Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #49401

From Data Trends to Privacy Insights in Mental Health Apps: an LLM-Powered Approach

Miss Xingyu You, Arcadia University

Xingyu You received a B.S. degree with a double major in Mathematics and Data Science from the Department of Computer Science and Mathematics at Arcadia University, and a B.S. in Mathematics from the Department of Mathematics and Applied Mathematics, School of Mathematical Sciences at Jiangsu University. Her research interests include artificial intelligence in healthcare, large language models (LLMs), and data science.

Mr. Wang Wang, Arcadia University

Wang Wang received a B.S. degree in Mathematics with a minor in Computer Science from the Department of Computer Science and Mathematics at Arcadia University, and a B.S. degree in Mathematics from the Department of Mathematics and Applied Mathematics, School of Mathematical Sciences at Jiangsu University.

Zhairui Shen
Dr. Yanxia Jia, Arcadia University

Dr. Yanxia Jia is a full Professor of Computer Science in the Department of Computer Science and Mathematics at Arcadia University. She earned a Ph.D. in Computing Science from the University of Alberta, Canada. Dr. Jia's research interests include artificial intelligence and machine learning, large language models (LLMs), data science, and computer science education.

From Data Trends to Privacy Insights in Mental Health Apps: an LLM-Powered Approach

Abstract

Objective and Motivation:

The use of mobile apps to manage mental health has grown significantly in recent years. With features ranging from mood tracking and guided meditation to virtual therapy sessions, these apps provide a convenient way for people to address mental health challenges in their daily lives. This study examines trends in mental health apps since 2009, including a comparison of app features and usage between pre- and post-COVID periods.

Privacy policies for mobile apps are crucial as they explain how user data is collected, used, stored, and shared. Given the sensitive nature of mental health data, privacy issues in mental health apps are particularly critical. However, these documents are often overlooked due to their length and complexity. This study leverages Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) techniques to analyze privacy policy documents from mobile apps available in the Apple App Store. The goal is to extract answers to key privacy and security-related questions, providing insights for stakeholders, including mental health app users, developers, and policymakers.

Methodology:

General app information—such as genres, user ratings, pricing, and release year—is collected from the Apple App Store using the Apple Store API and Python tools. An exploratory analysis is conducted to identify trends in mental health apps since 2009.

Privacy policy documents of mental health apps are also collected for analysis. ChatGPT is utilized to extract privacy-related metrics, such as the percentage of apps that reference privacy regulations like the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and the Children's Online Privacy Protection Act (COPPA). LLMs and RAG are employed to answer critical privacy and security-related questions from the dataset of privacy policy documents. These questions cover multiple categories, such as the types of user information collected, details on third-party data sharing, and whether users are given options to opt out of data collection.

Results:

Our analysis reveals a significant increase in both the number and quality of mental health apps in the post-pandemic era. Out of 4,764 English-language apps, 3,371 have privacy policies and 1,393 do not. Approximately 37.9% of the apps reference privacy regulations such as GDPR, HIPAA, or COPPA in their privacy policies, revealing gaps in awareness and compliance with privacy standards. In addition, the analysis of privacy policy documents reveals that the vast majority of mental health apps are transparent about their user data collection actions (91%) and purpose (96.7%). However, a smaller percentage (76.9%) disclose data sharing with third parties, and 81.3% provide users with options to opt out of data collection or to delete their data.

Conclusions:

As the adoption of digital mental health solutions continues to rise, ensuring compliance with privacy regulations is increasingly critical. Additionally, tools that simplify the understanding of complex privacy policies are also essential to enhance user trust and awareness. This study underscores the importance of transparency and regulation in the growing field of mental health apps, offering insights for improving user awareness and protection while promoting responsible app development.

Introduction

Over the past decade, there has been a significant increase in the provision of mental health services through mobile health (mHealth) systems, such as mobile applications, offering a wide range of functionalities from mood tracking and guided meditation to virtual therapy. A survey conducted in the United States revealed that 70% of 320 outpatient participants expressed a strong interest in using mental health tools to manage their psychological well-being [1]. This study examines trends in mental health apps since 2009, including a comparison of app statistics between the pre- and post-COVID periods. Additionally, analysis of app's privacy policy information is also conducted.

Privacy policies for mobile apps are crucial as they explain how user data is collected, used, stored, and shared. Studies have shown that many mobile Health applications have significant security vulnerabilities, jeopardizing the privacy of millions of users [2]. For mental health apps, it is particularly critical for users of to understand the privacy implications of these apps, given the sensitive nature of mental health-related data and the heightened vulnerability of individuals replying on these tools. Both the Apple App Store and Google Play Store have strict requirements for privacy policies, especially for apps that collect personal or sensitive data [3, 4]. However, these policies are often written in complex and obscure language, leading users to either skip reading them or struggle to fully understand their content. It was estimated that it would take an average user 201 hours per year to read all the privacy policies they encounter [5].

To address this issue, this study employs natural language processing (NLP) tools and techniques, such as large language models (LLMs) and retrieval-augmented generation (RAG), to automatically extract answers to privacy-related questions. These questions are based on the mental health evaluation model [6] developed by the American Psychiatric Association (APA), designed to help users make informed decision when considering mobile health apps. The questions cover topics such as whether the app collects user data, the types of data collected, whether user data is shared with third parties, and whether the user has an option to opt out of data collections. The answers obtained are further analyzed to identify patterns and data-handling practices in mobile app privacy policies.

In recent years, there has been a considerable work on automatic identifying, extracting, and analyzing privacy policies. As summarized in [7], the task of automatic privacy policy analysis falls into two categories: symbolic approaches and statistical approaches. Symbolic approaches rely on pre-defined rules, while statistical approaches leverage machine learning and deep learning techniques.

With the advent of powerful large language models, such as GPT-4, Llama 3, and Claude 2, research efforts [7, 8] have explored leveraging LLMs to perform automatic privacy policy analysis. These methods focus on prompt engineering and pre-processing, such as segmentation of the policy text. In contrast, this study employs RAG [9], a hybrid NLP technique that combines information retrieval with text generation, as illustrated in Figure 1. Compared to the prompt Engineering-based approaches, such as those using ChatGPT API, the additional retrieval step in RAG improves accuracy and relevance by retrieving the most contextually relevant information before calling LLM for answer generation. This approach reduces hallucination and improves overall performance, as demonstrated in Tables 1 and 2, by grounding the LLM's responses in the retrieved data.

Methodology

Data Collection

To conduct this study, we use the Apple Store API [11] and apply "mental health" as a filter to collect app data related mental health. The collected attributes include, but are not limited to, the app name, release date, primary genre, average user ratings, user rating count, price, recent update times, supported languages, and a link to the app's privacy policy [12]. The privacy policy links contain various data formats, e.g., HTML, PDF, Notion and Google Docs. Some are written in non-English languages, and some links lead to the HTTP "404" error. Therefore, pre-processing is performed to clean the data and ensure valid privacy policy documents. Various tools, including BeautifulSoup, Tesseract OCR, and command line scripts are used for this purpose. As a result, out of the total 5,039 apps collected, 4,764 English privacy policies are retained for analysis.

Research Questions and Methods

Based on the collected mental health app data, we investigate the following research questions (RQ) across two categories. To address these questions, we use exploratory analysis and develop an LLM and RAG-based question-answering program.

Category 1: research questions related to general information of the apps	
RQ1	How have the quantity and quality (measured by the app's average user ratings) of mental health apps evolved from 2009 to 2024?
RQ2	What changes, if any, have occurred in the quantity and quality of mental health apps between pre- and post-pandemic periods?

Category 2: research questions related to app privacy policies	
Prevalence of Privacy Policies among mobile apps	
RQ3	What proportion of apps have privacy policies?
Alignment with Key Privacy Laws and Standards	
RQ4	What proportion of apps reference key privacy laws or standards, such as the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and Children's Online Privacy Protection Act (COPPA)?
RQ5	What are the most frequently referenced privacy laws or standards referenced in the privacy policy documents?
Transparency in Data Collection, Usage and Sharing Practices	
RQ6	What proportion of apps declare the collection of user data in their privacy policies?
RQ7	For apps that declare the collection of user data, what proportion specify the purpose of data collection and use?
RQ8	For apps that collect user data, what proportion disclose sharing data with third parties?
Availability of Options for Users to Manage Their Personal Data	
RQ9	For apps that collect user data, what proportion provide users with options to opt out of data collection or to delete their data?

Category 2 examines privacy policies based on several research questions, each addressing a distinct aspect of app privacy.

- Prevalence of Privacy Policies among Mobile Apps (RQ3)
- Alignment with Key Privacy Laws and Standards (RQ4 and RQ5)
- Transparency in Data Collection, Usage and Sharing Practices (RQ6 ~ RQ8)
- Availability of Options for Users to Manage Their Personal Data (RQ9)

To address RQ3, the 4,764 English-based apps are processed and those lacking an accessible privacy policy document are identified.

For RQ4 and RQ5, references to privacy laws or regulations within privacy policy documents are identified through a two-step process. First, ChatGPT is prompted to generate a comprehensive list of privacy laws. Then, regular expressions are applied to detect matches for these regulations within the documents.

RQ6 through RQ9 are developed based on the privacy and security questions in the app evaluation model [16] proposed by the American Psychiatric Association (APA), which are designed to assist users in make informed decision when selecting mobile health apps. This model includes nine privacy and security-related questions, from which the following four are selected:

APA_Q1: Does the app declare the collection of data? What type of data does it collect?
APA_Q2: Does the app declare the purpose of data collection and use?

APA_Q3: Does the app share data with third parties? What third parties does it share data with?
APA_Q4: Can you opt out of data collection or delete data?

The remaining questions are excluded due to challenges with verification. For example, the question “If appropriate, is the app equipped to respond to potential harms or safety concerns?” can yield inconsistent responses among human evaluators. If a policy states that users are provided with a contact email for questions, some human evaluators interpret this as a “yes” while others say “no”. We find that LLMs can provide two different answers along with reasonable analysis in this case, further complicating verification.

For the APA questions, we develop a RAG-based question-answering(QA) program utilizing the haystack [13] framework and the ChatGPT API to automatically extract answers from the privacy policies. Specifically, the *gpt-3.5-turbo* model is used for generating answers based on the given questions. A retrieval stage is introduced before querying the model to extract the top three segments of the document most relevant to the question. As demonstrated in Tables 1 and 2, this approach improves overall performance compared to querying the policy document solely using the ChatGPT API.

The QA program consists of two components: an indexing pipeline and a RAG pipeline, as illustrated in Figure 1.

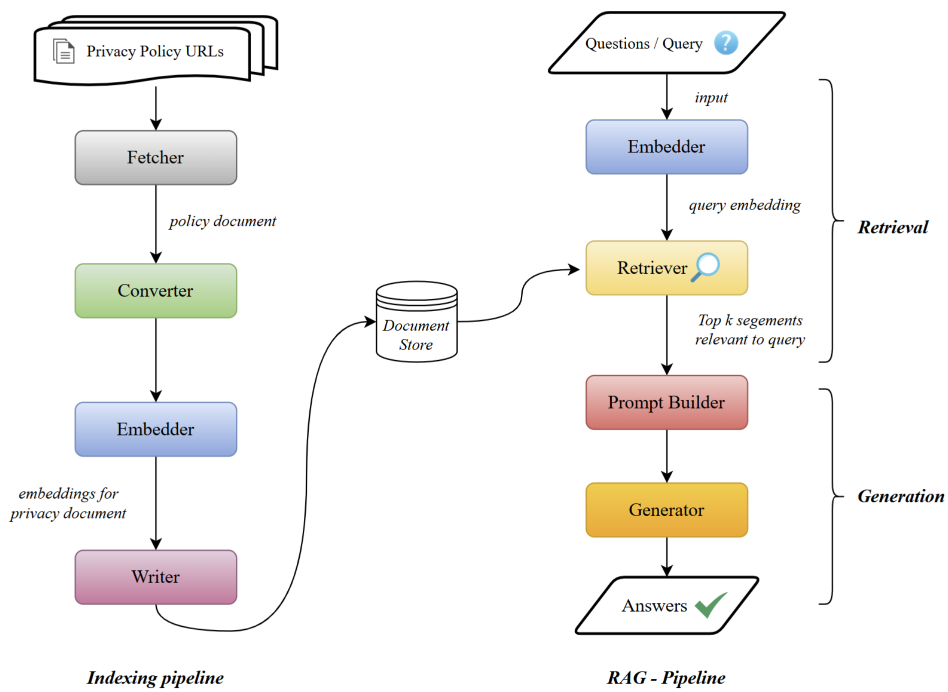

Figure 1: The QA System Leverages LLM and RAG.

The indexing pipeline is responsible for creating embeddings for the policy document and consists of the following four components:

- **Fetcher**: Fetches the privacy policy based on the policy URL.
- **Converter**: Converts the document into a suitable format with *HTMLToDocument*.

- **Embedder**: Creates embeddings for the privacy documents.
- **Writer**: Stores the document embeddings in a document store, which will be used during the retrieving stage.

The RAG pipeline consists of the following components:

- **Embedder**: Creates an embedding for the question/query.
- **Retriever**: Uses the query embedding to retrieve the top K most relevant segments from the previously saved document embeddings. This ensures that the query is executed against the retrieved segment most relevant to the query, instead of the whole document.
- **Prompt builder**: Build a prompt using the Chain-of-Thought prompting technique [14].
- **Generator**: Queries an LLM model, e.g., "gpt-3.5-turbo", to generate answers based on the question/query, the retrieved document, and the prompt.

The answers are stored in JSON format, with each record containing metadata, such as app id and URL, as well as a reply section including the question, generated answer, analysis and reference sections. The reference section contains original text from which the answer is derived, providing transparency. The analysis section contains the LLM's rationale behind the answer, aiding verification. These two sections allow tracing answers back to their sources and evaluate the reasoning behind them, thereby achieving accuracy and interpretability.

Performance Evaluation of the QA System

To ensure the accuracy and reliability of the program, we randomly selected 100 apps to manually verify the correctness of the QA program's Yes/No answers to APA_Q1 through APA_Q4. Each generated response is verified by two independent evaluators, and in case of discrepancies, a third reviewer's opinion is considered. Table 1 shows the results of performance evaluation based on performance metrics precision, recall and F1 score.

Question	Accuracy	Recall	Precision	F1 score
APA_Q1	0.98	**0.9890**	0.9890	0.9890
APA_Q2	0.95	0.9889	**0.9570**	0.9727
APA_Q3	0.99	**1.0000**	0.9861	0.9930
APA_Q4	0.84	1.0000	**0.8261**	0.9048

Table 1: Performance Results for Generating Yes/No Answers to APA_Q1 ~ Q4, using RAG-based method.

We also compare the performance of the QA program, which is based on RAG, and that of ChatGPT API-based QA to evaluate the effectiveness of the RAG-based approach, as opposed to the basic ChatGPT API-based method.

Question	Accuracy	Recall	Precision	F1 score
APA_Q1	0.95	**1.0000**	0.9479	0.9733
APA_Q2	0.92	1.0000	**0.9184**	0.9574
APA_Q3	0.92	**0.9718**	0.9200	0.9452
APA_Q4	0.80	1.0000	**0.7917**	0.8837

Table 2: Performance Results for Generating Yes/No Answers to APA_Q1~Q4, using ChatGPT API-based method.

As revealed in Table 2, the RAG-based method achieves higher accuracy and F1 scores across all four questions compared to the ChatGPT API-based method. For APA_Q1 and APA_Q3, recall -- an important measure for minimizing false negatives -- is the key metric. Both methods achieve strong recall performance. For APA_Q2 and APA_Q4, precision – an indicator of false positive rates, is more critical. In these cases, the RAG-based method achieves superior precision, demonstrating overall better performance than the ChatGPT API-based approach.

Given these performance results, the answers to APA_Q1 ~ APA_Q4 are further processed to address research questions RQ6 ~ RQ9, and the results are detailed in the Results section.

Results

In this section, we will address the research questions in the two categories mentioned before.

- <u>Category 1: research questions related to general app information of apps (RQ1 and RQ2)</u>

	UserRatingCount	URCPerYear	Average User Rating	Price	Language Count	Genre Count	ReleaseYear	UpdateYear
min	0	0.00	0.00	0	1	1	2009	2012
25%	0	0.00	0.00	0	1	1	2020	2022
50%	1	0.33	3.88	0	1	2	2022	2024
75%	10	3.00	4.88	0	1	2	2023	2024
95%	631	135.97	5.00	0	11	4	2024	2024
max	1784629	148719.08	5.00	129	41	4	2024	2024

Table 3: Descriptive Statistics: User Ratings, Pricing, Language Count and Release/Update Years.

Table 3 shows that 75% of the 5,039 mental health apps are released in 2020 or later and half of all of the apps have active updates continuing in 2024. Most apps receive very few reviews, with 75% receiving less than ten; however, average user ratings are generally high, with a median of 3.88.

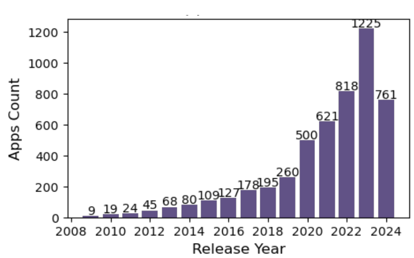

Figure 2: Trends in the number of Apps Released

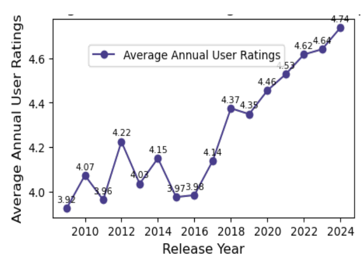

Figure 3a: Average Annual User Ratings for Released Apps

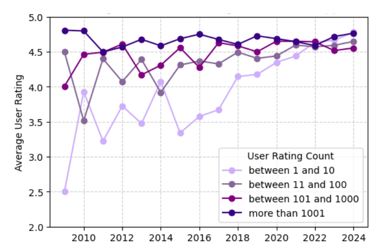

Figure 3b: Average Annual User Ratings based on Review Count

Figure 2, Figure 3a and Figure 3b address the previously mentioned RQ1. As shown in Figure 2, the number of mental health app releases has significantly increased since 2009, peaking in 2023.This trend reflects a growing attention on mental health issues. Figure 3a shows a general upward trajectory in the average user rating (AUR) over the years, especially in the past decade. This suggests that newer mental health apps tend to receive better ratings, possibly due to improved app quality, better user experiences, or more refined review mechanisms. Considering that 75% of apps receive less than ten reviews, we further analyze the trends in user rating across different levels of review volume to observe the impact of review count on ratings. Figure 3b reveals that apps with more user ratings (> 10 reviews) tend to have higher and more stable average rating over time, whereas those with fewer user ratings (1 ~ 10 reviews) generally exhibit lower but gradually increasing ratings.

RQ2, the quantity and quality comparison between pre- and post-pandemic mental health apps, is addressed by Figure 4 and 5. Figure 4 shows that mental health-related apps launched after the pandemic received higher average user ratings. Figure 5 illustrates pre- and post-comparison of the total count of apps releases in the top five major categories, with the most significant difference shown in the Health & Fitness and Lifestyle categories. Our analysis reveals a notable increase in both the quantity and user satisfaction level of mental health apps in the post-pandemic time, compared with pre-pandemic.

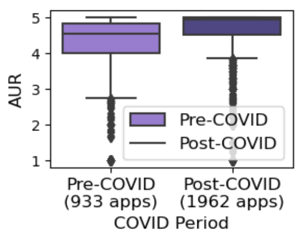

Figure 4: Average User Rating Pre- and Post-COVID (excluding Apps with Zero Ratings)

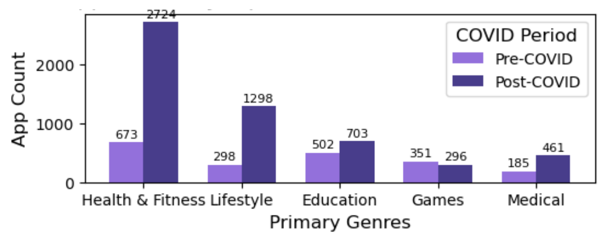

Figure 5: App count Pre- and Post- COVID-19 for the Top 5 Primary Genres

- Category 2: research questions related to app privacy policies (RQ3 ~ RQ9)

RQ3 examines the prevalence of privacy policies among mental health mobile apps. Our analysis reveals that among 4,764 English-based apps, 3,371 (70.8%) provide accessible privacy policies, while 1,393 (29.2%) do not. This finding indicates that although most apps have privacy policies, a significant proportion still lack basic safeguards for user privacy.

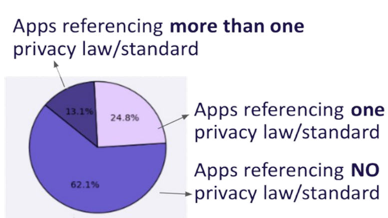

Figure 6: Distribution of Apps Based on Referencing Frequency of Privacy Laws or Standards

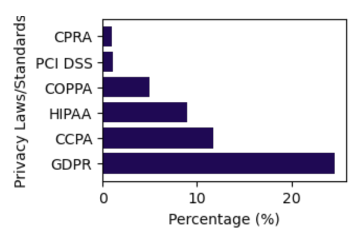

Figure 7: Most Frequently Referenced Privacy Regulations

Figure 6 shows that approximately 37.9% of the apps mentioned at least one privacy regulation, such as GDPR, HIPAA, COPPA [15], in their privacy policy, suggesting some level of compliance awareness.

RQ4 focuses the alignment of mental health apps with key privacy laws and standards. As depicted in Figure 7, the most frequently cited privacy regulations include GDPR, CCPA (California Consumer Privacy Act), HIPAA, and COPPA, reflecting a higher degree of adherence to these frameworks compared to other standards.

For RQ6 ~ RQ9, the following results are obtained: 91% of apps declare the collection of user data in their privacy policies. Among these, 96.7% specify the purpose of data collection and use, 76.9% disclose sharing data with third parties, and 81.3% provide users with options to opt out of data collection or to delete their data.

The word clouds in Figure 8 and 9 illustrate the most frequently mentioned phrases related to the types of data collected by apps and the third parties with which the apps share user data.

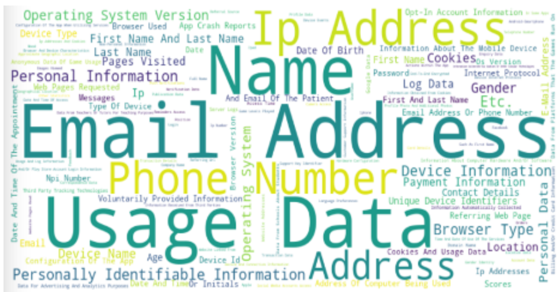

Figure 8: Word Cloud of User Data Types Collected by Apps.

Figure 9: Word Cloud of Third Parties with Which Apps Share User Data

Study Limitations and Future Work

An important objective of our study is to leverage LLM and RAG techniques to automatically extract answers to key privacy and security-related questions from privacy policy documents. While our program performed strongly on the four APA critical privacy questions, we did not apply this method to extract more open-ended questions due to challenges in evaluation methods and time constraints. In the future, we aim to expand the privacy question list to include more open-ended inquiries and explore effective methods to evaluate RAG performance on these complex queries [16].

Additionally, while our program performed well with the chosen sample size of 100 apps, we observed an increase in hallucination when applying the method to a significantly larger sample size of approximately 1,000 apps. To ensure high accuracy in addressing research questions RQ6 to RQ9, we prioritized a smaller, high-performing sample over a larger, less accurate one. In the future, we intend to explore advanced RAG techniques to mitigate hallucination and evaluate these methods on larger datasets. This approach will help balance scalability with the accuracy needed for robust insights into privacy policy analysis.

Lastly, our data collection currently focuses on apps from Apple's App Store. In the future, we plan to expand data collection to include apps from the Google play store, ensuring broader coverage and diversity in our analysis.

Conclusions

This study analyzes trends in mental health app development and user ratings from 2009 to 2024 while also conducting an in-depth evaluation of app privacy policies using LLM and RAG-based question answering. Our findings reveal a significant increase in both the quantity and quality of mental health apps, particularly in the post-pandemic period.

We also find that 29.2% of apps lack accessible privacy policies, raising concerns about transparency and compliance. Among apps with privacy policies, only 37.9% reference key privacy laws such as GDPR, HIPAA, or COPPA, indicating gaps in regulatory alignment. Further, while 91% of apps declare the collection of user data, only 76.9% disclose sharing this data with third parties, and 81.3% provide users with options to opt out or delete their data.

The LLM- and RAG-based analysis demonstrates high accuracy and reliability in answering privacy-related Yes/No APA questions, outperforming methods solely based on prompt engineering by grounding responses in relevant policy text. This approach offers a promising solution for automated privacy policy analysis, helping users, developers, and policymakers assess data practices more effectively.

Beyond its technical contributions, this study was conducted as part of an undergraduate research project, where student researchers actively participated in research problem identification, data analysis, and performance evaluation. The project provided hands-on experience in privacy policy analysis using LLMs and RAG techniques, while also contributing novel insights into mental health app privacy practices.

These findings underscore the growing need for transparency, regulatory compliance, and user awareness in digital mental health solutions. As privacy concerns continue to evolve, ensuring clearer, more accessible privacy disclosures will be critical in fostering trust and accountability in the mHealth ecosystem.

Appendix

A: Sample output file for the generated answer


```
{
  "meta": {
    "id": 1054458809,
    "url": "http://www.getmoodfit.com/privacy-policy",
    "title": ""
  },
  "reply": {
    "question": "1. Does the app declare the collection of data? What type of data does it collect?",
    "answer": {
      "full_answer": "The app does declare the collection of data. It collects User Provided Information such as email addresses, passwords, transaction-related information, and information entered into the system when using the Application. Additionally, it collects Automatically Collected Information including the type of mobile device, unique device ID, IP address, mobile operating system, mobile Internet browsers, and usage information.",
      "simple_answer": "Yes",
      "extented_simple_answer": {
        "comment": "data collected",
        "content": [
          "Email addresses",
          "Passwords",
          "Transaction-related information",
          "Information entered into the system",
          "Type of mobile device",
```

```
                    "Unique device ID",
                    "IP address",
                    "Mobile operating system",
                    "Mobile Internet browsers",
                    "Usage information"
                  ]
                }
              },
              "analysis": "The privacy policy clearly outlines the types of data that the Moodfit app collects under User Provided Information and Automatically Collected Information categories. This information includes a wide range of personal and device-related data, highlighting the extent of data collection by the application.",
              "reference": "The Application obtains the information you provide when you download and register the Application. When you register with us and use the Application, you generally provide (a) your email address and a password you create; (b) transaction-related information, such as when you download and use the Application and its features; (c) information you provide us if you contact us for help; (d) information you enter into our system when using the Application, such as entries you save in the Application. We may also use the information you provided us to contact you from time to time to provide you with important information, required notices, and marketing promotions. We keep your user provided data in encrypted form on servers located in the United States. URL: http://www.getmoodfit.com/privacy-policy."
            }
          },
```

B: Prompt template

```
prompt = """
You are a privacy policy expert. You are provided with {{app_url}}, which contains the privacy policy document for an app.
Your task is to:
   - answer the question based on the provided privacy policy document ONLY, instead of your previous knowledge
   - provide references for your answers based on the section in the privacy policy document from which your answer is generated.
   - Produce your results by strictly following JSON format below. Do not add any extra information beyond JSON.
 - Note that The 'url' in 'meta' section MUST be exactly the {{app_url}}).

{
  "meta": {
    "id": {{ app_id }},
    "url": {{ app_url }},
    "title": {{ app_name | tojson }}
  },
  "reply": {
    "question": "{{ question | escape }}",
    "answer": {
      "full_answer": "{{ full_answer | escape }}",
      "simple_answer": "{{ simple_answer | escape }}",
      "extented_simple_answer": {
        "comment": "{{ extended_comment | escape }}",
        "content": "{{ extended_content | escape }}"
      }
    },
    "analysis": "{{ analysis | escape }}",
    "reference": "{{ reference | escape }}"
  }
}
```

Instructions:
1. Approach the question systematically by following these steps:
a. **Understand the Question**: Break down the question into specific components or sub-questions if needed.
b. **Identify Relevant Context**: Highlight specific parts of the privacy policy document that directly relate to the question. You will need to output this in the reference part of the JSON output.
c. **Analyze the Context**: Explain how the identified Relevant Context addresses the question.
d. **Formulate the Answer**: Construct your answer logically, integrating the relevant context, your analysis, and the specific requirements for each JSON field.
e. **Generate Reference**:
- The Reference section should be composed of both the {{app_url}} and the exact original text from which the answer is obtained. Note that the original text must come from only the relevant context in page {{app_url}}
- The Reference section format should be like this: (original text from the url) + 'URL: ' + {{app_url}}
- The reference must not default only to the URL. If reference cannot be found, report: 'N/A. URL: ' + {{app_url}}

2. Output your reasoning and answer using the following structure:
a. **Analysis Section**: Provide a detailed explanation of your reasoning process and how you arrived at the answer.
b. **Full Answer Section**: Deliver a complete and detailed response to the question. Full answer must include at least information from the "Simple Answer Section" and the "Extended Simple Answer Section". The full answer section must not be empty.
c. **Simple Answer Section**: Generate this section strictly based on the requirements in the "The simple answer section should ..." part in the question. This section must not be empty.
d. **Extended Simple Answer Section**: Generate this section strictly based on the requirements in the "The extended simple answer section should .." part in the question. If the "The extended simple answer section should .." part in the question is missing, leave the "extended_simple_answer" field empty.
e. **Reference Section**:
- Include the original text from the privacy policy document that supports your answer.
- The context must come from the app URL {{app_url}}.
- Attach the {{app_url}} in the end.
- Ensure JSON compatibility by replacing double quotes with single quotes.

3. For output, adhere strictly to the JSON format provided above. Do not include any additional or irrelevant information outside this JSON format.

Ensure all fields adhere to the following requirements:
1. All responses must logically follow the chain of thought process described above.
2. Each section of the JSON must contain clear and valid information as outlined in the format.

Context:
{% for doc in documents %}
{{ doc.content }}
URL: {{ doc.meta['url'] }}
{% endfor %}

Question: {{ query }}
Answer:
"""

References

[1] J. Torous, H. Wisniewski, G. Liu, and M. Keshavan, "Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: Comparative survey study," *JMIR Ment. Health*, vol. 5, no. 4, pp. e11715, Dec. 2018.

[2] A. Papageorgiou, M. Strigkos, E. Politou, E. Alepis, A. Solanas and C. Patsakis, "Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice," in *Proc. IEEE Access*, vol. 6, pp. 9390-9403, 2018

[3] App Store. “App Privacy Details on the App Store.” App Store Developer. Accessed Jan. 2025. [Online]. Available: https://developer.apple.com/app-store/app-privacy-details/.

[4] Google. “Provide Information for Google Play's Data Safety Section.” Play Console Help. Accessed Jan. 2025. [Online]. Available: https://support.google.com/googleplay/android-developer/answer/10787469?hl=en.

[5] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies," *ISJLP*, vol. 4, pp. 543–565, 2008.

[6] American Psychiatric Association. “Mental Health Apps Evaluations.” American Psychiatric Association. Accessed Jan. 2025. [Online]. Available: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model.

[7] D. Rodriguez, I. Yang, N. Sadeh, and J. M. Del Alamo, "Large language models: A new approach for privacy policy analysis at scale," *arXiv*, 31 May 2024. Accessed Jan. 2025. [Online]. Available: https://arxiv.org/pdf/2405.20900

[8] C. Tang, Z. Liu, C. Ma, Z. Wu, Y. Li, W. Liu, D. Zhu, Q. Li, X. Li, T. Liu, and L. Fan, "PolicyGPT: Automated analysis of privacy policies with large language models," *arXiv*, 20 Sep. 2023. Accessed Jan. 2025. [Online]. Available: https://arxiv.org/pdf/2309.

[9] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," *arXiv*, 22 May 2020. Accessed Jan. 2025. [Online]. Available: https://arxiv.org/abs/2005.11401.

[10] LlamaIndex. “Introduction to RAG,” LlamaIndex. Accessed Jan. 2025. [Online]. Available: https://docs.llamaindex.ai/en/stable/understanding/rag/

[11] G. Facundo. “App-Store-Scraper (Version 0.12.0).” GitHub. Accessed Jan. 2025. [Online]. Available: https://github.com/facundoolano/app-store-scraper.

[12] Grand View Research. “Mental Health Apps Market Size, Share & Trends Analysis Report by Platform (Android, iOS), by Application (Meditation Management, Stress Management), by Region, and Segment Forecasts, 2024-2030 (Report No. GVR-4-68039-911-7).” Grand View Research. Accessed Jan. 2025. [Online].
Available: https://www.grandviewresearch.com/industry-analysis/mental-health-apps-market-report.

[13] Deepset. “The Production-Ready Open Source AI Framework.” Haystack by Deepset. Accessed Jan. 2025. [Online]. Available: https://haystack.deepset.ai/.

[14] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.” arXiv, 28 Jan 2022. Accessed Jan. 2025. [Online]. Available: https://arxiv.org/abs/2201.11903.

[15] L. H. Iwaya, M. A. Babar, A. Rashid, and C. Wijayarathna, "On the privacy of mental health apps: An empirical investigation and its implications for app development," *Empir. Softw. Eng.*, vol. 28, no. 1, p. 2, 2023.

[16] D. Rodriguez, I. Yang, N. Sadeh, and J. M. Del Alamo, "Large language models: A new approach for privacy policy analysis at scale," *arXiv*, May 2024. Accessed Jan. 2025. [Online]. Available: https://arxiv.org/abs/2405.07437.