Role of Data Privacy Towards Safe and Trustworthy Mental Health Assistants: Project Report

Surjodeep Sarkar[1]† , **Ekta Pandey**[1]† , **Rachit Saini**[1]† , **Bhavani Shankar Mahamkali**[1]†

Abstract

Mental health assistants(MHAs) have become increasingly popular as a means of providing support and treatment to individuals with mental health issues. However, the use of these apps raises significant concerns about data privacy. As users input sensitive personal information into these apps, there is a risk that this data may be misused, shared with third parties without consent, or breached by malicious actors. Additionally, the lack of regulation in the mental health app industry further exacerbates these concerns, as there are no clear guidelines or standards for how user data should be collected, stored, and protected. This has led to a growing need for greater transparency and accountability from app developers, as well as for stronger regulatory frameworks to protect user privacy. Ultimately, the protection of user data in mental health apps is crucial for ensuring that individuals can access the care they need without compromising their privacy .

This project aims to address this gap in the literature by examining the data privacy practices of a sample of mental health apps**(e.g., Woebot, HeadSpace, BetterHelp, TalkSpace, and Calm)**. Specifically, we will analyze the extent to which these apps comply with data privacy regulations, the types of user data they collect and how it is used, and the measures they have in place to protect user privacy. By shedding light on these issues, this study aims to inform the development of better privacy practices for mental health app**(e.g., TalkSpace)** following privacy by design. We shed information on the methodology behind our approach, results from our development and discuss on some of the pressing question related to this project. Finally, we provide conclusion and future scope of work beyond the current literature toward enriching the privacy concern of **TalkSpace** using privacy enhancing techniques(PETs) to make it safe and wholesomely trustworthy. For reproducibility and for reference, the code can be found in Github

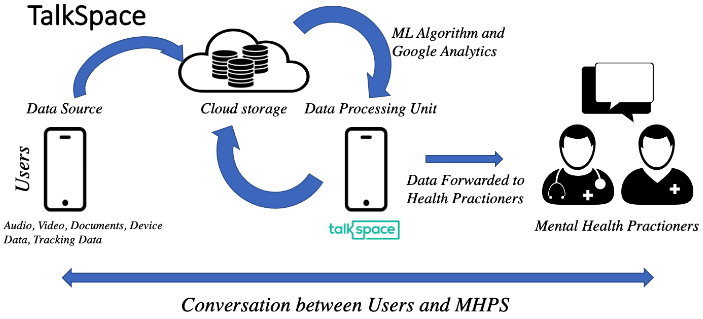

TalkSpace
ML Algorithm and Google Analytics
Data Source
Cloud storage
Data Processing Unit
Users
Audio, Video, Documents, Device Data, Tracking Data
talkspace
Data Forwarded to Health Practioners
Mental Health Practioners
Conversation between Users and MHPS

Figure 1: Illustration of the TalkSpace app. While connecting the dots how TalkSpace works, we have created a high level model architecture.

1 Introduction

Mental illness is a worldwide concern, constituting a significant cause of distress in people's lives and impacting society's health and well-being[Zhang et. al.(2022)]. According to the National Survey on Drug Use and Health, nearly one in five U.S. adults live with a mental illness (52.9 million in 2020) [1]. Reports released in August 2021[2] indicate that *1.6 million people* in England were on waiting lists to seek professional help with mental health care. Such an overwhelming rise in the number of patients as compared to MHPs necessitated the use of (i) public health forums (e.g., dialogue4health), (ii) online communities (e.g., r/depression subreddit on Reddit), (iii) Talklife, and (iv) Mental Health Assistants (MHAs), for informative healthcare. The anonymous functioning of (i), (ii), (iii) removed the psychological stigma in patients, which even refrained them from seeing an MHP [Hyman et al.(2008)].

In addition, the unavailability of interpersonal interactions from other pure information agents resulted in the need to develop Mental Health Assistants (MHAs).

MHAs: MHAs are artificial intelligence (AI)-based agents designed to provide emotional support through structured

[1] http://www.samhsa.gov/data/release/2020-national-survey-drug-use-and-health-nsduh-releases

[2] https://www.theguardian.com/society/2021/aug/29/strain-on-mental-health-care-leaves-8m-people-without-help-say-nhs-leaders

conversational sequences targeted to screen patients for mental health conditions and alert mental health professionals (MHPs) through *informed triaging*[3]. Despite the proliferation of research at the intersection of clinical psychology, artificial intelligence (AI), and Natural Language Understanding (NLU), MHAs missed an opportunity to serve as life-saving contextualized, personalized, and reliable decision support during COVID-19 under the *apollo* moment [Srivastava et al.(2021); Czeisler et al.(2020)]. MHAs' ability to function as simple information agents (e.g., suggest meditation, relaxation exercises, or give positive affirmations) *did not* bridge the gap between *monitoring the health condition* and *necessitating an MHP visit* for the patient.

Hence, MHAs are increasingly popularized and endorsed by MHPs as a cheap and accessible alternative to in-patient visits.However, with the increased popularity, app users are also suffering the loss of personal privacy due to data sharing practices to third parties [Parker and Halter(2019)]. Health apps usually share data with multiple third parties for internal research and app functionality but also for commercial use [Tangari and Gioacchino(2021)].

To holistically contribute towards *safe* and *trustworthy* behavior in mental health assistants, there is a need to critically examine *privacy policy*, *policy settings*, the use of clinical knowledge for *anonymization*, along with privacy enhancing techniques. **Our Contributions**: This report spans 4 major research dimensions: (i) What are privacy policies in MHAs? (ii) What are the current functionalities and limitations of MHAs?, (iii) What functionalities can be imagined in MHAs for which patient's privacy is intact? and (iv) What changes in functionalities is required with respect to privacy by design?

2 Related Work

In this section, we introduce the related research work on privacy by design techniques on MHAs.

Over last few years there has been a tremendous growth in mobile computing and the introduction of mental health apps for general users. However, with this outgrowth, privacy has also been a major concern with the users. The European Commission's analysis on data protection for citizens across all 28 EU Member States [4] reveals that over 50% percentage of participants from the surveyed countries expressed concern about their everyday activities being tracked via mobile phones or apps. This feedback underscores the widespread anxiety over unclear and insufficient protection of personal data in the age of pervasive computing. In response to these concerns, the European Commission put into effect a new, more rigorous legal structure in 2016, known as the General Data Protection Regulation (GDPR) [Magdziarczyk(2019)]. The GDPR, which replaced the outdated 1995 Data Protection Directive [Burkert(1996)], became directly enforceable in all EU Member States in May 2018. This move helped standardize the disparate national regulations across the EU. Despite the extended debates and rigorous discussions about its relevance in the era of big data and the Internet of Things (IoT) [De Hert and Papakonstantinou(2016)], the regulation paves the way for significant transformations in the data protection landscape. Notably, this includes the implementation of pseudonymisation and the right to data portability. However, apprehensions regarding mental health applications are not solely confined to the European Union. For instance, in the United States, many question the feasibility of applying national standards like the Health Insurance Portability and Accountability Act (HIPAA) [5], which establishes rules for safeguarding the privacy and security of personally identifiable health data. Notably, several experts contend that health apps don't always align with the regulatory protections provided by the HIPAA [He and Naveed(2014)].

In the light of this, there has been post by washingtonpost [6] which shares similar concerns about the regulatory protections provided by the HIPAA. This has led to companies violating privacy practices involved with health apps and their data collection practices. A review has been done by https://privacynotincluded.com to highlight the inconsistencies by several health apps. On the similar lines, [Papageorgiou and Strigkos(2018)] have highlighted in their report about the dangers of data mining practices and collecting user's health data with other companies.

The aforementioned studies shed light on significant issues stemming from how each mental health app collects, handles, and shares user's private data. For instance, trust issues emerge when an app gathers more data than necessary to deliver its services, thereby breaching the principles of data minimization and purpose limitation set out in contemporary data protection regulations.

The research we introduce in this article extends the previously referenced studies, focusing on the privacy evaluation of MHAs available in online marketplaces. We explore the privacy risks of the 4-5 most popular MHAs from https://privacynotincluded.com, with a particular focus on privacy and personal data protection when transmitting sensitive health information to third-party entities. By scrutinizing how the apps request, manage, and distribute sensitive personal information, we can ultimately identify the necessary steps for its protection.

3 Methodology

In this portion of the paper, we first introduce the MHA (for example, TalkSpace), along with its data collection methods and the features provided by TalkSpace. Next, we demonstrate how the data minimization technique has been implemented to ensure that only necessary data is collected for the anticipated features to function properly. Finally, we delve into the individual features that form part of this project, and the augmented functionalities that we aim to implement.

3.1 TalkSpace & Data Collection Methodology

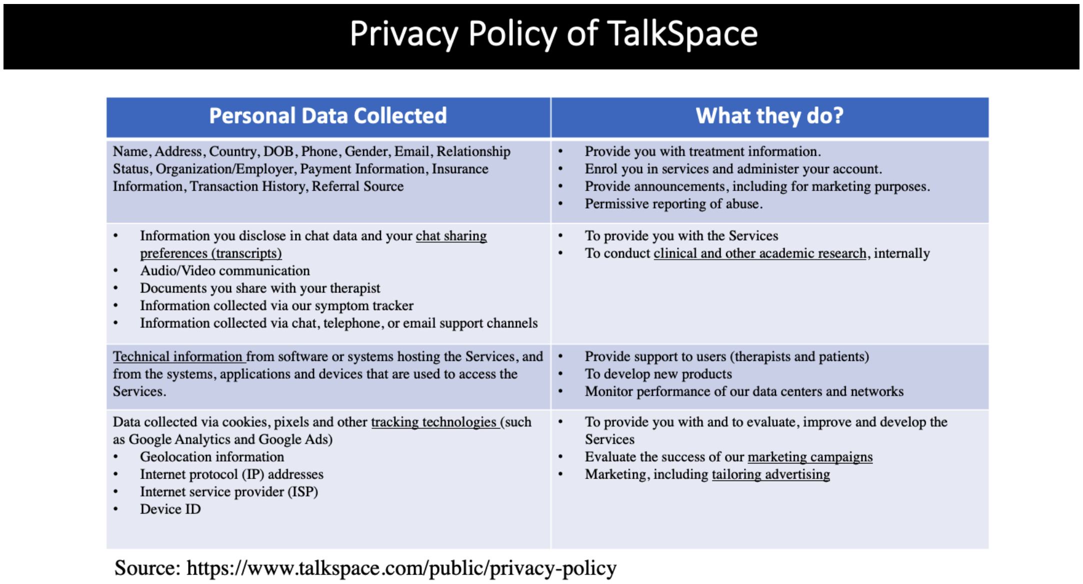
Our goal was to conduct a preliminary review of potential mobile health applications (MHAs) from the website https://privacynotincluded.com, which led us to identify **TalkSpace**

[3] https://code4health.org/chat-bot/

[4] https://data.europa.eu/data/datasets/?locale=en

[5] https://www.hhs.gov/hipaa/index.html

[6] https://www.washingtonpost.com/technology/2022/09/22/health-apps-privacy/



Privacy Policy of TalkSpace

Personal Data Collected	What they do?
Name, Address, Country, DOB, Phone, Gender, Email, Relationship Status, Organization/Employer, Payment Information, Insurance Information, Transaction History, Referral Source	• Provide you with treatment information. • Enrol you in services and administer your account. • Provide announcements, including for marketing purposes. • Permissive reporting of abuse.
• Information you disclose in chat data and your chat sharing preferences (transcripts) • Audio/Video communication • Documents you share with your therapist • Information collected via our symptom tracker • Information collected via chat, telephone, or email support channels	• To provide you with the Services • To conduct clinical and other academic research, internally
Technical information from software or systems hosting the Services, and from the systems, applications and devices that are used to access the Services.	• Provide support to users (therapists and patients) • To develop new products • Monitor performance of our data centers and networks
Data collected via cookies, pixels and other tracking technologies (such as Google Analytics and Google Ads) • Geolocation information • Internet protocol (IP) addresses • Internet service provider (ISP) • Device ID	• To provide you with and to evaluate, improve and develop the Services • Evaluate the success of our marketing campaigns • Marketing, including tailoring advertising

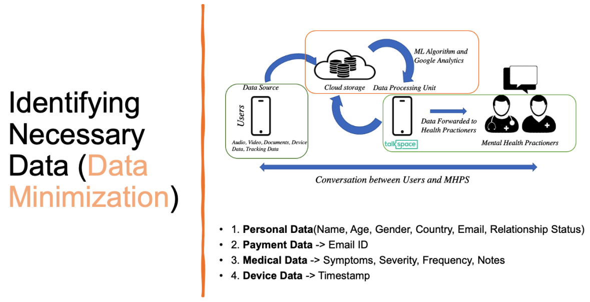
Source: https://www.talkspace.com/public/privacy-policy

Figure 2: Illustration of the TalkSpace app. While connecting the dots how TalkSpace works, we have created a high level model architecture.

as the ideal candidate for our project. As per a recent blog post by Mozilla [7], TalkSpace stood out as the app most frequently implicated in violating users' privacy and for our further development of having better privacy settings. Furthermore, we managed to decipher the internal mechanics of **TalkSpace**, which are depicted in the Figure 1.

- **TalkSpace:** A teletherapy app that connects users with licensed therapists for online therapy sessions. Talkspace claims to encrypts user data and complies with relevant privacy laws and regulations. However, this app have been a hot topic in newspapers and articles that data may be shared with third-party service providers for marketing campaigns.
- **Features:** We have identified the functionalities of TalkSpace for which it collects huge amount of personal data which can be referred from Figure 2 and are as follows:
  1. Psychoeducation
  2. Mindfulness
  3. Deep Breathing
  4. Coach/Therapist Connection
  5. Track Symptoms
  6. Track Medication
  7. Goal Settings
- **Data Collection:** During our survey we have found out that **TalkSpace** do extensively collect a huge amount personal and sensitive data about a user in order to pair up with an appropriate therapist. Please refer to Figure 2 which provides further information.

During the survey we came across few major privacy related question based on which we have examined **TalkSpace**:



Identifying Necessary Data (Data Minimization)

Data Source
Users
Audio, Video, Documents, Device Data, Tracking Data
Cloud storage
Data Processing Unit
ML Algorithm and Google Analytics
talkspace
Data Forwarded to Health Practioners
Mental Health Practioners
Conversation between Users and MHPS

- 1. **Personal Data**(Name, Age, Gender, Country, Email, Relationship Status)
- 2. **Payment Data** -> Email ID
- 3. **Medical Data** -> Symptoms, Severity, Frequency, Notes
- 4. **Device Data** -> Timestamp

Figure 3: Illustration of data minimization in TalkSpace app.

- What kind of data are being collected?
- How long is the data being stored?
- Whether the data can be deleted?
- What is the purpose of the data being collected?
- Third Party Access to Data?
- Informed consent?

3.2 Data Minimization

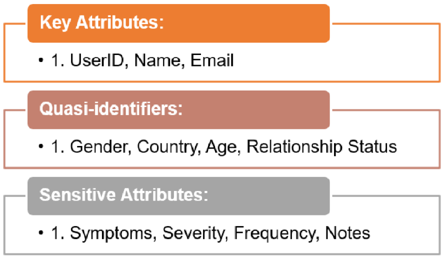
We have implemented data minimization which is a privacy principle and practice that encourages companies and organizations to collect, process, and retain only the minimum amount of personal data necessary to achieve their legitimate purpose. The process has been exemplified in the Figure 3

3.3 Features and Functionalities

Psychoeducation

Psychoeducation (also known as psychological education) is an evidence-based therapeutic intervention for patients and their loved ones that provides information and support

[7] https://www.techrepublic.com/article/mozilla-privacy-survey-finds-mental-health-and-prayer-apps-fail-privacy-test-pretty-spectacularly/


Key Attributes:

- 1. UserID, Name, Email

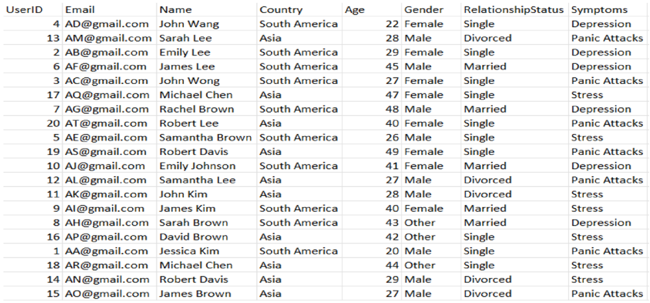
Quasi-identifiers:

- 1. Gender, Country, Age, Relationship Status

Sensitive Attributes:

- 1. Symptoms, Severity, Frequency, Notes

Figure 4: Classification of attributes



UserID	Email	Name	Country	Age	Gender	RelationshipStatus	Symptoms
4	AD@gmail.com	John Wang	South America	22	Female	Single	Depression
13	AM@gmail.com	Sarah Lee	Asia	28	Male	Divorced	Panic Attacks
2	AB@gmail.com	Emily Lee	South America	29	Female	Single	Depression
6	AF@gmail.com	James Lee	South America	45	Male	Married	Depression
3	AC@gmail.com	John Wong	South America	27	Female	Single	Panic Attacks
17	AQ@gmail.com	Michael Chen	Asia	47	Female	Single	Stress
7	AG@gmail.com	Rachel Brown	South America	48	Male	Married	Depression
20	AT@gmail.com	Robert Lee	Asia	40	Female	Single	Panic Attacks
5	AE@gmail.com	Samantha Brown	South America	26	Male	Single	Stress
19	AS@gmail.com	Robert Davis	Asia	49	Female	Single	Panic Attacks
10	AJ@gmail.com	Emily Johnson	South America	41	Female	Married	Depression
12	AL@gmail.com	Samantha Lee	Asia	27	Male	Divorced	Panic Attacks
11	AK@gmail.com	John Kim	Asia	28	Male	Divorced	Stress
9	AI@gmail.com	James Kim	South America	40	Female	Married	Stress
8	AH@gmail.com	Sarah Brown	South America	43	Other	Married	Depression
16	AP@gmail.com	David Brown	Asia	42	Other	Single	Stress
1	AA@gmail.com	Jessica Kim	South America	20	Male	Single	Panic Attacks
18	AR@gmail.com	Michael Chen	Asia	44	Other	Single	Stress
14	AN@gmail.com	Robert Davis	Asia	29	Male	Divorced	Stress
15	AO@gmail.com	James Brown	Asia	27	Male	Divorced	Panic Attacks

Figure 5: Original data for psychoeducation.

to better understand and cope with illness. Psychoeducation is most often associated with serious mental illness, including dementia, schizophrenia, clinical depression, anxiety disorders, psychotic illnesses, eating disorders, personality disorders, and autism, although the term has also been used for programs that address physical illnesses, such as cancer[PCE(2023)].

Psychoeducation offered to patients and family members teaches problem-solving and communication skills and provides education and resources in an empathetic and supportive environment. Results from more than 30 studies indicate psychoeducation improves family well-being, lowers rates of relapse and improves recovery[PCE(2023)].

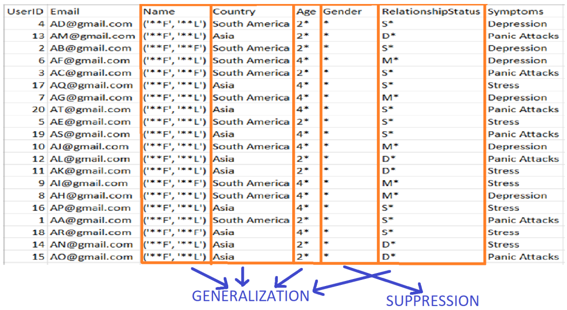
Some of the common benefits while using this feature:

- Reduces relapse rate and/or improves rehabilitation rate
- Significantly decreases burden and distress of families
- reduces medication nonadherence and encourages positive attitude toward medication
- reduces symptoms of mania
- improves quality of life
- improves social and global functioning
- improves problem-solving abilities
- improves self-management behaviors and self-care[THC(2023)]

Approach

Firstly, we aim to classify the attributes which is exemplified in figure 4 based on which we aim to perform data masking and anonymization on the original dataset for psychoeducation feature as well as for our enhanced feature 1 which can be referred from figure 5. As a part of anonimization, we perform the below techniques:

- **Data Generalization :** It allows users to replace a data value with a less precise one[DG(2023)].
- **Data Suppression :** It refers to the process of withholding or removing selected information — most commonly in public reports and data sets to protect the identities, privacy, and personal information of individuals. Data suppression is used whenever there is chance that the information contained in a publicly available report could be used to reveal or infer the identities of specific individuals[DS(2023)].



UserID	Email	Name	Country	Age	Gender	RelationshipStatus	Symptoms
4	AD@gmail.com	('**F', '**L')	South America	2*	*	S*	Depression
13	AM@gmail.com	('**F', '**L')	Asia	2*	*	D*	Panic Attacks
2	AB@gmail.com	('**F', '**F')	South America	2*	*	S*	Depression
6	AF@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
3	AC@gmail.com	('**F', '**F')	South America	2*	*	S*	Panic Attacks
17	AQ@gmail.com	('**F', '**L')	Asia	4*	*	S*	Stress
7	AG@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
20	AT@gmail.com	('**F', '**L')	Asia	4*	*	S*	Panic Attacks
5	AE@gmail.com	('**F', '**L')	South America	2*	*	S*	Stress
19	AS@gmail.com	('**F', '**L')	Asia	4*	*	S*	Panic Attacks
10	AJ@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
12	AL@gmail.com	('**F', '**L')	Asia	2*	*	D*	Panic Attacks
11	AK@gmail.com	('**F', '**L')	Asia	2*	*	D*	Stress
9	AI@gmail.com	('**F', '**F')	South America	4*	*	M*	Stress
8	AH@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
16	AP@gmail.com	('**F', '**L')	Asia	4*	*	S*	Stress
1	AA@gmail.com	('**F', '**L')	South America	2*	*	S*	Panic Attacks
18	AR@gmail.com	('**F', '**F')	Asia	4*	*	S*	Stress
14	AN@gmail.com	('**F', '**L')	Asia	2*	*	D*	Stress
15	AO@gmail.com	('**F', '**L')	Asia	2*	*	D*	Panic Attacks

Figure 6: Data Generalization and Suppression.

- **K-Anonymization :** K anonymity is a data anonymization technique that is used to protect individuals' privacy in a dataset. It involves data generalization, data masking, or replacing Personally Identifiable Information (PII) with a pseudonym to ensure no single individual can be identified[KA(2023)].

The table in figure 6 is created with the following fields : **UserID**, **Email**, **Name**, **Country**, **Age**, **Gender**, **RelationshipStatus**, **Symptoms**.

The IDs are assigned in range 1 to 20 in random order. The emails assigned to 20 unique IDs are also unique and are assigned in random order. The name column(consisting of first and last names) is masked in the table which was allocated in a unique manner. The countries are chosen from two continents - Asia and South America. The age ranges are chosen to be in 20s and 40s in a shuffled order. Two genders(male and female) are allocated in a shuffled order. Three relationship parameters - Single, Married and Divorced are arranged in a shuffled manner. The list of symptoms from Depression, Panic Attacks and Stress arranged in random order.

The **data masking** and **K-Anonymity** privacy preserving techniques are applied on the figure 6. The data masking is done on the **name column** and the **gender column**. As shown in figure 6, the **name column, ID column and email column**, are the **key** attributes. The **Symptoms column** is the **sensitive** attribute. The **country column, age column, gender column and relationship status column** are all **quasi-identifiers**. While sharing the data with the therapist from the user on the other end by the Talkspace app, the key attributes are suppressed and the quasi-identifiers with privacy technique applied along with sensitive attributes as it is are sent to the therapist. The therapist based on the symptoms provides the user with the relevant material to help them in coping with the mental illness they are suffering from.

3.4 Symptom Tracker

The symptom tracking feature is designed to help users monitor and manage their mental health symptoms effectively. With this feature, users can log and track their symptoms over time, gaining valuable insights into their well-being. By consistently tracking symptoms, users can identify patterns, triggers, or changes in their mental health. This valuable information can be shared with their therapist, facilitating more targeted discussions during therapy sessions and enhancing the treatment process.

Additionally, the symptom tracker empowers users by involving them in their own care, enabling them to set goals, track progress, and make informed decisions about their mental well-being. Ultimately, the symptom tracker feature promotes self-care, communication with professionals, and a deeper understanding of one's mental health.

Approach

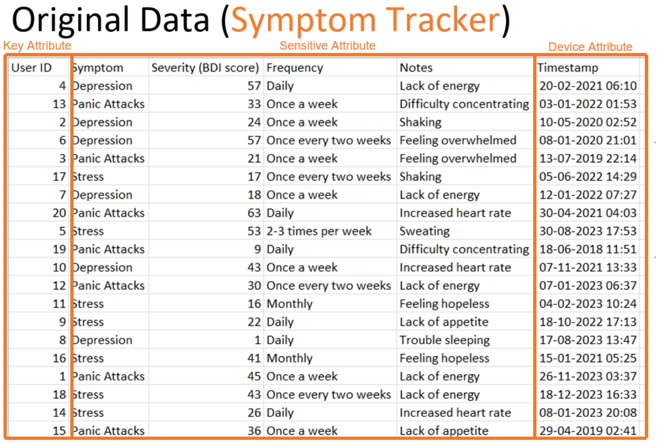
The symptom tracker feature typically collects several data points from users to support its functionality. This includes the type of symptoms experienced by the individual, the severity level, the frequency of occurrence, any accompanying notes or comments, and the timestamp of each entry. Instead of users rating the severity on a numerical scale, a scoring system like BDI is utilized to map the severity level.

The Beck Depression Inventory (BDI) is a 21-item, self-report rating inventory that measures characteristic attitudes and symptoms of depression [Beck et al.(1961)Beck, Ward, Mendelson, Mock, and Erbaugh].There will be a total of 21 questions and score ranges from [0-3] for each question. So, the BDI Score is in the range [0-63]. The BDI score is mapped to different leves of severity as below [8].

Table 1: Levels of severity based on BDI Score

Total Score	Levels of Severity
1-10	These ups and downs are considered normal
11-16	Mild mood disturbance
17-20	Borderline clinical depression
21-30	Moderate depression
31-40	Severe depression
over 40	Extreme depression

The selection of the Beck Depression Inventory (BDI) as the scoring system for severity data in symptom tracking ensures safety and reliability. The BDI follows a clinically grounded procedural guideline, providing a structured approach for individuals to accurately assess their mental state. By utilizing the BDI, patients can overcome challenges in accessing and expressing their mental well-being, enabling a comprehensive evaluation of depression symptoms over time. This choice aligns with established clinical practices, enhancing the credibility and effectiveness of the symptom tracking process.


Original Data (Symptom Tracker)
Key Attribute
Sensitive Attribute
Device Attribute

User ID	Symptom	Severity (BDI score)	Frequency	Notes	Timestamp
4	Depression	57	Daily	Lack of energy	20-02-2021 06:10
13	Panic Attacks	33	Once a week	Difficulty concentrating	03-01-2022 01:53
2	Depression	24	Once a week	Shaking	10-05-2020 02:52
6	Depression	57	Once every two weeks	Feeling overwhelmed	08-01-2020 21:01
3	Panic Attacks	21	Once a week	Feeling overwhelmed	13-07-2019 22:14
17	Stress	17	Once every two weeks	Shaking	05-06-2022 14:29
7	Depression	18	Once a week	Lack of energy	12-01-2022 07:27
20	Panic Attacks	63	Daily	Increased heart rate	30-04-2021 04:03
5	Stress	53	2-3 times per week	Sweating	30-08-2023 17:53
19	Panic Attacks	9	Daily	Difficulty concentrating	18-06-2018 11:51
10	Depression	43	Once a week	Increased heart rate	07-11-2021 13:33
12	Panic Attacks	30	Once every two weeks	Lack of energy	07-01-2023 06:37
11	Stress	16	Monthly	Feeling hopeless	04-02-2023 10:24
9	Stress	22	Daily	Lack of appetite	18-10-2022 17:13
8	Depression	1	Daily	Trouble sleeping	17-08-2023 13:47
16	Stress	41	Monthly	Feeling hopeless	15-01-2021 05:25
1	Panic Attacks	45	Once a week	Lack of energy	26-11-2023 03:37
18	Stress	43	Once every two weeks	Lack of energy	18-12-2023 16:33
14	Stress	26	Daily	Increased heart rate	08-01-2023 20:08
15	Panic Attacks	36	Once a week	Lack of appetite	29-04-2019 02:41

Figure 7: Original data collected for symptom tracker.

3.5 Enhanced Feature 1 : (Psychoeducation + Symptom Tracker)

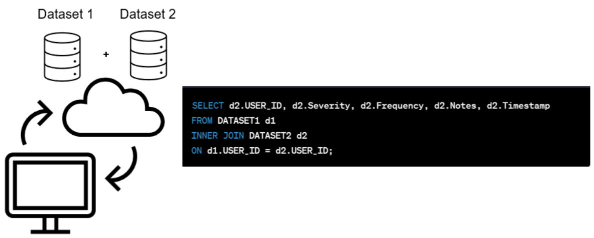
This feature provides not only a symptom tracker but also incorporates passive psychoeducation for users. Alongside monitoring their symptoms, users have access to a wealth of educational content related to mental health, which will be recommended by the therapist without even knowing user identifiable details. This passive psychoeducation component serves to enhance users' understanding of their symptoms, mental health conditions, and effective coping mechanisms. By integrating symptom tracking with educational resources, this feature empowers users to gain insight, deepen their knowledge, and actively participate in their therapeutic journey. It enables therapists to support users with valuable educational materials and facilitates collaborative discussions around symptom management and mental well-being.

Approach

The severity and timestamp columns in the Figure 7 are modified, allowing for the grouping of users with similar symptoms, as shown in Figure 15. This grouped information is provided to the therapist, who can then recommend relevant psychoeducation materials to these specific user groups. Additionally, the therapist can track changes in severity levels over time.

In order to extract data for the final table, an SQL query called ”inner join” will be performed as shown in Figure 8. This query combines data from both DATASET1 and DATASET2 using the USERID column as the common identifier. It selects the USERID, Severity, Frequency, Notes, and Timestamp columns from DATASET2, while ensuring that the corresponding USERID exists in both datasets. This allows for the grouping of users with similar symptoms and provides the therapist with the necessary information to recommend relevant psychoeducation materials and track changes in severity levels over time.

[8] https://www.ismanet.org/doctoryourspirit/pdfs/Beck-Depression-Inventory-BDI.pdf

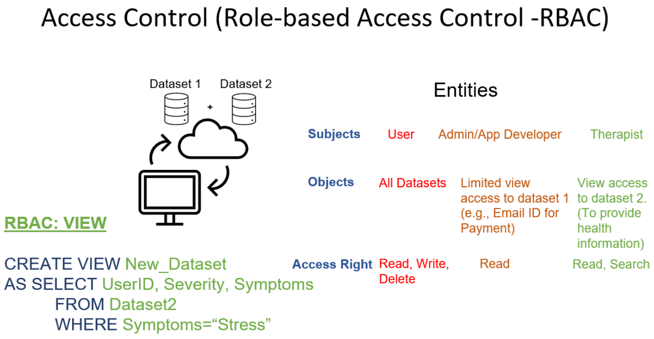


```sql
SELECT d2.USER_ID, d2.Severity, d2.Frequency, d2.Notes, d2.Timestamp
FROM DATASET1 d1
INNER JOIN DATASET2 d2
ON d1.USER_ID = d2.USER_ID;
```

Figure 8: SQL query to extract data for final table.

Incorporating Role-based Access Control (RBAC) can be instrumental in addressing privacy concerns. RBAC has been chosen for its ability to provide structured and efficient management of access rights, ensuring appropriate access permissions based on defined roles and reducing administrative overhead. It enhances security and privacy by restricting access to sensitive data and functionalities, minimizing the risk of unauthorized access and privacy breaches.

In Figure 9, the entities and their respective access rights are depicted for users, app developers, and therapists in a Role-based Access Control (RBAC) context. It includes a sample RBAC view specifically for therapists, enabling them to access all users who have reported symptoms related to "stress." This allows therapists to effectively monitor and provide appropriate support to individuals experiencing stress-related symptoms.


Access Control (Role-based Access Control -RBAC)

Entities

Subjects	User	Admin/App Developer	Therapist
Objects	All Datasets	Limited view access to dataset 1 (e.g., Email ID for Payment)	View access to dataset 2. (To provide health information)
Access Right	Read, Write, Delete	Read	Read, Search

RBAC: VIEW

```sql
CREATE VIEW New_Dataset
AS SELECT UserID, Severity, Symptoms
        FROM Dataset2
        WHERE Symptoms=“Stress”
```

Figure 9: Entites details and sample RBAC view for therapist.

3.6 Enhanced Feature 2: Crisis Management during Chat

Chat-based crisis management is a crucial feature of the application designed to address urgent situations involving patients. It comes into play when a patient experiences a high-stress emergency, such as an anxiety or depression attack, and the attending doctor is unavailable at that particular moment. The primary objective is twofold: firstly, to promptly connect the patient with a healthcare professional during emergencies, and secondly, to ensure the privacy of the patient while effectively communicating their condition to the doctor, thereby assisting in timely and appropriate care.

The importance of this feature lies in leveraging real-time communication tools within the application, patients can seek immediate assistance and support, even when their regular doctor is not accessible. Finally, it improves patient outcomes by allowing timely access to treatment, lowering wait times for assistance, and boosting communication between patients and healthcare personnel. It improves overall healthcare delivery quality, especially in emergency situations, and highlights the necessity of harnessing technology to meet vital healthcare demands.

Approach

We have tried to simulate a conversation between a chat and a patient using a general question *"How may I help you"* and then BDI questions. BDI refers to Beck Depression Inventory [Beck et al.(1961)Beck, Ward, Mendelson, Mock, and Erbaugh]. I have taken these questions as our baseline question as they are clinical asked questions. These are a set of self-report questionnaire which measure the severity of depression and similarly we have also taken other questions which corresponds to other symptoms. The BDI is used as an additional assessment tool to evaluate the severity of depression symptoms based on the user's responses. After determining the presence of anxiety and depression symptoms through specific questions.

We have used questions like : *"Do you often feel sad or down?"*, *"Do you feel guilty or worthless?"*, *"Have you lost interest in things you used to enjoy?"*. Following is one of the examples:

After processing the input ans calculating the total score we are training a model to asses the probability of mental health decline.We have included differential privacy to the BDI score ,to introduce the randomeness to protect the privacy.We are trying to provide an interpretation ranging from minimal depression to severe depression. If the severity is classified as severe depression which is the probability exceeds a certain threshold(in our case we have defines it as 70 percent) ,an alert is send to the doctor. Following are Demonstration we got while sending the notification:

Privacy Aspect in our Approach

When dealing with sensitive material, such as patient discussions, it is critical for an app developer to emphasize user privacy and maintain confidentiality. To comply to these values, we devised our strategy to ensure that we do not have access to the contents of the patient's dialogue while functioning as a trusted person. However, because this is an improved feature, we are implementing these ways over E2EE encryption, which was our baseline because talkspace already has it. To do this, we developed a scoring mechanism that evaluates the intensity of symptoms without exposing the details of the conversation. This scoring method employs strategies such as differential privacy, which introduces random noise into the dataset impact a single data point can have on the overall result. In the given code, sensitivity is calculated using the L1 norm. The L1 norm measures the absolute difference between each sentence in the dataset and counts how many times each sentence appears. The maximum count among all the sentences represents the sensitivity value. By calculating the sensitivity, we can determine the maximum amount of noise that should be added to the output in order to provide privacy guarantees. The noise is added in a way that

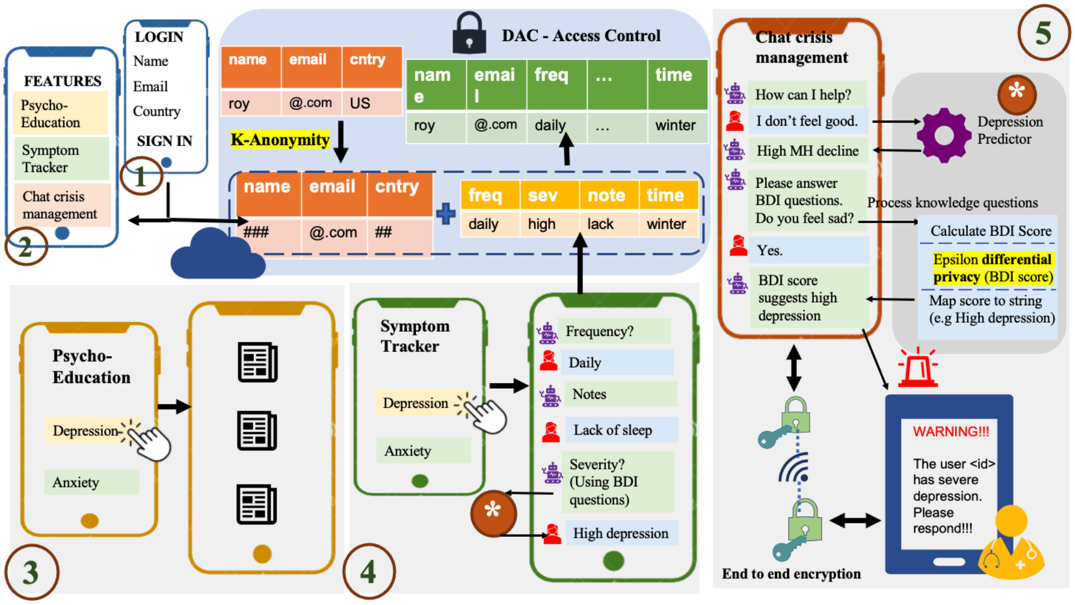
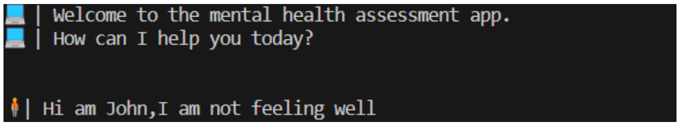
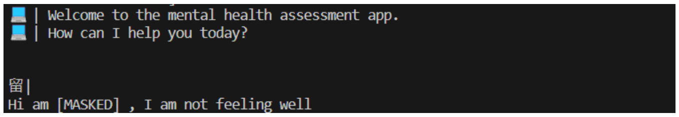


Figure 10: Illustration of the modified TalkSpace app.



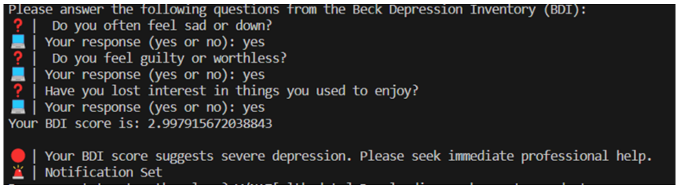
```
💻 | Welcome to the mental health assessment app.
💻 | How can I help you today?


🧍| Hi am John,I am not feeling well
```

Figure 11: Before we introduce masking for private details.



```
💻 | Welcome to the mental health assessment app.
💻 | How can I help you today?


留|
Hi am [MASKED] , I am not feeling well
```

Figure 12: After we introduce masking for private details.



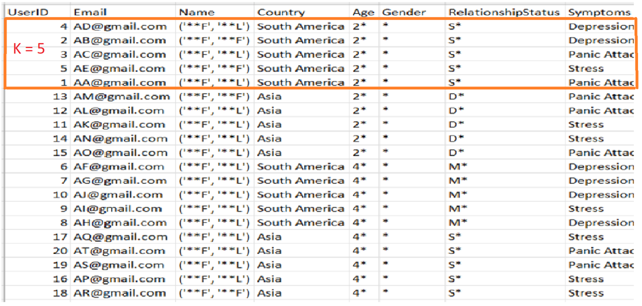
```
Please answer the following questions from the Beck Depression Inventory (BDI):
❓ |  Do you often feel sad or down?
💻 | Your response (yes or no): yes
❓ |  Do you feel guilty or worthless?
💻 | Your response (yes or no): yes
❓ | Have you lost interest in things you used to enjoy?
💻 | Your response (yes or no): yes
Your BDI score is: 2.997915672038843

🔴 | Your BDI score suggests severe depression. Please seek immediate professional help.
🚨 | Notification Set
```

Figure 13: Chat Demonstration.

ensures individual data points cannot be distinguished in the final result, thus protecting the privacy of the individuals in the data set., and masking of crucial data in spoken questions. We reduce the possibility of data leaking by using end-to-end encryption, providing consumers with a safe environment. In short, our primary goal was to protect user privacy by creating a system that measures mental health without jeopardizing the confidentiality of patient discussions. We attempt to establish a safe and trustworthy environment for consumers by utilizing techniques such as differential privacy, data masking, and encryption.

4 Results

Following the previously outlined methodology, tests were conducted for the health apps under review. In this section we present the outcomes of the analysis for the app after reviewing their privacy policies. In addition, we also present the the modified architecture of our proposed app after integrating all our envisaged features which can be referred from figure 10.

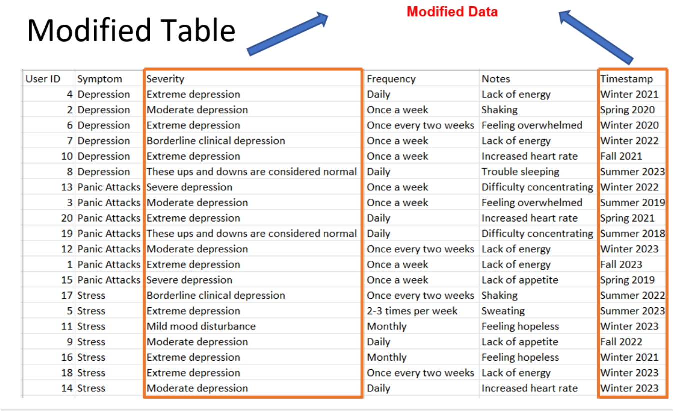


	UserID	Email	Name	Country	Age	Gender	RelationshipStatus	Symptoms
K = 5	4	AD@gmail.com	('**F', '**L')	South America	2*	*	S*	Depression
	2	AB@gmail.com	('**F', '**F')	South America	2*	*	S*	Depression
	3	AC@gmail.com	('**F', '**L')	South America	2*	*	S*	Panic Attac
	5	AE@gmail.com	('**F', '**F')	South America	2*	*	S*	Stress
	1	AA@gmail.com	('**F', '**L')	South America	2*	*	S*	Panic Attac
	13	AM@gmail.com	('**F', '**F')	Asia	2*	*	D*	Panic Attac
	12	AL@gmail.com	('**F', '**L')	Asia	2*	*	D*	Panic Attac
	11	AK@gmail.com	('**F', '**L')	Asia	2*	*	D*	Stress
	14	AN@gmail.com	('**F', '**L')	Asia	2*	*	D*	Stress
	15	AO@gmail.com	('**F', '**L')	Asia	2*	*	D*	Panic Attac
	6	AF@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
	7	AG@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
	10	AJ@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
	9	AI@gmail.com	('**F', '**L')	South America	4*	*	M*	Stress
	8	AH@gmail.com	('**F', '**L')	South America	4*	*	M*	Depression
	17	AQ@gmail.com	('**F', '**L')	Asia	4*	*	S*	Stress
	20	AT@gmail.com	('**F', '**L')	Asia	4*	*	S*	Panic Attac
	19	AS@gmail.com	('**F', '**L')	Asia	4*	*	S*	Panic Attac
	16	AP@gmail.com	('**F', '**L')	Asia	4*	*	S*	Stress
	18	AR@gmail.com	('**F', '**F')	Asia	4*	*	S*	Stress

Figure 14: Illustration of the final table for Psychoeducation with K-Anonymity.

4.1 Psychoeducation

After the aforementioned methodology, the data exemplified in the figure 14 highlights the k-anonymity which is 5 to have a set of records indistinguishable and further to be used along with the modified symptom tracker dataset to have a combined dataset to be used for our enhanced feature which we have proposed for our project.


Modified Table
Modified Data

User ID	Symptom	Severity	Frequency	Notes	Timestamp
4	Depression	Extreme depression	Daily	Lack of energy	Winter 2021
2	Depression	Moderate depression	Once a week	Shaking	Spring 2020
6	Depression	Extreme depression	Once every two weeks	Feeling overwhelmed	Winter 2020
7	Depression	Borderline clinical depression	Once a week	Lack of energy	Winter 2022
10	Depression	Extreme depression	Once a week	Increased heart rate	Fall 2021
8	Depression	These ups and downs are considered normal	Daily	Trouble sleeping	Summer 2023
13	Panic Attacks	Severe depression	Once a week	Difficulty concentrating	Winter 2022
3	Panic Attacks	Moderate depression	Once a week	Feeling overwhelmed	Summer 2019
20	Panic Attacks	Extreme depression	Daily	Increased heart rate	Spring 2021
19	Panic Attacks	These ups and downs are considered normal	Daily	Difficulty concentrating	Summer 2018
12	Panic Attacks	Moderate depression	Once every two weeks	Lack of energy	Winter 2023
1	Panic Attacks	Extreme depression	Once a week	Lack of energy	Fall 2023
15	Panic Attacks	Severe depression	Once a week	Lack of appetite	Spring 2019
17	Stress	Borderline clinical depression	Once every two weeks	Shaking	Summer 2022
5	Stress	Extreme depression	2-3 times per week	Sweating	Summer 2023
11	Stress	Mild mood disturbance	Monthly	Feeling hopeless	Winter 2023
9	Stress	Moderate depression	Daily	Lack of appetite	Fall 2022
16	Stress	Extreme depression	Monthly	Feeling hopeless	Winter 2021
18	Stress	Extreme depression	Once every two weeks	Lack of energy	Winter 2023
14	Stress	Moderate depression	Daily	Increased heart rate	Winter 2023

Figure 15: Illustration of the modified table for symptom tracker.

4.2 Symptom Tracker

For tracking symptoms, the data presented in Figure 7 represents the original data collected from the user. It is important to note that this information is not directly accessible to the therapist.

Privacy is a critical consideration when making modifications to the timestamp information and severity columns, as illustrated in Figure 15. These adjustments are implemented to safeguard the confidentiality of users' sensitive data while ensuring the provision of meaningful insights for symptom tracking. By altering the timestamp information and severity columns, the potential risks associated with directly knowing the exact timestamps of recorded symptoms and the specific Beck Depression Inventory (BDI) severity scores are mitigated. These modifications address privacy concerns by reducing the likelihood of predicting the specific answers provided by users to obtain their BDI scores. By proactively addressing these privacy considerations, the system maintains the integrity of user data while upholding the highest standards of privacy protection.

4.3 Crisis Management

The application's deployment of chat-based crisis management generates positive outcomes. For starters, it allows for quick contacts between patients and healthcare experts during high-stress situations, ensuring that patients receive instant treatment and support when their normal doctor is unavailable. Furthermore, the feature emphasizes patient privacy by utilizing tactics such as differential privacy and data masking to protect the anonymity of patient conversations while successfully transmitting their condition to the doctor. This not only protects user privacy but also creates a safe and trustworthy environment for customers. Furthermore, the application's real-time communication facilities allow for faster access to therapy, minimizing wait times for help and boosting patient results. The crisis management component illustrates the need of harnessing technology to offer great healthcare, especially in emergency situations, by satisfying critical healthcare demands. In Figure 16, the graphical representation demonstrates different probabilities for various cases and timelines. This visual representation provides a clear understanding of the probabilistic assessment of mental health decline and helps identify the severity of depression symptoms, ranging from minimal to severe depression.

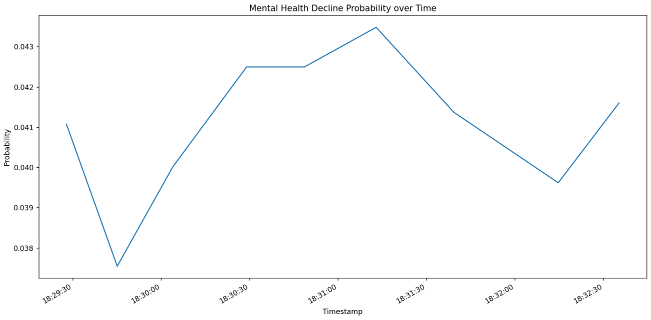


Figure 16: Graphical Results for the Crisis Management in Chat.

5 Discussion

In this section, we will highlight and discuss the results of the health app's testing and analysis, as well as the implications of its privacy policies. The analysis results give useful insights into the app's privacy policies and any hazards linked with user data. Furthermore, in accordance with the feedback received from Dr. Roberto Yus, we have articulated our discussion section accordingly. We have framed the question as Q_n asked us during the feedback and tried to answer them to the best of our knowledge.

5.1 Psychoeducation

Q1: Are the key attributes suppressed before sharing the information with the therapist?
The key attributes are suppressed before sharing the information with the therapist except the UserID.

Q2: Is there any sensitive attribute shared with the therapist after applying K-Anonymization technique on the dataset?
The symptoms attribute is the sensitive attribute shared with the therapist after applying K-Anonymization technique on the dataset.

5.2 Symptom Tracker

Q1: Why Role-based access control (RBAC) have been used over Discretionary access control(DAC)?
RBAC is chosen over DAC for our feature due to its structured and efficient approach to access control. With distinct roles for users, app developers/admins, and therapists, RBAC provides granular control and flexibility in assigning permissions. This enhances security, privacy, and minimizes the risk of unauthorized access. RBAC also reduces administrative overhead by centralizing access control management. Overall, RBAC offers a more streamlined and effective solution for our feature's access control requirements.

5.3 Crisis Management

Q1: How the value of epsilon being used and its effect on the feature?

In our approach epsilon is used in calculation of private counts,which is a list containing the private counts of occurrences for each sentence. By adding Laplace noise to the real counts, these private counts are generated.We have calculated the sensitivity using the L1 norm,which represents the maximum difference in counts when replacing a sentence with another. Each sentence count includes the Laplace noise. Divide the sensitivity by the value of epsilon to get the scale of the Laplace distribution. As epsilon is decreased, the scale grows, resulting in more noise being added to the counts. Increasing epsilon, on the other hand, lowers the quantity of noise generated. For training, private counts from noisy data are utilized as labels. Smaller epsilon values give more privacy but introduce uncertainty, whereas bigger values provide less privacy but more accuracy.

Q2: Why have we used chat system and BDI together? Ideally, a chat system with masking or anonymization would have sufficed, but in order to calculate the mental health decrease precisely, we needed certain clinically based inquiries.That's why, in addition to the Chat system, we have to integrate BDI questions.

Q3: Different scenarios (e.g., Epsilon values, Epochs) to test our Crisis Management service? We carried out a variety of experiments, including changes to the training and testing datasets, alterations to epsilon values, and investigations into other variables. In the course of our experiments, we applied three different epsilon (e.g., 0.1, 0.6, 0.9) and epochs (e.g., 50, 100, 150) values to find the optimal setup for evaluating the likelihood of mental health deterioration.

To further boost accuracy and maintain user safety, we deem it crucial to considerably broaden our term database. This expansion will equip the system to deal with a broader spectrum of situations and expressions, while also facilitating additional epoch runs to refine the results. In this code, Laplace noise is incorporated into each sentence's counts to achieve differential privacy. The scale parameter of the Laplace distribution is computed by dividing sensitivity by epsilon (sensitivity / epsilon).

It's important to remember that a smaller epsilon leads to a larger scale of the Laplace noise. This contributes to a higher volume of noise being added to the counts, enhancing privacy protection, but potentially introducing more distortion or inaccuracies in the data and vice versa.

Our observations indicated that the selection of epsilon had a considerable influence on the outcomes. When epsilon was assigned a value of 0.1, the system attained a 41% accuracy in evaluating the probability of mental health decline. Upon incrementing epsilon to 0.6, the accuracy rose to 53%, and a further increase to 0.9 yielded an accuracy of 81%. These percentages reflect the proportion of the maximum sensitivity added to the counts, with higher epsilon values indicating a larger amount of noise introduced.

6 Conclusion

In conclusion, this paper underlines the growing apprehensions regarding the accumulation, storage, and confidentiality of user data in mental health applications. Given the lack of definitive principles or standards in this domain, it is necessary to advocate for more transparency, accountability, and robust regulatory structures to protect user privacy. The study probes into the data privacy policies of a range of mental health apps, examining their adherence to regulations, methods of data collection and usage, and strategies to maintain user privacy. Overall, the two new improved features have been rolled out: an enhanced Psychoeducation+ symptom tracker and a crisis management tool within chat services. Previously, these functionalities operated without any specific privacy measures, but the updated features offer greater capabilities while significantly bolstering privacy. By implementing anonymization and differential privacy, they've added a trigger warning system for patients needing urgent medical help, ensuring that this alert system maintains both safety and privacy within the chat service.

7 Limitations and Future Work

Our current implementation of the machine learning model for assessing the decline in mental health, with the intervention of a doctor and a trusted third party, has certain limitations that we acknowledge. One such limitation is the need to enhance the accuracy of the model, considering its sensitivity to patient health. Although we have incorporated noise to address privacy concerns, we recognize the importance of further improving the accuracy of our system. This aspect will be addressed in our future work.

Furthermore, another limitation of our project is the limited scope of questions included, which are currently focused only on the Beck Depression Inventory (BDI). To provide a more comprehensive assessment, it is crucial to include questions related to additional symptoms and mental health indicators.

As part of our future work, we aim to expand our database to incorporate unseen keywords and enhance the range of questions to cover a broader spectrum of mental health symptoms. This expansion will enable us to provide a more comprehensive and accurate assessment of an individual's mental health condition.

Contribution

- **Surjodeep Sarkar:** I have reviewed various MHAs from https://privacynotincluded.com and https://mindapps.org/ViewApp to analyze the functionalities and their related privacy issues. After having done a systematic review, I was able to create Figure 1 which will help us to determine which PET techniques needs to be added. Additionally, the survey helped me to understand the inner working of our targeted app **(e.g., TalkSpace)** in order to accumulate the data collection process being done by TalkSpace which is illustrated in Figure 2. After discussing with my team and curating the idea to develop the project I was able to create the modified architecture of our app which is exemplified in figure 10 along with the dataset creation with my team mates. The presentation was done and curated by me which I will be sharing over the mail. I have also helped the team to curate and write the mid-report and final report for the submission. I have taken some content of the introduction from my own research paper as well and for reference: https://arxiv.org/abs/2304.13191.

- **Ekta Pandey:** I started by analyzing the functionalities and privacy issues of various Mental Health Apps (MHAs) from sources like https://privacynotincluded.com and https://mindapps.org/ViewApp along with my team mates. My work focuses on developing the optimal technology for privacy preservation, data reduction, and architectural implementation, with the fundamental approach for the **T**alkSpace being on achieving a balance between utility, privacy, and data security. Because data minimization is the initial stage, my first tactic was to restrict the data collecting process (e.g., instead of utilizing GPS data for meetups, only use static addresses) and decide who has access to what data and how it is acquired. In addition, I focused on a new emergency chat system function and its coding.This includes BDI questions and masks, as well as PET-like differential privacy.My future work will include working on improving the accuracy of code for a chatbot in order to reach maximum accuracy for the benefit of the patient while keeping the privacy element of the same and testing it in various scenarios.
- **Rachit Saini:** I looked into the apps from https://privacynotincluded.com which are Headspace, Woebot, Calm, TalkSpace, and Betterhelp and gathered information about them. Based on the input from all the above apps, I understood where the gap is in terms of privacy applied in these apps. I focused specifically on studying the psychoeducation feature of the Talkspace App because it is of immense use in helping users to cope with their illnesses and accompanying treatments. I explained the intervention of psychoeducation feature. Moreover, I discussed the conditions addressed by the psychoeducation feature along with potential benefits. I created synthetic dataset in python for studying psychoeducation feature and elaborated the privacy techniques applied on this feature. Original data table, table with data generalization and suppression and K-Anonymity applied table are presented in this report. The future work will include on making other features of Talkspace app privacy enhanced. Moreover, similar features in other mental health provider apps to be made privacy enhanced.
- **Bhavani Shankar Mahamkali:** I began by studying the top five mental health apps listed on https://privacynotincluded.com which are Headspace, Woebot, Calm, TalkSpace, and Betterhelp. My primary objective was to analyze the features offered by these apps and the data that they collect from users to provide their functionalities. I also investigated the privacy issues associated with the collected data. I looked into whether the app uses any privacy enhancement techniques to mitigate the privacy concerns related to the collected data. I used the insights gained from this study in our project discussions to better understand the functioning of mental health apps and how privacy concerns can be addressed.To initiate the development of the symptom tracker feature, I began by thoroughly studying and comprehending its functionalities. As part of this process, I wrote a Python script that generates a synthetic dataset for users. This dataset includes essential information related to symptoms, such as severity, frequency, notes, and timestamps. Additionally, I implemented modifications to the table structure to ensure the privacy of the users' sensitive data. By anonymizing certain columns and applying mappings, I safeguarded user privacy while still preserving the necessary information for effective symptom tracking. This initial step sets the foundation for further implementation and enhancement of the symptom tracker feature.

References

[Zhang et. al.(2022)] Zhang et al T. Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine* **5** (2022) 46.

[Hyman et al.(2008)] Hyman et al I. Self-disclosure and its impact on individuals who receive mental health services. *SAMHSA* (2008).

[Srivastava et al.(2021)] Srivastava et al B. Did chatbots miss their "apollo moment"? potential, gaps, and lessons from using collaboration assistants during covid-19. *Patterns* (2021).

[Czeisler et al.(2020)] Czeisler et al MÉ. Mental health, substance use, and suicidal ideation during the covid-19 pandemic—united states, june 24–30, 2020. *Morbidity and Mortality Weekly Report* (2020).

[Parker and Halter(2019)] Parker L, Halter V. How private is your mental health app data? an empirical study of mental health app privacy policies and practices. *International Journal of Law and Psychiatry* **64** (2019) 198–204. doi:https://doi.org/10.1016/j.ijlp.2019.04.002.

[Tangari and Gioacchino(2021)] Tangari, Gioacchino. Mobile health and privacy: cross sectional study. *bmj* **373** (2021).

[Magdziarczyk(2019)] Magdziarczyk M. Right to be forgotten in light of regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec. *6th International Multidisciplinary Scientific Conference on Social Sciences and Art Sgem 2019* (2019), 177–184.

[Burkert(1996)] Burkert H. Some preliminary comments on the directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Lex Electronica* **2** (1996).

[De Hert and Papakonstantinou(2016)] De Hert P, Papakonstantinou. The new general data protection regulation: Still a sound system for the protection of individuals? *Computer law & security review* **32** (2016) 179–194.

[He and Naveed(2014)] He D, Naveed. Security concerns in android mhealth apps. *AMIA annual symposium proceedings* (American Medical Informatics Association) (2014), vol. 2014, 645.

[Papageorgiou and Strigkos(2018)] Papageorgiou A, Strigkos. Security and privacy analysis of mobile health applications: the alarming state of practice. *Ieee Access* **6** (2018) 9390–9403.

[PCE(2023)] [Dataset] Psychoeducation. https://en.wikipedia.org/wiki/Psychoeducation (2023).

[THC(2023)] [Dataset] The human condition. https://thehumancondition.com/psychoeducation/ (2023).

[DG(2023)] [Dataset] Data generalization. www.privitar.com/blog/data-generalization-advanced-de-identification/ (2023).

[DS(2023)] [Dataset] Data suppression. https://www.edglossary.org/data-suppression/ (2023).

[KA(2023)] [Dataset] K-anonymity. www.k2view.com/blog/what-is-k-anonymity (2023).

[Beck et al.(1961)Beck, Ward, Mendelson, Mock, and Erbaugh] Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of General Psychiatry* **4** (1961) 561–571.