1–17.
- [72] S. Hart, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Human mental workload/Elsevier*, 1988.
- [73] J. Saldaña, "The coding manual for qualitative researchers," 2021.
- [74] S. Höltervenhoff, P. Klostermeyer, N. Wöhler, Y. Acar, and S. Fahl, "I wouldn't want my unsafe code to run my pacemaker": An Interview Study on the Use, Comprehension, and Perceived Risks of Unsafe Rust," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 2509–2525.
- [75] L. Liu, L. Gao, N. Soni, and Y. Yao, "Exploring design opportunities for family-based privacy education in informal learning spaces," *Proceedings on Privacy Enhancing Technologies*, 2024.
- [76] N. McDonald, S. Schoenebeck, and A. Forte, "Reliability and interrater reliability in qualitative research: Norms and guidelines for cscw and hci practice," *Proceedings of the ACM on human-computer interaction*, vol. 3, no. CSCW, pp. 1–23, 2019.
- [77] R. F. Woolson, "Wilcoxon signed-rank test," *Encyclopedia of Biostatistics*, vol. 8, 2005.
- [78] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [79] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.
- [80] K. N. Truong, G. R. Hayes, and G. D. Abowd, "Storyboarding: an empirical determination of best practices and effective guidelines," in *Proceedings of the 6th conference on Designing Interactive systems*, 2006, pp. 12–21.
- [81] T. R. Kelley and E. Sung, "Sketching by design: Teaching sketching to young learners," *International Journal of Technology and Design Education*, vol. 27, pp. 363–386, 2017.
- [82] M. Oates, Y. Ahmadullah, A. Marsh, C. Swoopes, S. Zhang, R. Balebako, and L. F. Cranor, "Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration," *Proceedings on Privacy Enhancing Technologies*, 2018.
- [83] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.
- [84] R. Stevens, D. Votipka, E. M. Redmiles, C. Ahern, P. Sweeney, and M. L. Mazurek, "The battle for new york: A case study of applied digital threat modeling at the enterprise level," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 621–637.
- [85] R. E. Thompson, M. McLaughlin, C. Powers, and D. Votipka, "'There are rabbit holes I want to go down that I'm not allowed to go down': An Investigation of Security Expert Threat Modeling Practices for Medical Devices," in *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 4909–4926.
- [86] D. Van Landuyt and W. Joosen, "A descriptive study of assumptions in stride security threat modeling," *Software and Systems Modeling*, pp. 1–18, 2022.
- [87] B. Shreeve, J. Hallett, M. Edwards, P. Anthonysamy, S. Frey, and A. Rashid, "'So If Mr Blue Head here clicks the Link...' Risk Thinking in Cyber Security Decision Making," *ACM Transactions on Privacy and Security (TOPS)*, vol. 24, no. 1, pp. 1–29, 2020.
- [88] F. J. Shull, "Evaluation of threat modeling methodologies," 2016, (Accessed: 2025-03-30). [Online]. Available: [https://insights.sei.cmu.edu/documents/4027/2016\\_017\\_001\\_474200.pdf](https://insights.sei.cmu.edu/documents/4027/2016_017_001_474200.pdf)
- [89] N. Shevchenko, T. A. Chick, P. O'Riordan, T. P. Scanlon, and C. Woody, "Threat modeling: a summary of available methods," *Software Engineering Institute—Carnegie Mellon University*, pp. 1–24, 2018.
- [90] J. Wen, Y. Zhao, W. Xu, Y. Yao, and H. Jin, "Teaching Data Science Students to Sketch Privacy Designs through Heuristics (Extended Technical Report)," 2025. [Online]. Available: <https://arxiv.org/abs/2504.04734>

## Appendix A.

### 15 Privacy-Related Design Decisions

TABLE 11: We developed a codebook of 15 privacy-related design decisions to assess the privacy-related design decisions covered in the sketches and the communication effectiveness between the sketch creator and interpreter.

Procedure	Design Decision	Explanation	Example
Data Collection	Collected Personal Data	Personal data collected to achieve goals	SSN, transaction history, photo, IP address
	Data Provider	Individuals from whom data is collected	User, attendee
	Collection Purpose	Purpose for data collection	Prediction, decision making
Data Retention	Stored Data	Personal data being stored	Transaction history, photo
	Storage Approach	Data storage implementation details	Data stored on cloud server for 3 months
	Post-Storage Action	Action taken after current data storage ends	Data transferred to an archive server after being deleted on the main server
Data Processing	Processing Input	Personal data to be processed	Transaction history, photo
	Processing Output	Output data after processing	Risk score derived from user’s history, facial features extracted from photo
	Processing Approach	Data processing implementation details	Facial recognition algorithm deployed on server
Data Access	Accessed Raw Data	Collected raw data to be accessed	SSN, transaction history, photo
	Accessed Output	Processed data to be accessed	Risk score, facial features
	Access Approach	Data access details	Meeting host views attendees’ attention scores in system-generated report
Choice & Notice	Choice Options	Options for the choice maker	User selects “accept” to allow data use for marketing or “reject” to deny it
	Choice Impacts	Effects of the user’s choice	“Accept” enables data scientist to use data in marketing analysis
	Choice Notification	System notification upon/after choice	Pop-up message such as “Your permission has been saved”

## Appendix B.

### Patterns Occurred in Creator-Interpreter Communication

TABLE 12: We summarized six patterns in the creator-interpreter communication regarding each design decision. We present sample cases to illustrate each pattern. For instance, if the creator annotated the content of a design decision (e.g., *Collected Personal Data*) as “SSN”, while the interpreter didn’t annotate, we would code the pattern as (E) Only the creator responded.

#	Pattern	Creator	Interpreter	#	Pattern	Creator	Interpreter
A	Full alignment	SSN, transaction history	SSN, transaction history	D	Only the interpreter responded	-	SSN
B	Partial alignment	SSN, transaction history	SSN, date of birth	E	Only the creator responded	SSN	-
C	No alignment	SSN	Transaction history	F	Neither responded	-	-

## Appendix C.

### Definition of Interpretation Recall and Precision in Section 5.3.2

TABLE 13: We define precision, recall, and F1 score based on the six patterns occurred in creator-interpreter communication. Specifically, the weight of partial alignment ( $\epsilon$ ) we used in Table 8 is 0.5. We also validated our results with  $\epsilon = 0$ . Due to space constraints, we discuss it in an extended technical report [90].

Term	Explanation	Formula
True Positive (TP)	Interpreter’s response fully or partially aligned with the creator’s response (partial alignment weighted by $\epsilon$ ).	$\#A + \epsilon \cdot \#B$
False Positive (FP)	Interpreter responded when the creator did not, including the unaligned $(1 - \epsilon)$ part of $B$ .	$\#D + (1 - \epsilon) \cdot \#B$
False Negative (FN)	The creator responded, but the interpreter either did not respond or misinterpreted the response.	$\#C + \#E$
True Negative (TN)	Neither the creator nor the interpreter responded.	$\#F$
<b>Precision</b>	The proportion of correctly aligned responses out of all responses made by the interpreter.	$\frac{\#A + \epsilon \cdot \#B}{\#A + \epsilon \cdot \#B + \#D + (1 - \epsilon) \cdot \#B}$
<b>Recall</b>	The proportion of correctly aligned responses out of all responses made by the creator.	$\frac{\#A + \epsilon \cdot \#B}{\#A + \epsilon \cdot \#B + \#C + \#E}$
<b>F1 Score</b>	The harmonic mean of precision and recall, balancing the two.	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$



## **Appendix D. Meta-Review**

The following meta-review was prepared by the program committee for the 2025 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

### **D.1. Summary**

The paper presents a user study of  $n=24$  data science students (both undergraduate and graduate) who were asked to sketch privacy designs of a set of privacy scenarios after being given three heuristics to follow at the beginning of the study. The paper describes two pre-studies. The first pre-study was a need-finding study ( $n=12$ ) that found that students lacked the vocabulary needed to sketch privacy designs effectively, the sketching may require multiple iterations, and the students had difficulty planning the sketch space in advance. The second pre-study is a teaching methods experiment that investigates 3 methods: object-oriented diagramming, vocabulary-based sketching, and heuristic-based sketching. The findings suggested that heuristic-based and vocabulary-based sketching would be the best for data science students, and those methods are examined in the main study. The main study ( $n=24$ ) compared the two methods in a sketching task, followed by a post-task interview. The findings suggest that the simple heuristic method performed best in sketching efficiency and broader coverage of privacy-related design decisions and had a reduced learning curve. A heuristic-based approach combined with the multi-layer representation is recommended for privacy design sketches conducted by data science students.

### **D.2. Scientific Contributions**

- Provides a Valuable Step Forward in an Established Field
- Establishes a New Research Direction

### **D.3. Reasons for Acceptance**

- 1) Provides a Valuable Step Forward in an Established Field. The paper contributes to the research on sketching privacy designs and offers insights into how we should teach data science students about data privacy design.
- 2) Establishes a New Research Direction. This paper presents the design and evaluation of a novel approach to prompting data science students to consider privacy in their design. The proposed heuristic-based prompting offers a novel direction of investigation for future work.