Evaluating LLMs Towards Automated Assessment of Privacy Policy Understandability

Keika Mori
Deloitte Tohmatsu Cyber LLC,
Waseda University
Tokyo, Japan
keika.mori@tohmatsu.co.jp

Daiki Ito
Deloitte Tohmatsu Cyber LLC
Tokyo, Japan

Takumi Fukunaga
Deloitte Tohmatsu Cyber LLC
Tokyo, Japan

Takuya Watanabe
Deloitte Tohmatsu Cyber LLC
Tokyo, Japan

Yuta Takata
Deloitte Tohmatsu Cyber LLC
Tokyo, Japan

Masaki Kamizono
Deloitte Tohmatsu Cyber LLC
Tokyo, Japan

Tatsuya Mori
Waseda University, NICT, RIKEN AIP
Tokyo, Japan

***Abstract*—Companies publish privacy policies to improve transparency regarding the handling of personal information. A discrepancy between the description of the privacy policy and the user's understanding can lead to a risk of a decrease in trust. Therefore, in creating a privacy policy, the user's understanding of the privacy policy should be evaluated. However, the periodic evaluation of privacy policies through user studies takes time and incurs financial costs. In this study, we investigated the understandability of privacy policies by large language models (LLMs) and the gaps between their understanding and that of users, as a first step towards replacing user studies with evaluation using LLMs. Obfuscated privacy policies were prepared along with questions to measure the comprehension of LLMs and users. In comparing the comprehension levels of LLMs and users, the average correct answer rates were 85.2% and 63.0%, respectively. The questions that LLMs answered incorrectly were also answered incorrectly by users, indicating that LLMs can detect descriptions that users tend to misunderstand. By contrast, LLMs understood the technical terms used in privacy policies, whereas users did not. The identified gaps in comprehension between LLMs and users, provide insights into the potential of automating privacy policy evaluations using LLMs.**

I. Introduction

A privacy policy is the primary means by which companies and organizations publicly disclose their objectives and procedures regarding how they collect, use, and protect personal information. Under the General Data Protection Regulation (GDPR), several sanctions have been imposed on companies failing to adequately inform users of their handling of personal information [1], [2]. In Japan, following the revision of the Personal Information Protection Act in April 2022, the standards for privacy compliance became more stringent. To avoid the risk of losing users' trust or incurring fines, companies and organizations are required to create appropriate privacy policies that ensure consistency between users' perceptions and actual practices.

The primary reasons privacy policies fail to effectively communicate with users are attributed to document quality and the inherent characteristics of privacy policies. Anca et al. [3] revealed that illogical document structures and complex expressions such as double negatives hinder user understanding. Tang et al. [4] indicated that the use of technical terms in privacy policies reduces comprehension among users without domain-specific knowledge. Given the rapidly evolving nature of various services, it is essential to accurately reflect the handling of personal information in privacy policies that often suffer from insufficient and inconsistent explanations [5], [6].

The ultimate goal of this study is to automatically evaluate the understandability of privacy policies for users. To achieve this, we measured the comprehension levels of both LLMs and users regarding the descriptions in privacy policies. By identifying the similarities and differences in privacy policy comprehension, we addressed the following research questions:

RQ1 To what extent do LLMs understand privacy policies?
RQ2 What gaps lie between LLMs and users in their understanding of privacy policies?

To answer these questions, we created custom privacy policies as analysis targets and performed a comparative analysis of the comprehension levels of three LLMs and 449 participants in our user study. A baseline policy was created based on app marketplace descriptions, intentionally incorporating 11 factors that reduce user understanding. Using comprehension questions that focused on statements related to these factors, we identified the factors that led to similar misunderstandings between LLMs and users along with factors that only LLMs or users can understand, further discussing methods for bridging the gap between LLMs and users, such as prompt engineering and personal settings, to facilitate automated assessments of understandability.

The contributions of this study are as follows:

Symposium on Usable Security and Privacy (USEC) 2025
24 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-5-9
https://dx.doi.org/10.14722/usec.2025.23009
www.ndss-symposium.org, https://www.usablesecurity.net/USEC/

TABLE I
KEY FACTORS HINDERING USER UNDERSTANDING OF PRIVACY POLICIES.

		Applicability to Policy					Questions Affected by Obfuscation				
Factor	Reference	A	B	C	D	E	A	B	C	D	E
Issues caused by writing											
Use of Double Negative	Yan et al. [7]	–	✓	✓	–	–		2	2		
Many Words per Sentences or Paragraphs	Yan et al. [7]	–	✓	–	–	–		3			
Illogical Presentation Order	Yan et al. [7]	–	✓	–	–	–		2			
Dispersed Information	Original	–	✓	–	✓	✓		2		1,2	2,5
Issues related to privacy policy characteristics											
Use of Technical Terms	Tang et al. [4]	–	–	✓	–	–			3		
Inconsistency between Paragraphs	Andow et al. [6]	–	–	–	✓	–				1,2,4	
Missing Information (Decoupling)	Hara et al. [8]	–	–	–	–	✓					1,4
Missing Information (Abstract Expression)	Matsuo et al. [9]	–	✓	✓	✓	✓		7	1,2,4,7	7	4,7
Omitting Information by Reference	Original	–	–	✓	–	✓			2,4		2,4
Description on Handling not Conducted	Original	–	–	✓	–	–			2,5		
Description on Unnecessary Information	Original	–	–	✓	✓	✓			6	6	6

- Based on a comparative analysis of privacy policy comprehension between three types of LLMs and 449 users, the correct answer rates in the study conducted in survey format were 85.2% for LLMs and 63.0% for users, indicating that LLMs outperformed users in comprehension.
- In case of dispersed or missing information, the comprehension levels of both LLMs and users decreased.
- User comprehension levels decreased because of a lack of technical term knowledge and overlooked information, whereas this tendency was not observed with LLMs.
- We conducted a case study applying the current LLMs to privacy policies in real world. We observed the same tendency as mentioned above, as well as new factors hindering LLM's understanding such as tabular format, referencing other pages, and differences in expression within privacy policies and question texts.
- Based on the comprehension gaps between LLMs and users, guidelines are provided for using LLMs to evaluate the understandability of privacy policies.

II. BACKGROUND AND RELATED WORK

A. *Users' Understanding of Privacy Policy*

Privacy policies serves as a primary channel for companies and organizations to convey their practices regarding the handling of personal information to users. However, various communication challenges have been identified, including users not reading the policy [10], absence of essential information [11], and users struggling to understand policy content [4]. In this study, focusing on user comprehension, we conducted a first study to explore the applicability of LLMs as an approach for automatically assessing whether users can understand a privacy policy.

The left-hand column of Table I summarizes the factors hindering users' understanding of privacy policies. From a writing perspective, Yan et al. [7] analyzed sentence structure and grammar that interfered with privacy policy readability and showed that over half of the policies contain the following aspects. Regarding the unique characteristics of privacy policies, Tang et al. [4] demonstrated that users often cannot understand the technical terms used in privacy policies. Andow et al. [6] indicated the existence of privacy policies that containing contradictory statements. Hara et al. [8] investigated the correspondence between data collection and purpose of use, and Matsuo et al. [9] argued that abstract descriptions lead to user misunderstandings. In addition, based on observations from our previous study on privacy policies, we have identified four new factors.

The factors summarized in Table I are primarily identified through user studies. Although user studies are the primary method for examining user understanding and perception, for companies and organizations that frequently release various services and make updates, recruiting users to assess the understandability of privacy policies each time is time-consuming and expensive.

B. *Utilization of LLMs for Privacy Policy*

LLMs are language models based on the transformer architecture and pre-trained on vast amounts of text data from sources such as web pages and books. Conventional studies using natural language processing technologies have focused on summarizing privacy policies [12] and checking compliance [11]. However, LLMs outperform these traditional approaches by demonstrating advanced performance in simulating user thought processes [13], [14] and conducting legal analyses [15].

The rapid advancement of LLMs has introduced new opportunities for automated analysis of privacy policies. Tang et al. [16] demonstrated that ChatGPT and GPT-4 can accurately classify privacy policy descriptions related to GDPR. Palka et al. [17] proposed a fully informative privacy policy format designed for LLM interpretation, showing that GPT-4 can assimilate the information written in the privacy policy with high precision. Goknil et al. [18] proposed PAPEL, a tool designed to streamline the extraction, annotation, and summarization of information from privacy policies. These studies aim for LLMs to exceed human capabilities in achieving the most accurate comprehension of privacy policies. In contrast, our objective was to leverage LLMs to evaluate and improve privacy policy understandability for users in a realistic setting, considering that privacy policies are documents written for users to read.

III. Privacy Policy for Analysis

A. Creating Baseline Privacy Policy

In this study, we created our own privacy policy for a fictional healthcare app, specifically a step-tracker app to collect data such as location, device, and health information. To meet the requirements of privacy policy content, we expanded the sample in a professional reference book authored by legal experts. Specific data handling and purposes of use were based on descriptions of the privacy policies of top-ranked healthcare apps on Google Play. The baseline Privacy Policy A (hereinafter referred to as PP-A), is illustrated in Figure 1.

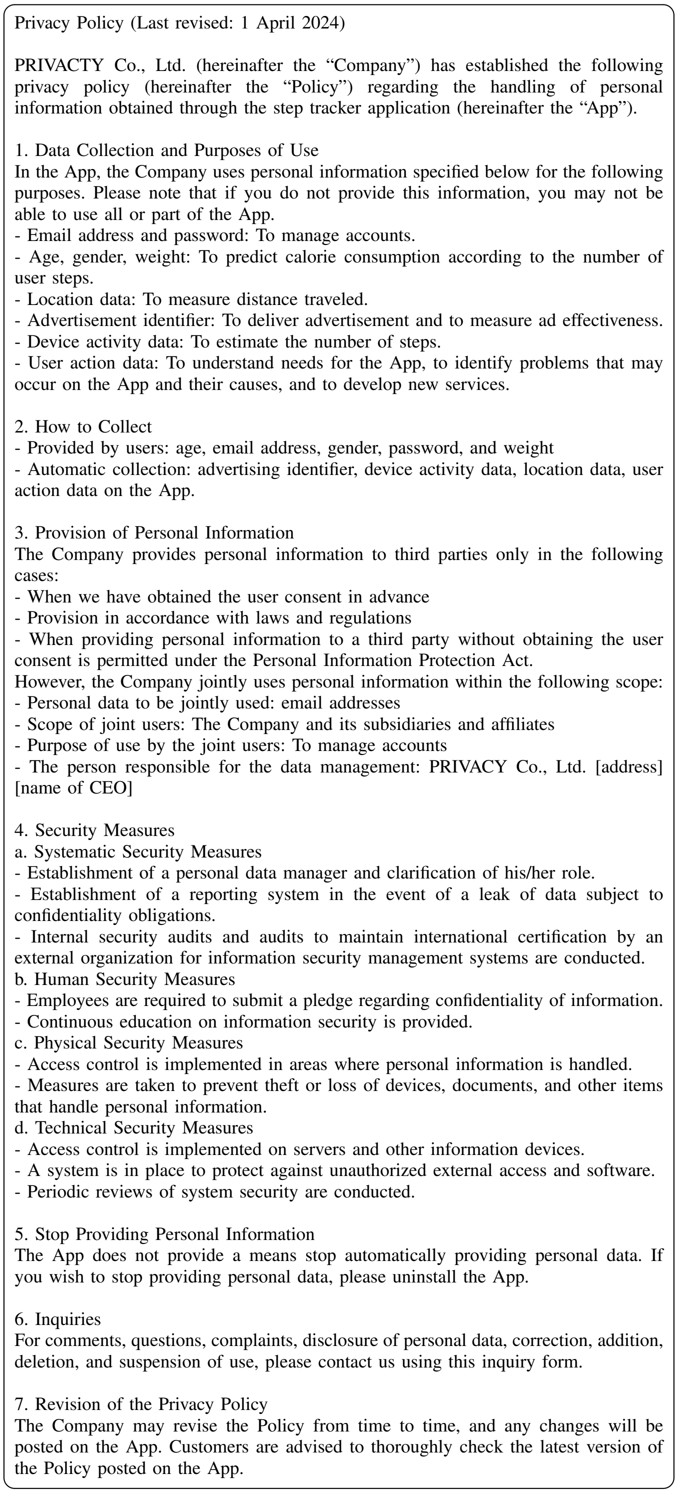
Throughout this paper, we use Japanese privacy policies for our experiments, as the participants primarily consisted of Japanese speakers. Please note that the policies and question texts were translated into English for presenting the methodology and results in each section. The impact of language differences on the tasks performed by users or LLMs is discussed in detail in Section VII.

B. Privacy Policy Obfuscation

The factors shown in Table I were intentionally applied to the baseline privacy policy (PP-A), for "obfuscation" as detailed in Appendix A. Applying all obfuscations to a single privacy policy significantly reduces readability. Therefore, as shown in the middle column of Table I, we created four privacy policies by applying several obfuscations: PP-B, PP-C, PP-D, and PP-E, as illustrated in Appendix Figures 4, 5, 6, and 7, respectively. The baseline PP-A contains no obfuscation. PP-B primarily involves writing-related obfuscations, whereas PP-C, PP-D, and PP-E involve obfuscations specific to privacy policies such as the use of technical terms, information dispersion, and information omission, respectively. Specifically, the application of an illogical presentation order can explain third-party data sharing and how to stop the data collection before detailing the collected data. Regarding the use of technical terms, we describe "international certification related to information security management systems by external organizations" in PP-A, using the technical term "ISO 27001 certification" in PP-C. The character counts for PP-A, PP-B, PP-C, PP-D, and PP-E were 1,526, 1,502, 1,696, 1,720, and 1,797 characters, respectively.

C. Comprehension Questions

Comprehension of privacy policies was measured by the responses to questions regarding descriptions. The question and answer options are listed in Table II. The correct answer options are different for each of PP-A, PP-B, PP-C, PP-D, and PP-E. As shown in the right-hand column of Table II, the obfuscations that affect each question also differ. They are assigned so that the impact can be evaluated by comparing responses to each question. We also asked the participants to identify the sections of the privacy policy to which they referred in their answers. Similarly, LLMs were prompted to indicate the specific parts of the privacy policy and the reasons for their answers. Based on these referred sections and reasons, we analyzed the aspects causing the misunderstandings.



Privacy Policy (Last revised: 1 April 2024)

PRIVACTY Co., Ltd. (hereinafter the "Company") has established the following privacy policy (hereinafter the "Policy") regarding the handling of personal information obtained through the step tracker application (hereinafter the "App").

1. Data Collection and Purposes of Use
In the App, the Company uses personal information specified below for the following purposes. Please note that if you do not provide this information, you may not be able to use all or part of the App.
- Email address and password: To manage accounts.
- Age, gender, weight: To predict calorie consumption according to the number of user steps.
- Location data: To measure distance traveled.
- Advertisement identifier: To deliver advertisement and to measure ad effectiveness.
- Device activity data: To estimate the number of steps.
- User action data: To understand needs for the App, to identify problems that may occur on the App and their causes, and to develop new services.

2. How to Collect
- Provided by users: age, email address, gender, password, and weight
- Automatic collection: advertising identifier, device activity data, location data, user action data on the App.

3. Provision of Personal Information
The Company provides personal information to third parties only in the following cases:
- When we have obtained the user consent in advance
- Provision in accordance with laws and regulations
- When providing personal information to a third party without obtaining the user consent is permitted under the Personal Information Protection Act.
However, the Company jointly uses personal information within the following scope:
- Personal data to be jointly used: email addresses
- Scope of joint users: The Company and its subsidiaries and affiliates
- Purpose of use by the joint users: To manage accounts
- The person responsible for the data management: PRIVACY Co., Ltd. [address] [name of CEO]

4. Security Measures
a. Systematic Security Measures
- Establishment of a personal data manager and clarification of his/her role.
- Establishment of a reporting system in the event of a leak of data subject to confidentiality obligations.
- Internal security audits and audits to maintain international certification by an external organization for information security management systems are conducted.
b. Human Security Measures
- Employees are required to submit a pledge regarding confidentiality of information.
- Continuous education on information security is provided.
c. Physical Security Measures
- Access control is implemented in areas where personal information is handled.
- Measures are taken to prevent theft or loss of devices, documents, and other items that handle personal information.
d. Technical Security Measures
- Access control is implemented on servers and other information devices.
- A system is in place to protect against unauthorized external access and software.
- Periodic reviews of system security are conducted.

5. Stop Providing Personal Information
The App does not provide a means stop automatically providing personal data. If you wish to stop providing personal data, please uninstall the App.

6. Inquiries
For comments, questions, complaints, disclosure of personal data, correction, addition, deletion, and suspension of use, please contact us using this inquiry form.

7. Revision of the Privacy Policy
The Company may revise the Policy from time to time, and any changes will be posted on the App. Customers are advised to thoroughly check the latest version of the Policy posted on the App.

Fig. 1. Privacy Policy A (PP-A).

TABLE II
QUESTIONS AND ANSWER OPTIONS TO CHECK COMPREHENSION LEVEL.

#	Question Text	Answer Options
Q1	Select all options that are correct for the information items to be collected and used by PRIVACY Co., Ltd. along with their purposes. If choosing option E, answer only option E without the others.	A. Use email addresses for account management. B. Use location data for advertisement. C. Use device data for step count estimation. D. Use email addresses for advertisement. E. Cannot be determined from the privacy policy (no description).
Q2	Select all options that are correct regarding the purpose of the personal data provided by PRIVACY Co., Ltd. to other companies. For option D, answer only option D without the others.	A. To measure advertising effectiveness. B. To develop new services. C. To contact with customers. D. Cannot be determined from the privacy policy (no description).
Q3	Select all options that are correct for security measures taken by PRIVACY Co., Ltd. For option D, answer only option D without the others.	A. Measures to prevent loss of devices that handle personal data. B. Maintaining security certification by external organization. C. Outsourcing supervision. D. Cannot be determined from the privacy policy (no description).
Q4	Select all options that are correct for the items and purposes of information that are jointly used by PRIVACY Co., Ltd. and its subsidiaries. For option D, answer only option D without the others.	A. Use email addresses for customer communication. B. Use email addresses for account management. C. Use age and gender for calorie consumption prediction. D. Cannot be determined from the privacy policy (no description).
Q5	Based on the privacy policy, is outsourcing conducted? Select one correct option.	A. Outsourcing is conducted. B. Outsourcing is not conducted. C. Cannot be determined from the privacy policy (no description).
Q6	Based on the privacy policy, where is the data center where personal information is stored? Select one correct option.	A. Japan B. Countries other than Japan. C. Cannot be determined from the privacy policy (no description).
Q7	Based on the privacy policy, where can users contact for data deletion? Select one correct option.	A. CEO of PRIVACY Co., Ltd. B. Inquiry form. C. Cannot be determined from the privacy policy (no description).

IV. ANALYSIS OF LLM'S UNDERSTANDING

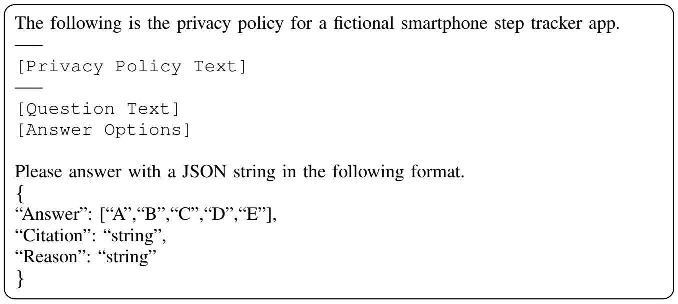
To answer **RQ1: To what extent do LLMs understand privacy policies?**, we input the privacy policies and comprehension questions created in the previous section as prompts into multiple LLMs and analyzed the outputs.

A. Models

We used the APIs of `GPT-4o-2024-05-13` [19], `Gemini-1.5-Pro` [20], and `Claude-3.5-Sonnet-20240620` [21], which are the latest models as of July 2024 released by OpenAI, Google, and Anthropic, respectively. To ensure the reproducibility of the LLM outputs, we set the temperature parameter [22] to 0.0 and reduced the variance. Note that we did not fix the seed values because Gemini and Claude do not support it.

B. Prompt Design

To conduct the analysis under the same conditions as those in the user study, we adopted zero-shot prompting [23], which provides no prior knowledge to LLMs. The input prompt to the LLMs uses the template shown in Figure 2. It consists of an explanatory text stating, "The following is the privacy policy for a fictional mobile app for step tracking," followed by the privacy policy text, a comprehension question, answer options, and instructions specifying the output format. Each input included one of the privacy policies created in Section III and one of the questions from Table II. The output consisted of the answers, sections of the privacy policy, and an explanation. Each question was asked ten times to account for the slight changes in LLM's answers, thereby measuring the comprehension level based on the correct answer rate for each question.



```
The following is the privacy policy for a fictional smartphone step tracker app.
——
[Privacy Policy Text]
——
[Question Text]
[Answer Options]

Please answer with a JSON string in the following format.
{
"Answer": ["A","B","C","D","E"],
"Citation": "string",
"Reason": "string"
}
```

Fig. 2. Prompt Template. `[Privacy Policy Text]` is replaced with privacy policies of PP-A, PP-B, PP-C, PP-D, and PP-E, and `[Question Text]` and `[Answer Options]` are replaced with question text and answer options listed in Table II, respectively.

C. Results

1) LLMs' Comprehension: The correct answer rates for each question by the LLM models are shown in Table III. For all models, the correct answer rates for PP-A (without obfuscation) and PP-B (with writing-related obfuscation) were 100%, indicating that the LLMs were not affected by writing-related obfuscations. In contrast, the correct answer rates for PP-C, PP-D, and PP-E, which contain obfuscations specific to privacy policies, were lower. The LLMs' answers were

TABLE III
PERCENTAGE OF CORRECT ANSWERS TO EACH QUESTION BY LLMS.(%)

PP	Model	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
A	GPT	100	100	100	100	100	100	100	100
	Gemini	100	100	100	100	100	100	100	100
	Claude	100	100	100	100	100	100	100	100
B	GPT	100	100	100	100	100	100	100	100
	Gemini	100	100	100	100	100	100	100	100
	Claude	100	100	100	100	100	100	100	100
C	GPT	0.0	80	100	0.0	100	100	100	69
	Gemini	0.0	0.0	100	0.0	100	100	0.0	43
	Claude	0.0	0.0	100	0.0	100	100	100	57
D	GPT	0.0	100	100	100	100	100	100	86
	Gemini	0.0	100	100	90	100	100	0.0	70
	Claude	100	100	100	0.0	100	100	100	86
E	GPT	0.0	100	100	0.0	100	100	100	71
	Gemini	0.0	100	100	10	100	100	0.0	59
	Claude	0.0	0.0	100	0.0	100	100	100	57

generally stable, and the correct answer rate for each question was either 0% or 100% with three exceptions.

2) LLMs' Tendencies in Incorrect Answers:

a) Effects of Dispersed Information: In obfuscation of PP-D, the description of data handling was dispersed across multiple sections. Therefore, Q1 on data handling required referring to the information written in section "Data Collection and Purposes of Use," which is the main part of the data handling, and section "Joint Utilization," which is supplementary. The correct answer was to select all the options such that in A :"Use email addresses for account management,"; in C :"Use device data for step count estimation,"; and in D :"Use email addresses for advertisement." Claude identified both sections and selected all three options. In contrast, GPT and Gemini referenced only the "Data Collection and Purposes of Use" section, selecting only options A and C. This suggests that even when multiple pieces of information are required to answer correctly, these models may refer only to the most relevant information while disregarding the rest.

b) Effects of Missing Information (decoupling): In the obfuscation of PP-E, the information items to be collected and their purposes were listed separately without clarifying the correspondence between specific items and their respective purposes. Therefore, the correct answer for Q1, which asks about the combination of data items and their purposes, is option E: "Cannot be determined from the privacy policy (no description)." However, GPT and Claude incorrectly answered, reasoning their choice by stating that "[the data item] is explicitly listed as collected information, and [the purpose] is included in the list of the purposes." This reveals that when the correspondence is not explicitly clarified, LLMs do not select "Cannot be determined (no description)" and instead mistakenly assume that all information listed in the data collection section is used for all the purposes described in the purposes section.

c) Effects of Missing Information (abstract expression): The obfuscation in PP-C contains an abstract expression regarding the usage purpose such as "To provide the basic functions of the service." In this study, we defined the "basic functions" of the step tracker app as "step count estimation" and "calorie consumption prediction." The correct answer options were A: "Use email addresses for account management" and C: "Use device data for step count estimation" for Q1, whereas B: "Use email addresses for account management" and C: "Use age and gender for calorie consumption prediction" for Q4. However, GPT (for Q1 and Q4), Claude (for Q1 and Q4), and Gemini (for Q1) answered "No clear statement" without interpreting the abstract expressions. This suggests that LLMs can detect a lack of information and potentially identify privacy policy statements that can hinder user understanding.

In the obfuscation of PP-B, PP-C, PP-D, and PP-E, the content of inquiries was described using the abstract expression "other," as in, "For comments, questions, complaints, or other inquiries related to the handling of personal information, please contact us through this inquiry form." Based on the assumption that "other inquiries related to the handling of personal information" includes data deletion, the correct answer for Q7 asking for contact information for data deletion was option B: "Inquiry form." Gemini selected the correct option B for PP-B but answered with "Cannot be determined (no description)" for PP-C, PP-D, and PP-E, without interpreting the scope of "Other." As PP-B primarily applied obfuscation related to writing, whereas PP-C, PP-D, and PP-E applied obfuscations specific to privacy policy characteristics, we found that the surrounding context affected the understanding of abstract expressions. Additionally, GPT and Claude could not interpret the abstract expression "To provide the basic functions of the service"; however, they did correctly interpret "other inquiries related to the handling of personal information." This suggests that the comprehension of abstract expressions by LLMs may vary depending on the degree of abstraction.

d) Effects of Omitting Information by Reference: In the obfuscation of PP-C and PP-E, the description of data handling in joint utilization was omitted by referring to the previous section. For Q2 and Q4, which are related to providing information to other entities, the correct response was based on the data-handling methods described in the "Data Collection and Purposes of Use" section with the purposes of the service provider serving as a reference source. However, GPT (PP-C-Q2) answered incorrectly in one out of ten trials, selecting only part of the necessary information from the reference source and failing to select all the required details. Additionally, Gemini (PP-C-Q2) did not refer to the correct source and did not select any of the options A, B, or C. In contrast, both GPT and Gemini answered PP-E-Q2 correctly. The major difference between the reference content of PP-C and PP-E is that the reference content of PP-C contains information both relevant and irrelevant to the question, whereas that of PP-E includes only relevant information. This indicates that when a privacy policy uses a format that refers to the previous sections, the manner in which the reference content is written affects LLMs' comprehension.

e) Effects of Technical Terms: We observed cases where Claude, GPT, and Gemini interpreted certain words differently. Specifically, regarding "providing information to other companies" in Q2, GPT and Gemini considered various forms of collaboration with other companies, such as third-party provision, joint utilization, and outsourcing. However, Claude responded, "Although join utilization is mentioned, it does not qualify as providing information to other companies [...omitted...]." Regarding the expression "account management," listed as one of the purposes for using personal information in the privacy policy, GPT and Gemini interpreted that account management does not include contacting users. In contrast, only Claude interpreted that "account management" includeds user communication, leading to an incorrect answer.

> **Key Takeaways**: The LLM's comprehension was not affected by writing-related obfuscations such as double negative and many words, however was affected by privacy-policy-specific obfuscations such as dispersed information and omitting information by reference.

V. Analysis of User Understanding

To answer **RQ2: What gaps lie between LLMs and users in their understanding of privacy policies?**, we designed and conducted a user study using privacy policies and comprehension questions. We compare these results with those from the previous section to derive the similarities and differences between LLMs and users in their understanding of privacy policies.

A. Survey Design

A user survey was performed using the crowdsourcing service Lancers [24], with 500 users participating as a result of recruitment. Recruitment was conducted on both weekdays and weekends in June and July 2024, to attract a diverse range of participants. The survey was expected to take 20 minutes to complete[1] and participants were compensated with 400 JPY for completing the survey, which exceeds the minimum wage standard in Japan.

The survey comprised four main sections: consent to participate, text on privacy policy, privacy policy questions, and demographic questions. Each participant read one of the privacy policies described in Section III and answered the seven questions listed in Table II. The demographic questions included device, gender, age, expertise, and experience related to privacy. A simple attention-check question was also included to ensure the quality of the users' answers. The full text of the survey is shown in Appendix B.

B. Results

1) Descriptive Statistics of the Participants: We recruited 100 participants for each PP-A, PP-B, PP-C, PP-D, and PP-E. After removing the invalid answers from the attention check, 88, 94, 82, 92, and 93 valid answers were collected for each policy, respectively. Table IV shows the demographics and experiences of the participants. Across all privacy policies, more than 65% of the participants completed the survey using a PC. The percentage of participants with expertise in "Administration/Law" was low at 1-4% and relatively high in "IT/Communication" with 8-18%. Most participants reported that they read the privacy policies "at the time of service registration," and reported "skimming" the policy rather than using keyword searches or tools.

TABLE IV
Demographics of Participants.(%)

	A	B	C	D	E
Device					
PC	72.7	74.5	74.4	68.5	79.6
Smartphone	22.7	24.5	23.2	23.9	18.3
Tablet	4.5	1.1	1.2	7.6	2.2
Other	0.0	0.0	1.2	0.0	0.0
Gender					
Male	55.7	53.2	57.3	53.3	61.3
Female	43.2	45.7	39.0	46.7	38.7
Other	0.0	0.0	1.2	0.0	0.0
Prefer not to answer	1.1	1.1	2.4	0.0	0.0
Age					
18–19	0.0	0.0	0.0	0.0	0.0
20–29	8.0	5.3	9.8	6.5	5.4
30–39	28.4	29.8	19.5	20.7	33.3
40–49	38.6	34.0	43.9	40.2	33.3
50–59	15.9	26.6	20.7	26.1	19.4
60–69	8.0	4.3	6.1	5.4	8.6
70 and over	1.1	0.0	0.0	1.1	0.0
Prefer not to answer	0.0	0.0	0.0	0.0	0.0
Expertise					
IT/Communication	18.2	11.7	15.9	7.6	11.8
Administration/Law	3.4	1.1	2.4	2.2	4.3
Other	68.2	84.0	76.8	83.7	72.0
Prefer not to answer	10.2	3.2	4.9	6.5	11.8
When to read privacy policies					
Registering for service	79.5	72.3	70.7	78.3	75.3
Receiving revision notice	34.1	24.5	28.0	22.8	22.6
Entering personal data	29.5	22.3	28.0	22.8	30.1
Not reading	17.0	21.3	20.7	12.0	24.7
Other	0.0	1.1	1.2	1.1	1.1
How to read privacy policies					
Reading thoroughly	8.0	3.2	3.7	5.4	12.9
Skimming	67.0	72.3	69.5	68.5	69.9
searching for keywords	15.9	7.4	11.0	8.7	10.8
looking at section headers	26.1	19.1	17.1	18.5	18.3
Using tools	1.1	0.0	2.4	1.1	0.0
Not reading	17.0	17.0	18.3	14.1	18.3
Other	2.3	5.3	0.0	2.2	0.0

2) Users comprehension: The correct answer rates of the participants for each question are shown in Table V. The rates ranged from 6% to 93%, with an average of 63%. The rates for PP-A, which contained no obfuscation, and PP-B, which included writing-related obfuscation, exceeded 70%. By contrast, the rates for PP-C, PP-D, and PP-E, which contained privacy-policy-specific obfuscations, were all below 60%.

3) User Tendencies in Incorrect Answers:

[1] Based on a pilot study conducted by the authors, the average response time was approximately 15 minutes.

TABLE V
Percentage of Correct Answers to Each Question by Users.(%)

PP	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
A	64.8	62.5	54.5	73.9	65.9	89.8	93.2	72.1
B	71.3	69.1	69.1	80.9	53.2	90.4	84.0	74.0
C	25.6	35.4	43.9	17.1	61.0	89.0	81.7	50.5
D	15.2	57.6	64.1	66.3	48.9	89.1	72.8	59.2
E	6.5	50.5	73.1	21.5	90.3	92.5	78.5	59.0

a) Effects of Dispersed Information: The correct answer rate for the question requiring participants to confirm the dispersed information (PP-D-Q1) was low at 15%. In terms of correct answer rates, options A and C had high rates of 97% and 90%, respectively, whereas option D dropped to 23%. Information about options A and C was located in sections closely related to the question, whereas information about option D was spread across less-related sections. From the participants' answers, those who failed to select option D did not refer to less-related sections. The participants tended to refer only to closely related sections when the required information was dispersed.

b) Effects of Missing Information (decoupling): For PP-E, with no correspondence between the information items to be collected and their purposes of use, the correct answer rate for the question on relationship was only 6.5%, that is, 6.5% of the participants answered correctly providing reasons such as "It is not specified for what purpose the collected information is used" and "From the descriptions of Sections 2. Data Collection and 3. Purposes of Use, there is no explicit connection indicating which information is used for which purpose." However, other participants added their own interpretation to answer the questions on correspondence between information items and their purposes. This indicates that when the necessary information is missing, users may add their own interpretation, leading to potential misunderstandings.

c) Effects of Missing Information (abstract expression): The correct answer rates for participants on questions PP-C-Q1 and PP-C-Q4, involving descriptions using abstract expressions, were 26% and 17%, respectively. For both Q1 and Q4, the correct answer rates for the options that did not involve abstract expressions were high at 96% and 76%, respectively, whereas the rates for options that did involve abstract expressions decreased to 52% and 30%, respectively.

d) Effects of Omitting Information by Reference: Q2 of PP-C and PP-E asked for information that was omitted by referring to the previous sections. Since the same phrase of the correct answer option was used in the privacy policy, correctly answering PP-C-Q2 and PP-E-Q2 using keyword search functions was easy. However, only 35% and 51% of the participants answered correctly. This result aligns with the findings shown in Table IV, which indicate that although over 70% of participants used a PC, only approximately 10% of participants utilized keyword searches when reading a privacy policy. Additionally, analyzing the comments provided by participants who answered incorrectly revealed that approximately 70% failed to refer to the previous section. This indicates that omitting detailed information by referring to other sections reduces user comprehension.

TABLE VI
Comparison of Comprehension between LLM(GPT) and Users.
○: Both understood, ◐: Only LLM understood,
●: Both misunderstood.

PP	Q1	Q2	Q3	Q4	Q5	Q6	Q7
A	○	○	○	○	○	○	○
B	○	○	○	○	○	○	○
C	●*1	◐*2	◐*3	●*2	○	○	○
D	●*2	○	○	○	◐*3	○	○
E	●*4	○	○	●*4	○	○	○

*1:Missing Information (Abstract Expression)
*2:Dispersed Information
*3:Use of Technical Terms
*4:Missing Information (Decoupling)

e) Lack of Knowledge of Technical Terms: PP-C used technical terms, leading to a lower correct answer rate for Q3 than PP-A, PP-B, PP-D, and PP-E, which did not use technical terms. This suggests that the use of technical terms in privacy policies can reduce user comprehension. Additionally, we conducted a chi-square test between the number of participants who answered correctly/incorrectly PP-C-Q3 and the participants with expertise in "IT/Communication" or "Administration/Law" versus those without this expertise. The results showed a significant difference ($p < 0.05$), indicating that participants with expertise in "IT/Communication" or "Administration/Law" tended to be able to understand the technical terms and answered correctly.

PP-B and PP-D did not mention outsourcing. However, in Q5, which asked about outsourcing, incorrect answers such as "Outsourcing is conducted" or "Outsourcing is not conducted" were observed at rates of 30% and 17% for PP-B and 39% and 12% for PP-D, respectively. Among the reasons provided for the former, 86% referred to joint utilization, 14% to provision to third parties, and 6% to security measures. Among the latter, 56% referred to joint utilization, 15% to provision to third parties, and 30% to security measures. This indicates that many participants confused outsourcing with joint utilization and provision to third parties.

f) Overlooked Information: The correct answer rate for Q3 in PP-A (without obfuscation) was low at 55%. The description of security measures in PP-A contained almost the same wording as the correct answer options A and B. Among the participants who answered incorrectly, 88% appropriately referred to the section on security measures, suggesting that they likely overlooked this information due to carelessness.

C. Understanding the Gap between LLM and Users

By comparing the results in Sections IV and V, we identified the similarities and differences in the tendency to misunderstand among LLMs and users. In this section, we focused on the results of GPT as a representative of the LLMs, as it had demonstrated the best performance as outlined in Section IV. Table VI compares the comprehension levels, where we classified an understanding as "achieved" when the correct answer rate exceeded 50%. In general, LLMs generally reached a higher correct answer rate than users.

Through detailed analysis, we observed three main categories: questions that both LLMs and users understood (○), those understood by neither (●), and those only understood by LLMs (◐). This categorization indicates that certain privacy policies (PP-A and PP-B) presented no issues in understandability for either LLMs or users, whereas in PP-E, LLMs could detect descriptions that were challenging for users to understand. However, comprehension gaps were evident in PP-C and D, where LLMs failed to detect discrepancies that affected user understanding.

Both LLMs and users were impeded by information dispersion and lack of clarity in certain privacy policy descriptions, which led to misunderstandings. Yet, some errors, particularly those stemming from users' lack of knowledge or overlooked information, were unique to users and not encountered by LLMs. Since LLMs do not exhibit "carelessness" or "knowledge gaps" typical of users, supplementary methods are essential to bridge these gaps. We discuss approaches to bridging these gaps in Section VII-A.

Key Takeaways: The users' comprehension was affected by both writing-related and privacy-policy-specific obfuscations. User-specific errors including lack of domain knowledge and overlooked information due to carelessness were observed.

VI. Evaluation of Privacy Policies in the Wild

In this section, as a first step in evaluating the understandability of privacy policies using LLMs, we focus on those of real-world applications. The objective is to demonstrate that LLMs can identify policy factors that may hinder user understanding in real-world contexts. Additionally, through manual inspection, we explored new factors not covered in our custom privacy policies. In this experiment, we employed GPT as the LLM for policy analysis, as it had demonstrated the best performance as outlined in Section IV.

A. Dataset

Since the data required for our evaluation was openly available, we collected the privacy policies to be evaluated from Android apps on Google Play. Currently, Google Play mandates that all apps include a link to their privacy policy and data safety section (DSS). A privacy policy, written in a natural language is required to provide legal and technical details regarding data access. In contrast, the DSS label is completed by developers according to a specific form, allowing users to quickly review the types and purposes of the data collected and processed by the app. Therefore, DSS is known to be more user-friendly [25], [26] and less prone to communication discrepancies. In our analysis, we treated the DSS as a pseudo ground truth indicating the data collection and purposes of use, and we have the LLM evaluate whether the app's privacy policy appropriately communicates this information to users. Note that we discuss the validity of using DSS labels in Section VII.

TABLE VII
Categories of collected apps.

Category Name
Health & Fitness
Finance
Maps & Navigation
Shopping
Social
Entertainment
Dating
Food & Drink
News & Magazines

The top five popular apps were selected from ten categories that are highly associated with personal data in the Google Play Store, as shown in Table VII. After fetching the privacy policy links of these apps, we excluded those that returned HTTP response errors, lacked a Japanese version of the privacy policy, or had privacy policies that were either too short or too long (fewer than 1,000 characters or more than 10,000 characters). This process helped eliminate vague and non-functional privacy policies from companies covering multiple services or apps. As a result, 19 apps were selected, and the corresponding DSS labels were collected.

B. Generating Questions and Answer Options

Templates for the prompts used in this experiment are shown in Figure 3. In this experiment, questions were generated based on Q1 in Table II. More precisely, the LLM read a privacy policy and selected the appropriate answer option considering the collected data and its purpose. To achieve this, we generated 210 answer options by combining the 30 data type labels defined in the DSS with seven purpose labels. All the labels are listed in Table VIII. Additionally, each prompt included a final answer option, indicating that none of the options apply, resulting in a total of 211 answer options in total for each prompt. The accuracy of the answer options selected by the LLM was verified by comparing them with the DSS settings for each app on Google Play. To mitigate the bias introduced by random variation, each prompt was executed 10 times. Since this process was repeated for 19 apps, the total of 190 prompts were executed.

C. Results

The correctness of the LLM answers was evaluated by comparing them with the pseudo ground truth based on the labels registered in the DSS. Therefore, when the LLM selects the data-handling practice registered in the DSS, this indicates true positive (TP); otherwise it indicates false negative (FN). When the LLM selects the data handling that is not registered in the DSS, it is a false positive (FP); otherwise, it is a true negative (TN).

TABLE VIII

LABELS FOR COLLECTED DATA TYPES AND PURPOSES AS DEFINED IN DSS [27]. A FEW LABELS LIKE "OTHER INFO" HAVE BEEN EXCLUDED. COMBINING THESE LABELS GENERATES A TOTAL OF 210 ANSWER OPTIONS.

30 labels for `[Collected Data]`	Approximate location, Precise location, Name, Email address, User IDs, Address, Phone number, Race and ethnicity, Political or religious beliefs, Sexual orientation, User payment info, Purchase history, Credit score, Health info, Fitness info, Emails, SMS or MMS, Photos, Videos, Voice or sound recordings, Music files, Files and docs, Calendar events, Contacts, App interactions, In-app search history, Installed apps, Web browsing history, Crash logs, Diagnostics
7 labels for `[Purpose]`	App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention & security & compliance, Personalization, Account management

```
The following is the privacy policy for a mobile app.
—–
[Privacy Policy Text]
—–
Select all options that are correct for the information items to be collected and used by [App Developer Name] and their purposes. If choose option A-211, answer only option A-211 without the others.

A-1. Use [Collected Data] for [Purpose].
A-2. Use [Collected Data] for [Purpose].
A-3. Use [Collected Data] for [Purpose].
...
A-210. Use [Collected Data] for [Purpose].
A-211. Cannot be determined from the privacy policy (no description).

The definition of the words used for the collected data and purposes in the options are as follows.
[Definitions]

Please answer with a JSON string in the following format.
{
"Answer": ["A-1","A-2","A-3",...,"A-211"],
"Citation": "string",
"Reason": "string"
}
```

Fig. 3. Prompt Template for Evaluation of Privacy Policy in the Wild. `[Privacy Policy Text]` is replaced with the privacy policy of each app, `[App Developer Name]` with the developer name listed on Google Play, `[Collected Data]` and `[Purpose]` are DSS labels as shown in Table VIII, and `[Definitions]` with the specific definitions of each collected data type or usage purpose as described in the official website [27], respectively.

1) Comprehension of Each Privacy Policy: Based on the 211 options selected or not selected per response, we calculated the F-scores, FP rate (FPR), and FN rate (FNR). The averages of the ten trials are listed in Table IX. FPR tended to be low across all privacy policies, whereas FNR tended to be high.

The F-scores for Med-2 and Fin-1 were relatively high, at 0.57 and 0.53, respectively. In the privacy policy of Med-2, the correspondence between the collected information and their purposes of use is clearly specified, resulting in high accuracy.

Conversely, in the privacy policy of Fin-1, although the purposes of use is clearly stated, the categories of the personal information collected are not explicitly indicated. The information about the categories was omitted by referring to the URL of a different page. However, the information items that are provided to third parties and jointly used were clearly described and it was natural that they were first obtained by the service provider in order to be provided to or shared. The LLM was able to reach the correct answer by referring to the sections on data provision and shared use.

For Map-2 and Dat-1, the correspondence between the collected information items and their purposes of use was specified. However, most of the items collected were described using abstract terms (e.g., "customer's personal information"), with limited use of concrete expressions (e.g., "location data" and "email address"). Consequently, only certain information items described with specific terms were linked to their purposes, leading to a lower FPR. Information items described using abstract expressions were not interpreted, resulting in not selecting the options. This aligns with the results in Section IV-C2c.

In five policies – Fin-3, Ent-1, Dat-2, Med-3, and Med-4 – the LLM did not select any of the correct options. For Dat-2 and Med-4, LLM selected "Cannot be determined from the privacy policy (no description)," which was judged to be incorrect based on DSS. These privacy policies only describe the policy for handling personal data, and not the specific practices. This outcome, in which the LLM's answer was marked as incorrect, indicates that these privacy policies are difficult to understand because of the lack of specific details. In the privacy policy of Fin-3, the correspondence between information items and purposes of use was specified by referring to the previous section. As observed in Section IV-C2d, the LLM was unable to refer to the previous section and thus did not arrive at the correct answer.

2) Comprehension of Each Data Handling Practice: To analyze what kind of data handling is easier for LLM to understand, namely the degree of difficulty of each option, we calculated the FNRs of each options. As the expression of each data handling varies depending on the privacy policies, we calculated the FPR across 190 trials (10 trials for all 19 privacy policies), which enable to absorb the fluctuations in difficulty. Table X shows the FNRs for data types and purposes with total number of FNs and TPs of 100 or more.

The LLM tended to have a better understanding of the handling of users' personal information, such as email addresses, phone numbers, and addresses, while demonstrating less understanding of items like in-app history and diagnostics. The numbers of privacy policies using the exact terms "email address," "phone number," and "address" in the section related to the collected data and its purpose of use were 10, 9, and 5, respectively. In contrast, the terms like "in-app search history,"

TABLE IX
F-SCORE, FPR, FNR OF EACH PRIVACY POLICY.

PP	F-score	FPR	FNR
Finance-1	0.53	0.17	0.26
Finance-2	0.39	0.07	0.57
Finance-3	NaN	0.02	1.00
Finance-4	0.28	0.03	0.81
Maps and Navigation-1	0.28	0.09	0.45
Maps and Navigation-2	0.36	0.00	0.78
Social-1	0.16	0.02	0.90
Entertainment-1	NaN	0.02	1.00
Dating-1	0.11	0.00	0.94
Dating-2	NaN	0.01	1.00
News and Magazines-1	NaN	0.02	0.95
Food and Drink-1	NaN	0.03	0.58
Food and Drink-2	0.19	0.03	0.86
Food and Drink-3	0.25	0.11	0.62
Medical-1	0.34	0.02	0.73
Medical-2	0.57	0.02	0.51
Medical-3	NaN	0.04	1.00
Medical-4	NaN	0.01	1.00
Medical-5	0.11	0.11	0.85

“diagnostics,” “app interactions,” and “crash logs” were not explicitly mentioned in any policies. Furthermore, regarding “approximate location,” the exact term “location information” was used in 8 policies, and the FNR was particularly low. This result indicates that when the expressions in privacy policies and those in answer options are closer, the LLM tend to understand the data handling.

Regarding purposes of use, the LLM understood the data handling for account management well. However, none of the privacy policies explicitly used the direct phrase “for account management.” It is likely that the LLM was able to identify collected data items used for account management, making it easier to understand the data handling of account management.

Although the terms “advertising,” “marketing,” and “analytics” were used in 11, 7, and 13 policies, respectively, the LLM’s understanding of these purposes was not high. This may be attributed to the fact that approximately half of the privacy policies explained information related to the analysis of browsing history within sections titled “About Cookies” or “Use of Information Collection Modules.” Therefore, the data handling for advertising and analytics purposes was mentioned across multiple sections (i.e., the information was dispersed), which decreases the LLM’s understanding.

D. New Factors Hindering Understanding

Not mentioned in Section IV are certain factors that were observed to reduce LLM’s understanding. The analysis target policy in Section IV comprised only plain text. However, privacy policies of Med-2, Map-2, and Dat-1 included not only text but also tabular formats. When converting the tabular format into plain text, sentences consisting of a list of unrelated words were generated, and the information corresponding to each column became unclear, which reduced LLM’s understanding. Using HTMLs or screenshots to input tabular data directly into LLM is a future work.

TABLE X
FNRS OF EACH OPTION.

Collected Data	FNR
Photos	1.0
In-app search history	1.0
Diagnostics	1.0
App interactions	0.99
Crash logs	0.99
Purchase history	0.81
User IDs	0.80
User payment info	0.80
Name	0.69
Email address	0.66
Phone number	0.62
Address	0.60
Approximate location	0.27
Purpose	FNR
Developer communications	1.0
Personalization	1.0
Analytics	0.90
App functionality	0.82
Advertising or marketing	0.76
Fraud prevention, security, and compliance	0.68
Account management	0.51

As mentioned in Section VI-C1, several privacy policies omit detailed descriptions of data handling practices by indicating the URL of a different page as a reference. This results in a lack of information in the privacy policy, which hinders LLM’s understanding. LLMs are required to analyze not only privacy policies but also other policies.

In Section III, we aligned the terms used in the privacy policies and those in the answer options. However, we found that the wording in privacy policies of real-world varied and the differences between the expressions of policies and answer options affected the level of understanding of LLM.

Key Takeaways: Our evaluation of privacy policies in real-world contexts showed low FPRs and high FNRs. Through the evaluation, we found new issues related to input data (i.e., prompts) to LLMs, such as accurate and comprehensive extraction of policy text and design of questions, answer options, and ground truth.

VII. DISCUSSION

A. Towards Automation of Understandability Assessment

1) Bridging the Gap between LLMs and Users: Based on the analysis results presented in Sections IV and V, we discussed approaches to leveraging LLMs for automating the assessment of privacy policy understandability and strategies for bridging the gap between LLMs and users.

The experimental results revealed that obfuscation caused by the dispersion and omission of information hinders understanding for both LLMs and users. For privacy policies exhibiting such shortcomings, LLMs can contribute to enhancing user comprehension by automatically identifying problematic sections that require revision. This process facilitates the improvement of descriptions in the policies that obstruct understanding.

Conversely, incorrect answers stemming from a lack of knowledge about technical terms were observed exclusively among users and not in LLMs. For privacy policies where LLMs instructed through zero-shot prompting fail to recognize such deficiencies, supplementary methods are necessary. For example, specialized classifiers could be employed to detect these issues, leveraging prior studies on users' understanding and misconceptions of such terms [28].

Nevertheless, the generally high understanding capabilities of LLMs do not guarantee that users will comprehend the sections that LLMs successfully interpret. User comprehension is influenced by individual attributes such as experience, age, and education, as well as by factors like limited knowledge of technical terms and overlooked information. To achieve the goal of automating user comprehension evaluation, additional efforts are required to enable LLMs to simulate the diverse comprehension levels of various users. We identify persona-based approaches [29] as a promising direction and intend to investigate their effectiveness as an alternative to user studies.

2) Generation of Questions for Evaluation: In this paper, we demonstrated a method for evaluating privacy policies by preparing preset questions and answer options, which were then given to LLMs. We discuss potential directions for refining this method further.

In the experiment detailed in Section VI, we asked a typical question to the LLM regarding the types of data collected by an app and their purposes. As shown in Table II of Section III, we also proposed several additional questions to evaluate privacy policy comprehension. These questions were designed based on typical use cases of apps and web services, as well as Japanese legal requirements. Expanding the range of questions and increasing their comprehensiveness would enable the identification of privacy policy deficiencies in greater detail. To this end, a potential approach could involve generating a large number of question texts directly using LLMs.

For the answer options, we provided the LLM with 211 candidates derived from the privacy labels included in Google Play's DSS. Other privacy labeling systems have been proposed, such as those used in the Apple App Store or prior studies [30], [31], each offering different categorizations and levels of abstraction. Incorporating a variety of such data sources to further enrich the answer options would also be beneficial for conducting more thorough evaluations of privacy policies. A distinct advantage of using LLMs lies in their ability to scale effectively, even when tasked with processing a large volume of questions and answer options.

Finally, in our experiment, we adopted the use of close-ended questions provided in multiple sets to the LLM. Exploring prompt engineering techniques to enable LLMs to handle open-ended questions, as well as analyzing the responses generated by such prompts, remains an important area for future work.

B. Time and Financial Costs in User Studies and LLM Tasks

Given the addition or updates of services operated by companies, it is essential for the evaluation process of privacy policies to be conducted both quickly and cost-effectively. Although LLMs in our study have not yet achieved the ability to accurately simulate user understanding, we provide a preliminary comparison of the time and financial costs between LLM-based evaluations and user studies. We note that this comparison relies on the hypothetical premise that LLMs could overcome the challenges described in the previous subsection and effectively simulate user understanding in future applications.

We examine the time required for the experiments described in Sections IV and V to highlight the potential benefits of using LLMs for this task. The task involves reading PP-A, as shown in Figure 1, and answering questions Q1–Q7 in Table II. Unfortunately, the crowdsourcing platform used in the user study did not allow us to measure the time taken by each participant from the start to the completion of their responses. Instead, we measured the total duration from initiating participant recruitment to obtaining the required number of responses. Including recruitment time is reasonable, as it is a necessary component of user studies. For this task, collecting responses from 100 participants took 46 hours and 48 minutes, with a total cost of 44,000 JPY. In contrast, performing the same task using GPT required only 31.5 seconds at a cost of approximately 15 JPY. Even when using multiple models or repeated prompts to generate variations in responses, LLMs demonstrated a significant advantage in both time and cost efficiency.

Additionally, a pilot study conducted by the authors found that a single task took approximately 20 minutes per participant, further highlighting the substantial speed advantage of LLMs. As the number of questions and answer options increases to provide a more comprehensive evaluation of privacy policies, this difference becomes even more pronounced. These results suggest that LLMs have the potential to dramatically streamline the iterative refinement process for privacy policies, significantly improving efficiency.

C. Limitations

1) Impact Degree of Each Obfuscation: In this study, multiple obfuscations were included in privacy policies to measure the comprehension of LLMs and users, to identify the factors contributing to misunderstandings. However, we did not investigate the magnitude of the impact of each obfuscation or their combined effect. Additionally, Section IV-C2 suggests that the effect of even the same obfuscation on LLMs may differ depending on the context. We believe that by clarifying the conditions about obfuscations that affect LLMs, we can reproducibly identify descriptions that are difficult for users to understand.

2) Use of AI by Crowdworkers: In our user study in Section V, the participants were instructed not to use AI tools. However, some answers were similar to those generated by LLMs. Distinguishing the answers of participants using LLMs is difficult based on just attention-check tests. Thus, a new verification test capable of detecting AI-generated answers should be introduced.

3) Reliability of DSS: In real-world applications, labeling the "true intent" that developers wanted to include in their privacy policies is highly challenging. In the experiments conducted in Section VI, we adopted the DSS in Google Play as the pseudo ground truth, representing the data collected and its intended use. The DSS labels provide explicit and structured disclosures, in contrast to the often-ambiguous nature of free-form privacy policies. Consequently, they facilitate a clearer communication between users and developers.

We recognize that, similar to other metadata channels such as permissions [32], descriptions [33], [34], and privacy policies [35], DSS labels can sometimes digress from the actual behavior of the application [26], [36]. However, our goal is to automatically evaluate the understandability of privacy policies, not to uncover inconsistencies with actual app behavior. Therefore, instead of analyzing application behavior, we chose DSS, which describes the practices developers intend to convey. In case of an inconsistency between DSS labels and users' understanding of the privacy policy, it should suggest a defect within the privacy policy. Additionally, Google states that if the DSS and the actual app behavior do not align, the app may be subject to blocking [37]. The apps evaluated in our study were popular and ranking within the top five in each category, with most of them developed by prominent brands. Therefore, we believe that the DSS of the evaluated apps has a certain level of reliability.

4) Language-Specific Factors: In Sections IV, V, and VI, we used Japanese privacy policies, question texts, and prompts because the majority of our participants were Japanese speakers. Since LLMs support multiple languages, we are confident that the primary contributions of this paper are globally applicable and not restricted to specific regions or languages. However, differences in LLM performance across languages and regional conventions in privacy handling practices may affect our observations.

According to the GPT-4 Technical Report [38], the performance of LLMs varies depending on the availability of resources in each language. While Japanese is one of the major languages used on the internet, alongside Spanish, German, and French, English has an overwhelmingly dominant presence [39]. As a result, LLMs may demonstrate better comprehension of privacy policies written in English than those written in Japanese. Additionally, Japan's personal data protection policies differ in some respects from Western standards such as GDPR and CCPA. This raises the possibility that our custom privacy policies may not fully address all potential obfuscations. A deeper investigation into how differences in language and regional practices affect tasks involving LLM comprehension of privacy policies is a topic that future works should address.

D. Ethics Consideration

In our user study, we assessed users' comprehension of privacy policies and did not collect sensitive personal data. Additionally, informed consent was obtained from all participants at the beginning of the survey. Participants were informed that they can quit the survey at any time and that the survey results would be handled as anonymous data and used solely for research purposes. The survey design was ethically reviewed and approved by our institutional review board.

VIII. Conclusion

In this study, five privacy policies containing 11 types of obfuscations were prepared, and a comparative analysis was conducted on the comprehension levels of three LLMs and 449 users. The results indicated that LLMs generally surpassed users in comprehension, and factors such as dispersed and missing information were identified as reducing the comprehension of both LLMs and users. Additionally, we identified the factors that led to user-specific misunderstandings, such as a lack of knowledge of technical terms and overlooked information. Through case studies of LLM evaluation using real-world privacy policies, we observed the LLM's tendency mentioned above and found new factors hindering the LLM's understanding. Although there are some descriptions that reduce understanding only by users, by using LLMs, we showed the possibility of automatically identifying descriptions that are misunderstood by both LLMs and users, i.e., a first step towards automation of understandability assessment by LLMs.

In the future, to fill in the comprehension gaps between LLMs and users, we will explore methods to classify technical terms or simulate user thought processes in LLMs by providing contextual information, such as assumptions and personas.

References

[1] Data Protection Commission, "Data Protection Commission Announces Decision in WhatsApp Inquiry." https://www.dataprotection.ie/en/news-media/press-releases/data-protection-commission-announces-decision-whatsapp-inquiry, 2021.

[2] Bloomberg News, "Amazon Given Record 888 Million EU Fine for Data Privacy Breach." https://www.bloomberg.com/news/articles/2021-07-30/amazon-given-record-888-million-eu-fine-for-data-privacy-breach, 2021.

[3] A. Micheti *et al.*, "Fixing Broken Doors: Strategies for Drafting Privacy Policies Young People Can Understand," *Bulletin of Science, Technology & Society*, vol. 30, no. 2, pp. 130–143, 2010.

[4] J. Tang, H. Shoemaker, A. Lerner, and E. Birrell, "Defining privacy: How users interpret technical terms in privacy policies," in *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2021.

[5] K. Mori, T. Nagai, Y. Takata, M. Kamizono, and T. Mori, "Analysis of Privacy Compliance by Classifying Policies Before and After the Japanese Law Revision," *Journal of Information Processing*, vol. 31, pp. 829–841, 2023.

[6] B. Andow, S. Y. Mahmud, W. Wang, J. Whitaker, W. Enck, B. Reaves, K. Singh, and T. Xie, "PolicyLint: Investigating Internal Privacy Policy Contradictions on Google Play," in *USENIX Security Symposium*, 2019.

[7] C. Yan, F. Xie, M. H. Meng, Y. Zhang, and G. Bai, "On the quality of privacy policy documents of virtual personal assistant applications," in *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2024.

[8] R. Hara, A. Hasegawa, J. Jamieson, and M. Akiyama, "Understanding the Consistency between Collected Data and its Objectives in Privacy Policies," in *SPT*, 2024.

[9] H. Matsuo, K. Tonomura, M. Tsujimoto, M. Mizukoshi, T. Nagai, A. Eri, and S. Tomohiro, "Creation and Interpretation of Terms of Use and Privacy Policy – Based on Domestic and International Transactions." SHOJIHOMU Co., Ltd., 2023.

[10] C. Jensen, C. Potts, and C. Jensen, "Privacy practices of Internet users: Self-reports versus observed behavior," *International Journal of Human-Computer Studies*, vol. 63, no. 1-2, pp. 203–227, 2005.

[11] K. Mori, T. Nagai, Y. Takata, and M. Kamizono, "Analysis of Privacy Compliance by Classifying Multiple Policies on the Web," in *IEEE International Conference on Computers, Software, and Applications (COMPSAC)*, 2022.

[12] R. N. Zaeem, R. L. German, and K. S. Barber, "Privacycheck: Automatic summarization of privacy policies using data mining," *ACM Transactions on Internet Technology (TOIT)*, vol. 18, no. 4, pp. 1–18, 2018.

[13] L. Salewski, S. Alaniz, I. Rio-Torto, E. Schulz, and Z. Akata, "In-context Impersonation Reveals Large Language Models' Strengths and Biases," in *Neural Information Processing Systems (NeurIPS)*, 2024.

[14] Y. Shao, L. Li, J. Dai, and X. Qiu, "Character-llm: A trainable agent for role-playing," *arXiv:2310.10158*, 2023.

[15] F. Yu, L. Quartey, and F. Schilder, "Exploring the Effectiveness of Prompt Engineering for Legal Reasoning Tasks," in *Association for Computational Linguistics (ACL)*, 2023.

[16] C. Tang, Z. Liu, C. Ma, Z. Wu, Y. Li, W. Liu, D. Zhu, Q. Li, X. Li, T. Liu, *et al.*, "PolicyGPT: Automated Analysis of Privacy Policies with Large Language Models," *arXiv:2309.10238*, 2023.

[17] P. Pałka, M. Lippi, F. Lagioia, R. Liepiņa, and G. Sartor, "No More Trade-Offs. GPT and Fully Informative Privacy Policies," *arXiv:2402.00013*, 2023.

[18] A. Goknil, F. B. Gelderblom, S. Tverdal, S. Tokas, and H. Song, "Privacy policy analysis through prompt engineering for llms," *arXiv:2409.14879*, 2024.

[19] OpenAI, "GPT-4o." https://platform.openai.com/docs/models/o1#gpt-4o, 2024.

[20] Google Deepmind, "Geimini Pro." https://deepmind.google/technologies/gemini/pro/, 2024.

[21] Anthoropic, "Claude 3.5 Sonnet." https://www.anthropic.com/news/claude-3-5-sonnet, 2024.

[22] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A Learning Algorithm for Boltzmann Machines," *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

[23] J. Wei, M. Bosma, V. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned Language Models are Zero-Shot Learners," in *International Conference on Learning Representations (ICLR)*, 2022.

[24] Lancers, "Lancers." https://www.lancers.jp/, 2024.

[25] R. Khandelwal, A. Nayak, P. Chung, and K. Fawaz, "Comparing privacy labels of applications in android and ios," in *Proceedings of the 22nd Workshop on Privacy in the Electronic Society*, pp. 61–73, 2023.

[26] Y. Lin, J. Juneja, E. Birrell, and L. F. Cranor, "Data safety vs. app privacy: Comparing the usability of android and ios privacy labels," *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2024.

[27] "Provide information for Google Play Data safety section." https://support.google.com/googleplay/android-developer/answer/10787469?hl=en. [Accessed 14-11-2024].

[28] S. Kanamori, M. Ikeda, K. Kameishi, and A. A. Hasegawa, "Japanese users' (mis)understandings of technical terms used in privacy policies and the privacy protection law," in *International Conference on HCI for Cybersecurity, Privacy and Trust*, pp. 245–264, Springer, 2024.

[29] C. Olea, H. Tucker, J. Phelan, C. Pattison, S. Zhang, M. Lieb, D. C. Schmidt, and J. White, "Evaluating Persona Prompting for Question Answering Tasks," in *International Conference on Artificial Intelligence and Soft Computing (AIS)*, 2024.

[30] C. C. Chen, D. Shu, H. Ravishankar, X. Li, Y. Agarwal, and L. F. Cranor, "Is a trustmark and QR code enough? the effect of iot security and privacy label information complexity on consumer comprehension and behavior," in *2024 CHI Conference on Human Factors in Computing Systems*, pp. 832:1–832:32, ACM, 2024.

[31] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," in *Symposium on Usable Privacy and Security (SOUPS)*, pp. 1–12, 2009.

[32] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner, "Android permissions demystified," in *ACM Conference on Computer and Communications Security (CCS)*, pp. 627–638, 2011.

[33] Z. Qu, V. Rastogi, X. Zhang, Y. Chen, T. Zhu, and Z. Chen, "Autocog: Measuring the description-to-permission fidelity in android applications," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 1354–1365, 2014.

[34] T. Watanabe, M. Akiyama, T. Sakai, and T. Mori, "Understanding the inconsistencies between text descriptions and the use of privacy-sensitive resources of mobile apps," in *Eleventh Symposium On Usable Privacy and Security (SOUPS)*, pp. 241–255, 2015.

[35] L. Yu, X. Luo, X. Liu, and T. Zhang, "Can we trust the privacy policies of android apps?," in *Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pp. 538–549, IEEE, 2016.

[36] R. Khandelwal, A. Nayak, P. Chung, K. Fawaz, A. Bianchi, Z. B. Celik, Y. Yarom, X. S. Shen, Z. Fang, S. Zhang, *et al.*, "Unpacking privacy labels: A measurement and developer perspective on google's data safety section," in *USENIX Security Symposium*, pp. 2831–2848, 2024.

[37] Google, "Provide information for Google Play Data safety section." https://support.google.com/googleplay/android-developer/answer/10787469?hl=en#zippy=%2Ccan-my-app-be-blocked-by-google-play-due-to-the-information-i-submit-in-my-data-safety-form. [Accessed 14-11-2024].

[38] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, *et al.*, "Gpt-4 technical report," *arXiv:2303.08774*, 2023.

[39] Q-Success, "Usage Statistics and Market Share of Content Languages for Websites, November 2024 — w3techs.com." https://w3techs.com/technologies/overview/content_language. [Accessed 19-11-2024].

Appendix

A. Custom Privacy Policy

We show our custom privacy policies PP-B, C, D, and E in Figures 4, 5, 6, and 7, respectively. PP-B was obfuscated from PP-A using the following factors hindering user understanding:

1) The factor *Use of Double Negative* was used for "The Company does **not** ... **except** in ..." in Section "1. Provision to Third Parties."
2) The factor *Many Words per Sentences or Paragraphs* was used in Section "5. Security Measures."
3) By using the factor *Illogical Presentation Order*, PP-A's Section "5. Stop Providing Personal Information" was moved to Section "2. Stop Providing Personal Information" in PP-B.
4) By using the factor *Dispersed Information*, PP-A's Section "3. Provision of Personal Information" was divided into Sections "1. Provision to Third Parties" and "6. Joint Utilization" in PP-B.
5) By using the factor *Missing Information (Abstract Expression)*, "disclosure of personal data, correction, addition, deletion, and suspension of use" was abbreviated to "other inquiries regarding the handling of personal information" in Section "7. Inquiries."

PP-C was obfuscated from PP-A using the following factors hindering user understanding:

1) The same factor *Use of Double Negative* as in PP-B (1) was used in Section "4. Provision of Personal Information."
2) As the factor "Use of Technical Terms," the term "ISO27001" was used in Section "5. Security Measure" instead of "international certification by an external organization for information security management systems."
3) The same factor *Missing Information (Abstract Expression)* as in PP-B (5) was used in Section "7. Inquiries."

4) By using the factor *Missing Information (Abstract Expression)*, “step count estimation and calorie consumption prediction” was abbreviated to “the basic functions of the service” in Section “2. Data Collection and Purposes of Use.”
5) By using the factor *Omitting Information by Reference*, data and purpose descriptions for joint utilization were replaced with references to other sections.
6) As the factor *Description on Handling not Conducted*, “(*) Note that There is **no** outsourcing ...” was added in Section “4. Provision of Personal Information.”
7) As the factor *Description on Unnecessary Information*, Section “1. Compliance with Act on the Protection of Personal Information” was added.

PP-D was obfuscated from PP-A using the following factors hindering user understanding:

- The same factor *Dispersed Information* as in PP-B (4) was used in Sections “4. Provision to Third Parties” and “5. Joint Utilization.”
- By using the factor *Inconsistency between Paragraphs*, purposes not mentioned in Section “2. Data Collection and Purposes of Use” are listed in Section “5 Joint Utilization.”
- The same factor *Missing Information (Abstract Expression)* as in PP-B (5) was used in Section “8. Inquiries.”
- The same factor *Description on Unnecessary Information* as in PP-C (7) was used in Section 1.

PP-E was obfuscated from PP-A using the following factors hindering user understanding:

- The same factor *Dispersed Information* as in PP-B (4) was used in Sections “5. Provision to Third Parties” and “7. Joint Utilization.”
- By using the factor *Missing Information (Decoupling)*, PP-A’s Section “1. Data Collection and Purposes of Use” was divided into Sections “2. Data Collection” and “3. Purposes of Use” in PP-E and the correspondence between them was deleted.
- The same factor *Missing Information (Abstract Expression)* as in PP-B (5) was used in Section “9. Inquiries.”
- The same factor *Omitting Information by Reference* as in PP-C (5) was used in Section “7. Joint Utilization.”
- The same factor *Description on Unnecessary Information* as in PP-C (7) was used in Section 1.

B. Questionnaire for User Survey

The full text of the survey is as follows. Please note that the survey was conducted in Japanese and that the following texts were translated into English.

We aim to investigate users’ perceptions and opinions about explanations related to online services and will not be used for purposes other than this study. Your answers will be treated as anonymous data, and no individuals will be identified from the answers. If you decide not to participate, you can abandon the task at anytime. If you decide to participate (and do not stop answering), your anonymous data may be used for our future studies.

I have read the above explanation and agree to participate in this questionnaire.

- Yes
- No

This questionnaire is for those who have not previously answered the survey titled “Questionnaire on Online Service A/B/C/D/E.” (If you do not meet this requirement, we will not be able to accept your answers.)

Do you meet the above requirement?

- Yes
- No

This questionnaire consists of 17 questions regarding the content of a fictitious privacy policy, and 9 questions about yourself. Below is the privacy policy of a fictional mobile app for step tracking. Please refer to the privacy policy as you answer the following questions.

[One of the privacy policies created in Section III.]

Q1. Select all options that are correct for the information items to be collected and used by PRIVACY Co., Ltd. along with their purposes. If choosing option E, answer only option E without the others.

A. Use email addresses for account management.
B. Use location data for advertisement.
C. Use device data for step count estimation.
D. Use email addresses for advertisement.
E. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q1.

Q2. Select all options that are correct regarding the purpose of the personal data provided by PRIVACY Co., Ltd. to other companies. For option D, answer only option D without the others.

A. To measure advertising effectiveness.
B. To develop new services.
C. To contact with customers.
D. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q2.

Q3. Select all options that are correct for security measures taken by PRIVACY Co., Ltd. For option D, answer only option D without the others.

A. Measures to prevent loss of devices that handle personal data.
B. Maintaining security certification by external organization.
C. Outsourcing supervision.
D. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q3.

Q4. Select all options that are correct for the items and purposes of information that are jointly used by PRIVACY Co., Ltd. and its subsidiaries. For option D, answer only option D without the others.

A. Use email addresses for customer communication.
B. Use email addresses for account management.
C. Use age and gender for calorie consumption prediction.
D. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q4.

Q5. Based on the privacy policy, is outsourcing conducted? Select one correct option.

A. Outsourcing is conducted.
B. Outsourcing is not conducted.
C. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q5.

Q6. Based on the privacy policy, where is the data center where personal information is stored? Select one correct option.

A. Japan
B. Countries other than Japan.
C. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q6.

Q7. Based on the privacy policy, where can users contact for data deletion? Select one correct option.

A. CEO of PRIVACY Co., Ltd.
B. Inquiry form.
C. Cannot be determined from the privacy policy (no description).

Please specify the sections of the privacy policy (e.g., chapter titles or text) you referred to when answering Q7.

Q8. Do you think the above privacy policy is long? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q9. Do you think the above privacy policy uses many technical terms? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q10. Do you think the above privacy policy lacks necessary explanations? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q11. Do you think related information scattered across multiple sections in the above privacy policy? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q12. Do you think the information in the above privacy policy is clearly stated? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q13. Do you think the order in which information is presented in the above privacy policy is appropriate? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q14. Do you think there are contradictions in the above privacy policy? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q15. Do you think the above privacy policy is easy to read? Please answer on a 5-point scale.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q16. If you have any reasons for finding the privacy policy easy or difficult to read, please describe them (other than those covered in Q8–-Q14).

Q17. This is a question to verify whether you are reading the questions. Please select the last option and proceed to the next question.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Q18. Please select your device to answer this questionnaire.

- Smartphone

- Tablet
- PC
- Other

If you selected “Other” in Q18, please describe the device. If you did not select “Other,” please enter “–”.

Q19. Please select your age.

- 18–19 years old
- 20–29 years old
- 30–39 years old
- 40–49 years old
- 50–59 years old
- 60–69 years old
- 70 years old or older
- Prefer not to answer

Q20. Please select your gender (self-identified).

- Male
- Female
- Other
- Prefer not to answer

Q21. Please select your current occupation.

- Student
- Company employee
- Company executive
- Public servant, educator, or nonprofit employee
- Part-time worker
- Homemaker
- Retired/unemployed
- Other occupation
- Prefer not to answer

Q22. Please select your highest education level.

- Secondary education
- High school diploma
- Technical college
- Undergraduate degree
- Graduate degree
- Doctorate degree
- Other
- Prefer not to answer

Q23. Please select the primary field of your occupation or study.

- Arts and Humanities
- Education
- Social Sciences
- Journalism
- Administration and Law
- Mathematics and statistics
- Information and Communication Technologies
- Manufacturing and construction
- Agriculture, forestry and fisheries
- Health and welfare
- Services
- Natural Sciences
- History
- Other
- Prefer not to answer

Q24. Please select all actions or experiences you have had regarding the handling of personal information.

- Inquiring with a service provider about the handling of personal information
- Submitting an opt-out request to stop the handling of personal information
- Choosing not to use a service due to concerns about the handling of personal information
- Sharing opinions on social media or forums regarding the handling of personal information of specific services
- None of the above

Q25. How do you read privacy policies?

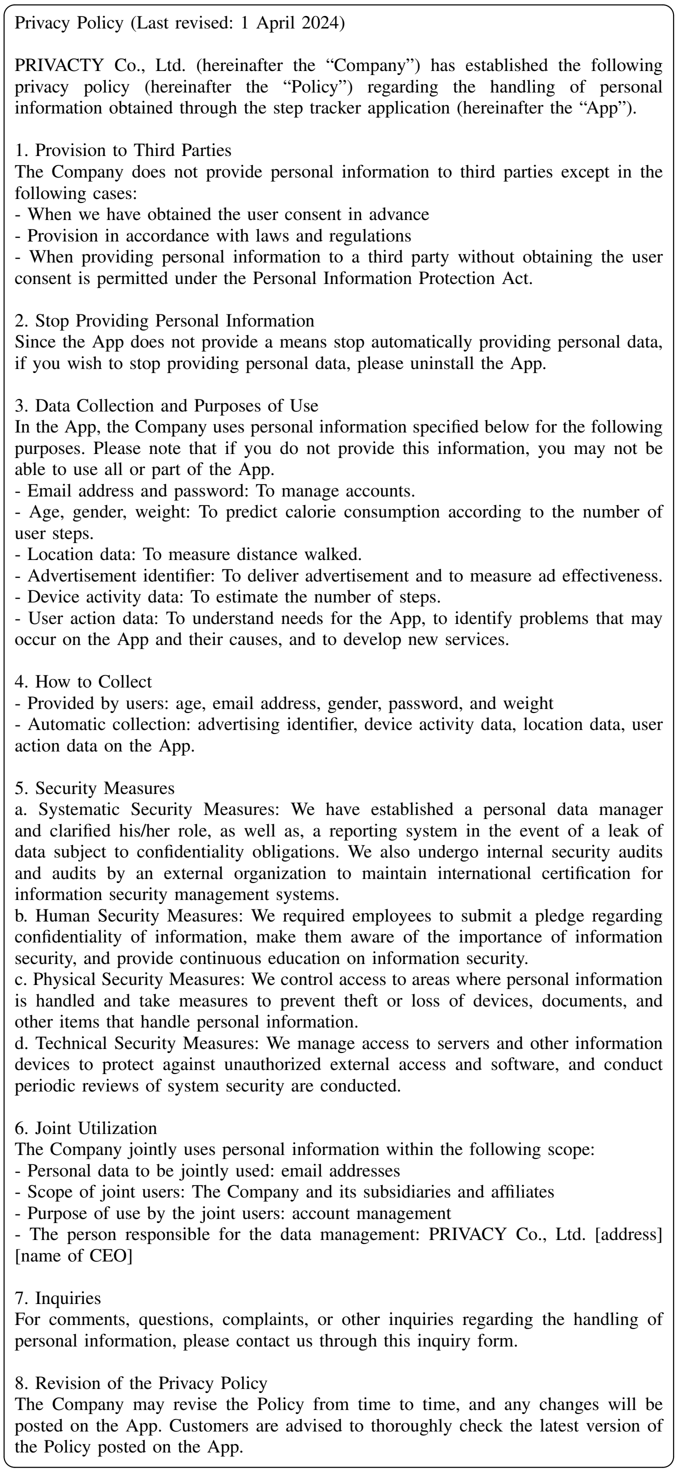
- Reading thoroughly
- Skimming
- searching for keywords
- looking at section headers
- Using tools
- Other
- Do not read privacy policies

If you selected “Other” in Q25, please describe how to read. If you did not select “Other,” please enter “–”.

Q26. When do you read privacy policies? Please select all that apply.

- When registering for a service
- Upon receiving a notification of privacy policy updates
- When entering personal data during service use
- Other
- Do not read privacy policies

If you selected “Other” in Q26, please describe when to read. If you did not select “Other,” please enter “–”.



Privacy Policy (Last revised: 1 April 2024)

PRIVACTY Co., Ltd. (hereinafter the “Company”) has established the following privacy policy (hereinafter the “Policy”) regarding the handling of personal information obtained through the step tracker application (hereinafter the “App”).

1. Provision to Third Parties
The Company does not provide personal information to third parties except in the following cases:
- When we have obtained the user consent in advance
- Provision in accordance with laws and regulations
- When providing personal information to a third party without obtaining the user consent is permitted under the Personal Information Protection Act.

2. Stop Providing Personal Information
Since the App does not provide a means stop automatically providing personal data, if you wish to stop providing personal data, please uninstall the App.

3. Data Collection and Purposes of Use
In the App, the Company uses personal information specified below for the following purposes. Please note that if you do not provide this information, you may not be able to use all or part of the App.
- Email address and password: To manage accounts.
- Age, gender, weight: To predict calorie consumption according to the number of user steps.
- Location data: To measure distance walked.
- Advertisement identifier: To deliver advertisement and to measure ad effectiveness.
- Device activity data: To estimate the number of steps.
- User action data: To understand needs for the App, to identify problems that may occur on the App and their causes, and to develop new services.

4. How to Collect
- Provided by users: age, email address, gender, password, and weight
- Automatic collection: advertising identifier, device activity data, location data, user action data on the App.

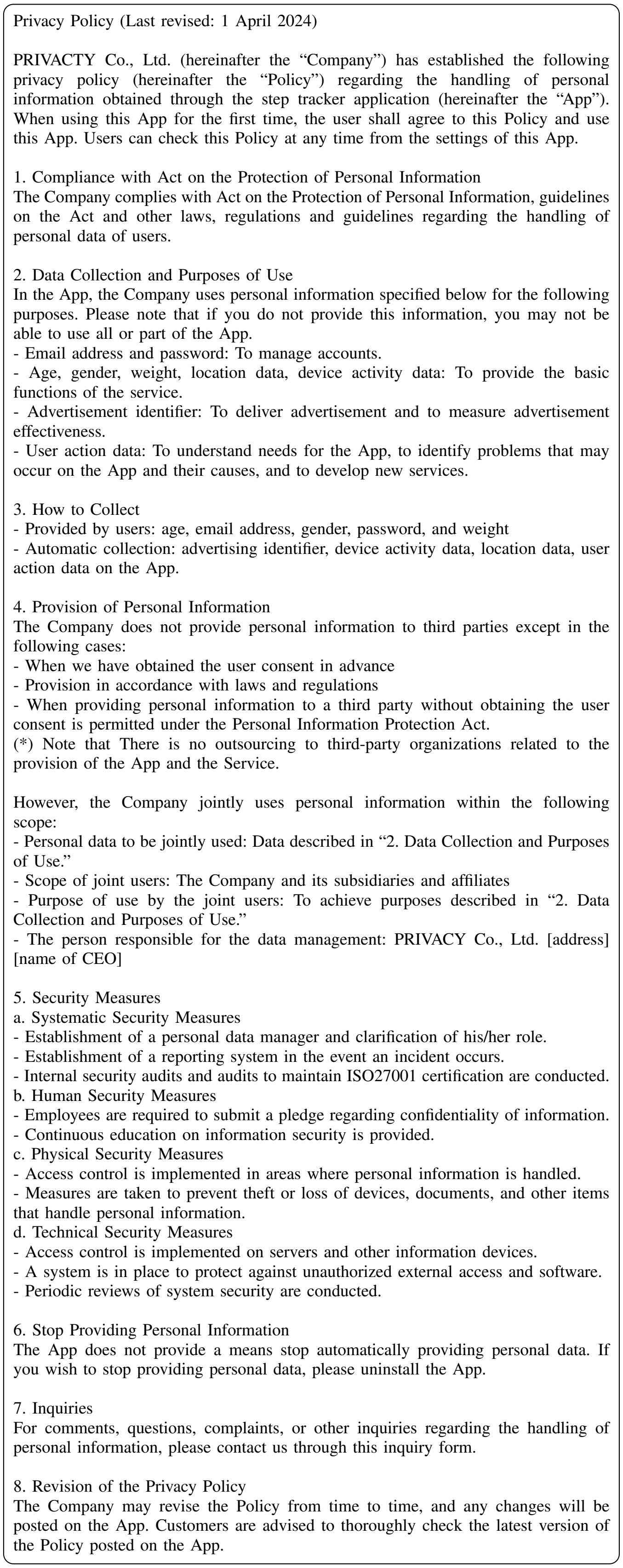
5. Security Measures
a. Systematic Security Measures: We have established a personal data manager and clarified his/her role, as well as, a reporting system in the event of a leak of data subject to confidentiality obligations. We also undergo internal security audits and audits by an external organization to maintain international certification for information security management systems.
b. Human Security Measures: We required employees to submit a pledge regarding confidentiality of information, make them aware of the importance of information security, and provide continuous education on information security.
c. Physical Security Measures: We control access to areas where personal information is handled and take measures to prevent theft or loss of devices, documents, and other items that handle personal information.
d. Technical Security Measures: We manage access to servers and other information devices to protect against unauthorized external access and software, and conduct periodic reviews of system security are conducted.

6. Joint Utilization
The Company jointly uses personal information within the following scope:
- Personal data to be jointly used: email addresses
- Scope of joint users: The Company and its subsidiaries and affiliates
- Purpose of use by the joint users: account management
- The person responsible for the data management: PRIVACY Co., Ltd. [address] [name of CEO]

7. Inquiries
For comments, questions, complaints, or other inquiries regarding the handling of personal information, please contact us through this inquiry form.

8. Revision of the Privacy Policy
The Company may revise the Policy from time to time, and any changes will be posted on the App. Customers are advised to thoroughly check the latest version of the Policy posted on the App.

Fig. 4. Privacy Policy B (PP-B) with Writing-based Obfuscation.



Privacy Policy (Last revised: 1 April 2024)

PRIVACTY Co., Ltd. (hereinafter the “Company”) has established the following privacy policy (hereinafter the “Policy”) regarding the handling of personal information obtained through the step tracker application (hereinafter the “App”). When using this App for the first time, the user shall agree to this Policy and use this App. Users can check this Policy at any time from the settings of this App.

1. Compliance with Act on the Protection of Personal Information
The Company complies with Act on the Protection of Personal Information, guidelines on the Act and other laws, regulations and guidelines regarding the handling of personal data of users.

2. Data Collection and Purposes of Use
In the App, the Company uses personal information specified below for the following purposes. Please note that if you do not provide this information, you may not be able to use all or part of the App.
- Email address and password: To manage accounts.
- Age, gender, weight, location data, device activity data: To provide the basic functions of the service.
- Advertisement identifier: To deliver advertisement and to measure advertisement effectiveness.
- User action data: To understand needs for the App, to identify problems that may occur on the App and their causes, and to develop new services.

3. How to Collect
- Provided by users: age, email address, gender, password, and weight
- Automatic collection: advertising identifier, device activity data, location data, user action data on the App.

4. Provision of Personal Information
The Company does not provide personal information to third parties except in the following cases:
- When we have obtained the user consent in advance
- Provision in accordance with laws and regulations
- When providing personal information to a third party without obtaining the user consent is permitted under the Personal Information Protection Act.
(*) Note that There is no outsourcing to third-party organizations related to the provision of the App and the Service.

However, the Company jointly uses personal information within the following scope:
- Personal data to be jointly used: Data described in “2. Data Collection and Purposes of Use.”
- Scope of joint users: The Company and its subsidiaries and affiliates
- Purpose of use by the joint users: To achieve purposes described in “2. Data Collection and Purposes of Use.”
- The person responsible for the data management: PRIVACY Co., Ltd. [address] [name of CEO]

5. Security Measures
a. Systematic Security Measures
- Establishment of a personal data manager and clarification of his/her role.
- Establishment of a reporting system in the event an incident occurs.
- Internal security audits and audits to maintain ISO27001 certification are conducted.
b. Human Security Measures
- Employees are required to submit a pledge regarding confidentiality of information.
- Continuous education on information security is provided.
c. Physical Security Measures
- Access control is implemented in areas where personal information is handled.
- Measures are taken to prevent theft or loss of devices, documents, and other items that handle personal information.
d. Technical Security Measures
- Access control is implemented on servers and other information devices.
- A system is in place to protect against unauthorized external access and software.
- Periodic reviews of system security are conducted.

6. Stop Providing Personal Information
The App does not provide a means stop automatically providing personal data. If you wish to stop providing personal data, please uninstall the App.

7. Inquiries
For comments, questions, complaints, or other inquiries regarding the handling of personal information, please contact us through this inquiry form.

8. Revision of the Privacy Policy
The Company may revise the Policy from time to time, and any changes will be posted on the App. Customers are advised to thoroughly check the latest version of the Policy posted on the App.

Fig. 5. Privacy Policy C (PP-C) with Obfuscation Specific to Privacy Policies, in particular, *Use of Technical Terms*.

Privacy Policy (Last revised: 1 April 2024)

PRIVACTY Co., Ltd. (hereinafter the “Company”) has established the following privacy policy (hereinafter the “Policy”) regarding the handling of personal information obtained through the step tracker application (hereinafter the “App”). When using this App for the first time, the user shall agree to this Policy and use this App. Users can check this Policy at any time from the settings of this App.

1. Compliance with Act on the Protection of Personal Information
The Company complies with Act on the Protection of Personal Information, guidelines on the Act and other laws, regulations and guidelines regarding the handling of personal data of users.

2. Data Collection and Purposes of Use
In the App, the Company uses personal information specified below for the following purposes. Please note that if you do not provide this information, you may not be able to use all or part of the App.
- Email address and password: To manage accounts.
- Age, gender, weight, location data, device activity data: To provide the basic functions of the service.
- Advertisement identifier: To deliver advertisement and to measure ad effectiveness.
- User action data: To understand needs for the App, to identify problems that may occur on the App and their causes, and to develop new services.

3. How to Collect
- Provided by users: age, email address, gender, password, and weight
- Automatic collection: advertising identifier, device activity data, location data, user action data on the App.

4. Provision to Third Parties
The Company provides personal information to third parties only in the following cases:
- When we have obtained the user consent in advance
- Provision in accordance with laws and regulations
- When providing personal information to a third party without obtaining the user consent is permitted under the Personal Information Protection Act.

5. Joint Utilization
The Company jointly uses personal information within the following scope:
- Personal data to be jointly used: age, email addresses, gender
- Scope of joint users: The Company and its subsidiaries and affiliates
- Purpose of use by the joint users: To contact with customers, to deliver advertisement, to measure ad effectiveness
- The person responsible for the data management: PRIVACY Co., Ltd. [address] [name of CEO]

6. Security Measures
a. Systematic Security Measures
- Establishment of a personal data manager and clarification of his/her role.
- Establishment of a reporting system in the event of a leak of data subject to confidentiality obligations.
- Internal security audits and audits to maintain international certification by an external organization for information security management systems are conducted.
b. Human Security Measures
- Employees are required to submit a pledge regarding confidentiality of information.
- Continuous education on information security is provided.
c. Physical Security Measures
- Access control is implemented in areas where personal information is handled.
- Measures are taken to prevent theft or loss of devices, documents, and other items that handle personal information.
d. Technical Security Measures
- Access control is implemented on servers and other information devices.
- A system is in place to protect against unauthorized external access and software.
- Periodic reviews of system security are conducted.

7. Stop Providing Personal Information
The App does not provide a means stop automatically providing personal data. If you wish to stop providing personal data, please uninstall the App.

8. Inquiries
For comments, questions, complaints, or other inquiries regarding the handling of personal information, please contact us through this inquiry form.

9. Revision of the Privacy Policy
The Company may revise the Policy from time to time, and any changes will be posted on the App. Customers are advised to thoroughly check the latest version of the Policy posted on the App.

Fig. 6. Privacy Policy D (PP-D) with Obfuscation Specific to Privacy Policies, in particular, *Dispersed Information*.

Privacy Policy (Last revised: 1 April 2024)

PRIVACTY Co., Ltd. (hereinafter the “Company”) has established the following privacy policy (hereinafter the “Policy”) regarding the handling of personal information obtained through the step tracker application (hereinafter the “App”). When using this App for the first time, the user shall agree to this Policy and use this App. Users can check this Policy at any time from the settings of this App.

1. Compliance with Act on the Protection of Personal Information
The Company complies with Act on the Protection of Personal Information, guidelines on the Act and other laws, regulations and guidelines regarding the handling of personal data of users.

2. Data Collection
In the App, the Company uses personal information specified below. Please note that if you do not provide this information, you may not be able to use all or part of the App.
- Email address
- Password
- Age
- Gender
- Weight
- Location data
- Device activity data
- Advertisement identifier
- User action data

3. Purposes of Use
In the App, the Company uses collected information for the following purposes.
- To provide the basic function of the service (to predict calorie consumption according to the number of user steps, to estimate the number of steps, and etc.)
- To manage accounts
- To deliver advertisements and to measure the effectiveness
- To understand needs for the App
- To identify problems that may occur on the App and their causes
- To develop new services

4. How to Collect
- Provided by users: age, email address, gender, password, and weight
- Automatic collection: advertising identifier, device activity data, location data, user action data on the App.

5. Provision to Third Parties
The Company provides personal information to third parties only in the following cases:
- When we have obtained the user consent in advance
- Provision in accordance with laws and regulations
- When providing personal information to a third party without obtaining the user consent is permitted under the Personal Information Protection Act.

6. Security Measures
a. Systematic Security Measures
- Establishment of a personal data manager and clarification of his/her role.
- Establishment of a reporting system in the event of a leak of data subject to confidentiality obligations.
- Internal security audits and audits to maintain international certification by an external organization for information security management systems are conducted.
- Measures are taken to require our subcontractors to take security measures, and necessary and appropriate supervision is carried out.

b. Human Security Measures
- Employees are required to submit a pledge regarding confidentiality of information.
- Continuous education on information security is provided.

c. Physical Security Measures
- Access control is implemented in areas where personal information is handled.
- Measures are taken to prevent theft or loss of devices, documents, and other items that handle personal information.

d. Technical Security Measures
- Access control is implemented on servers and other information devices.
- A system is in place to protect against unauthorized external access and software.
- Periodic reviews of system security are conducted.

7. Joint Utilization
The Company jointly uses personal information within the following scope:
- Personal data to be jointly used: Data described in “2. Data Collection”
- Scope of joint users: The Company and its subsidiaries and affiliates
- Purpose of use by the joint users: To achieve purposes described in “3. Purposes of Use”
- The person responsible for the data management: PRIVACY Co., Ltd. [address] [name of CEO]

8. Stop Providing Personal Information
The App does not provide a means stop automatically providing personal data. If you wish to stop providing personal data, please uninstall the App.

9. Inquiries
For comments, questions, complaints, or other inquiries regarding the handling of personal information, please contact us through this inquiry form.

10. Revision of the Privacy Policy
The Company may revise the Policy from time to time, and any changes will be posted on the App. Customers are advised to thoroughly check the latest version of the Policy posted on the App.

Fig. 7. Privacy Policy E (PP-E) with Obfuscation Specific to Privacy Policies, in particular, *Missing Information (Decoupling)*.