Privacy Control in Conversational LLM Platforms: A Walkthrough Study

Zhuoyang Li
Department of Industrial Design
Eindhoven University of Technology
Eindhoven, Noord-Brabant
Netherlands
z.li7@tue.nl

Yanlai Wu*
University of Central Florida
Orlando, Florida, USA
wuyanlai@gmail.com

Yao Li
University of Central Florida
Orlando, Florida, USA
yao.li@ucf.edu

Xinning Gui
The College of Information Science and Technology
The Pennsylvania State University
University Park, Pennsylvania, USA
xinninggui@psu.edu

Yuhan Luo†
Department of Computer Science
City University of Hong Kong
Hong Kong, China
yuhanluo@cityu.edu.hk

ABSTRACT

Large language models (LLMs) are increasingly integrated into daily life through conversational interfaces, processing user data via natural language inputs and exhibiting advanced reasoning capabilities, which raises new concerns about user control over privacy. While much research has focused on potential privacy risks, less attention has been paid to the data control mechanisms these platforms provide. This study examines six conversational LLM platforms, analyzing how they define and implement features for users to access, edit, delete, and share data. Our analysis reveals an emerging paradigm of data control in conversational LLM platforms, where user data is generated and derived through interaction itself, natural language enables flexible yet often ambiguous control, and multi-user interactions with shared data raise questions of co-ownership and governance. Based on these findings, we offer practical insights for platform developers, policymakers, and researchers to design more effective and usable privacy controls in LLM-powered conversational interactions.

CCS CONCEPTS

• **Human-centered computing → Walkthrough evaluations**.

KEYWORDS

Usable Security and Privacy, Large Language Model (LLM), Conversational User Interface (CUI), Walkthrough

ACM Reference Format:

Zhuoyang Li, Yanlai Wu, Yao Li, Xinning Gui, and Yuhan Luo. 2026. Privacy Control in Conversational LLM Platforms: A Walkthrough Study. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3772318.3791054

*The author participated in this work from January to June 2025.
†Corresponding author.

CHI '26, April 13–17, 2026, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791054

1 INTRODUCTION

As large language models (LLMs) become integrated into daily life—from productivity tools to emotional support and public health interventions—they gather a variety of data from our everyday activities and thus introduce new privacy challenges [129]. Unlike traditional platforms that collect user information in isolated and structured data fields, conversational LLMs process free-form user inputs via natural languages, which can include various types of sensitive data such as personally identifiable information, proprietary data, financial records, and medical histories [128]. Recent work has suggested that LLMs built on decode-only Transformer (e.g., ChatGPT, Gemini) are injective, which means that a user's exact text input can be reconstructed from the output [78]. In other words, the sensitive personal data sent to LLMs can be regenerated verbatim, either intentionally or unintentionally, leading to direct privacy breaches [124, 127]. Beyond direct data retention, LLMs can infer additional information from user input, such as the preferences and sentiment behind their language use, even when not explicitly shared [4, 102, 112], introducing additional privacy risks.

At the same time, conversational interactions that involve a virtual agent engaging with users to gather, process, and generate information are now the dominant mode on LLM platforms [19, 106]. Their human-like style can encourage more sensitive disclosure [49, 60]. While this enables the system to gather richer responses, it also amplifies privacy concerns [53, 70, 112, 128]. Moreover, new interaction features such as model customization and conversation sharing complicate how users can control and manage their own data [70].

Within previous work that investigated the privacy implications of conversational LLMs, researchers primarily focused on identifying the conversation scenarios of sensitive disclosure (e.g., the type of sensitive data) [128], examining users' privacy perceptions and mental models regarding emerging features on conversational LLM

platforms [70, 128] and highlighting the potential privacy risks associated with conversational LLMs [46, 52, 55, 102, 119, 120].

However, there remains a gap in understanding how conversational LLM platforms currently present their privacy policies and controls for users to manage how their data is being stored, trained, used, and shared. This understanding is crucial because control operations such as accessing, editing, retrieving, archiving, sharing, and deleting data are deeply intertwined with privacy matters, since they directly govern how personal data is handled, protected, and managed throughout its life cycle. Although technical implementations such as encryption and anonymization mechanisms are important, the interface ultimately determines whether users can recognize and make use of available privacy protections [28, 44, 58]. For this reason, we set out to examine the control options presented at the interface level: What end users can directly see and do. Understanding these interface-level options can guide the design of more user-centered privacy mechanisms, which aligns with broader calls on usable privacy (e.g., General Data Protection Regulation (GDPR) Art. 12 and 25 [93]; California Consumer Privacy Act of 2018 (CCPA) code 1798.130 [10]). In this light, we ask: **How do conversational LLM platforms govern data practices and provide controls for users to manage their data?**

To answer the research question (RQ), we conducted an application walkthrough [63] of six widely used consumer-facing conversational LLM platforms: Character.ai, ChatGPT, Claude, Gemini, Meta AI, and Pi. These platforms were identified through a three-stage screening process guided by predefined criteria, including the presence of a consumer-facing conversational user interface (CUI), general-purpose use, popularity, and built by organizations headquartered in the U.S. Following the application walkthrough method (an expert-driven approach focusing on examining the technical features and intended functions of a system [63]), we examined how privacy control mechanisms are governed, presented, structured, and made accessible to users. While this method does not incorporate user perspectives directly, it provides an empirical foundation for documenting current design practices and identifying areas where future user studies may be particularly valuable.

Our findings revealed that each platform offers distinct privacy control mechanisms concerning whether, what, who, and how user data can be accessed, edited, deleted, or shared. Specifically, we identified unique characteristics of data units, control options, and control execution mechanisms on these platforms that differ from the ones found on conventional platforms. Given the rapid and ongoing evolution of LLM platforms, the specific interface features and control mechanisms continue to change. Nevertheless, the patterns identified in this study capture an emerging paradigm in which platforms are experimenting with new ways to define data units, explore natural language-based controls, and negotiate multi-user data sharing. These trends indicate broader directions in how privacy management is being imagined in the age of conversational LLMs. Based on these insights, we propose empirically grounded implications to support platform developers, policymakers, and researchers in improving data practices and identifying opportunities to strengthen user data management and privacy protection. We also outline directions for future user-centered research building on these findings.

2 RELATED WORK

Here, we first cover the related work on existing privacy control mechanisms on conventional platforms, and then describe privacy risks on conversational LLM platforms identified by researchers.

2.1 In-platform Privacy Controls

Privacy is a concept deeply shaped by its roots in philosophical, legal, sociological, political and economic traditions [80]. In the digital age, it is typically known as the right to control personal data, determining when, how, and to what extent it is shared [113]. This conceptualization of privacy focuses on informational control, which is also directly operationalized in modern regulatory frameworks. For instance, the EU's General Data Protection Regulation (GDPR) grants users specific rights over their data, including the rights to access, rectify, erase, and restrict the processing of their information, as well as rights to data portability and to object to certain types of processing [93]. Similarly, the California Consumer Privacy Act of 2018 (CCPA) provides rights to know, delete, and opt-out of the sale of personal information [10]. These frameworks, despite their differences in geographical scope and legislative detail, share a common principle: empowering individuals with control over their personal data. For example, accessing ensures users can verify how their information is stored and used, and editing allows users to correct errors or limit oversharing. From users' perspectives, one common way to understand how their data is handled and to control their data is to refer to the privacy policies or privacy notices that explain a platform's data practices [24, 95, 110]. However, such information is often criticized for being long, vague, complex, misleading, and difficult to read, making it easy to be ignored and ineffective in truly helping users manage their privacy [87, 90, 94, 95, 110].

Besides policy statements, many platforms have implemented in-platform privacy control mechanisms for users to actively manage the use of their data, including permission settings (e.g., access to built-in sensors such as GPS and microphone) [27], information usage (e.g., third-party access) [20, 37, 41], consent interfaces [34, 38], and more granular options for editing, deleting, or sharing their data [47, 96, 126], etc. Extensive research has explored the design of these control features, such as the visual presentations and the timing of prompts, as well as how they influence the ways that users control their data [28, 37, 40, 41, 47]. Researchers noted that although these features provide users with a degree of control, there were several usability barriers that hinder users from configuring them effectively [39, 59, 75]. For example, Habib et al. [39] found that users struggled with locating the appropriate settings page and understanding the content of opt-out controls [39].

To address these usability barriers, prior work has proposed various solutions. For example, Liu et al. developed a "*personalized privacy assistant*" that provides recommendations and nudges to help users stay aware of the privacy choices they have previously consented to and encourage them to review and adjust these choices [66]. Other research has also explored supporting users in managing permission settings by profiling their privacy preferences [65]. Nevertheless, these strategies primarily operate on conventional platforms where user data is collected in isolated, structured fields [20, 37, 41], an interaction mode that was more

common before the rise of LLMs. On modern LLM platforms where conversational interfaces become the mainstream, the input of unstructured natural language creates a fundamentally different privacy dynamic, because users can no longer easily foresee what personal data is being revealed or derived when engaging in free-form conversation, as every utterance can become a data point for sensitive inference [128]. This shift highlights the need to investigate new privacy control paradigms tailored to conversational LLM platforms.

2.2 Privacy in Conversational LLM Platforms

While interactions with conversational agents (CA) powered by LLMs enabled rich and natural exchanges, researchers have highlighted the privacy risks associated with personal data disclosure in these interactions [4, 70, 128]. A primary concern is "memorization," which manifests in two ways. First, LLMs store conversation history to maintain context, often retaining a multitude of user data, including personally identifiable information (e.g., names, email addresses, age), sensitive experiences (e.g., health records, finances, emotions), and thoughts, posing significant privacy risks [128]. Despite the common belief that the black-box nature of LLMs inherently obscures user data, a recent work showed that user input to LLMs can be inverted, creating a risk of direct input recovery [97]. Second, LLMs may retain and use user-provided information for model training, improving the performance of future models but inadvertently increasing the risk of privacy breaches, as these models can unintentionally leak memorized user information in responses to others [2, 11, 69, 123, 128]. For instance, Zhang et al. documented a memorization leak where a participant using an LLM plugin for programming software experienced an unexpected disclosure: Typing a classmate's name triggered an auto-completion suggestion revealing their school ID [128].

Furthermore, as CAs on LLM platforms were often designed with contextual understanding and empathy, many users perceive them as companions and willingly share private and sensitive information, even though the platforms were not designed for such disclosures [49, 53, 60, 112]. This risk can be further amplified with LLMs' "reasoning" ability, which can derive personal information from users' inputs, potentially revealing details users did not explicitly disclose [4, 60, 102, 109, 125]. For instance, Staab et al. demonstrated that LLMs can infer personal location, income, and gender based on textual patterns in user-provided content [102]. These findings heightened the privacy risks, as users may unintentionally disclose more information than they realize [102].

As LLMs are increasingly integrated into various platforms to support our daily work and life [107], empowering users to control their privacy is more critical than ever. Despite knowing the potential privacy risks and users' concerns, there is a lack of an overview of how platforms implement data control mechanisms or how effective these mechanisms might be. To address this gap, we walked through data control features and mechanisms on six widely used conversational LLM platforms, marking a first step toward designing more effective privacy control support and establishing privacy guidelines for human-LLM interactions.

3 METHOD

To examine data control features provided by existing conversational LLM platforms and the mechanisms behind these features, we conducted an application walkthrough [63] covering six consumer-facing platforms from November 19th 2024 to January 2nd 2025. In the following, we first explain what an application walkthrough method is and the rationale of choosing this method. We then present how we searched and identified target platforms, conducted the walkthrough, and performed data analysis.

3.1 Methodological Considerations

3.1.1 Application Walkthrough as an Expert-driven Approach. Application walkthrough[1] is a widely adopted approach for researchers to explore the features and functionalities of digital interfaces, websites, and software applications through a socio-technical lens [74, 83, 89, 91]. This method is often led by domain experts, as they possess the necessary knowledge to identify potential usability issues and evaluate the effectiveness of data control mechanisms. Similar expert-driven inspection methods (e.g., heuristic evaluation, cognitive walkthrough) have also been widely applied in usable security and privacy research [1, 23, 37, 72, 99, 114], which can yield empirical contributions by generating structured observations of platform features and design patterns [117].

We chose application walkthrough to answer our RQ, because it enabled us to document exposed settings, fine-grained controls, and nuanced privacy terminology—elements that are often hidden, technically complex, or require expert interpretation to fully understand [63]. Therefore, an expert-driven walkthrough enables researchers to identify potential gaps or design opportunities by examining how interface features are presented to users. As suggested by Light et al. [63], such analytic groundwork is crucial for building a concrete understanding of current design trends and informing future user-centered investigations.

3.1.2 Analytic Lens Within Temporal Limits. While conducting the walkthrough, we were aware that the interface designs and control features on LLM platforms were evolving rapidly, and since then, new models such as Grok4 [121] have emerged. We are also aware that such walkthroughs often require in-depth analyses across several months. It is therefore natural that an analysis of multiple platforms spans over a long time, during which incremental updates inevitably occur [39]. Hence, we acknowledge that the walkthrough captures only a snapshot describing emerging design paradigms rather than an exhaustive, continuously updated catalog of all conversational LLM interfaces.

However, this temporal scope does not compromise the validity of our findings. Our goal was not to document every interface feature that updates over time, but to identify the underlying interaction paradigms through which platforms conceptualize and operationalize user data control, which persist even as individual interface elements evolve. By examining these broader design paradigms instead of version-specific features, the walkthrough

[1] The application walkthrough method employed in this study differs from the "cognitive walkthrough" method, which evaluates usability by examining how a first-time or infrequent user would carry out predefined **tasks** [71]. In contrast, our walkthrough follows Light et al.'s [63] approach, focusing on how platforms structure and present data control **features**, rather than on assessing usability.

uncovered design directions that continue to characterize contemporary conversational LLM platforms.

3.1.3 Researchers' Positionality. Given that application walkthrough is an expert-driven and time-bound approach, here we disclose our positionality and expertise for transparency [21, 42]. All authors are experienced qualitative researchers in human-computer interaction and "early adopters" of the emerging features brought by conversational LLM platforms in our daily lives. In particular, the first and the corresponding authors have conducted several research projects on conversational LLM platforms; the other authors have several years of research experience in usable privacy and security, making the team well-equipped to conduct this application walkthrough and data analysis.

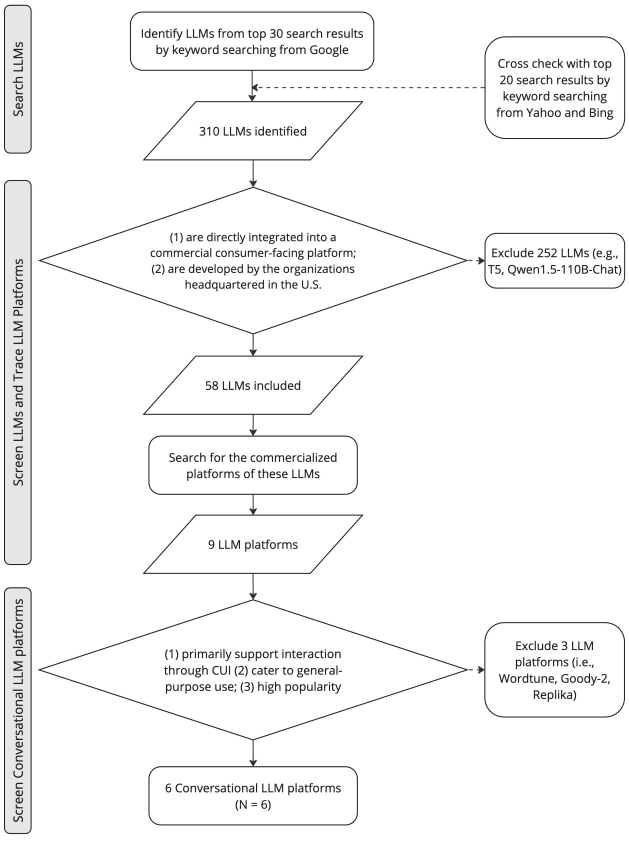


Figure 1: The three-stage data collection process.

3.2 Platform Identification and Selection

Our data collection process consisted of three stages (see Figure 1): (1) a search and screening of LLMs across three search engines; (2) tracing the models back to the consumer-facing platforms where the models are deployed and the user interaction data are directly managed by the model developers; and (3) screening conversational platforms to meet our inclusion criteria, which are detailed in the following section.

3.2.1 Searching LLMs. First, we identified a set of LLMs using the three most widely used search engines in the United States as of June 2024: Google (87.14% market share), Bing (7.81%), and Yahoo (2.58%) [103]. Before conducting the searches, we cleared the search history and enabled incognito mode in the Chrome browser and set the device's IP location to the United States to minimize location bias in search. We also set the time range for search results to between November 2022 (when LLMs gained significant attention following the launch of ChatGPT [13] in November 2022) and June 2024 (the month we conducted our searches). Our search process focused on identifying unique language models (e.g., GPT-3.5-Turbo), rather than platforms that use existing models. The search query was:

("Large language models" OR "LLMs" OR "Generative AI" OR "GenAI" OR "Chatbot" OR "Conversational AI") AND ("overview" OR "list" OR "survey" OR "review" OR "popular" OR "best" OR "top" OR "most used")

Given Google's dominant market share (87.14%) [103], we started from Google to look into the top 30 search results (highest-ranked links returned for the search query while excluding advertisements and sponsored content to maintain the objectivity and reliability of our data collection) and generated an initial list of LLMs mentioned. Then, we cross-checked the top 20 search results from Bing and Yahoo, to complement our list; during that process, we did not find any new LLMs beyond what had already been identified in the Google search results, indicating that data saturation had been reached [30]; we therefore concluded our search at that point.

The search results included academic papers, news articles, industry reports, leaderboards (i.e., systematic summaries of existing LLMs on open-source platforms such as Chatbot Arena [67]), and product reviews. The first author reviewed these results and extracted 310 LLMs, including both the latest versions of these language models and their historical iterations.

3.2.2 Screening LLMs and Tracing LLM Platforms. In this step, we traced the platforms that originally built and deployed these LLMs (e.g., ChatGPT [13]) rather than third-party platforms that integrate LLMs via APIs (e.g., Poe [86]). This is because the latter platforms lack full control over the model behavior, and thus are not directly accountable for the privacy risks, such as potential data breaches. Additionally, they often host multiple LLMs from different developers, introducing complexity that warrants further research. In addition, while LLMs are generally trained for multilingual tasks, English remains the dominant source of their training data [35]. Additionally, we focus on models developed by organizations headquartered in the U.S., because data practices of digital platforms are closely tied to regional regulatory and commercial contexts [64]. Exploring models built in other cultural and regulatory environments, such as Ernie, which must comply with China's Personal Information Protection Law (PIPL) [82], would require a broader cross-cultural and cross-regulatory analysis that is beyond the scope of this paper [15, 32, 62]. We point to Section 6 for how future work can take up this comparative perspective.

Based on these considerations, we further screened the 310 LLMs identified in the previous search results based on the following criteria:

Table 1: Conversational LLM platforms that are analyzed in this study and the time frames in which we conducted the application walkthrough.

Platform	Developer	Employed Models (Oct 2024)	Popularity	Walkthrough Time Frame
Character.ai	Character.ai	In-house model	"*Character.ai has over 20 million monthly active users in 2024.*" [76]	Dec 4–10, 2024
ChatGPT	Open AI	GPT-3.5, GPT-4, GPT-4o, GPT-4o mini, DALL·E	"*ChatGPT currently has over 180 million users.*" [26]	Nov 19–26, 2024
Claude	Anthropic	Claude 3.5 Sonnet, Claude 3 Sonnet, Claude 3 Opus, Claude 3 Haiku	"*The website sees nearly 54.4 million visitors every month.*" [111]	Dec 25–30, 2024
Gemini	Google	Gemini 1.5 Flash, Gemini 1.5 pro	"*Google Gemini has an average of 274.7 million monthly visits by September 2024.*" [77]	Dec 26, 2024–Jan 2, 2025
Meta AI	Meta	Llama 3.1- 405B	"*The assistant(Meta AI) reached 400 million monthly active users and 40 million daily active users in early August.*" [16]	Dec 23–26, 2024
Pi	Pi	Inflection-2.5	"*Our one million daily and six million monthly active users have now exchanged more than four billion messages.*" [45]	Dec 19–23, 2024

- Models should be deployed by a commercialized consumer-facing platform, where the user interaction data are directly managed by the model developers.
- Models should be developed by organizations headquartered in the U.S., meaning their primary regulatory obligations fall under U.S. frameworks. As a result, their default interface language is English, their official documentation and policies are primarily in English, and their developers provide support in English first.

During this process, we excluded 196 models that are not commercialized (e.g., people can only use T5 [2] by installing the API, but it is neither commercialized, nor designed for non-expert consumers). We also excluded 47 models because they are not developed by the organizations headquartered in the U.S. (e.g., ERNIE 3.0 Titan is developed by Baidu, headquartered in China; Mistral-7B is developed by Mistral AI, headquartered in France). For the remaining 58 models, we traced back to the developers of these models and their initial conversational interfaces, leading to 9 platforms, namely, Character.AI, ChatGPT, Claude, Gemini, Meta AI, Pi, Replika, Goody2, Wordtune.

3.2.3 Screening Conversational LLM Platforms. With the platforms identified in the previous step, we conducted further screening based on the following criteria:

- The platform should primarily support interaction through conversations, featuring a clear user interface that includes elements such as a user input field and LLM-generated responses displayed in chat bubbles (e.g., Wordtune [118] is excluded, because it is an LLM-powered browser extension for reading and writing, and its users can receive rewrite suggestions by highlighting specific text, but it does not have a conversational interface).
- The platform should explicitly state to provide general-purpose support across various domains such as productivity, journaling, and entertainment, rather than focusing on a specific domain. For example, Replika [92] was excluded because it is advertised as "*an empathetic friend*" in a 3D anime-style environment, positioning it primarily as an application focused on emotional support. This positioning aligns with the use scenarios identified in prior studies [22, 56, 88].By contrast, Character.ai was included because, despite its emphasis on character-based interactions, it allows users to create or engage with characters for a wide range of purposes. For example, there are some categories provided in the navigation bar, such as "*assistants*," "*anime*," "*learning*," "*lifestyle*," etc. This criterion ensures that our analysis captures conversational LLM platforms designed for diverse user interactions rather than niche applications.
- The platform is widely used with validated sources reporting its user base (e.g., Goody-2 [33] was excluded because we could not find validated information about its user base by the time of data collection). This criterion ensures that our walkthrough focuses on platforms with large real-world impact.

These criteria ensure that our findings apply to a broad range of LLM interactions rather than being restricted to specific use cases. They also help focus our analysis on platforms with real-world impact and widely adopted privacy control mechanisms. As such, we excluded 3 platforms and included 6 platforms for the final data analysis, namely Character.ai [12], ChatGPT [13], Claude [17], Gemini [31], Meta AI [73], and Pi [85], see Table 1 for more details. We only focused on the versions for individual use, rather than enterprise use, API integration, etc.

3.3 Walkthrough Procedure

We followed Light et al.'s application walkthrough method, which involves researchers navigating through the interface or system in a structured manner while documenting their observations and experience, taking notes of interface elements, navigation flows, content organization, and usability [63]. Specifically, application walkthrough includes two steps: analyzing the *expected environment of use* and *technical walkthrough*, which we detail below.

3.3.1 Analysis of Expected Environment of Use. This part refers to each platform's **vision** (i.e., the anticipated usage scenario and

[2] T5 is an LLM introduced by Google Research in 2019.

target users), **operating model** (i.e., business strategies, such as premium versions or in-app purchases), and **governance** (i.e., the strategies employed by the service provider to oversee and regulate user activity to support their operating model and fulfill their vision). Thus, we analyzed relevant institutional materials encountered during our navigation through the platforms, such as Terms of Service, Privacy Policies, FAQs, and other relevant documents.

3.3.2 Technical Walkthrough. We examined the mechanisms that allow users to control their personal data, such as opt-in/out, accessing, editing, updating, and deleting personal information. We also considered how these options can be executed by users (e.g., indicating their choice by using toggles), which is key to privacy control. To document any variations in data control options available to users based on their account status and subscription plan, we examined all the platforms as a signed-out user, a signed-in unpaid user, and a signed-in paid user, respectively. For features involving sharing among multiple users (e.g., ChatGPT allows users to share their conversation with other users through an auto-generated shared link), we used a second user account to examine the platform from the perspective of a user who receives shared content. The time frames of walking through each platform are specified in Table 1.

Initial Walkthrough Protocol. To develop a consistent walkthrough protocol across platforms, we began with an initial walkthrough of ChatGPT and Character.ai, because the former was the most widely used platform at the time of data collection, while the latter adopts a distinct vision and operational model. Examining both allowed us to cover different design considerations and interaction paradigms at the outset. The first author meticulously documented every interface deemed to contain data control features (e.g., pop-up windows, page refreshes) during the initial interaction with ChatGPT, resulting in 95 distinct screenshots accompanied by detailed field notes. These notes captured not only the interface elements but also contextual information about the interaction (e.g., whether the feature was accessed while signed in or signed out, and the interaction flow leading to the interface). Each screenshot could contain multiple data control options; for example, a single drop-down menu might simultaneously display edit, delete, and share functions.

After this, all authors reviewed the screenshots and notes together to determine which parts of the interface required further examination. For instance, in the initial walkthrough, the first author experimented with different types of personal information to see whether they would trigger ChatGPT's memory feature. However, through our discussions, we realized that the analysis should focus only on interface-level data control options, rather than backend or undocumented mechanisms, because such implementations cannot be reliably verified through a walkthrough method. As a result, any screenshots or notes outside this scope were excluded from further analysis.

The first author then conducted a second technical walkthrough of Character.ai, producing 49 screenshots with notes. Next, all authors jointly reviewed the materials and identified the relevant features for the study. These discussions led us to develop a walkthrough protocol (the interface terminology used throughout the paper is illustrated in Figure 2, using ChatGPT as an example). The final overall technical walkthrough flow and the specific steps are outlined chronologically:

(1) Boarding page: We reviewed the institutional materials, such as Privacy Policy, Terms of Use, and the linked materials.
(2) Conversational User Interface (CUI): We initiated several rounds of conversation using topics potentially containing personal information (e.g., "*My name is Johnny. I am an HCI researcher. My research is about usable privacy.*"; "*I live in city X and I want some trip advice for city Y.*"; and "*I had two meetings today. One is with my colleague A, and another is with my supervisor B. I now want to improve my presentation skills.*"). Their primary function was to serve as consistent stimuli for eliciting relevant interface elements across platforms, rather than to compare how different inputs were processed or to examine different conversational topics. We therefore used short, neutral conversations and made minor adjustments when necessary to trigger comparable data-control options across systems (e.g., we added "*Remember...*" to trigger memory-related features). During these conversations, we explored all the available control operations (e.g., sharing the conversation, regenerating response) and linked interfaces (e.g., Memory Widget, Memory Portal) that emerged in the CUI.
(3) Side panel (Chat history): We accessed, edited and deleted chat history. We also interacted with other control options (e.g., archive, rename the session).
(4) Side panel (Customization): We interacted with customization-related features (e.g., customize a conversational agent to be "*coach that can improve my presentation skills* or customize a "Project" for "giving feedback on a walkthrough research project" by uploading some prior literature). We also tried to access, edit, delete, and other linked controls (e.g., share the customized persona) on the customization records and customize items through all linked interfaces (e.g., Customization Store).
(5) Settings page: We performed all the control options available (e.g., access shared message, export data, opt-out on model training) and explored all the linked interfaces (e.g., Shared Links Portal).
(6) Other control options: We interacted with available options related to data control (e.g., Turn on the "*Temporary Chat*" mode, view the "*Customization settings*").
(7) Log out: We deleted all the data and logged out.

When testing control options that would influence the conversational interactions, we always returned to the CUI to initiate new conversations. For example, after customizing a CA persona, we started a new conversation with the customized CA and explored operations such as accessing, editing, deleting, and sharing the chat history. Similarly, after sharing a conversation, we used the second user account to continue the conversation and examine what control features were enabled for the receiving user.

Protocol Refinement. While the technical walkthrough structure remained consistent across platforms, slight adjustments were made to accommodate each platform's unique layout and feature set. For instance, on Character.ai, the CUI is flanked by two side panels: The left panel provides access to previously interacted "characters,"

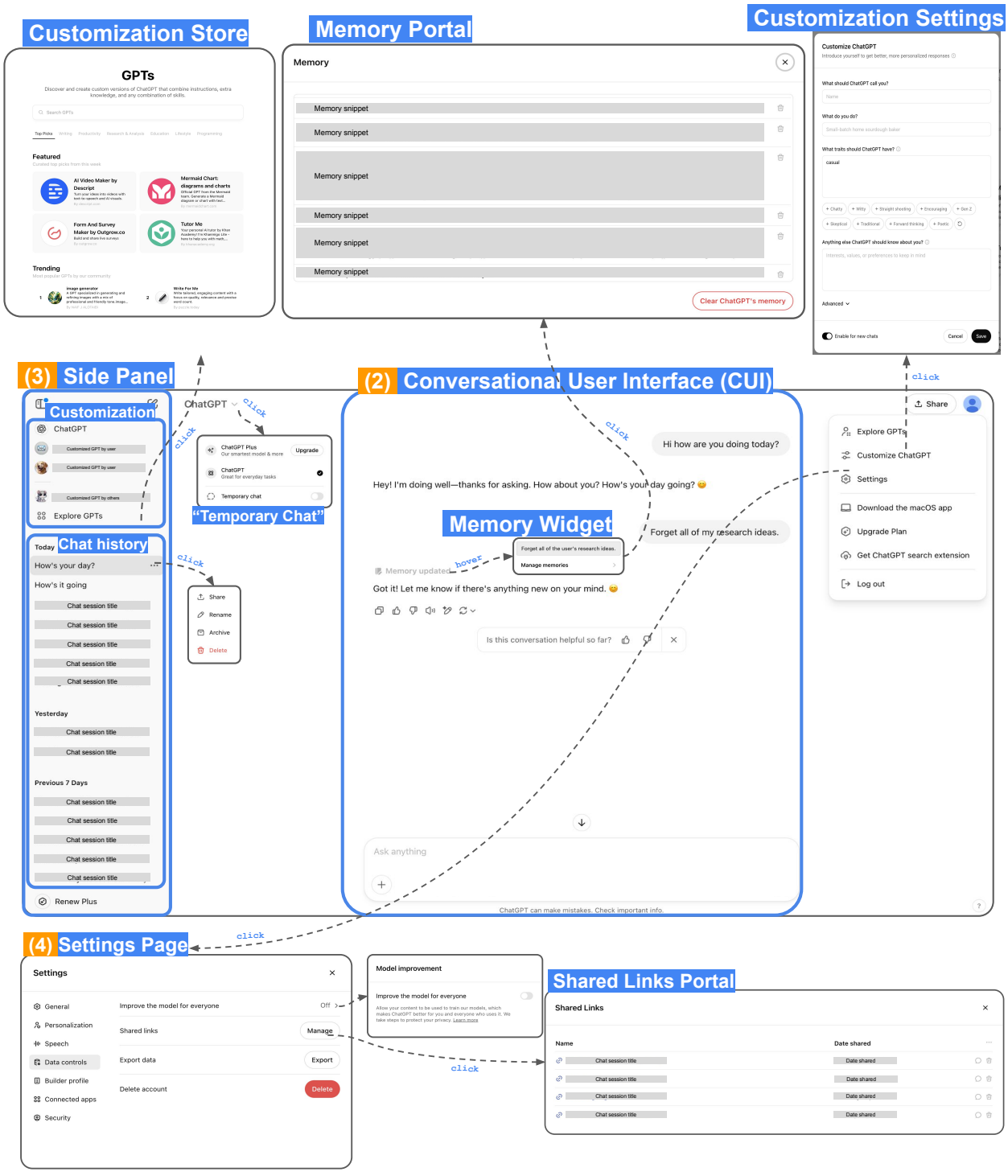


Figure 2: Interface examples of ChatGPT (free plan). Conversational User Interface (CUI) is the main interface where users interact with ChatGPT. Memory Widget appears when a user input triggers the memory-related feature, where users can view memory snippets and access the memory portal. Side Panel is typically located next to the chat window, which provides access to: chat history, organized by chat sessions; customization GPTs created by other users and Customized GPTs created by the user; customization Store for discovering and interacting with customized CAs created by other users. Settings Page is a centralized hub for managing various user preferences. Shared Links Portal provides access to and control on shared conversations. Customization Settings allows users to input descriptive information to customize their ChatGPT. Memory Portal is a space where users can manage all stored memory snippets.

option to create new ones, and settings, while the right panel displays details about the interacting character, including chat history with that character, pinned messages, and options for voice and style customization. We therefore conducted the technical walkthrough on Character.ai by the sequence: boarding page, CUI, right side panel, left side panel, settings page, and other control options.

Using this refined protocol, we retrospectively checked screenshots and field notes from ChatGPT and Character.ai, and discarded those that are not relevant. We then walked through the remaining platforms, collecting screenshots and writing field notes for Claude (38 screenshots), Gemini (33 screenshots), Meta AI (17 screenshots), and Pi (13 screenshots). The smaller number of screenshots for Meta AI and Pi reflects their limited data-control features, as they did not have customization and memory-related functions, and their sharing mechanisms were relatively simple. All the walkthrough logs are documented in an Excel file, with columns "*Platform Name*," "*Derive time*,""*Links*," "*Screenshots*," "*Field Notes*," and "*Initial Codes*," see Appendix A for some examples of our walkthrough logs. For presentation purposes, the screenshots included in this paper were post-processed with adjustments to color, contrast, and lightness to enhance readability, and the information potentially disclosing the authors' identities was redacted.

3.4 Data & Analysis

Our analysis focused on two main parts of the data: information related to data governance based on the platforms' expected environment of use, and data gathered through the technical walkthrough (see Section 3.3). We focused on data governance because it is a key aspect of understanding how platforms define and manage the privacy-related data practices and user control. Data Governance refers to the regulation regulating data availability, usability, integrity, and security to ensure consistent handling in compliance with policies and regulations [51]. The other two aspects of the expected environment of use (i.e., vision and operating model) were clearly stated by each platform; thus, we did not conduct thematic analysis on them.

The first author began the analysis by examining ChatGPT and Character.ai, as they embody different visions and expected usage, while both platforms cover a wide range of use cases. Character.ai emphasizes user-driven customization and character sharing, whereas ChatGPT emphasizes versatility for everyday life. All our analysis was done manually without any AI-assistance.

3.4.1 Governance-related information. For analyzing data governance, we employed a hybrid approach [29], combining deductive application of existing categories from the privacy policy annotation scheme developed by Wilson et al. [115] with inductive theme development. First, the first and corresponding authors reviewed institutional materials for ChatGPT and Character.ai (e.g., Privacy Policies and Terms of Use). Next, all authors collaboratively revised the privacy policy annotation scheme to suit the study context. This scheme, originally developed to annotate a substantial dataset (i.e., "*115 privacy policies (267K words) with manual annotations for 23K fine-grained data practices*") [115], has been widely adopted in usable privacy research [6, 116]. We tailored the scheme to the current context by incorporating insights from relevant literature.

During analysis, we identified an emergent theme, "data sharing among multiple users," in Character.ai's Privacy Policy. Acknowledging the relevance of this theme to data governance practices and its absence in the original scheme, we incorporated it into our revised analytical framework. The adapted scheme used in the final analysis is presented in Table 2. The first author then integrated governance-related codes from the remaining four platforms into this scheme.

3.4.2 Data from Technical Walkthrough. Our technical walkthrough started with an inductive approach to develop initial indices, followed by a deductive approach that applied these indices in subsequent analysis. In the inductive analysis phase, the first author applied an inductive approach to analyze the field notes and screenshots from the technical walkthrough of ChatGPT and Character.ai. This process led to the generation of initial codes, such as "*Granularity on controllable information (access, edit, deletion, and sharing)*." These preliminary codes were then discussed in meetings with all authors to synthesize insights from the technical walkthrough of these two platforms. Through this process, we observed that user data recorded on these platforms is not limited to chat history but also the memory stored about the users and their customized objects, such as descriptions of a conversational character on Character.ai. We also identified nuances in data control units (e.g., Character.ai allows users to customize both the conversational character and roles for users themselves to play on), interaction methods to perform controls (e.g., natural language commands for managing "memory" in ChatGPT), and data-sharing features (e.g., users can continue conversations shared by others in ChatGPT). Based on these preliminary findings, we developed three categories for subsequent analysis:

- **Data Types and Units:** Categories of user data recorded on the platforms (e.g., chat history, customization record, memory) and the ways they are structured as discrete elements for control (e.g., one chat session, one piece of memory snippet, one customized conversational agent).
- **Control Options:** Actions users can take on their data (e.g., access, retrieval, sharing, memorization) or choices they can make (i.e., opt-in and opt-out) over first/third-party data collection and usage.
- **Control Executions:** Methods and effects for executing control operations (e.g., graphical user interfaces (GUIs) in settings, natural language commands in chat).

Next, our analysis of the four remaining platforms followed a deductive approach, strategically guided by the categories derived from the walkthroughs of ChatGPT and Character.ai. To ensure consistency in the analyzing process, the first author developed a form documenting each data type and unit, the platform, and specific examples from each platform. All authors met regularly to review emerging examples and discuss ambiguous cases. For instance, after discussion, we treated Character.ai's "*pinned message*" as a form of *memory* rather than merely a conversational message, because pinned messages can be referred to across chat sessions (see Section 4.2.2). Similarly, we categorized Claude's "*Projects*" as a type of *customized object* rather than a new data type, since its primary function—like other customized objects—is to allow users to personalize their conversations (see Section 4.2.3). Upon finishing

Table 2: Adapted data practices scheme.

Category	Description
Data Ownership	Who is responsible for and fully controls the information, and what information do they own [3, 8].
First Party Collection/Use	Whether, how and why a service provider collects and uses information.
Third Party Sharing/Collection	Whether, how and why user information may be shared with or collected by the third party.
User Choice	Whether and how users can make choices regarding choice type, choice scope, personal information type, purpose, and user type.
User Access, Edit, and Deletion	Whether and how users may access, edit, or delete their *personal data.*
Data Retention	How long user information is stored.
Data Security	How user information is protected.
Policy Change	Whether and how users will be informed about changes to the privacy policy.
Do Not Track	Whether and how Do Not Track signals for online tracking and advertising are honored.
* Data sharing among multiple users	What and how can information be shared among multiple users. Whether and how users interact with and control the shared information.

Note: This scheme is primarily derived from Wilson et al.'s annotation scheme that captures the data practices included in privacy policies [115], and synthesized with other studies [3, 8, 9, 100]. Key adjustments include: (1) revised category wordings for clarity, for example, we changed "User Choice/Control" to "User Choice," because in our study's context, user control encompasses both user choice and user access, edit, and deletion; (2) added "Data Ownership" that is not part of the original scheme, but an important prerequisite for exercising data rights [3, 8]; (3) excluded "International & Specific Audiences" from the original scheme, since incorporating multilingual or international platforms would conflate heterogeneous regulatory, cultural, and socio-technical contexts. Such cross-contextual analysis warrants a dedicated comparative framework, which we leave for future work; (4) added new items as the walkthrough processed, marked with "*".

the analysis, no new categories emerged beyond those established in the initial walkthroughs. The final sub-themes and themes are reflected in the sub-sections and section titles presented in the Findings. The detailed collaborative and iterative analysis process is provided in Appendix A.

4 FINDINGS

In this section, we first summarize the visions, operating models, and data governance of the six platforms, and then detail what data can be controlled and how these controls are executed on the platforms. First, all platforms share a common **vision** of harnessing LLMs to humanize digital interactions across a wide spectrum of life. Yet, each of them prioritizes different aspects to achieve this vision: Character.ai emphasizes interactive customization and entertainment by featuring diverse user-created characters on its homepage; ChatGPT, Gemini, and Meta AI position themselves as versatile assistants to augment creativity and productivity (e.g., the tagline of Meta AI "*Ask Meta AI anything.*"); Claude branded itself more for productivity purposes, saying "*Claude is a next generation AI assistant built by Anthropic and trained to be safe, accurate, and secure to help you do your best work*;" and Pi highlights its emotional intelligence, branding itself as "*The first emotionally intelligent AI.*"

Regarding **operating models**, Character.ai, ChatGPT, Claude, and Gemini offer subscription plans for individual use with more advanced model capabilities (e.g., allowing uploading files, faster and more messages), early access to new features, and additional features that are not available to unpaid users (e.g., "*customize GPT*" in ChatGPT, "*Project*" in Claude, and "*Save Info*" in Gemini), while Meta AI and Pi do not have subscription plans.

With the understanding of each platform's vision and operating model, we present findings from the analysis of data governance and the technical walkthrough below.

4.1 Data Governance

As described in Section 3.4.1, we applied a revised version of Wilson et al.'s privacy policy annotation scheme [115] to analyze each platform's data governance practices based on their Privacy Policies and Terms of Service. Table 3 presents an overview of data governance on these aspects across the six platforms. Below, we describe the governance practice that stood out among conversational LLM platforms compared to traditional platforms, highlighting whether they grant users ownership of models' generated data, what controls are available, and how shared data is governed among multiple users.

4.1.1 Data Ownership. Except for Pi, which only mentions that users own their inputs, all other platforms explicitly grant users full ownership of both submitted content (e.g., input text, uploaded files, pictures, audio recordings) and platform-generated outputs, as stated in their Terms of Service or Terms of Use. Additionally, all platforms claim to offer users the right to access, update, correct, or delete their personal data, ensuring they have control over their information.

4.1.2 First/Third-Party Data Usage. All platforms provide detailed information on data collection, usage, and users' rights to data control by describing the types of data collected and the purposes of data processing by first (i.e., the platform itself) and third parties (e.g., analytics tools, external reviewers, and vendors). The collected data includes the information directly provided by users (e.g., the email address entered in the user's profile) as well as data automatically gathered through system interactions (e.g., the chat history). The primary purposes of both first and third-party data collection include service improvement, research and analysis, legal compliance, and personalized advertising.

4.1.3 User Choice. Except for Meta AI, which does not explicitly mention its practices about opt-out choice, other platforms claim that users have the right to opt out of data usage involving both first and third parties on these platforms. Character.ai and Claude allow opt-out of targeted advertising through the third-party analytics tools they employ. Similarly, Pi claims that they offer the opt-out option for online tracking by disabling third-party cookies. ChatGPT and Gemini claim they offer opt-out options that allow users to prevent their data from being used for model training.

Table 3: Data governance practices stated in the platforms' Privacy Policies or Terms of Use. ✓ indicates that the platform claims to offer the listed practice, whereas ✗ indicates that it does not.

Governance / Platform		Character.ai	ChatGPT	Claude	Gemini	Meta AI	Pi
Data ownership	Ownership of user input remains with the user	✓	✓	✓	✓	✓	✓
	Ownership of model output remains with the user	✓	✓	✓	✓	✓	✗
User access, modification, and deletion of their data		✓	✓	✓	✓	✓	✓
First-/Third-party data	Purpose statement	✓	✓	✓	✓	✓	✓
	User choice	✗	✗	✓	✓	✗	✗
Data sharing with other users		✓	✓	✗	✓	✗	✓

Table 4: Controllable units of different types of user data.

Data		Description	Applicable Platforms
Data Type	**Data Units**		
Chat History	Model-generated image	The image generated by Claude, which is included in a message	Claude
	Individual message	A piece of message sent by the user or generated by the model	Character.ai, Claude, Gemini, Pi
	Multiple messages	Two or more individual messages from the user, the model, or both of them	Pi
	Single conversational round	One message from the user followed by one reply generated by the model	Gemini, Meta AI
	Multiple conversational rounds	Two or more back-and-forth turns of messages between the user and the model	Gemini
	Chat session	A continuous message exchange between the user and the model, usually within a specific time frame or topic	All platforms
	Multiple chat sessions	Two or more chat sessions between the user and the model	Claude
	Messages with a "character"	All conversation messages that the user has had with a specific "character"	Character.ai
	All the chat history	All conversation messages between the user and all models saved on the platform	ChatGPT, Claude, Gemini, Meta AI
Memory	Memory snippet	A piece of information that the model derives from individual messages for personalizing interactions, usually appears as a rephrased form of the original input.	ChatGPT, Gemini
	"Pinned" Message	A piece of message that users can "pin" to make the character remember and personalize future interactions with that character	Character.ai
Customized Object	Conversational agent (CA)	The CA is customized by users with specified personas and communication styles—the "character" in Character.ai, the "Gem" in Gemini	Character.ai, ChatGPT, Gemini
	User persona	A virtual role that the user wants to play during the interaction	Character.ai
	Project	A specialized workspace for users to organize chats, upload documents, and create custom instructions to help streamline repetitive tasks and facilitate team collaboration	Claude

4.1.4 Data Sharing Among Multiple Users. Regarding data-sharing governance, Character.ai stands out as a community-driven platform, explicitly stating that "*popular characters*" created by users may be retained even after account deletion. They state in their Privacy Policy:

"*If a Character you (the user) create and set to 'Public' reaches a certain threshold of popularity, we (Character.ai) reserve the right to preserve that Character's characteristics and to keep that Character active on the Services, even if you otherwise delete your data and your account. We do this to avoid impacting the experience of other users, given that a highly popular Character by definition is having active conversations with many thousands of users. [...].*"

Gemini emphasizes user control over shared content but warns that public information may become searchable. ChatGPT allows users to share conversations and interact with third-party services, with shared data subject to external privacy policies. Other platforms did not explicitly mention their practices in the aforementioned documents.

Table 5: The control options offered by the six conversational LLM platforms. ✓ refers to that the platform supports the specific control operation, while ✗ refers to that the platform does not support the specific control operation; N/A means that no other control options are available or the user data is not integrated in the corresponding form.

Control / Platform		Character.ai	ChatGPT	Claude	Gemini	Meta AI	Pi
Data type	**Control option**						
Chat History	Access	✓	✓	✓	✓	✓	✓
	Retrieve	✓	✓	✓	✓	✓	✓
	Edit	✓	✗	✓	✓	✓	✗
	Delete	✓	✓	✓	✓	✓	✗
	Share	✗	✓	✓	✓	✓	✓
	Others	"Remove", export, and archive	Export and (un)archive	Export	N/A	Export	Export
Memory	"Memorize"	✓ GUI	✓ NL	N/A	✓ NL, GUI	N/A	N/A
	Access	✓ GUI	✓ GUI		✓ GUI		
	Retrieve	✓ NL	✓ NL		✓ NL		
	"Update memory"	✓ GUI	✓ NL		✓ NL, GUI		
	"Forget" (or delete)	✓ GUI	✓ NL, GUI		✓ NL, GUI		
Customized Object	Access	✓	✓	✓	✓	N/A	N/A
	Retrieve	✓	✓	✓	✓		
	Edit	✓	✓	✓	✓		
	Delete	✓	✓	✓	✓		
	Share	✓	✓	✗	✓		
	Others	"Remove"	N/A	Archive	N/A		

4.2 Controllable Data and Options

Through our walkthrough, we observed three types of data users can control: Chat history, memory, and customization objects. However, the data units are structured differently across platforms, as summarized in Table 4. As mentioned in Section 3.4.2, we define data units as discrete elements of data for control, which are building blocks for platforms to implement controls on users' data. The options that users can perform control on these data (e.g., access, edit, delete, share) also vary by platforms (see Table 5). Below, we describe these units in detail alongside the control options that can be performed over them. For ease of reading, data units appear in **BOLD SMALL CAPS**, control options are underlined, and *italics* text refers to words or phrasing directly quoted from the platforms' interfaces or their institutional materials.

4.2.1 Chat History: Varying Data Units From Message to Session. As the central component of user data, chat history refers to recorded conversational interactions between users and LLMs, which consist of users' inputs and the LLMs' responses. We found that each of the six platforms aggregates chat history differently, varying from message, conversational round to chat session, see all data units and their descriptions in Table 5.

The most basic unit is **A MESSAGE**, either a single user input or a single model output. Building on this, **A CONVERSATIONAL ROUND** contains a user input with corresponding model response, while **A CHAT SESSION** ("*Threads*" in Meta AI and Pi) encompasses a continuous sequence of exchanges within a particular context or time frame, see Figure 3 (a)→ (c). Sessions are typically initiated when users select "*Start a new chat session.*" At the broadest level, some platforms allow users to control **ALL CHAT HISTORY** collectively, treating them as a bulk dataset.

As summarized in Table 5, all platforms allow users to access and retrieve their chat history. Except for Gemini, they also support exporting all of the user's chat history. However, the availability of other control options, as well as the granularity of the data units over which these controls can be exercised, varies across platforms:

- Editing: All platforms except ChatGPT and Pi allow users to edit the most recent user message. Character.ai additionally permits users to edit the last response from the model. Yet no platform currently supports editing earlier messages or full chat sessions.
- Deletion: With the exception of Character.ai, all platforms do not allow users to delete individual messages. ChatGPT, Claude, Gemini, and Meta AI support the deletion of chat histories by chat sessions. Claude further provides the deletion option of multiple sessions via keyword search. Gemini stands out by offering the option to delete **SINGLE** or **MULTIPLE CONVERSATIONAL ROUNDS** within a selected time frame through its "*Gemini Apps Activity*" interface, see Figure 4 (d) and (b). At a global level, ChatGPT, Claude, Gemini, and Meta AI allow users to delete all chat history from account settings.
- Sharing: Character.ai does not allow users to share chat history at any level. ChatGPT and Gemini support sharing entire chat sessions. Pi provides more flexible sharing, allowing users to select and share multiple non-sequential messages (including either user inputs, model outputs, or both). Meta AI allows sharing by conversational round, meaning each

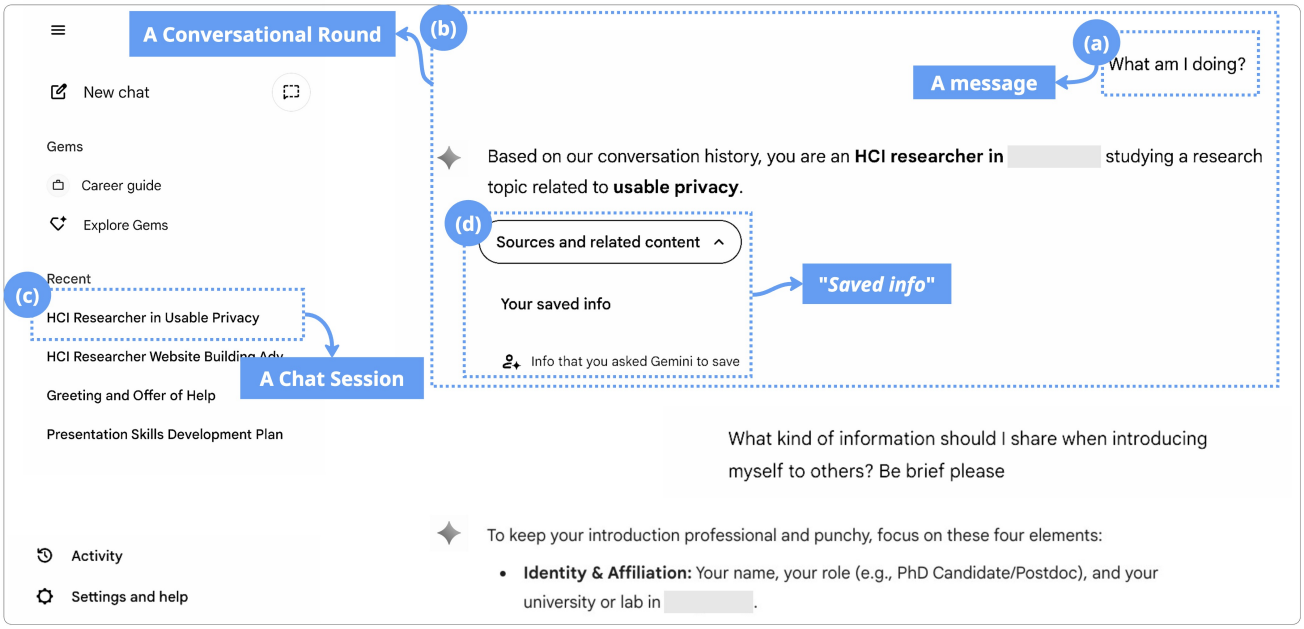


Figure 3: In Gemini, the data units of chat history include (a) a message, (b) a conversational round, and (c) a chat session. (d) When a message generated by Gemini is based on "*saved info*" (i.e., memory), a notification appears indicating which memory snippet was referenced.

shared unit must include a user input and the corresponding model response. In addition, Claude supports the sharing of generated artifacts (e.g., images created by the model) independently of the chat context in which they were produced. These are treated as standalone outputs, suggesting Claude differentiates between conversational content and generative media.

4.2.2 Memory: A Type of Derived User Data. Memory is a unique type of data on conversational LLM platforms, available on ChatGPT and Gemini. Unlike the original user input, memories are "derived" data that the model considers meaningful for personalizing user interaction. As ChatGPT's "*Memory FAQ*" states, "*Memory works similarly to Custom instructions, [...] when you share information that might be useful for future conversations, we've trained the model to add a summary of that information to its memory.*" In this sense, memory represents a more complex integration of information accumulated over time, rather than a simple collection of individual messages or chat sessions.

Our analysis found that there are no distinct, pre-defined units of memory. Instead, both ChatGPT and Gemini store memory as pieces of information reflecting the models' evolving understanding of a user, which we term "**MEMORY SNIPPET**." In short, a memory snippet is a specific data unit that captures what the model decides to keep and use from a user-input message, usually in a shortened or rephrased form. We also found that each memory snippet is derived from a single message but may contain one or multiple distinct facts or attributes, such as "*I like blue*" and "*My name is ...*", depending on what information is considered important to remember by the model. Memory snippets function as contextual information across chat sessions. In other words, snippets generated in one session can be retrieved, edited, and deleted in other sessions.

As shown in Figure 5 (a) and (b), memory is created and controlled in the form of memory snippets on ChatGPT and Gemini, where users can access, edit, delete these memory snippets. On Character.ai, while the term memory is not used, we observed a feature similar to memory: users can "*pin*" specific messages and ask the character to "memorize" them, see Figure 6 (a). These "**PINNED MESSAGES**" will then be the controllable data unit of memory on Character.ai, see Figure 6 (b). Vice versa, they can also "*unpin*" these messages so that the CA will "*forget*" the information and update its memory accordingly.

Beyond simply reviewing the saved memory snippets, Character.ai, ChatGPT, and Gemini allow users to retrieve information from memory, which refers to the model actively bringing up previously "remembered" information to an ongoing conversation. For example, if a user sends a message, "*My name is Johnny*," and later inquires "*What is my name?*," the model can then retrieve from its memory and send a response like "*Your name is Johnny.*" Note that such retrieval is not limited to direct user inquiries but can also occur implicitly to support personalization. In another example, if a user once shared, "*I prefer concise answers*," the model may adjust its response style in future interactions without needing an explicit prompt.

4.2.3 Customized object: An Encapsulated Data Unit. On these conversational LLM platforms, "customization" refers to user-initiated adjustments in how the model responds to their input, which differs from the customization of GUI components such as color mode or font size on traditional platforms. We use the term customized

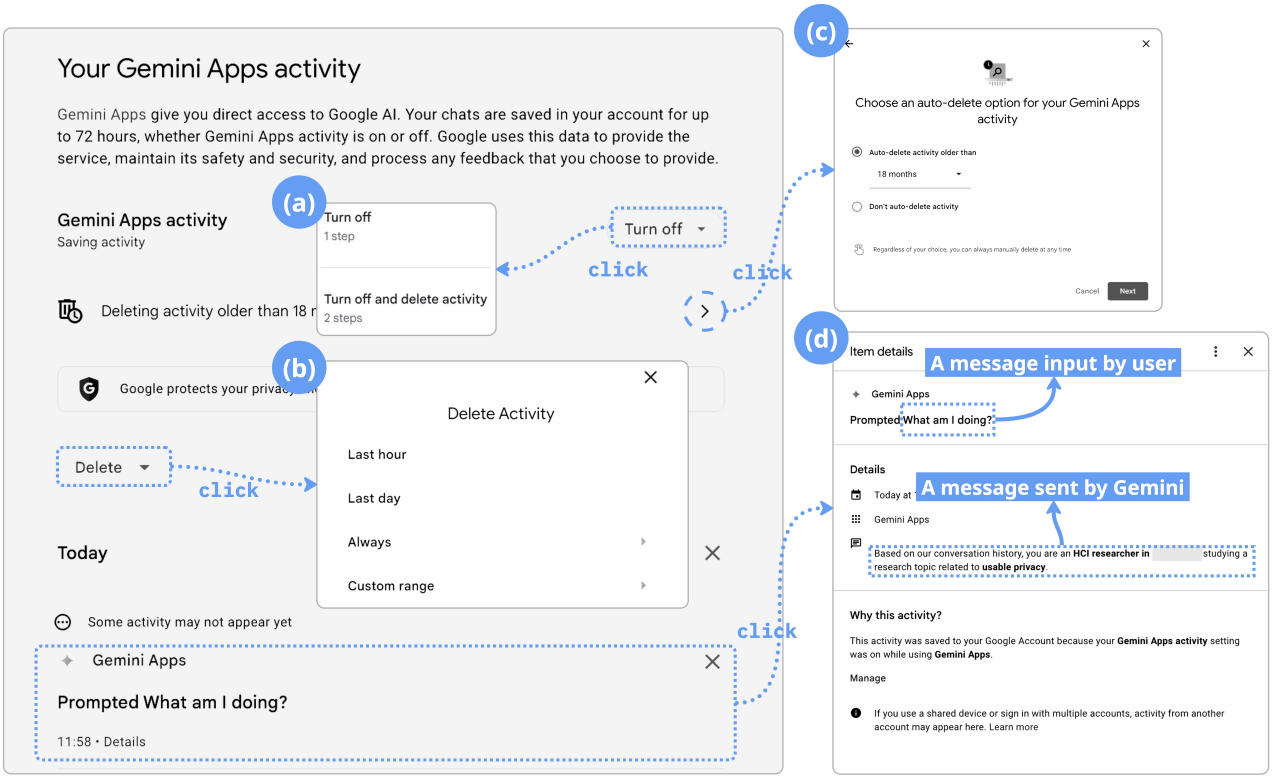


Figure 4: The "*Gemini Apps Activity*" page, where users can manage their interaction history with Gemini. (a) Users can choose to "*Turn off*" or "*Turn off and delete activity*". Turning off activity prevents future chats from appearing in the activity log and from being used to train the models. Choosing "*Turn off and delete activity*" also deletes all chat history. (b) Users can manually delete individual conversational rounds or multiple rounds within "*Last hour*," "*Last day*," "*Always*," or a "*Custom range*." (c) Users can enable auto-deletion within the Gemini app, with options for "*3 months*," "*18 months*" (default), "*36 months*," or "*Not auto-delete activity*." (d) Chat history is grouped under "*Item Details*," which represents a conversational round containing one user input and one Gemini output.

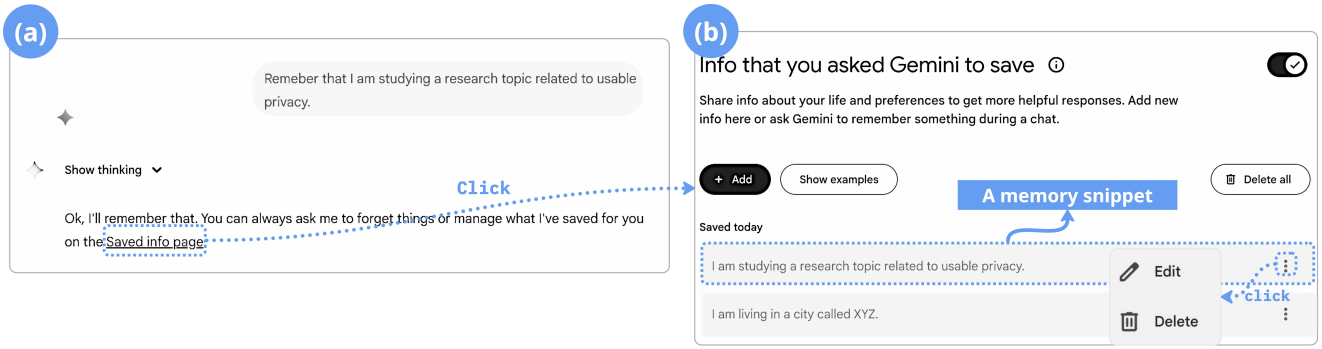


Figure 5: Gemini's memory snippets and Saved Info portal. (a) Using a prompt such as "*Remember ...*," users can instruct Gemini to save information as memory. (b) Saved information is displayed in the "*Info that you asked Gemini to save*" page, where the memory is organized by snippets (pieces of information). Users can access, edit, or delete individual snippets, and can also delete all snippets from this interface.

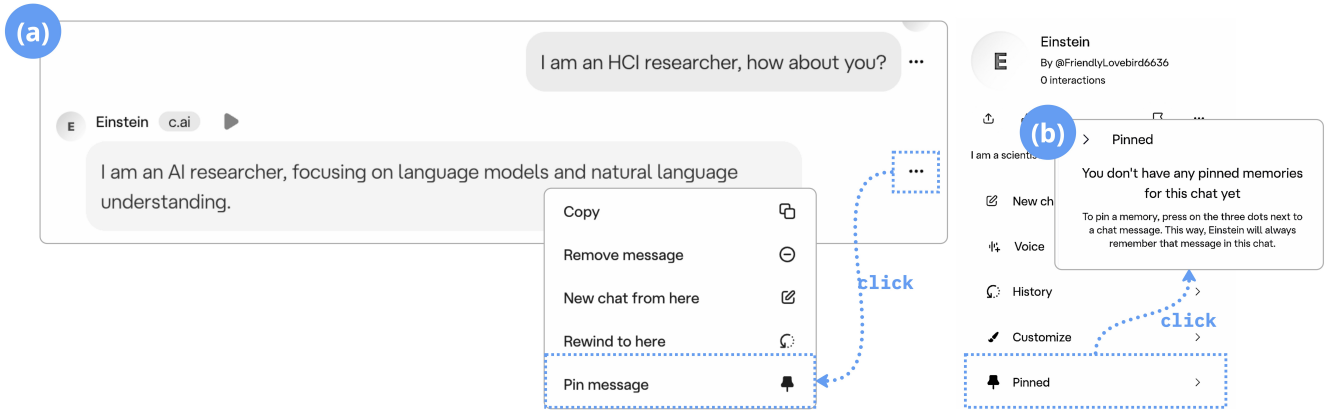


Figure 6: On Character.ai, (a) users can "*pin*" messages, then the character "*will always remember that message in this chat.*" (b) These pinned messages will be displayed in the right side panel beside the CUI.

object for a set of platform-specific features that share the same structure: they encapsulate user-provided textual instructions, documents, images, or audio, to form a single data unit, with unique instructions to shape the model responses. Across platforms, these customized objects manifest in three primary forms: conversational agents (CA), user personas, and projects.

Among the six platforms, Character.ai, ChatGPT, Claude, and Gemini support users to customize model responses, often in the form of specifying the identity and communications styles of the **CUSTOMIZED CA**, including the *character* in Character.ai (see Figure 7 (a) and (b)), *GPT* in ChatGPT, *Gem* in Gemini. Character.ai also offers customization for **USER PERSONAS**, a virtual role that users want to play during their interactions with the character, see Figure 7 (b) and (d).

For example, on ChatGPT, a customized GPT may include a role-defining instruction (e.g., "*You are my personal writing assistant.*"), a supplementary file (e.g., a writing sample), a profile image, and a selected voice. It is noteworthy that, for each customized object, we see its properties such as avatar, voice, textual descriptions, and uploaded files as part of the object, rather than controllable units.

Besides customized CAs and user personas, users can customize **PROJECTS** on Claude; projects are specialized workspaces for users to organize chats, upload documents, and create custom instructions, which helps streamline repetitive tasks and facilitate team collaboration, see Figure 8 (a)→(d). All of these customized objects can contain rich information about users. For instance, customized CAs and user personas may reveal sensitive information about users' identities and behaviors, while projects can include confidential documents and instructions.

Regarding controls, the four platforms that allow customization of model responses all support users in accessing and editing the content in their customized objects (avatar, voice of customized characters, files in customized projects), with other varying control options described below.

- Deletion: ChatGPT, Claude, and Gemini allow users to delete the customized object (e.g., CA persona, projects, and response style). Character.ai, however, only allows deletion of customized user personas but not customized "*Characters*," see Figure 7 (b) and (c). Additionally, ChatGPT allows users to delete historical versions of customized CAs that are chronologically organized.
- Sharing: Character.ai, ChatGPT, and Gemini allow users to share customized CAs but not customized user personas; Claude does not provide a sharing option for customized projects on personal accounts. These platforms also offer different sharing mechanisms, which will be further detailed in Section 4.3.4.
- Retrieval: Similar to memory, the information about customized objects can be retrieved. For example, if a user customizes a CA as "*You are the philosopher Aristotle. You have wisdom on observing the world and providing provocative insights*," users can retrieve the information by inquiring "*Who are you?*" and "*How can you help me?*" during the conversations with the customized character.

4.3 Control Execution

Building on the above section on what can be controlled and the associated control options, this section elaborates on how users could execute these control options (interaction method), when they can execute the control (before or after data input), the consequences of their control execution (scope of effects), and control executions with shared data (rights of the sharers and sharees).

4.3.1 Graphic User Interface vs. Natural Language Control. Our walkthrough showed that all the control options mentioned above (except for retrieving memory or information about customized objects) can be executed at the graphic user interface (GUI) level. Similar to controls on traditional digital platforms [7], these controls are embedded as visible interface elements such as in-conversation menus, settings pages, and side panels, and are enacted through familiar mechanisms, such as direct text-field edits, toggles, checkboxes, drop-down menus, buttons, etc. On ChatGPT, for example, users can access memory snippets through the memory portal and directly edit or delete these snippets (see Figure 2).

Figure 7: Creating, accessing, editing, deleting, and sharing customized objects on Character.ai. (a) Users can create a customized character by entering a description and choosing its visibility settings "*private*," "*unlisted*," or "*public for others to discover and chat with*." (b) Customized characters can be edited through the "*Characters section*" in Settings. (c) Through the Settings page under the section "*Personas*," users can edit, set or clear a persona as default for new chats, or remove the customized persona. (d) Users can "*create multiple personas to change their background info between chats*," and "*characters will remember the user's persona information to improve the conversations*."

It stands out that ChatGPT and Gemini also enable natural language (NL) control for memory and customized objects, most commonly, by instructing the model to "*remember*" or "*forget*" certain information in a chat session. On ChatGPT, there is explicit guidance on this approach under the"Personalization" tab of the Settings page: "*To understand what ChatGPT remembers or teach it something new, just chat with it: 'Remember that I like concise responses.' 'I just got a puppy!' [...].*" Gemini provides similar instructions, with prompt examples to guide users on how to influence the CA's memory. As indicated by their instructions, the prompts are rather intuitive without fixed commands to follow. As mentioned in Section 4.2.2 and 4.2.3, besides the explicit snippets in the memory portal, users can also retrieve the information from their previous interactions by NL.

To inform users what is being remembered during the conversation, both ChatGPT and Genimi incorporate visual indicators. For instance, in ChatGPT, when actions such as "*memorizing*," "*editing memory*," or "*forgetting*" are triggered, a small widget labeled "*memory updated*" would appear at the top of the model response, which can be clicked and show the detailed memory snippet, see Figure 2. In Gemini, when the message sent by Gemini is referenced from memory, there is a label attached at the end of the message "*Sources and related content*," where users are made aware of the referred memory, see Figure 3 (d).

4.3.2 Retrospective vs. Proactive Control. All the platforms granted the right of retrospective control, such as access, edit, and deletion (see Table 3), meaning users can act on their data **after** it has already been entered or generated. When it comes to memory control, these retrospective options give users a chance to "correct" or delete the data that exists, but they cannot prevent it from being created in the first place. In other words, users do not know in advance whether or what memory snippets will be created.

In contrast, proactive control refers to mechanisms that allow users to set boundaries on data collection or usage **before** their data is processed.

On ChatGPT, there is a "*Temporary Chat*" mode (see Figure 2), where user inputs are neither stored in memory nor used for model training, as stated in the conversation page: "*You're in control of ChatGPT's memory. You can reset it, clear specific or all memories, or turn this feature off entirely in your settings. If you'd like to have a conversation without memory, use Temporary Chat.*" Gemini offers auto-deletion options, allowing users to configure the interaction history to be automatically deleted after a certain period of time, see Figure 4 (c). Gemini also provides a combined form of retrospective and proactive control through the "*Gemini Apps Activity*" page, see Figure 4 (a). Users may choose "*Turn off*" activity, which proactively prevents future chats from being logged or used for model training, or "*Turn off and delete activity*," which additionally executes a retrospective deletion of all existing chat history.

Additionally, as reported in Section 4.1.3, on ChatGPT and Gemini, users can opt out of first-party uses for model training under their account settings, although it would not prevent the model from "memorizing" information deemed useful for personalizing further conversations. However, the other four platforms do not provide similar control mechanisms for users to proactively manage their data, while they allow users to opt out of third-party data sharing, such as for advertising or analytics purposes.

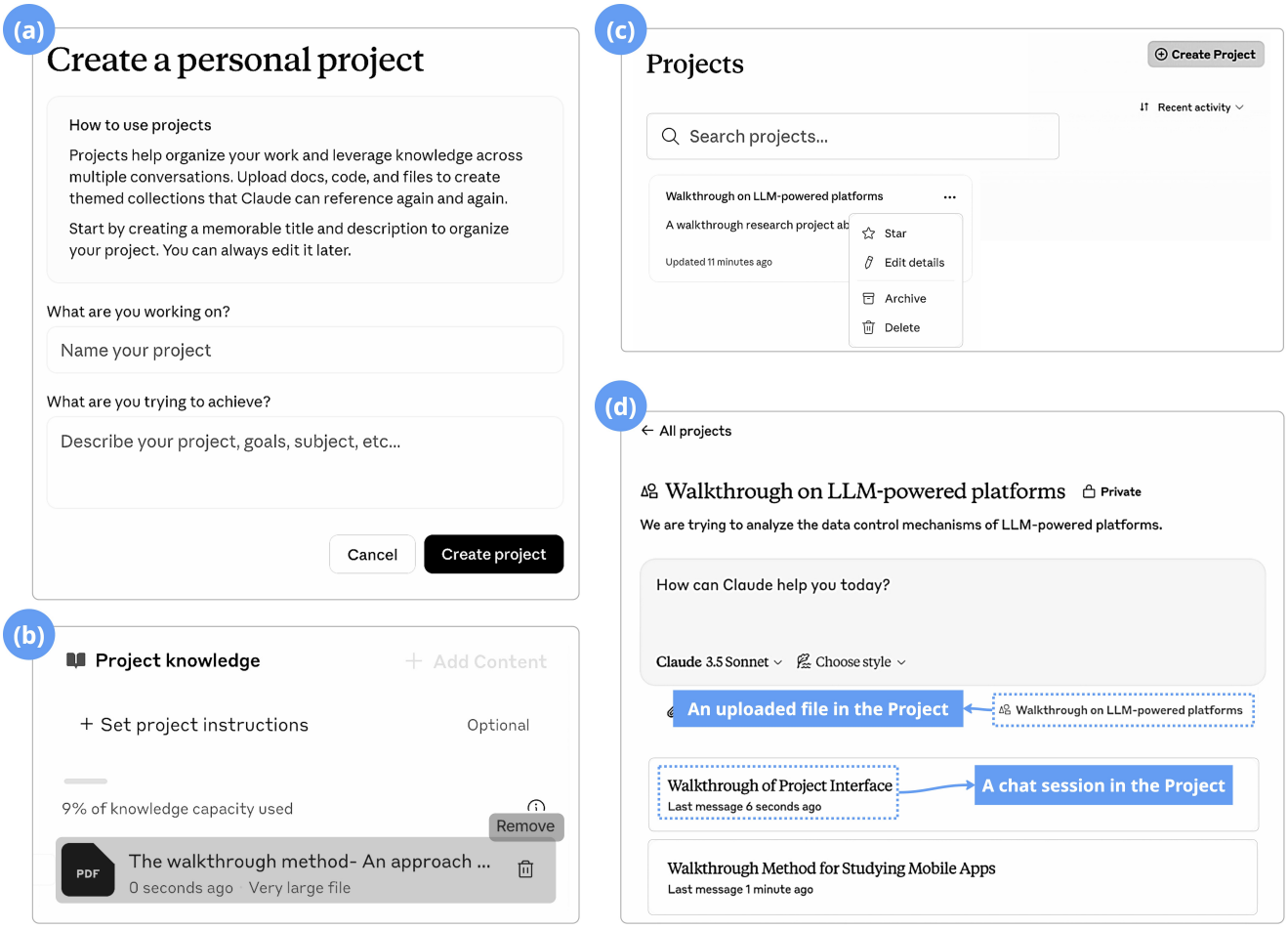


Figure 8: Claude allows users to create, access, edit, delete, and share projects. (a) Users can create a project by providing project title and description. (b) Users can customize projects by "*uploading documents, code, and other files* to a project's knowledge base for "*creating themed collections that Claude can reference again and again.*" (c) Users can access customized projects through the "Projects Portal." (d) Users can initiate multiple chat sessions within a project.

4.3.3 Local vs. Global Effects. With the introduction of memory features, conversational LLM platforms no longer treat control operations as uniform. The seemingly same commands—such as delete, forget, or remove—can apply at different layers of the system, producing either ***local effects*** (affect only the current chat session or information presented on the current interface) or ***global effects*** (changes across all user data stored by the platform). In our walkthrough, only one type of control action consistently produces global effects, which is the "delete all chat history" option available in the account settings of ChatGPT, Claude, Gemini (see Figure 4 (a)), and Meta AI. By contrast, most other controls produce only local effects.

In ChatGPT, for example, deleting a chat session has only a local effect: the conversation disappears from the side panel, but any information stored in memory remains. Similarly, when a user deletes a memory snippet through the memory portal, the stored information is removed from memory, while the original chat transcript remains accessible unless deleted separately. When a user attempts to delete a chat session, there will be a pop-up message: "*This will delete <automatically generated chat session title>. To clear any memories from this chat, visit your settings.*" Vice versa, deleting a snippet from memory leaves the chat history intact unless it is explicitly deleted. Gemini follows a similar logic: when a user deletes a chat session, the platform shows a warning message "*You'll no longer see this chat here. This will also delete related activity like prompts, responses, and feedback from your Gemini Apps Activity.*" However, this action will not affect the information in "*saved info.*" Similarly, Character.ai uses the term "*Remove*" that carries only local effects: The operation hides the character and associated chat history from the side panel but does not erase the conversation from storage; the data can be retrieved later.

4.3.4 "Sharer" vs "Sharee" Control the Shared Data. Character.ai, Claude, ChatGPT and Gemini allow users to manage the visibility of their shared information, stop sharing, or update the shared

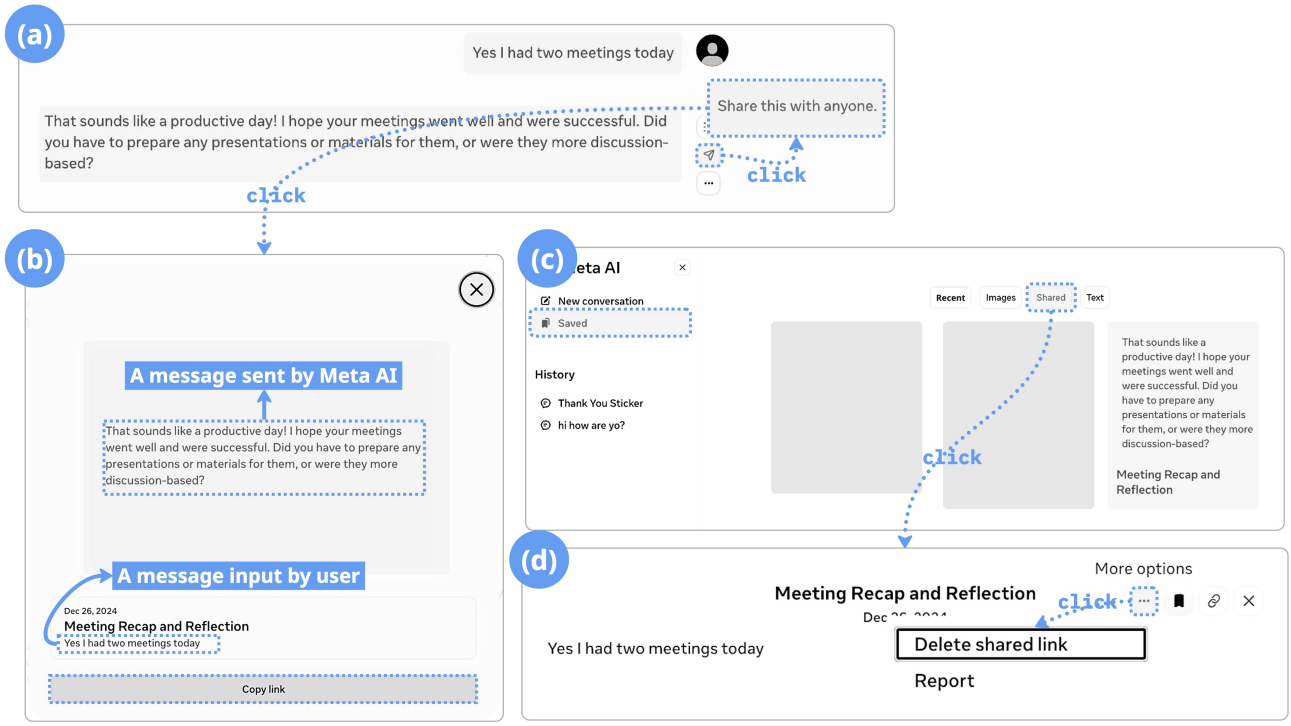


Figure 9: Meta AI allows users to share a conversational round by generating a shareable link. (a) Users can share a conversational round by selecting "*Share this with anyone*" next to the model's output. (b) When a round is shared, Meta AI automatically generates a link and the shared conversational round is saved. (c) Users can access shared conversational rounds through the "*Saved*" section in the sidebar of Meta AI's conversational interface. (d) Users can delete a shared link by opening the corresponding conversational round in the "*Saved*" section.

content. Unlike data sharing on traditional platforms, which merely makes the data available for others to view, data sharing on these conversational LLM platforms enables other users to directly reuse the shared data to generate new outputs of their own. Here, we define users who generate and share original data as "**sharers**," and users who access and interact with that shared data as "**sharees**." In what follows, we first report findings regarding chat history sharing and then detail those in customized CA sharing.

Chat History Sharing. All platforms except Character.ai allow users to share their chat history (see Section 4.2.1), which typically involves creating a link that points to a specific chat session, conversational round, message, or generated artifact. As shown in Figure 9 (a) and (b), Meta AI provides a "*Share this with anyone*" option next to each generated message. When selected, the platform creates a shareable link containing both the model's response and the corresponding user input—that is, the entire conversational round. On ChatGPT, users can also configure whether a shared conversation is publicly visible in web searches by selecting a checkbox labeled "*Make this chat discoverable*," whereas other platforms do not provide such visibility settings.

On the sharers' side, how they can manage and control the shared content differs by platform. ChatGPT is the only platform that allows sharers to update shared content (e.g., editing messages or generating new responses) with the same link. To stop sharing, users of ChatGPT, Claude, and Gemini can switch the visibility of the shareable link off or directly delete the link (e.g., clicking the "*Unpublish*" button on Claude). By contrast, on Meta AI, sharers need to delete a shared link through the "*Saved*" portal in the sidebar next to the CUI, see Figure 9 (c) and (d). In Gemini, shareable links will expire after six months by default unless users manually remove the expiration date. However, the consequences of deleting the shared chat session also vary across platforms. On ChatGPT, if the sharer deletes a shared chat session from their interface or deletes their user account, the shareable link automatically becomes unavailable, but on Gemini, deleting a shared chat session does not automatically remove the shared link. To do so, the sharer must manually delete the link under the "*Public links*" tab in "*Settings*."

On the sharees' side, once a chat is shared, ChatGPT, Gemini, and Meta AI allow them to view the chat history and continue engaging in the conversation from where it left off. When a sharee continues a shared conversation, the entire session (including the original messages and any new ones added later) is saved to their own account. Similarly, Gemini allows sharees to continue the shared chat by clicking the button "*Go to Gemini*", see Figure 11 (c). On Meta AI, sharees have the option to engage with shared prompts through a button labeled "*Try this prompt on Meta AI.*" It is important to note that continuing a conversation does not affect the original

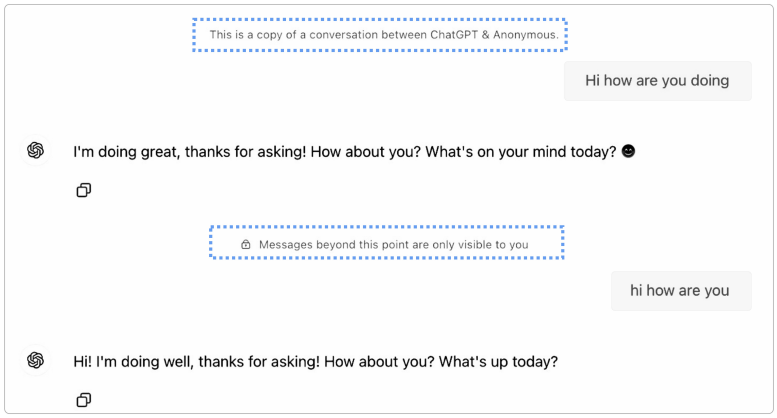


Figure 10: On ChatGPT, shared conversations can be accessed by both signed-in and signed-out users, accompanied by a fixed statement at the top of the page: "*This is a copy of a conversation between ChatGPT & Anonymous.*" When a recipient continues the shared chat, any new messages appear below a divider indicating that "*Messages beyond this point are only visible to you.*"

shared version on the sharer's side. For example, ChatGPT makes this clear by displaying a visual divider, see Figure 10.

Customized CA Sharing. As described in Section 4.2.3, Character.ai, ChatGPT, and Gemini support the sharing of customized conversational agents (CAs). Users on these platforms can create their own personalized agents and make them available for others to interact with (sharers). Once shared, these CAs can be accessed by other users (sharees), who may then engage with the CAs and control related chat history.

All three platforms allow sharers to control visibility settings upon sharing. For example, ChatGPT and Gemini provide three choices: "*Only me*," "*Anyone with the link*," and publishing in the "*GPT/Gem Store*." similar to Character.ai: "*Private*," "*Unlisted*" (accessible only via link), and "*Public*" (discoverable by anyone). However, as noted in Section 4.1.4, Character.ai explicitly states that "*popular characters*" created by users may be retained even after account deletion in their privacy policy. Furthermore, the extent to which customization details of these customized CA are disclosed upon sharing varies. On Character.ai, sharers may choose whether to reveal the "*definition*" of the CA (customization instruction) by toggling the option "*Keep character definition private*", see Figure 7 (a). Likewise, Gemini allows users to choose "*Share conversation and Gem instructions*," see Figure 11 (a). ChatGPT, by contrast, does not allow customization instructions of the CAs to be visible when shared. Similar to sharing chats, Character.ai and ChatGPT allow sharers to keep editing their customized CA after they are shared. Updates can be applied through a "*Save the changes*" button on Character.ai or an "*Update*" button on ChatGPT, and the changes will automatically carry over for sharees who continue interacting through the same link.

When interacting with a shared CA, sharees may also see information about the sharer, such as the sharer's username, profile page, or other published personas, while the sharers cannot access sharees' information. To clarify this boundary, ChatGPT explicitly notifies sharees that the sharers cannot view their data, see Figure 11 (c). Unlike sharers, however, the sharees cannot further edit the Shared CA. Gemini also provides the sharer an option to **share their chat histories along with the associated customized CA's instruction**, as illustrated in Figure 11 (a) and (b).

To sum up, our walkthrough highlights the emerging paradigm of data control on conversational LLM platforms around what data can be controlled, how data units are defined, and how these controls can be executed. In the next section, we build on these observations to discuss their broader implications for designing usable, transparent, and scalable privacy controls in conversational LLM platforms.

5 DISCUSSION

Our walkthrough reveals several unique aspects of data control on mainstream conversational LLM platforms: (1) Since the primary interaction on conversational LLM platforms is natural language (both input and output), the boundaries of what constitutes "user data" are less fixed and often emerge through the interaction itself. These interaction-derived data (e.g., memory, customized objects) differ from those on conventional platforms, which are typically structured and isolated fields [112]; (2) Relatedly, the introduction of natural language control offers an intuitive yet ambiguous way for data control, bringing up new questions on how to control the rich interaction-derived data in a more user-friendly and privacy-preserving way; (3) Multiple users can interact with shared data (e.g., chat history and customized CAs), turning sharers and sharees into "co-owners" and raising new implications for data governance for both parties. Drawing from the findings, we discuss how these emerging mechanisms inform the design of privacy control on conversational LLM platforms and directions for future research.

5.1 Re-conceptualizing Controllable Data Units

On traditional digital platforms, when new types of user data are introduced, privacy control mechanisms are typically implemented by adding new toggles—binary switches that allow users to enable or disable access to that specific data point (e.g., location,

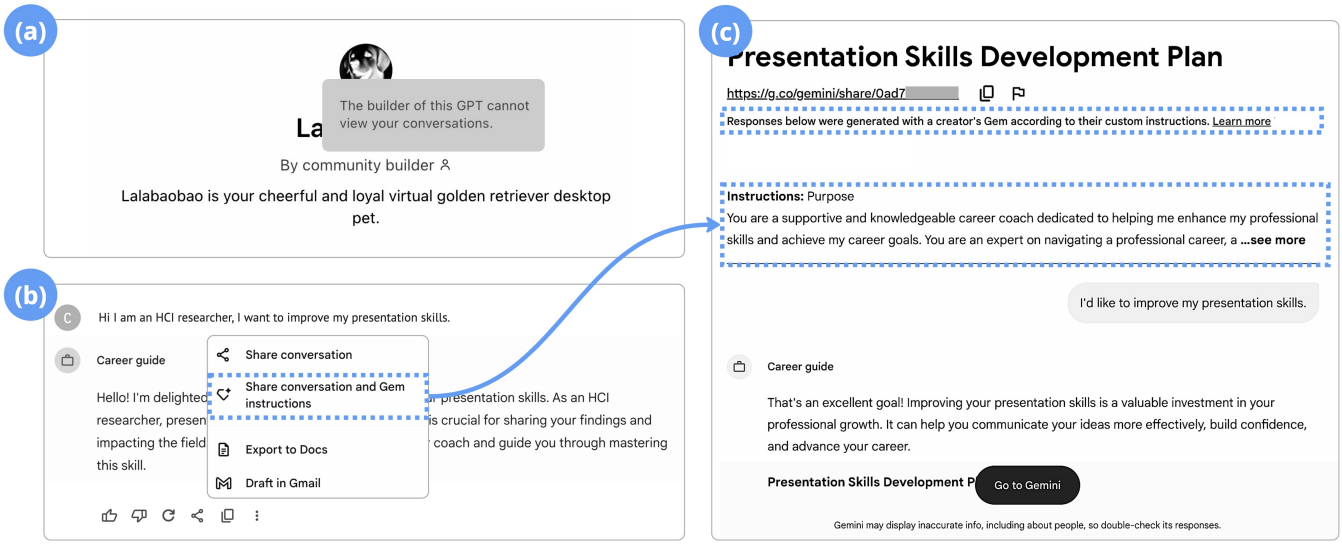


Figure 11: Share CA personas on Gemini and ChatGPT. (a) On ChatGPT, if the sharee hovers on the GPT builder's name, there will be a message saying "*the builder of this GPT cannot view your conversations.*" (b) On Gemini, users can share chat history or chat history combined with the CA persona customization records. (c) On Gemini, when a conversation is shared along with its Gem instructions, sharees can view both, but they can only continue the conversation—they cannot edit the Gem instructions. The interface also informs sharees that "*Responses below were generated with a creator's Gem according to their custom instructions.*"

contacts, and activity logs) [20, 37, 41]. While this model enables granular control, it scales poorly in environments where the form and amount of data are continually evolving, not to mention the extra burden imposed on users to perform control [39, 59, 61, 75].

On conversational LLM platforms, this challenge becomes even more pronounced. As our walkthrough shows, the types of data are not fixed, but are often emergent and constructed dynamically through interaction. For instance, a single user message might contain multiple factual assertions, emotional expressions, or preferences, any of which may be summarized and stored as a memory snippet by the model. Similarly, the customized objects such as conversational CAs, user personas, or projects may encapsulate identity traits, instructions, personal documents, etc. These units of data do not correspond to static form fields or discrete entries, and they often lack clearly defined boundaries.

To tackle this dilemma, designers of the platforms need to first understand the fundamental privacy needs of users—not just in terms of what data is collected, but how that data is interpreted, contextualized, and reused by the platforms. Prior research has shown that users' privacy concerns are often less about the type or format of data and more about what the data reveals, that is, its implications for identity, social perception, and future use [50, 79]. As such, adding more control options to each data point is neither sufficient nor desirable for governing what gets remembered, reused, or referenced by the model. Instead, platforms should ***re-conceptualize controllable units as interaction-derived constructs***. For example, instead of configuring controls solely for specific attributes such as name, gender, or age, platforms can make it flexible for users to turn on or off protections for broader categories of data, such as "personally identifiable information," "professional conversations," "medical related information," etc, without needing to manually adjust settings for each individual data point. Likewise, control interfaces might shift from toggle-styles to topic-oriented filters or conversation-aware dashboards that mirror the lived experience of interacting with LLMs.

Moreover, our walkthrough revealed that current platform designs often treat data controls in a local scope (e.g., conversation session-based), but users' data, including chat messages, customizations, and memory snippets, circulates beyond these boundaries. This disconnect makes it difficult for users to track and manage information that may be retained, referenced, or reused beyond its original context. As a result, users may experience fragmented oversight, where they lose visibility into how their data is being used, remembered, or shared over time [70, 112, 128], which not only undermines transparency but also erodes user trust. Thus, platforms should consider ***implementing cross-contextual control mechanisms***, such as unified memory dashboards, longitudinal data timelines, or system-prompted review summaries that allow users to view, revise, or remove stored data regardless of its origin. Such mechanisms would better align data governance with the evolving nature of conversational interactions and help restore a sense of control in long-term engagement with LLMs.

5.2 Toward Efficient and Scalable Natural Language (NL) Control

Our walkthrough found that users can use NL to specify privacy preference regarding the memory derived by the model, such as "forget" or "do not remember this" (see Section 4.3.1). Compared

to GUI-based controls on traditional platforms, NL-based control lowers interaction barriers for being more flexible and intuitive [54, 84]. However, NL control can create a tension between control freedom and clarity. Users might feel they have more control over their interactions when, in reality, the inherent ambiguity of NL control can affect their ability to foresee, verify, and review how their data is being handled [68]. Prior work has shown that even in conventional GUI-based privacy interfaces, small mismatches between user intent and system interpretation, such as unclear labels, confusing navigation paths, or insufficient feedback, often confuse users regarding which action actually governs data deletion or other privacy rights [39]. Below, we discuss implications for making NL control more efficient, reliable, and scalable.

5.2.1 Clarifying Control Implications. As shown in prior research, users often expressed confusion and uncertainty when confronted with inconsistent or ambiguous terminology in data control interfaces [25, 36], not to mention on conversational LLM platforms, where control actions involve natural language, which is inherently ambiguous. To improve the clarity of available NL controls, platforms could ***support multi-turn privacy clarification*** rather than treating NL control as a one-shot command. This means enabling users to enter clarifying back-and-forth dialogues with the model to better define what they want to protect, in what way, and for how long. For instance, when users are not clear about what a control option means, they should be supported with the ability to ask questions and receive real-time, intelligible responses. Unlike traditional platforms where users are limited to static documentation or help pages [87, 90, 94], it is technically feasible for LLM-based interfaces to offer in-situ, natural language inquiries, such as "*What does 'forget' mean in this context*?" or "*If I delete this, will it still be remembered later*?" Furthermore, this clarification process can become negotiable and iterative: the model may ask follow-up questions when user intent is ambiguous, confirm actions before execution, or offer options for partial deletion. As such, platforms can dynamically clarify the scope, function, and implications of control actions, responsive to individual users and contexts.

By the time of writing, we noticed that ChatGPT and Gemini have begun to implement similar features by introducing memory widgets that visualize memory-related changes. In particular, ChatGPT now allows users to proactively manage whether saved memory and chat history can be referenced in future interactions through a toggle in the Settings [14]. These interfaces represent a promising step toward greater transparency and legibility and could be further supported through contextual explanations on why certain information is being remembered, and how it may be used in future interactions. For example, platforms may allow users to review, confirm, or undo memory-related changes before they take effect, either through GUI interactions (e.g., clicking an "*undo*" button) or natural language commands (e.g., "*undo memory updates*").

5.2.2 Extending NL Control Beyond Memory. While current implementations of NL control primarily focus on memory-related operations, we see opportunities to ***expand the scope of these NL commands to support a wider range of privacy control operations*** without being limited to memory management functions. For example, users could issue commands like "*Please do not use this chat to train your model*" to opt out of data use for model training or "*Please delete the character I created yesterday morning*" to quickly delete the customized CA. Moreover, LLMs themselves offer a path toward more responsive privacy mechanisms. These models are increasingly able to detect sensitive content and contextual cues during interaction [5]. Rather than relying entirely on user-initiated commands, platforms could ***surface inferred data units that may warrant review or deletion to create a context-aware, system-assisted privacy control environment***, where users and systems collaboratively shape what is stored, remembered, or removed.

Additionally, the platforms can ***proactively suggest privacy options that users might not explicitly mention***, such as not retaining inferred preferences, thereby expanding the scope of privacy control in an intelligent, context-sensitive manner. Such proactive suggestions also serve an important educational role: they can make users aware of the existence and capabilities of NL-based privacy controls. Although LLM providers may gain technical advantages from using user data for model training, they can ultimately strengthen user trust and long-term retention by prioritizing and demonstrating responsible data practices [122]. Therefore, we advocate for coordinated action between companies and regulators: Platforms should operationalize NL controls with clear execution boundaries and verifiable outcomes, and regulators should establish standards that require interpretability, auditability, and minimum functional guarantees for privacy-related NL commands. This collaborative approach will position LLMs as not just passive executors of user commands but active partners in privacy management.

5.3 Governance of "Co-owned" Data

As presented in Table 3, the Privacy Policies of the studied platforms largely follow the established frameworks of other digital platforms. However, their institutional materials do not thoroughly address data-sharing practices or issues associated with data ownership. For instance, Character.ai provides ambiguous statements about the "*popular characters*," as it does not define what counts as a "*popular character*."

To fully understand and design for the mechanisms of sharing among multiple users, it is necessary to ***consider the layers of shared data***. Shared data can include chat sessions, customization instructions, embedded memory, modification histories, generated artifacts, and other derivative content. Actions that users can perform on shared data may involve viewing, continuing interactions, editing, deleting, or re-sharing content. Control over these actions is distributed among sharers, sharees, the platforms, and even third parties, while visibility may extend through shared links, GPT Store listings, web searches, or other channels. Moreover, shared customized objects can become living artifacts: As recipients build upon them, extensions propagate across users, making ownership, attribution, and privacy management increasingly complex.

Our analysis also reveals that shared data occurring on conversational LLM platforms is not solely owned by the sharers (i.e., data creators). Instead, sharees who build upon shared data also acquire ownership stakes, creating overlapping boundaries of control and complicating the allocation of data rights. Ownership and

control are further blurred because models, platform systems, and even third parties are involved in these sharing mechanisms This challenge, to some extent, resonates with those identified in social networking contexts, where interactions and data sharing are inherently multi-user and collaborative [101]. Addressing shared ownership requires more than technical control mechanisms; it also demands transparent governance frameworks tailored for multi-user scenarios to prevent disputes and ensure that all parties understand their rights and responsibilities [43, 70, 81, 104, 105]. The frameworks may include ***collaborative privacy agreements that allow users to negotiate and agree on the shared ownership*** [57]; policies and tools for clear attribution of data ownership and rights, which define and display the ownership of shared data and the rights of each participant [70, 98]; and guidelines [105] that help users to resolve conflicts regarding data use and control.

Moreover, drawing an analogy to GitHub repositories, where sharees can directly view and modify the underlying code [48], customized objects on LLM platforms often conceal the underlying instructions. This invisibility protects the original creator's privacy and creative logic but also limits recipients' understanding and agency over the object, increasing uncertainty regarding data attribution and responsibility. Therefore, LLM platforms need to explore ways to ***provide transparent traceability and controllability while protecting the underlying instructions***. For example, traceable mechanisms could help users track the evolution of objects [18], and configurable permission interfaces could allow creators to specify which aspects of their objects are accessible and extensible by other users. Such designs not only reduce misunderstandings and conflicts arising from control asymmetry but also help users make more informed trade-offs between collaboration and privacy.

6 LIMITATION AND FUTURE WORK

Our study, grounded in an expert-driven application walkthrough, focuses on identifying user data control mechanisms across six widely used conversational LLM platforms between November 2024 and January 2025. That said, at the time of screening, some models had not yet launched their consumer-facing interfaces (e.g., Grok). As such, our findings provide a comparative overview of six popular platforms with temporal limits. However, as mentioned in Section 3.1.2, our goal was not to produce an exhaustive or up-to-date catalog or version-specific of interface features. Rather, we aimed to examine the emerging interaction paradigms implemented by the mainstream platforms to structure user control. Thus, such temporal limits do not compromise the validity of our findings, which hold values in informing design directions that continue to characterize contemporary conversational LLM platforms.

Our analysis focused on platforms primarily developed by organizations headquartered in the U.S. due to considerations around maintaining analytical consistency across platforms. Currently, most major LLM platforms are developed and maintained in the U.S., Europe, and East Asia [67], each shaped by distinct regulatory frameworks (e.g., GDPR in the European Union [93], CCPA in the United States [10], and PIPL in China [82]) and broader debates around AI sovereignty [108]. Such cross-cultural analysis demands much more extensive data collection, translation expertise, and a multi-layered comparative framework to accurately account for the variations in platform design, regulatory compliance, and user expectations across different regions. As shown in prior cross-cultural privacy research, people's expectations of data control vary substantially across cultural contexts, which are shaped by not only regulation but also by local norms, histories of technology governance, and differing interpretations of privacy-related concepts [15, 32, 62]. Therefore, we see a walkthrough of LLM platforms across primary languages, regions, and cultures, an important direction for future research.

Furthermore, we acknowledge that follow-up user studies can provide valuable insights into how people understand and interact with these platform features. However, designing meaningful tasks for such studies first requires a detailed understanding of each platform's settings, options, and data practices. Without this groundwork, user studies risk overlooking important features or privacy-related features or overwhelming users with intensive interaction tasks. As Light et al. noted, walkthroughs "*serve as a foundation for further user-centered research that can identify how users resist these arrangements and appropriate app technology for their own purposes*" [63]. Our walkthrough, therefore, builds the necessary framework to inform opportunities for user studies. For example, future work can examine how NL control aligns with user intentions, explore privacy concerns in multi-user interactions in shared conversations or personas, etc.

7 CONCLUSION

This study walked through privacy control features across six conversational LLM platforms, uncovering unique mechanisms that distinguish them from traditional digital platforms—varying and unique data units, use of natural language commands for control, and shared data ownership. The findings advance our understanding of the usable privacy challenges inherently embedded in human-LLM interactions, and pave the way for future research to develop more transparent and user-friendly control mechanisms.

ACKNOWLEDGMENTS

We thank anonymous reviewers for their thoughtful suggestions. The project was supported by City University of Hong Kong (#9220150 and #7020106). GPT-5 contributed to improving the readability of this manuscript through grammatical corrections and enhancements in language fluency.

REFERENCES

[1] Ruba Abu-Salma and Benjamin Livshits. 2020. Evaluating the End-User Experience of Private Browsing Mode. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376440

[2] Harshvardhan Aditya, Siddansh Chawla, Gunika Dhingra, Parijat Rai, Saumil Sood, Tanmay Singh, Zeba Mohsin Wase, Arshdeep Bahga, and Vijay K Madisetti. 2024. Evaluating Privacy Leakage and Memorization Attacks on Large Language Models (LLMs) in Generative AI Applications. *Journal of Software Engineering and Applications* 17, 5 (2024), 421–447.

[3] Jad Asswad and Jorge Marx Gómez. 2021. Data ownership: a survey. *Information* 12, 11 (2021), 465. https://doi.org/10.3390/info12110465

[4] Sumit Asthana, Jane Im, Zhe Chen, and Nikola Banovic. 2024. "I know even if you don't tell me": Understanding Users' Privacy Preferences Regarding AI-based Inferences of Sensitive Information for Personalization. In *Proceedings*

of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 782, 21 pages. https://doi.org/10.1145/3613904.3642180

[5] Eugene Bagdasarian, Ren Yi, Sahra Ghalebikesabi, Peter Kairouz, Marco Gruteser, Sewoong Oh, Borja Balle, and Daniel Ramage. 2024. AirGapAgent: Protecting Privacy-Conscious Conversational Agents. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (Salt Lake City, UT, USA) *(CCS '24)*. Association for Computing Machinery, New York, NY, USA, 3868–3882. https://doi.org/10.1145/3658644.3690350

[6] Susanne Barth, Dan Ionita, and Pieter Hartel. 2022. Understanding Online Privacy—A Systematic Review of Privacy Visualizations and Privacy by Design Guidelines. *ACM Comput. Surv.* 55, 3, Article 63 (Feb. 2022), 37 pages. https://doi.org/10.1145/3502288

[7] Florian Bemmann, Helena Stoll, and Sven Mayer. 2024. Privacy Slider: Fine-Grain Privacy Control for Smartphones. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 272 (Sept. 2024), 31 pages. https://doi.org/10.1145/3676519

[8] Anne Wells Branscomb. 1994. *Who Owns Information? From Privacy to Public Access.* Basic Books, Inc., USA.

[9] Duc Bui, Brian Tang, and Kang G. Shin. 2022. Do Opt-Outs Really Opt Me Out?. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (Los Angeles, CA, USA) *(CCS '22)*. Association for Computing Machinery, New York, NY, USA, 425–439. https://doi.org/10.1145/3548606.3560574

[10] California Legislature. 2018. California Consumer Privacy Act of 2018 (CCPA). https://cppa.ca.gov/regulations/pdf/ccpa_statute.pdf

[11] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2021. Extracting Training Data from Large Language Models. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Virtual, 2633–2650.

[12] Character.ai. 2024. Character.ai. https://character.ai/. Accessed: 2024-12-01.

[13] ChatGPT. 2024. ChatGPT. https://chatgpt.com/. Accessed: 2024-12-01.

[14] ChatGPT. 2025. Memory FAQ. https://help.openai.com/en/articles/8590148-memory-faq#h_b8745c1ae1 Retrieved in February 2025.

[15] Hichang Cho, Bart Knijnenburg, Alfred Kobsa, and Yao Li. 2018. Collective Privacy Management in Social Media: A Cross-Cultural Validation. *ACM Trans. Comput.-Hum. Interact.* 25, 3, Article 17 (June 2018), 33 pages. https://doi.org/10.1145/3193120

[16] Chris McKay. 2024. Meta AI Hits 400 Million Monthly Users. https://www.maginative.com/article/meta-ai-hits-400-million-monthly-users/. Accessed: 2024-12-01.

[17] Claude. 2024. Claude. https://www.anthropic.com/claude. Accessed: 2024-12-01.

[18] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 1277–1286. https://doi.org/10.1145/2145204.2145396

[19] Sumit Kumar Dam, Choong Seon Hong, Yu Qiao, and Chaoning Zhang. 2024. A Complete Survey on LLM-based AI Chatbots. arXiv:2406.16937 [cs.CL] https://arxiv.org/abs/2406.16937

[20] danah boyd and Eszter Hargittai. 2010. Facebook Privacy Settings: Who Cares? *First Monday* (2010). https://doi.org/10.5210/fm.v15i8.3086

[21] Breno Felix de Sousa, Ronnie de Souza Santos, and Kiev Gama. 2025. Integrating Positionality Statements in Empirical Software Engineering Research. In *Proceedings of the 2025 IEEE/ACM International Workshop on Methodological Issues with Empirical Studies in Software Engineering* (Ottawa, Ontario, Canada) *(WSESE '25)*. IEEE Press, 28–35. https://doi.org/10.1109/WSESE66602.2025.00012

[22] Iliana Depounti, Paula Saukko, and Simone Natale. 2023. Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture & Society* 45, 4 (2023), 720–736. https://doi.org/10.1177/01634437221119021

[23] Michael Dieter and Nathaniel Tkacz. 2020. The Patterning of Finance/security: A Designerly Walkthrough of Challenger Banking Apps. *Computational Culture: A Journal of Software Studies* 7 (2020). http://computationalculture.net/the-patterning-of-finance-security/

[24] Nico Ebert, Kurt Alexander Ackermann, and Björn Scheppler. 2021. Bolder is Better: Raising User Awareness through Salient and Concise Privacy Notices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 67, 12 pages. https://doi.org/10.1145/3411764.3445516

[25] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. 2017. Large-scale Readability Analysis of Privacy Policies. In *Proceedings of the International Conference on Web Intelligence* (Leipzig, Germany) *(WI '17)*. Association for Computing Machinery, New York, NY, USA, 18–25. https://doi.org/10.1145/3106426.3106427

[26] Fabio Duarte. 2024. Number of ChatGPT Users (Oct 2024). https://explodingtopics.com/blog/chatgpt-users. Accessed: 2024-12-01.

[27] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android Permissions: User Attention, Comprehension, and Behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. https://doi.org/10.1145/2335356.2335360

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

[29] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International journal of qualitative methods* 5, 1 (2006), 80–92. https://doi.org/10.1177/160940690600500107

[30] Patricia I Fusch Ph D and Lawrence R Ness. 2015. Are we there yet? Data saturation in qualitative research. *The Qualitative Report* (2015). https://doi.org/10.46743/2160-3715/2015.2281

[31] Gemini. 2024. Gemini. https://gemini.google.com/app. Accessed: 2024-12-01.

[32] Reza Ghaiumy Anaraky, Yao Li, and Bart Knijnenburg. 2021. Difficulties of Measuring Culture in Privacy Studies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 378 (Oct. 2021), 26 pages. https://doi.org/10.1145/3479522

[33] Goody2. 2024. Goody2. https://www.goody2.ai/chat. Accessed: 2024-12-01.

[34] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. https://doi.org/10.1145/3411764.3445779

[35] Yanzhu Guo, Simone Conia, Zelin Zhou, Min Li, Saloni Potdar, and Henry Xiao. 2024. Do Large Language Models Have an English Accent? Evaluating and Improving the Naturalness of Multilingual LLMs. arXiv:2410.15956 [cs.CL] https://arxiv.org/abs/2410.15956

[36] Hana Habib. 2021. *Evaluating the Usability of Privacy Choice Mechanisms.* Ph. D. Dissertation. Carnegie Mellon University, USA.

[37] Hana Habib and Lorrie Faith Cranor. 2022. Evaluating the Usability of Privacy Choice Mechanisms. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 273–289. https://www.usenix.org/conference/soups2022/presentation/habib

[38] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 621, 27 pages. https://doi.org/10.1145/3491102.3501985

[39] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. "It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376511

[40] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 387–406. https://www.usenix.org/conference/soups2019/presentation/habib

[41] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 2021. Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 63, 25 pages. https://doi.org/10.1145/3411764.3445387

[42] Andrew Gary Darwin Holmes. 2020. Researcher Positionality—A Consideration of Its Influence and Place in Qualitative Research—A New Researcher Guide. *Shanlax International Journal of Education* 8, 4 (2020), 1–10. https://doi.org/10.34293/education.v8i4.3232

[43] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2013. Multiparty Access Control for Online Social Networks: Model and Mechanisms. *IEEE Transactions on Knowledge and Data Engineering* 25, 7 (2013), 1614–1627. https://doi.org/10.1109/TKDE.2012.97

[44] Jane Im, Ruiyi Wang, Weikun Lyu, Nick Cook, Hana Habib, Lorrie Faith Cranor, Nikola Banovic, and Florian Schaub. 2023. Less is Not More: Improving Findability and Actionability of Privacy Controls for Online Behavioral Advertising. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 661, 33 pages. https://doi.org/10.1145/3544548.3580773

[45] Inflection AI. 2024. Inflection-2.5: meet the world's best personal AI. https://inflection.ai/blog/inflection-2-5. Accessed: 2024-12-01.

[46] Umar Iqbal, Tadayoshi Kohno, and Franziska Roesner. 2025. *LLM Platform Security: Applying a Systematic Evaluation Framework to OpenAI's ChatGPT Plugins.* AAAI Press, 611–623.

[47] Benjamin N Jacobsen. 2022. 'You can't delete a memory': Managing the Data Past on Social Media in Everyday Life. *Sociological Research Online* 27, 4 (2022), 1003–1019. https://doi.org/10.1177/13607804221110237

[48] Jing Jiang, David Lo, Jiahuan He, Xin Xia, Pavneet Singh Kochhar, and Li Zhang. 2017. Why and How Developers Fork What from Whom in GitHub. *Empirical Software Engineering* 22, 1 (2017), 547–578. https://doi.org/10.1007/s10664-016-9436-6

[49] Eunkyung Jo, Yuin Jeong, Sohyun Park, Daniel A. Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 440, 21 pages. https://doi.org/10.1145/3613904.3642420

[50] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of The Internet and Implications for Privacy and Security. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. Usenix Association, Ottawa, Canada, 39–52. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

[51] Vijay Khatri and Carol V Brown. 2010. Designing data governance. *Commun. ACM* 53, 1 (2010), 148–152. https://doi.org/10.1145/1629175.1629210

[52] Sunder Ali Khowaja, Parus Khuwaja, Kapal Dev, Weizheng Wang, and Lewis Nkenyereye. 2024. Chatgpt Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. *Cognitive Computation* (2024), 1–23. https://doi.org/10.1007/s12559-024-10285-1

[53] Youjeong Kim and S Shyam Sundar. 2012. Anthropomorphism of Computers: Is It Mindful or Mindless? *Computers in Human Behavior* 28, 1 (2012), 241–250.

[54] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. 2021. Data@Hand: Fostering Visual Exploration of Personal Data on;Smartphones Leveraging Speech and Touch Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 462, 17 pages. https://doi.org/10.1145/3411764.3445421

[55] Jan H. Klemmer, Stefan Albert Horstmann, Nikhil Patnaik, Cordelia Ludden, Cordell Burton, Carson Powers, Fabio Massacci, Akond Rahman, Daniel Votipka, Heather Richter Lipford, Awais Rashid, Alena Naiakshina, and Sascha Fahl. 2024. Using AI Assistants in Software Development: A Qualitative Study on Security Practices and Concerns. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (Salt Lake City, UT, USA) *(CCS '24)*. Association for Computing Machinery, New York, NY, USA, 2726–2740. https://doi.org/10.1145/3658644.3690283

[56] Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illenčík, and Celeste Campos-Castillo. 2024. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society* 26, 10 (2024), 5923–5941. https://doi.org/10.1177/14614448221142007

[57] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3217–3226. https://doi.org/10.1145/1978942.1979420

[58] Hyunsoo Lee, Yugyeong Jung, Hei Yiu Law, Seolyeong Bae, and Uichin Lee. 2024. PriviAware: Exploring Data Visualization and Dynamic Privacy Control Support for Data Collection in Mobile Sensing Research. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 787, 17 pages. https://doi.org/10.1145/3613904.3642815

[59] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. 2012. Why Johnny can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 589–598. https://doi.org/10.1145/2207676.2207759

[60] Tianshi Li, Sauvik Das, Hao-Ping (Hank) Lee, Dakuo Wang, Bingsheng Yao, and Zhiping Zhang. 2024. Human-Centered Privacy Research in the Age of Large Language Models. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 581, 4 pages. https://doi.org/10.1145/3613905.3643983

[61] Yao Li, Xinning Gui, Yunan Chen, Heng Xu, and Alfred Kobsa. 2018. When SNS Privacy Settings Become Granular: Investigating Users' Choices, Rationales, and Influences on Their Social Experience. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 108 (Nov. 2018), 21 pages. https://doi.org/10.1145/3274377

[62] Yao Li, Eugenia Ha Rim Rho, and Alfred Kobsa. 2022. Cultural Differences in the Effects of Contextual Factors and Privacy Concerns on Users' Privacy Decision on Social Networking Sites. *Behaviour & Information Technology* 41, 3 (2022), 655–677. https://doi.org/10.1080/0144929X.2020.1831608

[63] Ben Light, Jean Burgess, and Stefanie Duguay. 2018. The Walkthrough Method: An Approach to the Study of Apps. *New media & society* 20, 3 (2018), 881–900. https://doi.org/10.1177/1461444816675438

[64] Sungjin Lim and Junhyoung Oh. 2025. Navigating Privacy: A Global Comparative Analysis of Data Protection Laws. *IET Information Security* 2025, 1 (2025), 5536763. https://doi.org/10.1049/ise2/5536763

[65] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. 2014. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security* (Menlo Park, CA) *(SOUPS '14)*. USENIX Association, USA, 199–212. https://www.usenix.org/conference/soups2014/proceedings/presentation/lin

[66] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Zhang, Norman Sadeh, Alessandro Acquisti, and Yuvraj Agarwal. 2016. Follow My Recommendations: A Personalized Privacy Assistant For Mobile App Permissions. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security* (Denver, CO, USA) *(SOUPS '16)*. USENIX Association, USA, 27–41.

[67] Lmarena.AI. 2024. Chatbot Arena (formerly LMSYS): Free AI Chat to Compare and Test Best AI Chatbots. https://lmarena.ai/. Accessed: 2024-12-01.

[68] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[69] Nils Lukas, Ahmed Salem, Robert Sim, Shruti Tople, Lukas Wutschitz, and Santiago Zanella-Béguelin. 2023. Analyzing Leakage of Personally Identifiable Information in Language Models. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, 346–363.

[70] Rongjun Ma, Caterina Maidhof, Juan Carlos Carrillo, Janne Lindqvist, and Jose Such. 2025. Privacy Perceptions of Custom GPTs by Users and Creators. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 237, 18 pages. https://doi.org/10.1145/3706598.3713540

[71] Thomas Mahatody, Mouldi Sagar, and Christophe Kolski. 2010. State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions. *International Journal of Human–Computer Interaction* 26, 8 (2010), 741–785. https://doi.org/10.1080/10447311003781409

[72] Lisa Mekioussa Malki, Ina Kaleva, Dilisha Patel, Mark Warner, and Ruba Abu-Salma. 2024. Exploring Privacy Practices of Female mHealth Apps in a Post-Roe World. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 576, 24 pages. https://doi.org/10.1145/3613904.3642521

[73] Meta AI. 2024. Meta AI. https://www.meta.ai/. Accessed: 2024-12-01.

[74] Sam Moradzadeh and Yubo Kou. 2024. "Wow another fake game from YouTube ad": Unpacking Fake Games Through a Mixed-Methods Investigation. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY, Article 350 (Oct. 2024), 36 pages. https://doi.org/10.1145/3677115

[75] Ambar Murillo, Andreas Kramm, Sebastian Schnorf, and Alexander De Luca. 2018. "If I press delete, it's gone" - User Understanding of Online Data Deletion and Expiration. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Baltimore, MD, 329–339. https://www.usenix.org/conference/soups2018/presentation/murillo

[76] Naveen Kumar. 2024. Character AI Statistics (2024) — 20 Million Active Users. https://www.demandsage.com/character-ai-statistics/. Accessed: 2024-12-01.

[77] Naveen Kumar. 2024. Google Gemini Statistics 2024 — Active Users Data. https://www.demandsage.com/google-gemini-statistics/. Accessed: 2024-12-01.

[78] Giorgos Nikolaou, Tommaso Mencattini, Donato Crisostomi, Andrea Santilli, Yannis Panagakis, and Emanuele Rodola. 2025. Language Models are Injective and Hence Invertible. *arXiv preprint arXiv:2510.15511* (2025). https://doi.org/10.48550/arXiv.2510.15511

[79] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[80] Kobbi Nissim and Alexandra Wood. 2018. Is Privacy Privacy? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2128 (2018), 20170358. https://doi.org/10.1098/rsta.2017.0358

[81] Farzad Nourmohammadzadeh Motlagh, Seyed Ali Alhosseini, Feng Cheng, and Christoph Meinel. 2023. An Approach to Multi-Party Privacy Conflict Resolution for Co-owned Images on Content Sharing Platforms. In *Proceedings of the 2023 8th International Conference on Machine Learning Technologies* (Stockholm, Sweden) *(ICMLT '23)*. Association for Computing Machinery, New York, NY, USA, 264–268. https://doi.org/10.1145/3589883.3589923

[82] National People's Congress of the People's Republic of China (NPC). 2021. Personal Information Protection Law (PIPL). https://personalinformationprotectionlaw.com/.

[83] Daisy O'Neill, Max V. Birk, and Regan L Mandryk. 2024. Unpacking Norms, Narratives, and Nourishment: A Feminist HCI Critique on Food Tracking Technologies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 514, 20 pages. https://doi.org/10.1145/3613904.3642600

[84] Somin Park, Xi Wang, Carol C Menassa, Vineet R Kamat, and Joyce Y Chai. 2024. Natural Language Instructions for Intuitive Human Interaction with Robotic Assistants in Field Construction Work. *Automation in Construction* 161 (2024), 105345. https://doi.org/10.1016/j.autcon.2024.105345

[85] Pi. 2024. Pi. https://pi.ai/. Accessed: 2024-12-01.

[86] Poe. 2024. Poe. https://poe.com/. Accessed: 2024-12-01.

[87] Irene Pollach. 2007. What's Wrong with Online Privacy Policies? *Commun. ACM* 50, 9 (Sept. 2007), 103–108. https://doi.org/10.1145/1284621.1284627

[88] Luca M Possati. 2023. Psychoanalyzing artificial intelligence: The case of Replika. *Ai & Society* 38, 4 (2023), 1725–1738. https://doi.org/10.1007/s00146-021-01379-7

[89] Ananya Reddy and Priya C. Kumar. 2024. 'A Teaspoon of Authenticity': Exploring How Young Adults BeReal on Social Media. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 907, 14 pages. https://doi.org/10.1145/3613904.3642690

[90] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, Rohan Ramanath, et al. 2015. Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding. *Berkeley Tech. LJ* 30 (2015), 39. https://heinonline.org/HOL/P?h=hein.journals/berktech30&i=51

[91] Lara Reime, Vasiliki Tsaknaki, and Marisa Leavitt Cohn. 2023. Walking Through Normativities of Reproductive Bodies: A Method for Critical Analysis of Tracking Applications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 658, 15 pages. https://doi.org/10.1145/3544548.3581450

[92] Replika. 2024. Replika. https://replika.com/. Accessed: 2024-12-01.

[93] The European Parliament and The Council of The European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2016/679/oj#

[94] André de Lima Salgado, Patrick C. K. Hung, and Renata P. M. Fortes. 2024. Six Usable Privacy Heuristics. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems* (Maceió, Brazil) *(IHC '23)*. Association for Computing Machinery, New York, NY, USA, Article 43, 11 pages. https://doi.org/10.1145/3638067.3638111

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

[96] Theodor Schnitzler, Christine Utz, Florian M Farke, Christina Pöpper, and Markus Dürmuth. 2020. Exploring User Perceptions of Deletion in Mobile Instant Messaging Applications. *Journal of Cybersecurity* 6, 1 (2020), tyz016. https://doi.org/10.1093/cybsec/tyz016

[97] Yashothara Shanmugarasa, Ming Ding, Chamikara Mahawaga Arachchige, and Thierry Rakotoarivelo. 2025. SoK: The Privacy Paradox of Large Language Models: Advancements, Privacy Risks, and Mitigation. In *Proceedings of the 20th ACM Asia Conference on Computer and Communications Security (ASIA CCS '25)*. Association for Computing Machinery, New York, NY, USA, 425–441. https://doi.org/10.1145/3708821.3733888

[98] Mina Sheikhalishahi, Gamze Tillem, Zekeriya Erkin, and Nicola Zannone. 2019. Privacy-Preserving Multi-Party Access Control. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society* (London, United Kingdom) *(WPES'19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3338498.3358643

[99] Justin Smith, Lisa Nguyen Quang Do, and Emerson Murphy-Hill. 2020. Why Can't Johnny Fix Vulnerabilities: A Usability Evaluation of Static Analysis Tools for Security. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. Usenix Association, 221–238. https://www.usenix.org/conference/soups2020/presentation/smith

[100] Jeff Sovern. 1999. Opting In, Opting Out, or No Options At All: The Fight for Control of Personal Information. *Wash. L. Rev.* 74 (1999), 1033. https://heinonline.org/HOL/P?h=hein.journals/washlr74&i=1047

[101] Anna C Squicciarini, Mohamed Shehab, and Joshua Wede. 2010. Privacy Policies for Shared Content in Social Network Sites. *The VLDB Journal* 19 (2010), 777–796. https://doi.org/10.1007/s00778-010-0193-7

[102] Robin Staab, Mark Vero, Mislav Balunović, and Martin Vechev. 2023. Beyond Memorization: Violating Privacy via Inference with Large Language Models. *arXiv preprint arXiv:2310.07298* (2023). https://doi.org/10.48550/arXiv.2310.07298

[103] Statcounter. 2024. Search Engine Market Share United States Of America. https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america/. Accessed: 2024-12-01.

[104] Jose M. Such and Natalia Criado. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Transactions on Knowledge and Data Engineering* 28, 7 (2016), 1851–1863. https://doi.org/10.1109/TKDE.2016.2539165

[105] Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-scale Empirical Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3821–3832. https://doi.org/10.1145/3025453.3025668

[106] Guangzhi Sun, Xiao Zhan, and Jose Such. 2024. Building Better AI Agents: A Provocation on the Utilisation of Persona in LLM-based Conversational Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 35, 6 pages. https://doi.org/10.1145/3640794.3665887

[107] Chunliang Tao, Xiaojing Fan, and Yahe Yang. 2024. Harnessing LLMs for API Interactions: A Framework for Classification and Synthetic Data Generation. *arXiv preprint arXiv:2409.11703* (2024).

[108] Paul Timmers. 2019. Ethics of AI and Cybersecurity When Sovereignty is at Stake. *Minds and Machines* 29, 4 (2019), 635–645. https://doi.org/10.1007/s11023-019-09508-4

[109] Batuhan Tömekçe, Mark Vero, Robin Staab, and Martin Vechev. 2024. Private Attribute Inference from Images with Vision-Language Models. arXiv:2404.10618 [cs.AI] https://arxiv.org/abs/2404.10618

[110] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 973–990. https://doi.org/10.1145/3319535.3354212

[111] Viraj Mahajan. 2024. 80+ Important Claude Statistics to Know in 2024. https://www.notta.ai/en/blog/claude-statistics. Accessed: 2024-12-01.

[112] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 214–229. https://doi.org/10.1145/3531146.3533088

[113] Alan F Westin. 1968. Privacy and Freedom. *Washington and Lee Law Review* 25, 1 (1968), 166. https://scholarlycommons.law.wlu.edu/wlulr/vol25/iss1/20

[114] Alma Whitten and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8* (Washington, D.C.) *(SSYM'99)*. USENIX Association, USA, 14.

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

[116] Maximiliane Windl, Niels Henze, Albrecht Schmidt, and Sebastian S. Feger. 2022. Automating Contextual Privacy Policies: Design and Evaluation of a Production Tool for Digital Consumer Privacy Awareness. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 34, 18 pages. https://doi.org/10.1145/3491102.3517688

[117] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[118] Wordtune. 2024. Wordtune. https://www.wordtune.com/. Accessed: 2024-12-01.

[119] Fangzhou Wu, Ning Zhang, Somesh Jha, Patrick McDaniel, and Chaowei Xiao. 2024. A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems. arXiv:2402.18649 [cs.CR] https://arxiv.org/abs/2402.18649

[120] Xiaodong Wu, Ran Duan, and Jianbing Ni. 2024. Unveiling Security, Privacy, and Ethical Concerns of ChatGPT. *Journal of Information and Intelligence* 2, 2 (2024), 102–115. https://doi.org/10.1016/j.jiixd.2023.10.007

[121] xAI. 2025. Grok 4. https://x.ai/news/grok-4. Accessed: 2025-09-01.

[122] Meihe Xu, Arianna Rossi, and Aurelia Tamò-Larrieux. 2025. The Future of Personalized Privacy Assistants: Gathering of Expert Opinions. *Digital Society* 4, 3 (2025), 1–32. https://doi.org/10.1007/s44206-025-00232-4

[123] Xinyu Yang, Zichen Wen, Wenjie Qu, Zhaorun Chen, Zhiying Xiang, Beidi Chen, and Huaxiu Yao. 2024. Memorization and Privacy Risks in Domain-Specific Large Language Models. https://openreview.net/forum?id=KmW8WkCKRx

[124] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. 2024. A Survey on Large Language Model (LLM) Security and Privacy: The Good, The Bad, and The Ugly. *High-Confidence Computing* 4, 2 (2024), 100211. https://doi.org/10.1016/j.hcc.2024.100211

[125] Hanna Yukhymenko, Robin Staab, Mark Vero, and Martin Vechev. 2024. A Synthetic Dataset for Personal Attribute Inference. *Advances in Neural Information Processing Systems* 37 (2024), 120735–120779.

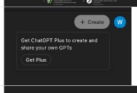
[126] Günce Su Yılmaz, Fiona Gasaway, Blase Ur, and Mainack Mondal. 2021. Perceptions of Retrospective Edits, Changes, and Deletion on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 15, 1 (May 2021), 841–852. https://doi.org/10.1609/icwsm.v15i1.18108

[127] Xiao Zhan, William Seymour, and Jose Such. 2024. Beyond Individual Concerns: Multi-user Privacy in Large Language Models. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 34, 6 pages. https://doi.org/10.1145/3640794.3665883

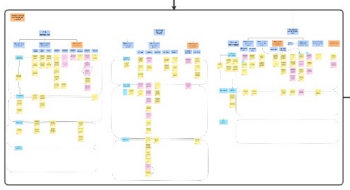
[128] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 156, 26 pages. https://doi.org/10.1145/3613904.3642385

[129] Yanjie Zhao, Xinyi Hou, Shenao Wang, and Haoyu Wang. 2024. LLM App Store Analysis: A Vision and Roadmap. arXiv:2404.12737 [cs.SE] https://arxiv.org/abs/2404.12737

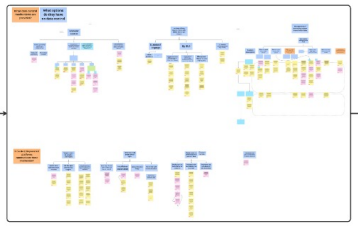
A WALKTHROUGH LOG EXAMPLES AND HYBRID THEMATIC ANALYSIS FOR THE DATA FROM THE TECHNICAL WALKTHROUGH.

	Platform	Short feature description	Derive time	Links of the sources	Screenshots / Sources	Field notes	(Initial) Codes
							Sometimes, I use the exact wording from the field notes as codes when rephrasing doesn't add clarity. Each field note can be linked to several codes.
187	ChatGPT (Website)	GPTs	Nov 25 202	chatgpt.com		They provide GPTs function, where users can "Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills." - Users can search the GPTs they want through the searching bar. Or they can browse the featured or trending GPTs. - In the GPTs section, they provide some tabs labeling the main purposes of those GPTs, including writing, productivity, research&analysis, education, lifestyle, programming	ChatGPT provides users with customized GPTs for different usage/purposes that customized by other users or ChatGPT team. This is for offering GPT that combines extra instructions, knowledge and skills.
188	ChatGPT (Website)	GPTs	Nov 25 202	chatgpt.com		"Get ChatGPT Plus to create and share your own GPTs."	ChatGPT Plus users can creat and share their own GPTs
189	ChatGPT (Website)	GPTs	Nov 25 202	chatgpt.com		When chating with GPTs, the initial interface will show the name of this GPTs and its creator. There's a globe icon next to the creator's name, and when the cursor hovers on it, it says "The builder of this GPT cannot view your conversations. "	The builder of GPT cannot view its users' conversations

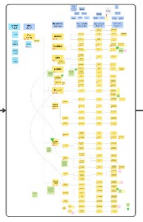
(a) Walkthrough log, using some screenshots from ChatGPT as examples. The first author conducted the application walkthrough following the protocols and documenting the platform name, a short decription of the feature, derive time, link of the source, screenshot, field notes and initial codes.



(b) First round of inductive process of coding and clustering for the data from ChatGPT and Character.ai. We identified themes: "*What can users opt-in/opt-out/access/edit/delete/ share*," "*how can users opt-in/out/edit/delete/share*," and "*how did first-party communicate these mechanisms*."



(c) Second round of inductive process of coding and clustering for the data from ChatGPT and Character.ai. We identified themes: "*What options do users have on data control (e.g., granular control, control by graphic or natural language interfaces, management of interaction snippet shared with others),*" and "*how do LLM-powered platforms communicate those mechanisms.*"



(d) All authors then met and discussed three categories for the next iteration of coding and clustering: "*Data type and unit*," "*Control operations*," and "*Control Executions*." The first author then clustered the codes for the remaining platforms. The third round of coding and clustering, conducted deductively using the established categories, is shown on the right. And we did not discover any new categories comes up.

[All] All the six platforms, [1] Character.ai, [2] ChatGPT, [3] Claude, [4] Gemini, [5] Meta AI, [6] Pi

How to control	Data Unit	Example
"Memorize", "Forget", and **"Update the memory"** (a) by using natural language across all the chat sessions; (b) by "pinning"/ "unpinning"/ "editing" a message; (c) by directly entering/editing the information in the "saved info" portal	(a) Memory snippet [2,4] (pieces of information that are deemed to be important about the user); (b) Message [1]; (c) Memory snippet [4]	(a) Users can use natural language commands (e.g., "remember ...," "I prefer ...," "forget ...," "... changed") to manage memory in any chat [2,4], but it is uncertain whether these attempts will succeed.
"Access memory" or **"Forget"** (a) by entering "Settings" or "Saved info" portal; (b) by entering "Pinned" in the side panel of the chat window	(a) Memory snippet [2,4]; (b) Message [1]	(a) Users can directly access or delete "memory", by entering "Settings" > "Personalization" > "Memory - Manage" [2].

[All] All the six platforms, [1] Character.ai, [2] ChatGPT, [3] Claude, [4] Gemini, [5] Meta AI, [6] Pi

How to control	Data unit	Example
Access by clicking "edit" in the customization portal (e.g., "public profile"[1], "My GPTs"[2,3], "Projects"[3], "Gem Manager"[5])	CA's persona [1,2,4], User's persona [1], Project [3], Response style [3], Version history of customized persona [2], CA's voice [1]	The version history of GPT customizations can be accessed through the user's profile by following these steps: navigate to the top-right corner, click on "My GPTs," select the "edit" icon next to the desired custom GPT, click the three-dot icon in the top-right corner, and then choose "Version history." [1]
Retrieve (a) by selecting the customized items and using natural language to chat with them in the side panel of the chat window; (b) by selecting the customized items in the side panel	(a) Extracted information in Response style [2], CA's persona [1,2,4], Project [2]; (b) Chatbot's voice [1], User's persona [1]	(a) For example, if a user customizes a chatbot persona (i.e., character) as "You are the philosopher Aristotle. You have wisdom on observing the world and providing provocative insights." Users can retrieve the information from such customization records by using natural language queries, such as "Who are you?" and "How can you help me?" by selecting this character and chatting with it [1].
Edit by clicking "edit" in the customization portal	CA's persona [1,2,4], User's persona [1], Customized project [3], Version history of customized persona [2], CA's voice [1]	Users can edit the customization records of the project by entering "Projects" in the side panel of the chat, clicking the three-dot icon in the top-right corner, and clicking "Edit details." [3]
Delete in the customization portal	Chatbot's persona [2,4], File in Project [3], Chat session in Project [3], Response style [2], Version history of customized CA persona [2], CA's voice [1]	To delete a customized response style, users can click the downward arrow below the text input field, select the "Create & Edit Styles" button, and then click the trash bin icon [2].
Remove in the customization portal	User's persona [1]	By clicking the three-dot next to the user's persona in "public profile", users can remove the customized user's persona from the public profile [1].
Archive by clicking the three-dot icon in the project window	Customized project [3]	To archive the customized project, the user can click the three-dot icon in the project window and select "archive." [3]

(e) All authors then met to review and discuss the coding results. The first and corresponding authors then organized all codes and themes into tables, labeled the platforms, and included specific examples.

[All] All the six platforms, [1] Character.ai, [2] ChatGPT, [3] Claude, [4] Gemini, [5] Meta AI, [6] Pi

How to control	Data Unit	Example
Access (a) in the side panel next to the chat window; (b) by keyword searching in the chat session panel's search bar	(a) Chat session/thread [All], Chat partner [1]; (b) Chat session [3]	(a) Users can access the latest chat session with a chat character [1]. (b) Users can find the specific chat sessions by searching keywords [3].
Retrieve by using natural language (a) across all chat sessions; (b) across all chat sessions with the same chat partner; (c) within a chat session	(a) Extracted information [2,4]; (b) Extracted information [3], (c) Extracted information [All]	For example, if a user sends the message, "My name is Johnny, a frequent user of LLMs," they can retrieve the extracted information (e.g., the name "Johnny") using natural language queries, such as "What is my name?" (b) within chats with the same chat partner, if the user has "pinned" the original message they sent [1]. But it is uncertain whether these attempts will succeed.
Edit by clicking the pencil icon next to the corresponding item	(a) Message [1,3,4]; (b) Image [5]	(a) Users can edit the messages they sent in the chat sessions by clicking the edit icon next to the corresponding message [3,4].
Delete (a) by clicking the option "delete" of the corresponding message in the chat window; (b) by clicking the option "delete" in the chat session panel; (c) in account settings	(a) Message [1]; (b) Chat session [2,3,4,5], Multiple chat sessions [3], All the chat history [3]; (c) Conversational round [4], All the conversational rounds within a period [4], All the chat history [2,4,5]	(c) In the "Gemini Apps Activity" page, users have the option to auto-delete activity in the Gemini app, with timeframes of "3 months", "18 months" (default), "36 months", or the choice to "Not auto-delete activity". Additionally, users can manually delete the chat history by selecting "Last hour", "Last day", "Always", or a "Custom range" [4].
"Remove" from the side panel next to the chat window (Note that "remove" does not mean deleted, users can still access the message later.)	CA [3]	To remove a chat history from the sidebar, users can click the three-dot icon next to the chat history with a specific chat partner in the left-side panel, then select "Remove" [3].
Export the chat history through the account settings	All the chat history [1,2,3,5,6]	By clicking the user profile picture, entering "Settings" > "Data Controls" and clicking "Export", users can export all the chat history [2].
Archive (a) by clicking the "archive" option; (b) in account settings	(a) Chat session [1,2], Project [3]; (b) All the chat history [3]	(b) By clicking the user profile picture, entering "Settings" > "General" and clicking "Archive all chats - Archive, " users can archive all the chat history [2].
Unarchive or **delete archived chat** in account settings	Chat session [2]	Users can manage their archived chat sessions by unarchiving them using the "Out" icon or permanently deleting them by clicking the trash bin icon in the settings [2].

The data control mechanisms (i.e., control operations, control interfaces, and data units that can be controlled) of chat history, along with corresponding examples. Each indexed symbol (e.g., (a), (b)) in the first column aligns with the corresponding data units and examples in the second and third columns, respectively. Superscripts denote different platforms, as specified at the beginning of the table. Note that, due to space constraints, examples are selectively presented.

(f) All authors then met to review and discuss the organized tables, and subsequently developed the final themes and sub-themes presented in the Findings section.