Analysis of Third-Party Data Leaks on Finnish Mental Health Websites

Sampsa Rauti*, Esko Vuorinen*, Panu Puhtila*, Robin Carlsson*
* University of Turku, Turku, Finland
sjprau@utu.fi

***Abstract*—Mental health websites process private and sensitive personal data, and it is essential to prevent this confidential data from being compromised. However, the increasing practice of using various third-party services on modern websites poses a threat to online privacy, including online services focusing on mental well-being. We present a study on the privacy of 10 Finnish mental health websites and conduct a network traffic analysis to see whether these online services inadvertently share sensitive data with third-party entities. Our findings indicate all of the studied websites leak sensitive contextual data (such as the visited URLs) to third parties. The current paper analyzes the characteristics of these data leaks, and gives suggestions to avoid such privacy concerns in future.**

***Index Terms*—mental health websites, data leaks, data concerning health, web privacy, third-party services**

I. Introduction

Mental health websites serve as essential digital services, enabling individuals to access resources, support, and information related to mental well-being. A wide range of services is provided by these digital platforms, including online therapy and counseling, self-help resources and materials on topics surrounding mental health, and peer support communities. Such online services allow citizens, particularly many vulnerable groups, to use them in an equal and secure manner.

The COVID-19 pandemic also accelerated this digital transformation, increasing the demand for essential digital services. Furthermore, the pandemic contributed to an increase in mental health problems and made existing issues more pronounced [1], which highlights the importance of confidential mental health services online. The challenge is that state-of-the-art websites are created using various third-party services, like tools for tracking user activity and measuring performance [2]. Often the software developers and organizations responsible for the websites are not aware that users' sensitive personal data is transmitted to these third parties. Such data leakages have the potential to harm users of mental health services.

In this study, we examine the privacy implications of 10 Finnish mental health websites with network traffic analysis. We investigate whether these online platforms unintentionally disclose sensitive data to third parties. The current study also delves into the characteristics of the found data leaks and gives recommendations to mitigate privacy concerns in web-based health services. The dark patterns that potentially lead users to accept data collection and the privacy policies of the selected websites were also studied. Instead of selecting a large set of websites, the current study aims to take a more in-depth look at the chosen Finnish web services.

The remainder of the paper is structured as follows. Section II presents a review of related work. Section III provides an overview of the study setting and methodology. Section IV comprehensively discusses the results of our network traffic analysis and provides a detailed account of the data leaks identified and the relevant third parties. Section V focuses on the potential implications of sensitive data leaks on mental health websites and provides guidance on enhancing the software development process with respect to privacy concerns. Finally, Section VI concludes the paper.

II. Related work

There are not many papers discussing privacy issues of mental health websites. Surani et al. [3] studied mental health websites and came to a conclusion that majority of the studied websites had severe problems in their privacy policy statements.

Although the privacy of mental health websites has not received that much attention, there is some research on privacy and user data collection in mental health related mobile applications. A study by Parker et al. [4] inspected 61 mental health applications in the market between 2016-2018, and came to a conclusion that nearly half of them did not have any privacy policies at all and that many of the surveyed apps also actively encouraged the user to share their personal data. A study by Robillard et al. [5] details an even more severe situation. In their analysis, only 4% of the Android apps and 18% of the iOS apps did provide privacy policy at all, and that a majority of the policies presented just stated that the users data may be shared with third parties.

Jiang [6] found out that a sizable majority of mental health mobile apps did either not sufficiently inform, or even actively misinformed, the user of the personal data to be collected by third parties. Sagar and Singh [7] concluded that 50% of the studied applications did not provide privacy policy for the user to read before accessing the service, and of those that did, many had privacy policies which were written in ways that were hard or even impossible

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

to comprehend for layperson not accustomed to technical and legal terms. Thus any consent to data collection by these apps was misinformed, at best. A similar paper by Iwaya et al. [8] surveyed the 27 most popular mental health apps in Google App Store, and found very similar problems as Singh and Sagar. In other words, privacy policies, if presented at all, were opaque and hard to understand, causing users to give misinformed consent to data collection. Iwaya et al. also studied other potential privacy and security concerns in mental health apps in their paper, and found that severe problems such as inadequate cryptography and dangerous permissions requested by the applications were commonplace.

Huo et al. [9] studied the leaks of health data and the presence of third-party analytics in 459 online patient portals. Google Analytics was detected in 14% of these portals. Notably, sensitive data concerning health was found to leak to third parties on 9 websites. The leaks contained details such as prescribed medications and laboratory results. The outcomes from Huo et al.'s study, much like our current research, highlight the absence of a privacy-by-design approach in many essential online services. There is a great need to educate website developers and data protection officers about potential privacy issues caused by third parties. The discourse on the utilization of third-party services on medical websites is extensive [10]–[13]. The message in these publications is quite unanimous: third-party analytics collect sensitive personal data without adequately informing users or obtaining their consent. Recent studies show that this threat to user privacy is present on hospital websites as well [14], [15].

The findings of previous studies paint a uniform picture of a situation, in which the landscape of online mental health services is severely plagued by either missing, hard-to-comprehend or outright misleading privacy policies. As these earlier studies mostly concern privacy policies and regulation, not much is said about the actual data leakages in digital mental health services. Our study aims to bridge this gap by presenting a network traffic analysis of the data leaks present in web-based mental health services.

III. Study Setting and Methods

In this study, we analyzed 10 mental health service websites, how they collected personal data, and whether this data was leaked to third parties. All of the studied websites are operated by mental health service providers based in Finland. In this study, we have chosen not to mention the selected mental health service providers by their actual names. Instead, we refer to them as MH1, MH2, etc. The service providers included private companies (MH6, MH7), services run by public sector bodies (MH1, MH2, MH4), and non-profit associations (MH3, MH5, MH9, MH8, MH10). Since we found no official list of mental healthcare providers in Finland, we performed a keyword search using the Google search engine. We used "Mielenterveyspalvelu" (mental health service) as a keyword to find suitable websites.

The exact test sequence used to analyze the websites varied between different services, since there was little uniformity between them, and their sub-pages were named differently. Hence we refer to these sub-pages only by their general category, such as help-seeking pages. The basic pattern of conducting the test consisted of the researcher navigating to the main page of the mental health web service, after which they consented to all cookies and data collection. The researcher then performed a search and navigated to the first link provided on the result page. The search terms we used were all mental health related Finnish words, including anxiety, depression and loneliness. The choice of words varied between websites since not all websites generated search results for the same search term. If the website in question did not have search functionality, the study was conducted by navigating to the "information about the services" page. In addition to these actions, the researcher accessed a help-seeking page or an appointment booking page.

During the described test sequence, all network traffic was recorded with the Google Chrome Developer Tools (devtools). The produced traffic log files were then filtered for web requests to external domains to determine whether the website leaked data to third parties. Furthermore, the network traffic was carefully inspected for potentially sensitive personal data leaks. In particular, we wanted to see whether the following data items leaked to third parties:

- The URLs of the pages visited by a user.
- The search terms used by a user, potentially revealing mental health conditions.
- The fact that the user accessed the help-seeking page, revealing a probable intent to seek help for mental health problems.

Apart from studying the actual data leakages, we also examined the privacy policies of these websites, and studied whether dark patterns were found in their cookie consent banners. The privacy policies were evaluated on the principle of whether all third-party analytics services were named, whether it was mentioned that the user can be identified from the collected data, whether it was mentioned that the information about the visited pages was collected, and whether the information about the used search terms was collected. In studying the dark patterns, we used the "Report of the work undertaken by the Cookie Banner Taskforce"[1] produced by the European Data Protection Board, which outlines several practices in cookie banner design and implementation which should be considered as intentionally misleading to the user. Of the dark patterns proposed in the report, we chose four to be studied: 1) no rejection button in the first layer of the cookie consent banner, 2) pre-ticked consent boxes, 3) deceptive button colors and 4) deceptive button contrasts. These four dark patterns are easy to understand and study, and also act as a consistent gauge whether the user is visibly being deceived by the cookie banner.

[1] https://edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-cookie-banner-taskforce_en

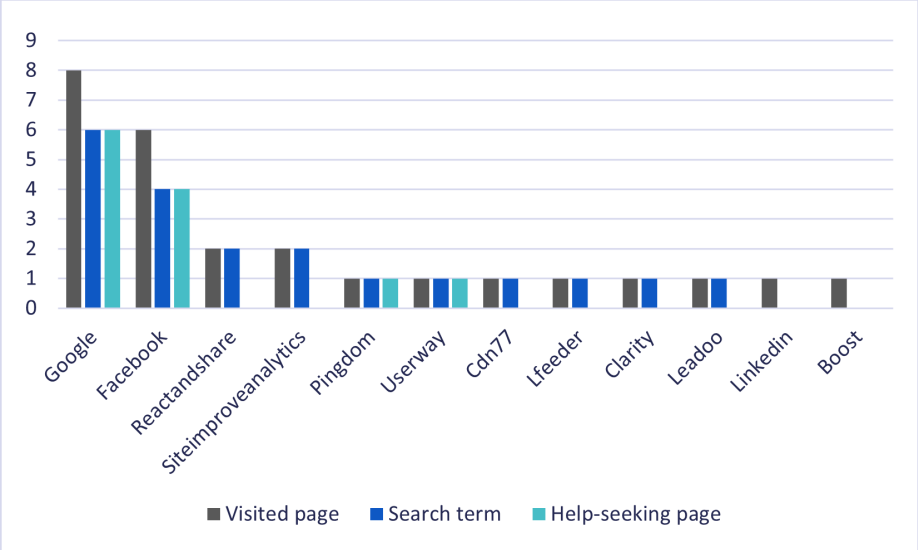

Fig. 1. The third-party services receiving sensitive personal data from mental health websites.

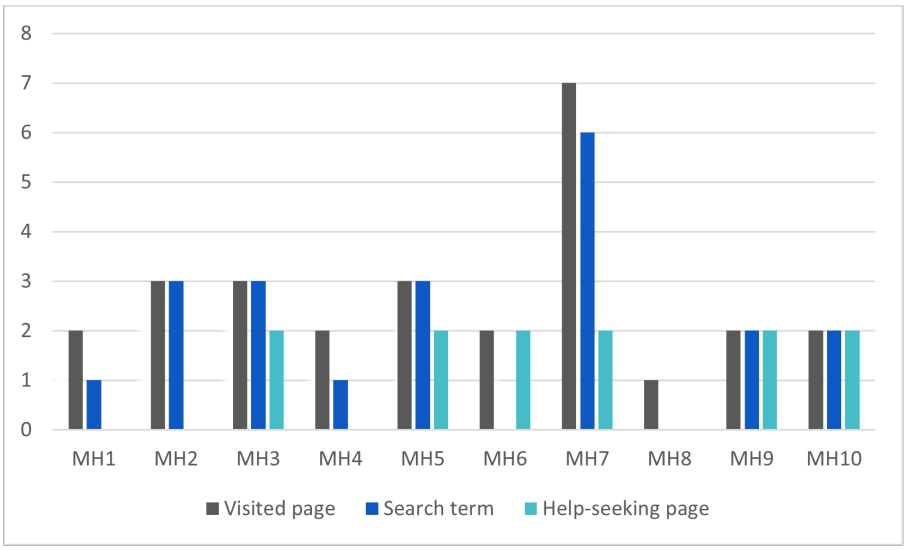

Fig. 2. The numbers of data leaks to distinct third parties on mental health websites by leak category.

Since ”personal data” is a central concept in this paper, we will briefly define the term here. The definition of the term given in GDPR and also used by the Finnish Office of the Data Protection Ombudsman is sufficient for this purpose. Hence, personal data is defined as ”all data related to an identified or identifiable person” [2]. By this definition, technical information such as IP addresses, device identifiers, accurate location data or any data point that identifies the user of the website counts as personal data. It must also be noted that while many technical details such as device type or screen resolution alone are not sufficient to identify someone, together these data items can be used for assisting in identification of a specific person. Therefore, they must be regarded as belonging under the concept of personal data.

[2]https://tietosuoja.fi/en/processing-of-personal-data

IV. Results

Figure 1 shows the third-party services receiving sensitive personal data on the studied mental health websites. Not surprisingly, Google Analytics and Meta Pixel analytics services received the largest amounts of sensitive personal data. Figure 1 also shows the types of data leaks for each third party. Google and Meta/Facebook, for example, were involved in all three studied types of data leaks, receiving information on visited URLs, used search terms, and accessed help-seeking pages. Google received sensitive data from 8 out of 10 mental health websites, while Meta/Facebook got data from 6 websites. Other third parties played a significantly smaller role – React & Share (user feedback and engagement tool) and Siteimproveanalytics (website analytics tool) received data from 2 websites. Pingdom (performance and availability monitor-

ing tool), UserWay (web accessibility solution), CDN77 (content delivery network), Leadfeeder (lead generation tool), Clarity (Microsoft's analytics service), Leadoo (lead generation tool), LinkedIn (professional networking platform) and boost.ai (AI chatbot platform) only received data from one website each.

Figure 2 shows the numbers of data leaks on each studied mental health website, classified into leak categories. Our network traffic analysis shows that every single one (10) of the studied websites leaked the URLs the user had visited to at least one third party. Search terms, potentially containing sensitive mental health conditions or symptoms, were leaked on 8 websites. The two remaining websites (MH6 and MH8) did not have a search box, which means none of the studied websites had a confidential search function! In 6 of the studied websites, the fact that the user accessed the help-seeking page was revealed to third parties.

These numbers paint a grim picture of the privacy of the analyzed mental health websites. Although in this experimental setup the consent to cookies and data collection has been given, sharing sensitive data associated with the user's mental health is ethically questionable. Even when users accept data collection on the website, they are unlikely to fully understand the potential consequences of leaking sensitive URLs to third-party services.

It is also noteworthy that in all services except for MH8, there were at least 2 distinct third parties. MH7, a private mental health clinic, had 7 third-party services tracking visited pages and 6 services tracking search terms! Such extensive sharing of potentially sensitive personal data significantly heightens the risk of data misuse. Other than that, mental health services from different sectors seemed to be quite even in terms of third-party data leaks. It is worth noting, though, that the public sectors websites (MH1, MH2, M4) did not have any data leaks on their help-seeking pages.

Table II shows how well the studied privacy policies informed the user about the actual data collection taking place on the websites. As can be seen by the sparse matrix, the results were far from satisfactory. To begin with, only two of the websites (Mh6, MH7), the private sector services, appropriately informed about the possibility of identifying a specific user. Similarly, two services (M2, M7) named all third parties collecting data. This is a problem because users should know who processes their data. Three services (MH1, MH6, MH9) mentioned that visited pages (URLs) are collected. The consequence of collecting URLs is very often the fact that search terms are also collected as a part of the URL address. Collecting search terms was never mentioned (it is worth noting MH6 and MH8 did not have a search function, but all other services did leak search terms).

Table I shows the dark patterns on cookie banners of the studied mental health websites. The consent column indicates whether the website had a cookie banner asking for consent, and the three columns. In the table, green indicates a positive outcome (consent was asked or a dark pattern was not present) and red indicates a negative outcome (consent was not asked or a dark pattern was present). Websites MH8–MH10 did not have consent banners, so dark patterns could not be studied (black cells).

Mental health websites MH1–MH7 asked for consent for cookies and data collection. They also had a reject button on the first layer in the cookies banner and did not contain any pre-ticked consent boxes. However, deceptive colors and contrast were found in 4 cases (MH3, MH5, MH6, MH7). Deceptive colors and contrast are usually used to clearly highlight the accept button to lure users into clicking it [16]. Websites MH8–MH10 failed to ask consent for data processing and therefore, violated the GDPR. According to the GDPR, consent must be freely given, specific, informed, and unambiguous [17]. These last three websites in Table I were all websites of mental health associations. On the other hand, the public sector mental health websites (MH1, MH2, MH4) had no dark patterns at all.

V. Discussion

Sensitive mental health related information needs special protection due to its potential to unveil an individual's medical conditions. It is important to acknowledge, however, that we cannot assert with certainty whether analytics providers actually store and utilize this sensitive data. Regardless, simply sharing of data with third parties is unacceptable. While these third parties may not necessarily exploit the leaked personal data, collecting and aggregating substantial amounts of data concerning the use of mental health services exposes individuals to potential unethical exploitation. This may include malicious targeting and profiling, which highlights the need for a serious approach to data privacy.

If users learn of data leaks, this can erode trust in the mental health platforms and lead to a potential loss of users. In worst cases, individuals may face discrimination or social stigma when their mental health data is exposed. This, in turn, can affect their personal and professional lives. Data leaks may also lead to stress and anxiety, worsening existing mental health conditions. When it comes to organizations that fail to protect personal data concerning mental health, they may face legal consequences. Organizations responsible for data leaks may also suffer reputational damage.

To prevent these kinds of repercussions, several considerations web developers and data protection officers should take into account regarding the privacy and third-party services on mental health websites:

- *Removing third-party analytics*: On mental healthcare websites, using third-party analytics and collecting of personal data should be avoided. When analytics are considered absolutely necessary, self-hosted solutions like Matomo [18], [19] should be deployed to maintain full control over the collected data.

TABLE I
DARK PATTERNS ON COOKIE BANNERS OF THE STUDIED MENTAL HEALTH WEBSITES.

Website	Consent	No reject button on the first layer	Pre-ticked consent boxes	Deceptive colors/ contrast
MH1				
MH2				
MH3				
MH4				
MH5				
MH6				
MH7				
MH8				
MH9				
MH10				

TABLE II
COMPARING THE PROVISIONS OF PRIVACY POLICIES ON THE STUDIED MENTAL HEALTH WEBSITES.

Website	Identifying individual user mentioned	All third parties named	URL address collection mentioned	Search term collection mentioned
MH1			X	
MH2		X		
MH3				
MH4				
MH6	X		X	
MH7	X	X		
MH8				
MH9			X	
MH10				

- *Careful evaluation of third-party services*: The third-party services used on mental health websites have to be thoroughly assessed in terms of user privacy. Their use has to be strongly justified, especially in the web services in which sensitive data is involved. Some trusted services like chat and appointment booking systems can be used after careful review.
- *Awareness of third-party services*: Many content management systems and off-the-shelf platforms used to build modern websites include third-party analytics by default or at least allow these services to be added very easily through extensions, and may not be best choices for building mental health websites without proper configuration. Developers should analyze the outgoing network traffic to ensure that visitors' personal data is not being sent to third parties, particularly on pages that handle sensitive personal data.
- *Familiarity with the healthcare sector*: Developers need to have a good understanding of regulations on processing health data. To figure out the data security requirements for a mental healthcare service, developers should also continuously communicate with stakeholders.
- *External audits*: Currently, websites in healthcare sector seem to be a blind spot when it comes to addressing privacy issues caused by the use of third-party services. Because of this, privacy audits for mental health web services processing sensitive personal data would not be a bad idea.
- *Transparent privacy policies*: Privacy policies should transparently reveal how third parties collect personal data. A privacy policy document should explicitly name the third parties that receive personal data, and list the processed data items. Using templates or checklists might also help in creating privacy policy documents [20].
- *Avoiding dark patterns*: Dark patterns lure users into accepting information collection by deceptive means and are particularly problematic on healthcare websites. Developers should take care to remove such patterns, also recognizing that they may be a part of pre-built cookie management systems.

VI. CONCLUSION

In the current study on Finnish mental health websites, we found that out of 10 sites examined, all leaked at least some type of personal and potentially sensitive data. The fact that users' potentially sensitive search terms are leaked is especially concerning. This result highlights the fact that there are significant privacy challenges in Finnish web-based mental healthcare services. Also, the problem likely extends beyond the sites we studied. In future work, we aim to analyze the privacy of other medical web resources, such as websites for medical condition support associations and Finnish healthcare district websites.

Our findings will hopefully encourage software developers and data protection officers of web-based healthcare services to prioritize the assessment of third-party services and adopt the privacy-by-design approach. It is important for developers and maintainers to understand their responsibility for protecting sensitive user data. Furthermore, users should be transparently informed about data processing practices and possible third-party involvement. In web-based mental health services, however, relying on external services for data collection is not justifiable. When healthcare service providers fail to address serious data leaks, user trust and privacy may be lost, and vulnerable user groups in particular can be affected. Developers should aim to build web-based mental health services that instill the same level of trust as traditional on-site healthcare settings do for users.

REFERENCES

[1] T. Wu, X. Jia, H. Shi, J. Niu, X. Yin, J. Xie, and X. Wang, "Prevalence of mental health problems during the covid-19 pandemic: A systematic review and meta-analysis," *Journal of affective disorders*, vol. 281, pp. 91–98, 2021.

[2] T. Heino, R. Carlsson, S. Rauti, and V. Leppänen, "Assessing discrepancies between network traffic and privacy policies of public sector web services," in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 2022, pp. 1–6.

[3] A. Surani, A. Bawaked, M. Wheeler, B. Kelsey, N. Roberts, D. Vincent, and S. Das, "Security and privacy of digital mental health: An analysis of web services and mobile applications," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2023, pp. 319–338.

[4] L. Parker, V. Halter, T. Karliychuk, and Q. Grundy, "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*, vol. 64, pp. 198–204, 2019.

[5] J. M. Robillard, T. L. Feng, A. B. Sporn, J.-A. Lai, C. Lo, M. Ta, and R. Nadler, "Availability, readability, and content of privacy policies and terms of agreements of mental health apps," *Internet Interventions*, vol. 17, p. 100243, 2019.

[6] K. Jiang, "Mental health mobile apps and the need to update federal regulations to protect users," *Mich. Tech. L. Rev.*, vol. 28, p. 421, 2021.

[7] S. Singh and R. Sagar, "Time to have effective regulation of the mental health apps market: Maximize gains and minimize harms," *Indian Journal of Psychological Medicine*, vol. 44, no. 4, pp. 399–404, 2022.

[8] 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," *Empirical Software Engineering*, vol. 28, no. 2, 2023.

[9] M. Huo, M. Bland, and K. Levchenko, "All eyes on me: Inside third party trackers' exfiltration of phi from healthcare providers' online systems," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, ser. WPES'22. New York, NY, USA: Association for Computing Machinery, 2022, p. 197–211.

[10] J. Burkell and A. Fortier, "Consumer health websites and behavioural tracking," ser. Proceedings of the Annual Conference of CAIS / Actes Du congrès Annuel De l'ACSI., 2013.

[11] ——, "Privacy policy disclosures of behavioural tracking on consumer health websites," ser. Proceedings of the American Society for Information Science and Technology, 2014.

[12] M. Huesch, "Privacy threats when seeking online health information," *JAMA Internal medicine*, vol. 173, 2013.

[13] K. Masters, "The gathering of user data by national medical association websites," *The Internet Journal of Medical Informatics*, vol. 6, 2012.

[14] X. Yu, N. Samarasinghe, M. Mannan, and A. Youssef, "Got sick and tracked: Privacy analysis of hospital websites," ser. IEEE European Symposium on Security and Privacy Workshops, 2022.

[15] A. B. Friedman, R. M. Merchant, A. Maley, K. Farhat, K. Smith, J. Felkins, R. E. Gonzales, L. Bauer, and M. S. McCoy, "Widespread third-party tracking on hospital websites poses privacy risks for patients and legal liability for hospitals," *Health Affairs*, vol. 42, 2023.

[16] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(un) informed consent: Studying gdpr consent notices in the field," in *Proceedings of the 2019 acm sigsac conference on computer and communications security*, 2019, pp. 973–990.

[17] S. Breen, K. Ouazzane, and P. Patel, "Gdpr: Is your consent valid?" *Business Information Review*, vol. 37, no. 1, pp. 19–24, 2020.

[18] A. Chandler and M. Wallace, "Using Piwik instead of Google analytics at the Cornell university library," *The Serials Librarian*, vol. 71, no. 3-4, pp. 173–179, 2016.

[19] J. Gamalielsson, B. Lundell, S. Butler, C. Brax, T. Persson, A. Mattsson, T. Gustavsson, J. Feist, and E. Lönroth, "Towards open government through open source software for web analytics: The case of matomo," *JeDEM-eJournal of eDemocracy and Open Government*, vol. 13, no. 2, pp. 133–153, 2021.

[20] M. Rowan and J. Dehlinger, "A privacy policy comparison of health and fitness related mobile applications," *Procedia Computer Science*, vol. 37, pp. 348–355, 2014.